Title

The Mortgage-Cash Premium Puzzle

Authors

REHER, MICHAEL
VALKANOV, ROSSEN

Publication Date

2024

DOI

10.1111/jofi.13373

The Mortgage-Cash Premium Puzzle

Michael Reher and Rossen Valkanov *

ABSTRACT

All-cash homebuyers account for one-third of U.S. home purchases between 1980 and 2017. We use multiple data sets and research designs to robustly estimate that mortgaged buyers pay an 11% premium over all-cash buyers to compensate home sellers for mortgage transaction frictions. A dynamic, representative-seller model implies only a 3% premium, which would suggest an 8% puzzle. Accounting for heterogeneity in selling conditions explains half of this difference, but a puzzle holds in conditions with high transaction risk. An experimental survey of U.S. homeowners replicates these patterns and suggests that belief distortions can explain the puzzle in these high-risk states.

Keywords: House Prices, Cash Buyers, Asset Pricing Puzzles, Affordability

JEL Classification: G31, R30, G12, G21, G41

---

*Reher and Valkanov are with the University of California San Diego, Rady School of Management. We thank On Amir, Carlos Avenancio-Leon, Paul Chiou, Andra Ghent, Sonia Gilbukh, Uri Gneezy, Lu Han, Jun Liu, Gregor Matvos, Peter Maxted, Will Mullins, Alberto Plazzi, Alessandro Previtero, Alberto Rossi, Yuval Rottenstreich, Gregor Schubert, Charles Sprenger, Alexis Toda, Rick Townsend, Neng Wang, Wei Xiong, an anonymous Associate Editor and referees, and seminar participants at the SFS Cavalcade, EFA Annual Meeting, AREUEA, PBFEAM, NFA Annual Meeting, FMA Annual Meeting, UC San Diego, Santa Clara University, University of Wisconsin, Indiana University, Aarhus University, Chapman University, City University of New York, the FDIC, the European Winter Finance Summit, and the FHFA for comments. Ziyi Chen and Shenyan Zhu provided excellent research assistance. We are grateful to the following data vendors for providing their data under academic terms: Zillow, Recursion Co, and the California Association of Realtors. We thank Redfin for the opportunity to analyze their data, and we especially thank Taylor Marr who facilitated the process. The views expressed in this paper are those of the authors and do not reflect the position of Zillow, Recursion, the California Association of Realtors, or Redfin. This study is exempt from Institutional Review Board requirements under category 45 CFR 46.104(d)(2). We have read the Journal of Finance's disclosure policy and have no conflicts of interest to disclose.

Correspondence: Rossen Valkanov, Rady School of Management, Department of Finance, University of California, San Diego, 9500 Gilman Drive, MC0553, Otterson Hall, La Jolla, CA 92037; e-mail: rvalkanov@ucsd.edu.

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 should be zero (Modigliani and Miller (1958)). More realistically, the premium should be positive to compensate sellers for frictions in the mortgage origination process, namely, the risk of transaction failure and a longer time to close. We find that the mortgage-cash premium averages 11% over the past 40 years. In dollar terms, a seller would be indifferent between a $500,000 all-cash offer, the average in our sample, and a mortgaged offer that is $55,000 larger. The magnitude of this premium is puzzlingly large relative to the degree of friction in the typical home purchase. In policy terms, U.S. taxpayers subsidize $8 trillion of mortgages to promote homeownership (Federal Reserve (2019)); reducing the mortgage-cash premium would enable a smaller subsidy to accomplish the same goal.

To make things more 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 (National Association of Realtors (2020)). If the transaction fails, the seller relists the home in one month after reducing the price by 10%, the average price cut during the 2008 crisis (Trulia (2009)). Supposing the seller is indifferent and has a discount rate of zero over this short horizon, the mortgaged offer should require a price premium that solves

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

The resulting premium equals 0.6%, far below what we estimate empirically.

While this back-of-envelope calculation omits many important considerations of real-life home sellers, in Section I we employ a more serious model that features a risk-averse seller with a short position in her current home (i.e., outstanding mortgage) and a down payment on her next one, a dynamic tradeoff between accepting a given buyer's offer versus waiting for another buyer to make an offer, option value from the possibility that buyers engage in an endogenous "bidding war," carrying costs from maintaining the home on the market, and many other realistic details regarding the micro-structure of home sales, including home-sale contingencies. We calibrate this model to match the conditions of a representative (i.e., average) U.S. home seller, using a rich collection of data sets. This exercise yields a 3% premium, which we take as a starting point going into our empirical analysis.

We begin the empirical analysis in Section II by documenting 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 most research focuses on mortgage-financed purchases. We then document that the average price of an all-cash purchase is 30% lower than the average price of a mortgaged purchase. At a descriptive level, these facts suggest that sellers prefer cash financing and, accordingly, require that mortgage-financed buyers pay a premium. Of course, we cannot reach this conclusion on the basis of summary statistics alone. For example, cheap properties may attract all-cash buyers, which would be the opposite of the causal chain we seek to estimate.

Using our model as a guide, in Section III we estimate the premium that a buyer with mortgage financing must pay to purchase the same home relative to using cash financing: the mortgage-cash premium. This exercise is challenging because we do not observe a property's counterfactual sales price under the alternative method of financing. We take a transparent approach to this challenge in our baseline estimation. Using a large data set on U.S. home purchases, we estimate a joint 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 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 together 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 and hedonic pricing estimator. We obtain similar estimates when we exclude homes that are subsequently flipped, when we exclude transactions involving an institution, non-U.S. party, or underwater seller, or when we restrict the sample to newly built homes. This last filter implies that the estimated premium does not confound a premium for adverse selection, since almost all new construction is high-quality housing (e.g., Rosenthal (2014)). Consistent with the result's external validity, we estimate a similar premium of 12.2% based on a purchase-level data set from an entirely different provider (CoreLogic) than our baseline provider (Zillow).

Next, in Section IV we take a systematic approach to selection bias. Such bias can take four forms. The first three forms concern internal validity. They stem from correlation between the method of financing and price-relevant characteristics of the buyer, the seller, or the property's condition. The fourth form of bias concerns external validity. Using five different data sets and 10 different estimators designed to assess selection bias, we consistently estimate a mortgage-cash premium between 8.6% and 16.9%, straddling our baseline of around 11%.

To address bias from buyer characteristics, we use novel data on nonaccepted offers from a large real estate brokerage to construct a counterfactual. An offer-level research design allows us to include two important controls related to strategic interaction between buyers and sellers. First, we can control for differences in the offer price of mortgaged versus all-cash buyers. This control addresses concerns that mortgaged buyers have a higher private valuation, possibly due to optimistic expectations about the housing market, or that all-cash buyers target properties with low quality or a "motivated" seller. Second, we can control for the price discount of winning offers. This control addresses concerns that all-cash buyers possess bargaining power, in the form of negotiation skill, that enables them to win at a lower price. We estimate an 8.6% premium using this offer-level approach, which, given its minimal identification assumptions, provides a credible lower bound.

We address bias from seller characteristics using three exercises. First, we return to our baseline data set and control for a battery of variables that govern the seller's joint motivation to list at a low price and to prioritize all-cash offers (e.g., moving propensity) and the buyer's ability to identify such sellers (e.g., institutional status, financing in other transactions). Second, we instrument for the method of financing using a regulatory discontinuity around the price at which bank-originated mortgages must come with an appraisal, which makes them much less attractive from the seller's perspective. We construct this instrument using information about past sales prices. Thus, it is plausibly uncorrelated with temporary urgency on the part of home sellers that could affect both the current sales price and the method of financing. Third, we use listing-level data from a major state realtor association to control for the seller's list price. These exercises result in an estimated premium between 12% and 14%.

The exercises described in the previous two paragraphs already address many issues related to time-varying property characteristics, such as the propensity of all-cash buyers to purchase properties in poor condition. We further investigate the scope for such bias through the Bajari et al. (2012) semi-structural estimator, which reduces bias by introducing a Markov structure for property condition. We then apply propensity score matching (e.g., Abadie and Imbens (2006)), which allows for a nonlinear functional form. These exercises yield premiums of 14.9% to 16.9%.

Lastly, we assess whether the sampling restrictions required for internal validity jeopardize the results' external validity. We estimate a premium of 16.1% when including properties that do not experience a repeat sale, consistent with the property fixed effects reducing bias in our baseline sample. We estimate a premium of 10.3% when weighting transactions by their inverse probability of appearing in our baseline repeat-sales sample, according to observed characteristics (e.g., Solon, Haider, and Wooldridge (2015)). Along with the previous exercises, these findings support the internal and external validity of a premium around 11%.

Recalling the 3% premium implied by our representative-seller model, the 11% premium obtained from our exhaustive empirical analysis would seem to suggest a large 8% price puzzle. The model-implied premium is a function of parameters that capture selling conditions – such as seller net total wealth, seller down payment, likelihood of transaction failure, and number of offers – calibrated from the data. These parameters vary considerably in the data, indicating significant heterogeneity in selling conditions. Our analysis up to this point overlooks how this heterogeneity might impact the theoretical premium. Accordingly, in Section V we revisit our model by recalibrating the theoretical mortgage-cash premium across the full distribution of the underlying parameters, again disciplined by the data. Here, we recover the average premium across sellers, versus the premium for the average seller. The two differ insofar as the premium depends nonlinearly on home-selling conditions. We obtain a 7% premium based on this methodology, which explains half of the apparent puzzle. Similar results obtain when we employ increasingly rich versions of the model with, for example, nonfinancial contingencies.

The previous results highlight the importance of heterogeneity in selling conditions, which we examine further in Section VI. We semi-structurally estimate the empirical mortgage-cash premium across the distribution of the model's parameters and compare it pointwise with its theoretical analogue. This allows us to identify conditions under which a puzzle actually exists. In terms of slope, both the empirical and theoretical premiums increase monotonically in the probability of mortgage transaction failure. In terms of level, however, the empirical premium far exceeds the theoretical premium in states with moderate or high risk. Consistent with a role for risk-bearing capacity, both the empirical and theoretical premiums decrease nonlinearly in the seller's net total wealth. We find no evidence of a price puzzle in conditions favorable to sellers, such as in markets with loose loan approval standards and thus low risk of mortgage transaction failure or in thick markets with multiple offers.

Since we rely only on fairly canonical economics this far, in Section VII we ask whether belief distortions may help explain the puzzlingly large premium that obtains under high transaction risk. We do so by posing our motivating thought experiment to a survey of 3,400 U.S. homeowners, administered in three waves over two years. We recruit respondents through a well-established crowdsourcing platform (Prolific). This helps us obtain a sample that represents the overall distribution of homeowners well in terms of geography and income, although we oversample the college-educated. We compensate respondents at three times the minimum wage to incentivize effort and minimize response bias.

The survey replicates our main results, in terms of both the average mortgage-cash premium and its distribution across the model's parameters. Following the standard in the experimental literature, we infer the premium through a multiple price list presented in dollars and percent. The average premium equals 10.6%, 10.4%, and 10.7% across respondents in the three waves of the survey, respectively. This finding jointly supports the validity of the baseline estimates and the relevance of the survey. Also consistent with our main results, survey respondents who are randomized into experiments with greater mortgage transaction risk require a higher premium.

A key advantage of conducting a survey relative to studying observational data is the ability to perform randomizations that evaluate specific belief distortions. We find the strongest evidence in favor of ambiguity-aversion, which refers to the idea that investors prefer to know the distribution of outcomes than to face an unknown distribution, even when they hold the same prior as the known distribution. Specifically, survey respondents require a much higher premium when the probability of transaction failure is ambiguous, as in reality, relative to when it is clearly stated, as in our model. In fact, respondents in the more realistic experiment with ambiguity exhibit strikingly similar behavior to our main results: their required premium matches their model-implied premium at low levels of risk, but they require a much higher premium as the probability of failure increases. By contrast, respondents facing a clearly stated distribution require a premium that is indistinguishable from the model's prediction. Since selling a home in real life can be both a risky and ambiguous process, this finding supports ambiguity-aversion as a relevant amplification mechanism in explaining the large empirical premium.

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

Related Literature. We contribute to three literatures. First, our work relates broadly to research on how canonical models struggle to explain large risk premia across a variety of settings. For example, viewing the mortgage-cash premium as analogous to a "credit spread," we parallel papers on large credit spreads in bond markets (e.g., Huang and Huang (2012), Chen (2010), Chen, Collin-Dufresne, and Goldstein (2009), Almeida and Philippon (2007)). In our setting, the transaction failure rate plays the role of the default rate. Just as the credit spread puzzle is largest in periods of crisis, so too the mortgage-cash premium puzzle is largest when the probability of transaction failure ("default") is high, home sellers ("investors") have limited risk-bearing capacity, and the expected number of home purchase offers ("market liquidity") is low. Unlike these papers, however, we show that belief distortions also play a role. In particular, the evidence supports ambiguity-aversion as the most relevant distortion.[1]

Second, we contribute to the overall mortgage and housing literatures by quantifying an alternative channel through which credit markets affect house prices. This channel operates through frictions in the microstructure of real estate transactions and therefore complements channels that operate through the overall demand for housing, on which there is much research, as summarized by Glaeser and Sinai (2013) and Piazzesi and Schneider (2016). We find, however,

[1]There are various ways to model ambiguity-aversion, including robust control theory (e.g., Cagetti et al. (2002), Hansen et al. (2002), Anderson, Hansen, and Sargent (2003), Maenhout (2004), Hansen and Sargent (2011), Strzalecki (2011), Barnett, Buchak, and Yannelis (2021)), maxmin optimization (e.g., Gilboa and Schmeidler (1989)), smooth decision making (e.g., Klibanoff, Marinacci, and Mukerji (2005), Caskey (2008)), and others summarized by Machina and Siniscalchi (2014). We do not attempt to distinguish between these different setups. At the end of the paper, we propose a model in the tradition of robust control theory to explain the mortgage-cash premium and back out the implied degree of ambiguity-aversion.

that these frictions are "overpriced," which offers a novel example of mispricing in real estate.[2] 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.[3] Lastly, our focus on sellers' behavior after listing their home complements the Guren (2018) model of how sellers determine list prices.

Third, we contribute to an emerging literature on the role of all-cash buyers in real estate.[4] In this vein, we relate to contemporaneous and independent work by Han and Hong (2023), Seo, Holmes, and Lee (2021), and Buchak et al. (2020), who respectively document price discounts for all-cash buyers in Los Angeles, in Tallahassee, and for a specific type of all-cash buyer, iBuyers. Hansz and Hayunga (2016) instead find a negative all-cash discount over 2002 to 2004 in Pinehurst, North Carolina. We differ from these papers by forcefully estimating this price discount over broader samples of all-cash buyers, but our main distinction lies in quantifying the theoretical mechanisms that drive the discount and experimentally testing the role of belief distortions. After accounting for differences in sample, our estimates agrees with the estimates in these papers, as we show in Section II.E.4 of the Internet Appendix.[5]

The most closely related paper within this set is Han and Hong (2023). Relative to our model, these authors adopt a more stylized approach that does not feature (i) seller utility over total wealth, (ii) an endogenous problem for buyers, including buyers' choice of leverage, (iii) a full dynamic problem for sellers that includes an endogenous choice of whether to decline an offer and the possibility of receiving multiple offers of different methods of financing, and (iv) other realistic features, like nonfinancial contingencies.[6] Empirically, Han and Hong (2023) estimate

---

[2]Optimistic growth is another example (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)).

[3]See Barwick and Pathak (2015), Han and Hong (2011), Hsieh and Moretti (2003), Levitt and Syverson (2008), and Gilbukh and Goldsmith-Pinkham (2023) for examples.

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

[5]The Internet Appendix is available in the online version of this article on the *Journal of Finance* website.

[6]Features (i) and (ii) are critical for semi-structurally tracing out the distribution of the premium across the model's parameter space. Feature (iii) is critical for generating the seller's option value from waiting; in Han and Hong (2023), the seller's continuation value after a failed mortgaged transaction is equal to receiving a price of zero, which is counterfactual and inflates the calibrated mortgage-cash premium as we show in Section II.E.4 of

the premium using geographically limited data, while we do so using transaction-level data that span most of the United States, using offer-level data, and using an experimental survey. Lastly, our calibration differs from Han and Hong (2023), who parameterize the failure rate using the average difference in time-to-close between mortgaged and all-cash transactions (16%), while we use the mortgage application denial rate. Parameterizing the failure rate using the median denial rate lowers the mortgage-cash premium in the model of Han and Hong (2023) from 7.9% to 2.2%, as also described in the appendix.

I. Model of the Mortgage-Cash Premium

We anchor our empirical analysis on a tractable model of the mortgage-cash premium. The model consists of households with preferences over housing services and numeraire consumption. Time is discrete, and t indexes the month. Housing services are produced by homes (i). Home sellers (s) are endowed with a home that produces a stream of housing services worth v_i units of numeraire. Home buyers (b) purchase homes from sellers, paying all-cash or with mortgage financing. To reduce notation, we suppress the subscripts (i, t, s, b) when possible.

A. *Home Sale Process*

Home sales involve a high degree of execution risk. We describe the key institutional details here and how we model them. Broadly, there are three sources of risk for home sellers: whether and when they receive an offer; for accepted mortgaged offers, whether and when the lender approves the loan application; and, for any offer, whether the buyer terminates the transaction for reasons unrelated to mortgage financing.

A.1. Offer Arrival Risk

There is substantial empirical variation in the number of offers received by home sellers and hence in the time it takes to sell their home. For example, in 2018 60% of homes sold within

the Internet Appendix. Our calibration also differs from Han and Hong (2023) in that we calculate the average premium across the full parameter distribution, as distinct from the premium evaluated at the average parameter values, which accounts for nonlinearities.

one month, while in 2011 only 24% sold that quickly and 32% required over six months to sell (see Internet Appendix Figure IA.1). Accordingly, home sellers in our model wait for offers to arrive at the monthly Poisson arrival rate λ. This implies that a seller receives no offers in a given month with probability $\varphi_0 \equiv e^{-\lambda}$, exactly one offer with probability $\varphi_1 \equiv \lambda e^{-\lambda}$, and multiple offers with complement probability.[7] Sellers must also account for whether a given offer is mortgage-financed, which occurs with probability m, versus all-cash.

In our model, offers arrive during a specific window in month t. At the end of this window, sellers who have received an offer decide whether to accept it, to decline it, or, in the case of multiple offers, to invite a tie-breaking bid as we describe in Section I.C. This simplification matches the industry's convention that courteous sellers decide upon offers within three days. Sellers who do not receive offers or decline them must wait until the next arrival window in $t+1$.

A.2. Mortgage Financing Risk

Sellers who accept a mortgage-financed offer face additional risks surrounding the outcome and the timing of loan underwriting. Fundamentally, these risks stem from agency frictions between the buyer and an unmodeled lender. These frictions incentivize the lender to conduct a thorough screening process to verify (i) the home's value as collateral, as performed through an appraisal, (ii) the legality of the transaction ("title"), and (iii) the buyer's income, assets, and credit. This process lasts Δ months. The lender then issues its decision.

The lender denies the buyer's loan application with probability q, usually for one of the three reasons mentioned in the previous paragraph. In this case, the transaction fails and the seller must find another buyer.[8] Otherwise, the lender approves the applicant. As in reality, the lender then deposits the origination balance into an escrow account. From this account, which

[7]We abstract from the role of the list price in directing buyer search, consistent with the fact that only 30% of offers come in at the list price in our offer-level data set. This assumption can be motivated by the Han and Strange (2016) result that the list price has little signalling value outside of busts.

[8]This outcome obtains for mortgaged offers with a financing or appraisal contingency, which describe 80% of mortgaged offers in our offer-level data set. For noncontingent offers, the seller can still recover the buyer's earnest money, typically around 5% of the offer price. Corroborating the importance of loan denial in triggering transaction failure, 83% of home sellers with a failure cite appraisal, title, hazard insurance, buyer job loss, contingencies, or other issues associated with mortgage financing as a reason (National Association of Realtors (2018, 2020)).

already includes the down payment, the buyer pays for the home.

Lenders typically take one month to issue their decision, but delays occur a quarter of the time (National Association of Realtors (2018)). Conditional on delay, the average time to close equals around two months, as we describe in Section III.D.2 of the Internet Appendix. Therefore, we approximate Δ as a Bernoulli variable that equals two with probability q^d ("d" for "delay") and equals one otherwise.

Given our focus on the approval process, we model the mortgage contract itself rather simply. Briefly, the loan amount equals ℓP, where P denotes the offer price and ℓ denotes the requested loan-to-value (LTV) ratio. Let $D\ell P$ denote the present discounted cost of the loan from the borrower's perspective, after adjusting for embedded options (e.g., prepayment, default). The borrower therefore values the loan at less than par if $D < 1$. We refer to D as the "disutility of borrowing."

A.3. Other Transaction Risk

Both all-cash and mortgaged offers may have contingencies unrelated to mortgage financing. The two most common nonfinancial contingencies require a satisfactory home inspection and the buyer selling his current home (National Association of Realtors (2018)). According to our offer-level data set, described shortly, the home inspection contingency appears in 83% of both all-cash and mortgage-financed offers. In our model, therefore, both offer types fail at the same rate, q^c, due to a failed home inspection. None of our data sources reports the distribution of the home-sale contingency across buyer types. We obtain an upper bound on the theoretical mortgage-cash premium by assuming that this contingency accompanies a share h of mortgaged offers and never for an all-cash offer.

Incorporating nonfinancial contingencies improves the model's realism, but it significantly complicates our final expressions. Given this section's emphasis on analytic intuition, we present our main results for the case $q^c = h = 0$. We relax this assumption in Section V.

B. Seller Problem

Sellers have preferences over numeraire consumption and housing. They begin to consume the numeraire once they have sold their home, using their wealth at the time of sale. Until that time, they incur monthly housing disutility δ maintaining their home. Since δ has units of utils, let $\tilde{\delta}$ denote the equivalent maintenance cost expressed as a share of the home's fundamental value.

At the completion of her home sale, a seller's wealth equals $v\omega + P$, where P denotes the home's sale price and

$$\omega \equiv w - \ell_s - \xi, \tag{2}$$

denotes the seller's remaining net wealth, where w equals her financial wealth entering the home sale process (e.g., checking account), ℓ_s equals her current LTV ratio, and ξ equals the down payment that she has already made on a subsequent home purchase, normalized by the value of her current home. Sellers have constant relative risk-aversion (CRRA) indirect utility over wealth at sale described by

$$u\left(v\omega + P\right) = \frac{\left(v\omega + P\right)^{1-\gamma}}{1-\gamma}. \tag{3}$$

Combining equations (2 and (3) implies that the seller's short position, $\ell_s + \xi$, will amplify the effect of her baseline risk-aversion, γ. We focus primarily on risk-aversion $\gamma > 1$.

Sellers choose whether to accept or decline any offers they receive during the offer arrival window. They optimally decide by comparing the values of an all-cash offer at price P, denoted $V^{\mathcal{S}}(C, P)$, a mortgaged offer, denoted $V^{\mathcal{S}}(M, P)$, and no offer, denoted $V^{\mathcal{S}}(\emptyset)$. Mor explicitly,

$$V^{\mathcal{S}}(C, P) = u\left(v\omega + P\right), \tag{4}$$

$$V^{\mathcal{S}}(M, P) = -\delta\left(e^{-\alpha} + q^d\right) + e^{-\alpha-\rho}\left[(1-q)u(v\omega + P) + qV^{\mathcal{S}}(\emptyset)\right], \tag{5}$$

$$V^{\mathcal{S}}(\emptyset) = -\delta + \varphi_0 e^{-\rho} V^{\mathcal{S}}(\emptyset) + (1-\varphi_0)\mathbb{E}\left[\max_{n\in\mathcal{N}}\left\{V^{\mathcal{S}}(F_n, P_n)\right\} | \mathcal{N} \neq \{\emptyset\}\right], \tag{6}$$

where ρ is the seller's subjective rate of time preference and $\alpha \equiv -\log(1 - q^d(1 - e^{-\rho}))$ adjusts this rate to account for the possibility of a late approval decision.

Equations (4) to (6) have a straightforward interpretation. First, consider the value of accepting an all-cash offer, $V^{\mathcal{S}}(C,P)$. Our setup implies that an all-cash buyer will deliver his contract price P without delay or risk. Thus, $V^{\mathcal{S}}(C,P)$ simply equals indirect utility over total wealth: $u\left(v\omega+P\right)$. In reality, all-cash offers are neither risk-free nor immediate. However, per the previous subsection, this simplification approximates the actual institutional details rather well. We allow for risky all-cash offers in Section V.

Second, consider the value of accepting a mortgaged offer, $V^{\mathcal{S}}(M,P)$. The seller waits for the approval decision, meanwhile incurring the maintenance disutility associated with listing her home, δ. With probability q^d, the decision arrives with a delay and the seller must maintain the home for another month, and hence the coefficient $e^{-\alpha}+q^d$. Once the lender makes its decision, the seller receives the buyer's contract price with probability $1-q$, which yields indirect utility $u\left(v\omega+P\right)$ discounted by $e^{-\alpha-\rho}$ to account for delays. If the buyer is denied credit, she receives the discounted value of having no offer, $e^{-\alpha-\rho}V^{\mathcal{S}}(\emptyset)$. Importantly, $V^{\mathcal{S}}(C,P) > V^{\mathcal{S}}(M,P)$ for any given price P and realistic parameter values. Thus, sellers strictly prefer all-cash offers to mortgaged offers unless the latter comes with a price premium.

Third, a seller with no offer incurs the disutility δ as she waits for offers to arrive. With probability $\varphi_0 \equiv e^{-\lambda}$, she receives no offers and thus enters next month with discounted value $e^{-\rho}V^{\mathcal{S}}(\emptyset)$. Otherwise, she receives offers from the nondegenerate set $\mathcal{N}$, where $N \equiv |\mathcal{N}|$ is the total number of offers. She chooses the best offer $n \in \mathcal{N}$. She could also decline all offers. Finally, sellers take the expectation in equation (6) over all combinations of all-cash and mortgaged offers that may arrive. Indeed, the possibility that multiple buyers competitively bid up the home's price raises the seller's option value of waiting. Going forward, it will be convenient to define the certainty-equivalent price of having no offer, expressed as a discount, κ, relative to the home's fundamental value,

$$V^{\mathcal{S}}(\emptyset) = u\left(v\omega + ve^{-\kappa}\right). \tag{7}$$

The discount κ plays an important role in the seller's decision. Anticipating some of the analytic results, it depends endogenously on other parameters (e.g., λ), to be shown in Lemma 1.

C. Buyer Problem

Buyers commit to a quality segment $v \in \mathcal{V}$ and method of financing $F \in \{C, M\}$, interpreted as pre-approval.They then search for homes with the help of a real estate agent who knows the home's fundamental value, v, and the seller's resources, ω. This timing implies that buyers conduct a specific, or "narrow" Piazzesi, Schneider, and Stroebel (2020) home search. Upon finding a home, they choose an offer price P, and, for mortgaged buyers, a loan amount ℓP. We focus on a simplified case in which buyers search for only one period and always find a home, but this simplification does not materially impact the resulting mortgage-cash premium.[9] Neither buyers nor their agents know the number of competing offers that the seller receives during the arrival window, and thus whether they will win the home.

Buyers maximize an indirect utility function of the same form as equation (3), evaluated over total wealth at the end of their home search. Explicitly, they solve

$$V^{\mathcal{B}}(Y) = \max_{F,v,P,\ell} \{Win(F,P)(1-q(F))u\left((Y+v-P\left[1-\ell\left[1-D\right]\right])\right) + ... \tag{8}$$

$$... \qquad (Win(F,P)q(F) + (1-Win(F,P)))\, u\left(Y\right)\}$$

$$s.t. \tag{9}$$

$$\ell \leq \bar{\ell} \cdot \mathbb{1}\left[F = M\right], \tag{10}$$

$$Y \geq P\left[1-\ell\right], \tag{11}$$

where Y denotes the buyer's liquid assets and $q(F) \equiv q \cdot \mathbb{1}\left[F = M\right]$. The function $Win(F,P)$ describes the probability that a buyer offering P under financing F wins. The seller value functions (4) to (6) imply that $Win(C,P) > Win(M,P)$ for any P and $\frac{\partial Win}{\partial P} > 0$. For buyers who choose to pay in cash, constraints (10) and (11) simply require enough money to pay for the home. For mortgaged buyers, these equations impose two standard constraints. First, $\ell \leq \bar{\ell}$ requires the buyer to put down equity to, say, overcome agency frictions. To match the bunching in the LTV ratio in Figure 1, we consider the case $D \leq 1$, which implies buyers have sufficient credit demand as to choose $\ell = \bar{\ell}$. Second, $Y \geq P\left[1-\ell\right]$ ensures buyers can cover the down

[9] The case in which buyers find homes at rate $\sigma < \infty$ would affect the mortgage-cash premium through their willingness-to-pay, $\bar{p}_b$. Perturbing $\bar{p}_b$ has little impact on the calibrated premium, as Internet Appendix Table IA.V shows.

payment.

C.1. Optimal Choice of Financing and Empirical Implications

Our setup delivers two empirically useful predictions about the buyer's choice of financing, F. First, less-wealthy buyers use mortgage financing. Second, buyers use mortgage financing to purchase higher-quality homes. We formally state and prove these predictions in Internet Appendix Lemma IA.1, and Section II.A below provides empirical support for them. Thus, our model recommends two important regression controls: location-by-time fixed effects, to hold Y fixed, and property fixed effects, to hold v fixed. Numerous robustness exercises in Section IV address instances of within-neighborhood variation in Y and within-property variation in v.

C.2. Optimal Offer Price

The choice of offer price requires that we account for strategic interactions. We use $\tilde{P}$ to distinguish the offer price from the sales price, P. As a first step, let $\bar{p}_b(F)$ denote the maximum price that buyer b using method of financing F is willing to pay, normalized by v. In particular, Internet Appendix Lemma IA.2 shows that

$$\bar{p}_b(C) = \min\{y, 1\}, \tag{12}$$

$$\bar{p}_b(M) = \min\{\underbrace{\frac{y}{1-\bar{\ell}}}_{\equiv L}, \frac{1}{1-\bar{\ell}\,[1-D]}\}, \tag{13}$$

where $y \equiv \frac{Y}{v}$. We often refer to L as the "leverage capacity" of mortgaged buyers, since it defines the most they can afford given their liquid assets (yv) and financial constraints ($\bar{\ell}$). In particular, relaxing these constraints raises their willingness-to-pay, $\bar{p}_b(M)$. For all-cash buyers, we focus on the empirically relevant case in which $p_b(C) = 1$ because they have ample cash-on-hand (e.g., Table III). Raising a buyer's willingness-to-pay will not directly affect the mortgage-cash premium, as it does not appear in the difference between $V^{\mathcal{S}}(C, \cdot)$ and $V^{\mathcal{S}}(M, \cdot)$. It will, however, lower the mortgage-cash premium indirectly through $V^{\mathcal{S}}(\emptyset)$: raising $\bar{p}_b$ lowers sellers' cost of failure, κ.

While $\bar{p}_b$ provides a useful upper bound, the actual offer price solves a complicated game between the seller, the buyer, and an unknown number of all-cash or mortgaged competitors. We simplify this game by approximating the home purchase as a sealed-bid, first-price auction. This setup has three advantages: it captures strategic interaction among anonymous bidders, it accords with how homebuyers pay their offer price in reality, and it allows us to borrow insights from auction theory.

Under this setup, which we describe in Section III.B of the Internet Appendix, buyers optimally bid

$$\tilde{P}^C = \begin{cases} \min\left\{\tilde{v}(C)e^{-\kappa}, v\bar{p}_b(C)\right\}, & \text{if} \quad N = 1 \\ v\bar{p}_b(C), & \text{if} \quad N > 1 \end{cases}, \tag{14}$$

$$\tilde{P}^M = \begin{cases} \min\left\{\tilde{v}(M)e^{-\kappa}, v\bar{p}_b(M)\right\}, & \text{if} \quad N = 1 \\ v\bar{p}_b(M), & \text{if} \quad N > 1 \end{cases}, \tag{15}$$

where $\tilde{v}$ denotes an offer price with certainty equivalence v, from the seller's perspective.[10] The case $N = 1$ reflects the basic concept that buyers in a first-price auction optimally bid less than their true willingness-to-pay, $\bar{p}_b$. However, they raise their offer up to $\bar{p}_b$ when necessary to win the home, shown in the case $N > 1$.

The auction setup is a useful modelling device, but other setups would deliver the same basic insights. For example, one can interpret prices in our setting as determined through Nash bargaining, whereby the buyer has full bargaining power when $N = 1$ while the seller has full bargaining power when $N > 1$. In particular, equations (14) and (15) imply the sales price

$$P = \underbrace{v\bar{p}_b}_{\substack{\text{Buyer} \\ \text{Willingness-to-Pay}}} - \underbrace{(1-\eta)}_{\substack{\text{Buyer} \\ \text{Bargaining Power}}} \times \underbrace{\left(v\bar{p}_b - \tilde{v}e^{-\kappa}\right)}_{\substack{\text{Total} \\ \text{Surplus}}}, \tag{16}$$

where $\eta \equiv \mathbb{1}\left[N > 1\right]$ indicates the presence of multiple offers. Equation (16) implies that the buyer's effective bargaining power in our model depends on the presence of a competitor, and

[10]Equations (14) and (15) constitute a Nash equilibrium (Internet Appendix Lemma IA.3). The certainty equivalence of an all-cash offer at v is simply the offer price itself: $\tilde{v}(C) = v$. The value of a mortgaged offer with certainty equivalence v requires scaling according to the mortgage-cash premium: $\tilde{v}(M) = ve^{\mu}$. Thus, we use "certainty equivalence" in a slightly loose sense that includes compensation for frictions unrelated to risk (e.g., δ).

thus does not differ by the method of financing. More generally, however, all-cash buyers may have greater bargaining power insofar as they possess negotiation skill. Our offer-level research design in Section IV.A will account for this possibility.

D. *The Mortgage-Cash Premium: Definition and Analytics*

Our goal is to understand the price premium that makes a seller indifferent between a mortgaged offer and an all-cash offer, the mortgage-cash premium.

DEFINITION 1 (Mortgage-Cash Premium): *Consider the sale of property i by home seller s to home buyer b. Let $\succ^{\mathcal{S}}$ denote the seller's preference relation. Let $P^C_{i,s,b}$ and $P^M_{i,s,b}$ denote the price when the buyer pays all-cash or with mortgage financing, respectively. Define the mortgage-cash premium, $\mu_{i,s,b}$ such that*

$$C_{i,s,b} \succ^{\mathcal{S}} M_{i,s,b}, \quad \text{if } P^M_{i,s,b} < e^{\mu_{i,s,b}} P^C_{i,s,b}, \tag{17}$$

$$M_{i,s,b} \succ^{\mathcal{S}} C_{i,s,b}, \quad \text{if } P^M_{i,s,b} > e^{\mu_{i,s,b}} P^C_{i,s,b}. \tag{18}$$

The mortgage-cash premium is distinct from the average difference in log prices between mortgaged and all-cash purchases. This distinction requires a careful empirical analysis that holds fixed the characteristics of the home, the seller, or the buyer in question, as best we can. By Definition 1, μ solves $V^{\mathcal{S}}(C, vp^C) = V^{\mathcal{S}}(M, vp^C e^{\mu})$, where $p^C \equiv \frac{P^C}{v}$. The next proposition summarizes the resulting closed-form expression for the mortgage-cash premium. For convenience, we use the model's homogeneity in v to normalize $v = 1$, unless otherwise stated.

PROPOSITION 1 (Deriving the Mortgage-Cash Premium): *Holding the property (i), seller (s), and buyer (b) fixed, the mortgage-cash premium equals*

$$\mu\left(\Theta\right) = \log\left(\left[(1-\gamma)\left(\frac{(e^{\alpha+\rho}-1)u_C + \Gamma\delta + u_C - u_\kappa}{1-q} + u_\kappa\right)\right]^{\frac{1}{1-\gamma}} - \omega\right) - \log(p^C), \tag{19}$$

where $u_C \equiv \frac{\left(\omega+p^C\right)^{1-\gamma}}{1-\gamma}$, $u_\kappa \equiv \frac{\left(\omega+e^{-\kappa(q,\omega,\lambda,\cdot)}\right)^{1-\gamma}}{1-\gamma}$, $\Gamma \equiv (1+q^d e^{\alpha})e^{\rho}$, *and* $\alpha \equiv -\log(1-q^d(1-e^{-\rho}))$. *The premium is a function of parameters* $\Theta = \{\Theta_1; \Theta_2\}$, *where* $\Theta_1 = \{q, \lambda, w, \xi, l_s, y, \bar{l}, q^d, m, \tilde{\delta}\}$ *contains exogenous, measurable quantities and* $\Theta_2 = \{\gamma, \rho, D\}$ *contains preference parameters.*

The continuation value, κ, is a function of the parameters in Θ, as expressed in Lemma 1.

Three frictions generate the mortgage-cash premium: transaction delay, risk of failure, and the cost failure.[11] First, sellers dislike the delay generated by the loan underwriting process because it requires an additional month of maintaining the home: $\Gamma\delta$. In addition, time discounting implies that they simply like how all-cash transactions close more quickly: $(e^{\alpha+\rho}-1)u_C$.

Turning to the second friction, sellers dislike how mortgaged offers come with a higher probability of failure, q. Once we turn toward quantitative implications, this friction will explain a larger share of the premium than will transaction delay. In particular, equation (19) shows how the effect of q on the premium interacts with the utility loss from failure: $u_C - u_\kappa$. The seller's net wealth, ω, plays an important role in amplifying this loss. Intuitively, poor sellers or those with negative equity have a high marginal utility of wealth. Consequently, they exhibit high effective risk-aversion and therefore require greater compensation for the risk of costly failure.

The third friction arises because the seller solves a dynamic problem and thus incurs a cost of failure, κ. Recalling its definition in equation (7), κ parameterizes the continuation value of a seller with no offer, $V^{\mathcal{S}}(\emptyset)$. Characterizing κ therefore requires that we first solve the corresponding Bellman equation (6). Lemma 1 summarizes the solution.

LEMMA 1 (Deriving the Seller's Cost of Failure): *The continuation value, κ, equals*

$$\kappa(\Theta) = -\log\left(\left[(1-\gamma)\left(\frac{\mathbb{E}\left[u|u>V^{\mathcal{S}}(\emptyset)\right]Pr\left[u>V^{\mathcal{S}}(\emptyset)\right]-\delta\Xi}{e^{\rho}-Pr[u=V^{\mathcal{S}}(\emptyset)]}\right)\right]^{\frac{1}{1-\gamma}}-\omega\right). \tag{20}$$

The magnitude of $\bar{p}_b(M)$ determines the expressions for the terms in equation (20). There are two cases. The first case obtains when $\bar{p}_b(M) \geq \overline{p_M}$. In this case,

$$\mathbb{E}\left[u|u>V^{\mathcal{S}}(\emptyset)\right]Pr\left[u>V^{\mathcal{S}}(\emptyset)\right] = \Phi_C u(\omega+\bar{p}_b(C)) + \Phi_M[1-q]e^{-\alpha-\rho}u(\omega+\bar{p}_b(M)), \tag{21}$$

$$Pr\left[u=V^{\mathcal{S}}(\emptyset)\right] = e^{-\alpha-\rho}\Phi_M q + e^{-\lambda}(1+\lambda), \tag{22}$$

$$\Xi = e^{\rho} + \Phi_M(1+q^d e^{-\rho}). \tag{23}$$

[11]The $[\cdot]^{1-\gamma}$ functional contains the effects of these frictions. The other terms are mechanical: holding the indirect utility of a successful mortgaged transaction fixed, a lower premium obtains when either ω or p^C are high simply because such sellers are already well-off and thus require a lower premium to meet that utility.

The second case obtains when $\overline{p_M} > \bar{p}_b(M)$*. In this case,*

$$\mathbb{E}\left[u|u > V^{\mathcal{S}}(\emptyset)\right] Pr\left[u > V^{\mathcal{S}}(\emptyset)\right] = \Phi_C u(\omega + \bar{p}_b(C)), \tag{24}$$

$$Pr\left[u = V^{\mathcal{S}}(\emptyset)\right] = e^{-\lambda(1-m)}[1 - e^{-\lambda m}(1 + \lambda m)] + e^{-\lambda}(1 + \lambda), \tag{25}$$

$$\Xi = e^{\rho}. \tag{26}$$

The term Φ_C *equals the probability that the seller receives multiple offers and that she accepts an all-cash offer, and* Φ_M *equals the analogous probability of accepting a mortgaged offer. The condition* $\bar{p}_b(M) < \overline{p_M}$ *implies that the seller optimally declines all mortgaged offers. Internet Appendix Lemma IA.4 derives the expressions for* $\{\Phi_C, \Phi_M, \overline{p_M}\}$*.*

Equation (20) inherits a similar form as the expression for μ. The numerator contains the expected per-period payoffs from having no offer. The most basic payoff is the cost of maintaining the home while waiting for offers: $\Xi\delta$. In addition, sellers have the option value of receiving multiple competitive and ultimately successful offers: $\mathbb{E}\left[u|u > V^{\mathcal{S}}(\emptyset)\right] \Pr\left[u > V^{\mathcal{S}}(\emptyset)\right]$. However, that value may be small when the probability of receiving multiple offers is low, corresponding to low values of the compound probabilities Φ_C and Φ_M that, in turn, depend on the fundamental offer arrival rate, λ, and the share of mortgaged buyers in the economy, m.

The denominator of equation (20) amplifies the payoffs from having no offer by the probability of continuing to experience these payoffs in the future, $\Pr\left[u = V^{\mathcal{S}}(\emptyset)\right]$. This occurs when the seller receives no offers, receives one offer from a buyer who holds her to her reservation value per the optimal strategy in equations (14) and (15), accepts a competitive mortgaged offer that subsequently fails, or, if $\bar{p}_b(M)$ is low enough to trigger the lemma's second case, receives multiple mortgaged offers and optimally declines all of them.

At a high level, Lemma 1 distinguishes between parameters that (i) directly affect μ by altering the static tradeoff between mortgaged and all-cash offers, such as q and ω, which directly appear in Proposition 1, and (ii) indirectly affect μ through the seller's continuation value, κ. For example, parameters that raise buyers' housing demand, such as $\bar{p}_b$, can lower the mortgage-cash premium by reducing the cost of failure. By the same logic, liquid markets with a high λ will also feature a lower premium. Section V verifies these statements numerically.

D.1. Implications for Identification

From an empirical standpoint, Proposition 1 also sheds light on how we can identify the mortgage-cash premium using observational data. In particular, we can recover the mortgage-cash premium from a regression of log price on the method of financing, after carefully controlling for the home's quality, the seller's cost of failure, and the buyer's surplus. Corollary 1 presents this result.

COROLLARY 1 (Identifying the Mortgage-Cash Premium): *Consider the sale of property i by home seller s to home buyer b. The log of the sales price, $Price_{i,b,s}$, has the form*

$$\log\left(Price_{i,b,s}\right) = \bar{\mu} \times Mortgaged_b \quad + \underbrace{\log(v_i)}_{\substack{\text{Property}\\ \text{Quality}}} + \underbrace{(1-\eta)\left[-\kappa_s\right]}_{\substack{\text{Effect of Seller}\\ \text{Outside Value}}} + \quad \underbrace{\eta\left[\log(\bar{p}_b) - \log(\tilde{v}_i)\right]}_{\substack{\text{Effect of Buyer}\\ \text{Surplus}}}, \tag{27}$$

where $Mortgaged_b \equiv \mathbb{1}\left[F_b = M_b\right]$ indicates whether b uses mortgage financing, $\eta \equiv \mathbb{1}\left[N > 1\right]$ indicates the presence of multiple offers, $\bar{\mu}$ is the average mortgage-cash premium, and $\tilde{v}_i$ is an offer price with certainty equivalence $v_i = 1$, from the seller's perspective.

E. Calibrating the Theoretical Premium of a Representative Seller

We calibrate the theoretical mortgage-cash premium derived in Proposition 1, $\mu(\hat{\Theta})$, using values of $\hat{\Theta}$ in Table I. The elements of $\hat{\Theta}$ equal their average across the empirical distribution.

We find a model-implied premium of 3.3%, as shown in Table I. Thus, this is the magnitude of the premium that a representative home seller requires for bearing mortgage transaction frictions. As discussed above, these frictions can be decomposed into three components: transaction delay, risk of failure, and the cost of failure. To eliminate one or more of these components, one can set a subset of the parameters in Θ to zero. For example, we obtain a premium of only 0.9% under a simplified setup in which: there is no loan approval delay, the seller has no existing debt or down payment, transaction failure results in an almost-certain all-cash offer next month, and there is little cost of maintaining the home on the market.[12] Hence, 3.3% is the largest premium

[12]Under this low-friction parameterization, $\xi = \ell_s = q^d = m = 0$ while $\lambda = 20$, which implies that the seller receives multiple offers with probability greater than 99.9%, $\tilde{\delta} = 0.01$, which is one fifth of its empirical value, and the other parameters are as reported in Table I. This parameterization also gives $\kappa = 0.9\%$.

that our model can deliver in the context of a representative home seller.

Of course, considering a representative home seller may be inappropriate given the nonlinearities in equations (19) and (20). In particular, a representative-seller approach ignores how extreme values of quantities in Θ can potentially drive the average mortgage-cash premium to very high levels. Whether there is significant heterogeneity in the exogenous variables in Θ and the extent to which that affects the empirical and theoretical premiums are important questions we tackle in subsequent sections.

II. Data and Motivating Facts

Our empirical analysis relies on six observational data sets and one experimental data set. Table II summarizes the two most important observational data sets, which we now describe.

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

The most important variables in the ZTRAX data set are the sales price and 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 IV and to study heterogeneity in Section VI. Lastly, we observe various hedonic characteristics of the property, which, along with time-varying geographic and property fixed effects, help us achieve an R^2 above 90% in our baseline regression.

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

The offer-level data set includes information on the offer's price and method of financing, the geographic location of the property, the date on which the offer was made, the number of competing offers, and other variables described in Section I of the Internet Appendix. Out of concern for client privacy, we do not directly view the microdata and instead analyze the offer-level data set by submitting a program with our desired calculations to Redfin.

Lastly, we use several additional data sets to assess selection bias and to calibrate our model. These include a transaction-level data set from CoreLogic that is analogous to ZTRAX, mortgage application data from the Home Mortgage Disclosure Act (HMDA), an aggregate time series of characteristics of home purchase transactions from the National Association of Realtors' Realtor Confidence Index (RCI), a zip code panel of income from the IRS Statement of Income (SOI), a highly detailed cross-section of household balance sheets from the Survey of Consumer Finances (SCF), Zillow's Observed Rent Index (ZORI), and a survey of California home sellers in 2019 by the California Association of Realtors (CAR).

A. Motivating Facts

We motivate our empirical analysis with three facts about all-cash purchases: sizeable market share; a price discount relative to mortgaged purchases; and correlation with characteristics related to transaction risk and other model parameters.

First, Figure 1, Panel A, shows how such purchases account for 35% of all home purchases over 1980 to 2017, based on the ZTRAX data set. More specifically, we plot the distribution of 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.

Second, Figure 1, Panel B, shows that the average all-cash purchase over 1980 to 2017 has a lower real sales price relative to the average mortgage-financed purchase. The average difference equals 35 log points. When 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, we obtain a real price difference between 33 and 46 log points.

Third, Table III documents significant correlation between the probability that a purchase is mortgaged and characteristics related to key parameters of the model. Columns (1) and (2) predict the method of financing using time-varying market characteristics. Consistent with the model, the probability of an all-cash purchase is higher in markets with a high mortgage application denial rate (q), lower transaction volume, which plausibly maps to higher κ, and a lower average price (v). These time-varying market characteristics explain a sizeable 30% of the variation in the method of financing without any geographic fixed effects, supporting the model's relevance.

Column (3) includes zip code-by-month fixed effects, which allows us to focus on variation within time-varying market conditions. The probability of an all-cash purchase is higher in sales with institutional and foreign buyers, consistent with these buyers having higher liquid assets (y), another sale within 12 months ("flips"), consistent with the industry's common view that flipping is most profitable on low-quality homes (v), foreign sellers, who have a greater cost of transaction failure (κ), and individual sellers, who plausibly have less nonfinancial resources than institutions (w). Finally, these relationships obtain after including property fixed effects in column (4). This specification most closely matches that of our main empirical analysis, and therefore provides a clear picture of the sources of identifying variation.

III. Empirical Mortgage-Cash Premium

We estimate the average mortgage cash premium using a repeat sales and hedonic pricing approach. This approach has the advantage of transparency and a long tradition in the literature. As its name implies, this approach requires controlling for property fixed effects ("repeat sales") and an exhaustive set of additional controls ("hedonic"). These ancillary parameters serve to absorb as much variation as possible, such that purchases effectively differ only in their price and method of financing. The regression equation is

$$\log\left(Price_{i,t}\right) = \mu Mortgaged_{i,t} + \psi X_{i,t} + \zeta_{z(i),t} + \alpha_i + \epsilon_{i,t}, \tag{28}$$

where $Mortgaged_{i,t}$ indicates whether 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 various hedonic characteristics.

Equation (28) identifies the mortgage-cash premium, μ, under the following assumption,

$$\mathbb{E}[Mortgaged_{i,t} \times \epsilon_{i,t} | \mathbf{X}_{i,t}] = 0, \tag{29}$$

where we use abbreviated notation $\mathbf{X}_{i,t} = \{\alpha_i, \zeta_{z(i),t}, X_{i,t}\}$. Assumption (29) states that, conditional on the fixed effects and controls ($\mathbf{X}_{i,t}$), the method of financing ($Mortgaged_{i,t}$) does not correlate with unobserved, price-relevant characteristics of the transaction ($\epsilon_{i,t}$). In particular, applying assumption (29) to the theoretical regression equation from Corollary 1 implies

$$\log\left(Price_{i,t}\right) = \mu Mortgaged_{i,t} + \underbrace{\log(v_{i,t}) - \kappa_{s(i,t)}\left(1 - \eta_{i,t}\right) + \eta_{i,t}\left[\log(\bar{p}_{b(i,t)}) - \log(\tilde{v}_{i,t})\right]}_{\psi X_{i,t} + \zeta_{z(i),t} + \alpha_i} + \epsilon_{i,t}.$$

For example, it seems reasonable that the home's fundamental value ($v_{i,t}$) is described by the property fixed effects (α_i) and time-varying price of hedonic characteristics ($\psi X_{i,t}$). Moreover, the zip code-by-month fixed effect ($\zeta_{z(i),t}$) plausibly describes the effect of the seller's outside value ($\kappa_{s(i,t)}\left(1 - \eta_{i,t}\right)$) which, per Lemma 1, depends on the liquidity of the local market (λ). Similarly, the effect of the buyer's surplus ($\eta_{i,t}\left[\log(\bar{p}_{b(i,t)}) - \tilde{v}_{i,t}\right]$) also depends on market conditions, such as credit constraints ($\bar{\ell}$). Of course, there may exist variation within $\mathbf{X}_{i,t}$ that correlates with both the sales price and the method of financing. Therefore, we devote Section IV to a rigorous assessment of assumption (29).

A. Results

We estimate equation (28) using the ZTRAX data set and report the results in Table IV. The estimated mortgage-cash premium over 1980 to 2017 equals 11.7%, as shown in column (1).

The regression features an R^2 of 91%, due largely to the 187,000 property fixed effects. This high R^2 means that there is little remaining variation in unobserved characteristics to constitute a violation of our identification assumption (29). Indeed, consistent with the property fixed effects reducing bias, we estimate a larger premium of 16.1% when including properties without a repeat sale, as shown in column (0).

The remaining columns of Table IV report the estimated premium within various subsamples. In columns (2) through (5), we obtain a consistent premium after partitioning the sample by time period. The higher premium over 2005 to 2010 corresponds to elevated transaction frictions during that period. Indeed, Panel A of Figure 2 reveals a strong correlation between the estimated premium and the mortgage application denial rate from HMDA, our key measure of q. Panel B shows a positive correlation between the estimated premium and the all-cash share of the market prior to 2010, suggesting movements in seller demand along the "supply curve" for all-cash buyers. The elevated all-cash share after the Great Recession may reflect sustained balance sheet impairment of households who would otherwise borrow.

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

In columns (8) and (9), we consider transactions between noninstitutional and nonforeign parties, respectively. The respective premiums equal 11.2% and 10.8%. These findings imply that we do not confound how institutional investors tend to transact in illiquid submarkets (e.g., Mills, Molloy, and Zarutskie (2019)) or how out-of-town buyers tend to have informational disadvantages (e.g., Chinco and Mayer (2016)). Lastly, we estimate a premium of 9.9% after restricting the sample to transactions in which the seller likely has positive equity, as shown in column (10). As in columns (6) and (7), such transactions have less asymmetric information about the property's condition.

Taken together, the stable estimates between 8% and 12% in columns (6) to (10) support the internal validity of an 11% premium.

IV. Selection Bias

We take the question of selection bias seriously. Accordingly, we perform over 10 exercises using five different data sets to assess internal and external validity. Table V summarizes the resulting estimates of the mortgage-cash premium. The estimates range from 8.6% to 16.9%, straddling our baseline estimate (11.7%) and hence supporting its validity.

We organize this section using Corollary 1, which states that the pricing error $\epsilon_{i,t}$ equals a weighted average of three innovations: variation in the property's condition that market participants observe but econometricians do not, the buyer's offer price relative to this condition, and, similarly, the seller's reservation value. Importantly, these innovations are defined relative to the value predicted by our rich set of conditioning variables, $\mathbf{X}_{i,t}$,

$$\epsilon_{i,t} = \underbrace{\Delta_{\mathbf{X}}\left[\log(v_{i,t})\right]}_{\text{Shocks to Property Quality}} + \underbrace{\Delta_{\mathbf{X}}\left[(1-\eta_{i,t})\left[-\kappa_{s(i,t)}\right]\right]}_{\text{Shocks to Seller Outside Value}} + \underbrace{\Delta_{\mathbf{X}}\left[\eta_{i,t}\left[\log(\bar{p}_{b(i,t)}) - \log(\tilde{v}_{i,t})\right]\right]}_{\text{Shocks to Buyer Surplus}}, \tag{30}$$

where $\Delta_{\mathbf{X}}\left[A_{i,t}\right] \equiv A_{i,t} - \mathbb{E}[A_{i,t}|\mathbf{X}_{i,t}]$. Violations of internal validity occur when the average of these innovations covaries with the method of financing. Columns (6) to (10) of Table IV already account for several specific violations of internal validity. Sections IV.A to IV.D further assess the scope for such bias. Section IV.E evaluates external validity.

A. Information on Nonaccepted Offers

An offer-level research design allows us to include two important controls related to strategic interaction between buyers and sellers. First, we can control for differences in the offer price of mortgaged versus all-cash buyers. This control addresses concerns that mortgaged buyers have a higher private surplus ($\log(\bar{p}_b) - \log(\tilde{v}_{i,t})$), possibly due to optimistic expectations about the housing market, and that all-cash buyers target properties with low quality or a "motivated" seller (low $v_{i,t}$, high κ_s). Second, we can control for the price discount of winning offers. This control addresses concerns that all-cash buyers possess bargaining power, in the form of negotiation skill, that enables them to win at a lower price. While some of these concerns do not fit neatly within our model, they could nevertheless be empirically relevant. Therefore, an offer-level research design provides an important test of the magnitude of the premium.

The model provides useful structure that helps guide our offer-level research design. Let $\tilde{P}^F_{i,t}$ now denote the price associated with a nonaccepted offer to purchase property i in month t under method of financing F. Let $\pi^M_{i,t}$ and $\pi^C_{i,t}$ denote the log price premium paid by an accepted offer under mortgaged and all-cash financing, respectively. That is, $\log(P^F_{i,t}) = \pi^F_{i,t} + \log(\tilde{P}^F_{i,t})$. Using buyers' optimal bidding equations (14) and (15) as a guide,

$$\log(P^F_{i,t}) = \pi^F_{i,t} + \log(\tilde{P}^F_{i,t}) = \pi^F_{i,t} + \log(v_{i,t}) + \log(\bar{p}_b(F)). \tag{31}$$

Applying Definition 1 requires holding $v_{i,t}$ and $\bar{p}_b(F)$ fixed so that

$$\mu_{i,t} = \log(P^M_{i,t}) - \log(P^C_{i,t}) = \pi^M_{i,t} - \pi^C_{i,t}. \tag{32}$$

Equation (32) implies that the mortgage-cash premium equals the difference-in-difference between an accepted versus nonaccepted offer price between mortgaged versus all-cash financing.

Applying these ideas to our main empirical analysis, we can rewrite equation (28) as

$$\log(Price_{i,t}) = \hat{\mu} Mortgaged_{i,t} + \psi X_{i,t} + \zeta_{z(i),t} + \alpha_i + \epsilon_{i,t}, \tag{33}$$

$$\hat{\mu} = \mu + \underbrace{\mathbb{E}[(v^M_{i,t} - v^C_{i,t}) + (\log(\bar{p}_b(M)) - \log(\bar{p}_b(C)))|\mathbf{X}_{i,t}]}_{\mathbb{E}[\log(\tilde{P}^M_{i,t}) - \log(\tilde{P}^C_{i,t})|\mathbf{X}_{i,t}]}, \tag{34}$$

where $v^F_{i,t} \equiv \mathbb{E}[\log(v_{i,t})|F]$. To interpret, our baseline regression equation suffers bias if the controls and fixed effects do not absorb differences in property quality v or private valuation $\bar{p}$ between mortgaged and all-cash buyers. However, with an offer-level approach, we can correct for this bias by controlling separately for the difference in offer prices: $\log(\tilde{P}^M_{i,t}) - \log(\tilde{P}^C_{i,t})$.

Motivated by this concern, we use the offer-level data set to estimate

$$\log(Price_{i,j,t}) = \mu(Mortgaged_{i,j,t} \times Winning_{i,j,t}) + ... \qquad (35)$$
$$... + \psi_0 Winning_{i,j,t} + \psi_1 Mortgaged_{i,j,t} + \zeta_{z(i)} + \tau_t + \upsilon_{i,j,t},$$

where i, j, and t index property, offer, and month, $Mortgaged_{i,j,t}$ indicates whether j is a mortgage-financed offer, $Winning_{i,j,t}$ indicates whether j is the accepted offer, $Price_{i,j,t}$ is the price offered by j, and $\zeta_{z(i)}$ and τ_t are zip code and month fixed effects.[13] Using equations (33) and (34),

$$\psi_1 = \mathbb{E}[\log(\tilde{P}^M_{i,t}) - \log(\tilde{P}^C_{i,t})|\mathbf{X}_{i,t}]. \qquad (36)$$

Thus, conditioning on $Mortgaged_{i,j,t}$ allows us to identify the mortgage-cash premium even if all-cash buyers make low offer prices because they target "low-ϵ" listings that, in our model, correspond to listings with a low v or high κ. The coefficient ψ_1 absorbs this effect.

Importantly, we condition on $Winning_{i,j,t}$ to allow for the possibility that all-cash buyers negotiate better, and thus can win at a lower price even without mortgage transaction frictions. While our model does not separately parameterize how certain skilled negotiators may win at a lower (certainty-equivalent) price, equation (35) nonparametrically allows for this possibility. In particular, the parameter ψ_0 absorbs this effect.

Table VI reports the results. We estimate a premium of 8.1% in column (1). In column (2), we estimate a premium of 8.6% after controlling nonparametrically for competitiveness, as measured by the number of competing offers. These estimates are close to those in Table IV, supporting the latter's internal validity.

It is important to reiterate how little we have assumed in this offer-level research design. For example, we can identify the mortgage-cash premium even if all-cash buyers select different types of properties or have different private valuations (ψ_1), or if they differentially win deals due to bargaining power (ψ_0). Thus, the smallest estimate of around 8% from Table VI provides

[13]We weight observations by the total number of offers on the listing to address the fact that we do not observe all offers. The unweighted estimates are very similar to the weighted estimates. The limited sample size forces us to restrict the set of fixed effects and controls relative to equation (28).

a credible lower bound on the mortgage-cash premium.

B. *Controlling for Buyer and Seller Characteristics*

Since many sources of bias may be observed or have a close empirical proxy, we also follow the more basic approach of simply controlling for buyer and seller characteristics that may jointly affect equilibrium price and financing. Appendix II.A describes the methodology. We estimate a premium between 10.0% and 12.9% as shown in Internet Appendix Table IA.II.

C. *Instrumental Variables*

An instrumental variables approach provides a clear and transparent source of identifying variation. We propose two instruments; their respective details are in Sections II.B and II.C of the Internet Appendix.

The first instrument exploits a regulatory discontinuity that determines whether a mortgaged buyer would require an appraisal. Specifically, federal law requires an appraisal on all bank-originated loans with a sales price above $250,000. Federal regulators introduced this requirement in the wake of the savings and loan crises of the 1980s as part of the Financial Institutions Reform, Recovery, and Enforcement Act of 1989. This federally mandated appraisal introduces significant risk for home sellers, since it occurs after the seller has already accepted the buyer's offer. Moreover, lenders often cannot originate the loan if the LTV obtained from the appraisal exceeds certain values. In the context of our model, transactions with a price just above $250,000 have a higher value of q relative to transactions with a price just below $250,000.

We implement this strategy similarly to Loutskina and Strahan (2015), who also construct an instrument using bunching in house prices around a regulatory threshold. Briefly, we use the previous sales price to forecast whether the price in the current sale exceeds the $250,000 threshold. This instrument correctly predicts a lower likelihood of mortgage financing, supporting its first stage. Moreover, to support the exclusion restriction, we limit the sample to a narrow bandwidth around the threshold. So, in effect, the instrument identifies the premium using variation in q that is otherwise orthogonal to the sales price

The second instrument equals the share of homes sold by the seller over our sample period that are to cash buyers, excluding the sale in question (a "leave-one-out mean"). It reflects the seller's persistent preference for cash financing (e.g., belief about transaction risk), since it is uncorrelated with idiosyncratic features of the transaction in question (e.g., urgency).

Using the first instrument, the regulatory discontinuity in appraisals, we estimate a premium of 13.6% (see Internet Appendix Table IA.III). Using the second instrument, the seller's propensity to accept all-cash offers, we estimate 13.9% (see Internet Appendix Table IA.IV). Comparing the two instruments, the regulatory discontinuity provides a more transparent source of variation, but we can construct the second instrument for a much larger share of the sample. Importantly, the instruments draw on two fundamentally different sources of variation: exogenous variation in q that comes from the buyer's side, and the seller's intrinsic preference for all-cash offers. The very similar estimates obtained from the two instruments provides strong cross-validation of our IV results.

D. Property Condition

Section IV.A already addresses many concerns related to bias from the property's condition, which buyers and sellers observe but we as econometricians do not. Section IV.D of the Internet Appendix pursues three additional exercises that address these concerns. In particular, we use the semi-structural hedonic estimator of Bajari et al. (2012), we use a nonparametric matching estimator, and we control for list price and time on market using the CAR data set. The estimated premiums range from 14% to 16%.

E. External Validity

Our emphasis on internal validity often requires us to restrict the variation used to identify the mortgage-cash premium. Section IV.E of the Internet Appendix assesses the external validity of our main results. We estimate a premium of 16.1% after including properties without a repeat sale, 12.2% when applying our repeat sales and hedonic pricing approach to a nationally representative, transaction-level data set from CoreLogic, and 10.3% when weighting purchases by their inverse probability of appearing in the baseline sample. We further show that our results agree with contemporaneous papers studying the price of cash-financed purchases.

V. Calibrating the Theoretical Premium: Heterogeneity

Recalling the 3% theoretical premium obtained from the stylized calibration in Table I, the robust empirical premium of around 11% would seem puzzlingly large. However, it is premature to conclude that a puzzle exists without first pursuing a more serious calibration. In particular, our stylized calibration simply calculated the premium at the average values of the parameter space: $\mu(\mathbb{E}[\Theta])$. Since the empirical premium obtains from variation in the data, a suitable calibration should instead calculate the average of the premium across the parameter space: $\mathbb{E}[\mu(\Theta)]$. We describe our methodology in Section V.A and report results in Section V.B.

A. Calibration Procedure

We approximate the distribution of the parameter space, Θ, using tercile midpoints of the empirical distribution of each parameter. More specifically, for each parameter in the vector

$$\Theta_1 = \left(q, \lambda, w, \xi, \ell_s, y, \bar{\ell}, q^d, m, \tilde{\delta}\right), \tag{37}$$

we approximate the parameter's support using the 15^{th}, 50^{th}, and 85^{th} percentiles of its empirical distribution, with each value occurring one-third of the time.[14]

Table VII summarizes the resulting parameter distributions. Obtaining these distributions is straightforward for parameters that have a clear empirical measure with ample cross-sectional and time-series variation. For parameters that lack such variation, we prioritize empirical measures with a clear conceptual connection to the target parameter. Since we cannot reliably calibrate the joint distribution of the parameters in Θ, we instead work under the simplifying assumption of statistical independence. Calibration details are in Section III.D.2 of the Internet Appendix.

Probability of Transaction Failure. We measure q using the mortgage application denial rate among pre-approved, first-lien, mortgages for the purchase of an owner-occupied, single-family home. The data come from the HMDA data set aggregated to the zip code-by-year level. The

[14] Internet Appendix Figure IA.3 replicates our main findings with quintile discretization, for robustness. Note that our discretization is nonparametric, unlike more formal quadrature rules (e.g., Tauchen and Hussey (1991)), which is appropriate as we lack reasonable priors for the distributions of some parameters.

grid points in Table VII obtain from the distribution of q across zip codes and years. Section III.D of the Internet Appendix describes potential sources of measurement error in the HMDA-based measure and recalculates the theoretical premium using an alternative measure from the NAR.

Offer Arrival Rate. We measure λ using the average number of offers received conditional on receiving an offer, $\mathbb{E}[N|N > 0]$, and making use of the relationship for a Poisson random variable, $\mathbb{E}[N|N > 0] = \frac{\lambda}{1-e^{-\lambda}}$. We obtain the distribution of $\mathbb{E}[N|N > 0]$ from the Offer-Level data set, aggregated to the zip code by month level. Since λ is a relatively challenging parameter to calibrate, we recalibrate it under various other data sources in Section III.D of the Internet Appendix.

Seller Financial Wealth-to-Housing. We measure w using the ratio of financial wealth, including retirement plans, to the value of the home. We obtain this ratio from the SCF data set, which describes household balance sheets with a high degree of detail. We use the 2016 version of the SCF, restricting to homeowners who have sold at least one home.

Seller Loan-to-Value Ratio and Down Payment. We measure ℓ_s and ξ using the ZTRAX data set. Specifically, ℓ_s is the ratio of mortgage balance outstanding, imputed by a straight-line amortization using the loan's initial term, to the sale price. We calibrate ξ as the difference in price between the home to which the seller moves next and her mortgage on that home, either of which may equal zero. We then normalize by the sale price on her current home.

Buyer Liquid Assets-to-Housing. We obtain y at the zip code by year level from the IRS SOI data set to calculate the numerator, and the baseline ZTRAX data set to calculate the denominator. The numerator equals the ratio of total adjusted gross income to total tax returns in the zip code and year. The denominator equals the average sales price. We calculate the distribution of the ratio across zip codes and years to obtain the grid points in Table VII.

Loan-to-Value Constraint. Given the bunching of LTV ratios at regulatory limits in Figure 1, we measure $\bar{\ell}$ using the LTV ratio conditional on lying between the GSE regulatory limit and 100%. We calculate the regulatory limit as the minimum of 80% and the ratio of the applicable conforming loan limit to the average sales price.

Additional Parameters. Section III.D.2 of the Internet Appendix describes the calibration of the probability of transaction delay (q^d), the mortgage offer share (m), and the maintenance cost ($\tilde{\delta}$). Our baseline calibration does not feature nonfinancial contingencies: $q^c = h = 0$. We only observe these parameters as short monthly time series of less than three years in the NAR RCI data set. So, in the associated extensions, we simply use the average values. Accordingly, $q^c = 1\%$. We set $h = 7\%$ to match the share of all offers that come with a home-sale contingency.[15]

Preferences. Unless otherwise specified, we work with the following values of preference parameters in Θ_2: coefficient of relative risk-aversion of $\gamma = 5$, an annualized subjective discount rate of $\rho = 0.04$, and a disutility of borrowing of $D = 1$. The values of γ and ρ are standard when using CRRA preferences. Since D is unique to our model, in Internet Appendix Table IA.V we assess sensitivity to it.

B. Calibration Results

We obtain an average theoretical mortgage-cash premium of 6.9%, summarized in Figure 3. This value far exceeds the 3.3% premium obtained for a representative seller, shown in Table I and reproduced in this figure. The difference points to an important role for heterogeneity. In particular, risk-aversion introduces a nonlinear dependence of μ on the model's parameters, shown in Proposition 1. Consequently, sellers facing a high degree of transaction risk require far more than one-for-one compensation.

Consistent with this view, the average premium across the parameter space equals 1.9% under risk neutrality, shown in the second bar of Figure 3. A natural question, therefore, is whether simply increasing γ could explain the empirical premium. However, the fifth bar shows how explaining the most conservative estimate of 8.6% in Table V requires at least $\gamma = 12$, which exceeds the typical ceiling of plausibility of around 10 (e.g., Mehra and Prescott (1985)).

Nonfinancial Contingencies. The question of risk leads us to revisit Section I.A.3 and recalculate the theoretical premium with nonfinancial contingencies. Repeating the setup, both all-cash and mortgaged offers fail with probability q^c due to the inspection contingency. In addition, a share h of offers from mortgaged buyers come with a home-sale contingency. These

[15] If we divide by the share of offers that are mortgaged to obtain a conditional probability, then h rises to 11.3% and the resulting premium rises from 7.9% to 8.1%.

sale-contingent buyers will terminate the sale in question if the sale of their own home fails. Section III.B of the Internet Appendix shows that sale-contingent offers fail with combined probability

$$q^h = q + q^c + \frac{mq + q^c}{1 - mh}, \tag{38}$$

up to an approximation for small probabilities. The sum of first two terms equals the probability that the buyer himself does not obtain financing (q) or backs out because of the inspection (q^c). The third term reflects the probability of failure through the home-sale contingency. The numerator equals the probability that the buyer's sale falls through because of financing (mq) or inspection (q^c). This probability is amplified according to the prevalence of sale-contingent offers in the economy (mh): the buyer of the buyer's home may invoke a sale contingency, which may have been triggered by another sale contingency, etc.[16]

In the presence of nonfinancial contingencies, the calibrated theoretical premium increases from 6.9% to 7.9%, as shown in Figure 3. It is admittedly difficult to reliably estimate how pervasive nonfinancial contingencies are. The best available data, the NAR RCI data set, implies that only 7% of buyers come with a home-sale contingency ($h = 0.07$). Yet, even under this low value, the 1% increase in the theoretical premium suggests that home-sale contingencies introduce important amplification that could at least partially explain the magnitude of the empirical premium.

VI. Heterogeneity in Selling Conditions

Connecting our empirical and theoretical results requires that we estimate the empirical premium across the distribution of the model's parameters and then test for a difference relative to the theoretical premium at that point in the parameter space. This allows us to pursue the aforementioned conjecture that heterogeneity drives the mortgage-cash premium, and a puzzle

[16] A seller must now separately consider the probability of receiving various combinations of offers that are all-cash and mortgaged with and without the sale contingency in solving her dynamic problem. The resulting expression for κ conveys the same basic intuition as in Lemma 1, but it contains many additional terms that reflect these cases (Internet Appendix Proposition IA.1). Studying nonfinancial contingencies adds an important degree of realism to the model, and we thank the Editor for encouraging us to do so.

may exist only in certain states of the world.

We estimate the following regression equation using the ZTRAX data set,

$$\log\left(Price_{i,t}\right) = \mu_0 Mortgaged_{i,t} + \sum_{\theta}\sum_{p} \mu_{\theta,p}\left(Mortgaged_{i,t} \times Tercile(\hat{\theta}, p)_{z,t}\right) + ... \tag{39}$$
$$... + \psi X_{i,t} + \zeta_{z(i),t} + \alpha_i + \epsilon_{i,t},$$

where $\hat{\theta}_{z,t}$ is an empirical measure for model parameter θ at the zip code-by-year level, and $Tercile(\hat{\theta}, p)_{z,t}$ indicates whether $\hat{\theta}$ lies in the p^{th} tercile of the empirical distribution. For some parameters, $\hat{\theta}_{z,t}$ corresponds exactly to the measure used to calibrate the model in Table VII. Section III.D.4 of the Internet Appendix describes how we construct a suitable proxy to address cases where the measure in Table VII lacks enough variation within the ZTRAX data set. Importantly, it suffices for $\hat{\theta}_{z,t}$ to preserve the ranking of the parameter in question across zip codes and years, even if it lacks the accuracy of magnitude of the measures in Table VII.

Estimating equation (39) allows us to construct confidence intervals for the empirical premium at each value of a given parameter, holding other parameters fixed at their average values. Mirroring this approach within the model, we estimate a regression analogous to equation (39) across the model's parameter space (Internet Appendix Section III.D.3). We then calculate the theoretical premium across the distribution of each parameter. We perform this exercise on the subset of parameters with the most economic intuition: q, ω, λ, and L.

A. Heterogeneity Results

We find significant heterogeneity in the empirical and theoretical mortgage-cash premiums, and this heterogeneity helps explain the apparent divergence between the two suggested by Figure 3. We report our results on heterogeneity in Figure 4. Internet Appendix Table IA.VI tabulates the estimates of equation (39). We reason on the success rate, $1 - q$, so that moving along the horizontal axis in each figure corresponds to a more favorable market for sellers.

A.1. Main Heterogeneity Results

Panel A of Figure 4 summarizes one of the paper's key ideas: the mortgage-cash premium puzzle arises primarily in markets with moderate or high transaction friction. Specifically, the solid line shows the average empirical premium across terciles of the transaction success rate, $1-q$, from estimating equation (39). The negative slope with respect to the success rate matches the model's predictions, shown by the red dashed line. Quantitatively, however, the steepness of the empirical slope far exceeds that of the theoretical slope. Consequently, at low success rates (i.e., high q), the empirical premium lies 9 percentage points (pps) above the theoretical premium. This gap falls until the two premiums converge to around 6% within the highest tercile of $1 - q$.

For reference, the light gray lines in Figure 4 show the premium obtained from more naive calculations of the model that would overstate the magnitude of the puzzle. First, the dotted line denoted by "Representative Agent" shows the premium when ignoring heterogeneity, as in Table I. Even in low-friction markets with high success rates, a representative agent approach would predict a puzzle because it ignores important variation in other model parameters, such as the seller's net wealth. Second, the solid line denoted by "Fixed κ" shows the premium when fixing the cost of failure at its value from Table I, but allowing variation in parameters that directly affect the premium shown in Proposition 1. This oversight would overstate the magnitude of the puzzle by 1.5 pps (16%) in markets with a high degree of transaction risk. In such markets, sellers particularly dislike failure because they must then wait for new offers, and the best such offer may be another mortgaged buyer who, as with the first, fails to obtain financing. Thus, dynamic amplification is especially important in high-friction markets.

Moving to Panel B of Figure 4, the slope of the empirical and theoretical premiums with respect to the seller's net wealth, ω, align quite well. The seller's net wealth in our model functions similarly to investor risk-bearing capacity in more general asset pricing models: low net wealth raises the seller's effective risk-aversion, and so she requires a high premium for bearing transaction risk. Thus, leverage plays an important role in explaining the mortgage-cash premium through sellers' portfolio problem. By contrast, a naive model that focuses on housing wealth and omits total wealth would overlook this channel. The buyer's leverage also impacts the mortgage-cash premium, as we shall see shortly, but through a more indirect and subtle channel.

A.2. Key Heterogeneity Extensions

The first key extension recalculates the theoretical premium after incorporating nonfinancial contingencies and a higher coefficient of risk-aversion. Panel A of Figure 5 shows how nonfinancial contingencies reduce the puzzle's magnitude in high-friction states. The intuition, discussed previously in Section V.B, reflects how transaction failures elsewhere, which may be due to either financial or nonfinancial contingencies, can trigger a chain of home-sale contingencies that lead the transaction in question to fail. This channel reduces the puzzle by 1.7 pps (18%) in markets with the highest friction. As before, home-sale contingencies cannot entirely explain the empirical premium simply because they occur rather infrequently. Even so, incorporating them has a similar impact as raising risk-aversion to $\gamma = 12$, per Panel B of Figure 5.

Second, Internet Appendix Figure IA.4 resembles Figure 4 in terms of the offer arrival rate, λ, and the buyer's leverage capacity, L. The empirical premium exhibits a negative slope with respect to both parameters, consistent with the model. Like in Figure 4, it lies above the theoretical premium in markets that are less favorable to sellers, namely illiquid markets (low λ) and markets in which borrowers have limited ability to use leverage to increase their offer price (low L).[17]

Overall, a proper treatment of heterogeneity reveals that the wedge between the empirical and theoretical premiums obtains primarily under conditions that are unfavorable to sellers, especially in states with high transaction risk. This finding motivates us to evaluate belief distortions that may amplify the baseline effect of transaction risk in these states.

[17]The buyer's leverage has a weaker impact on both the empirical and theoretical premium than does the seller's leverage. In the model, this occurs because the buyer's leverage capacity only impacts the seller through the possibility that the current transaction fails and the seller then receives multiple offers, the best of which is a mortgaged buyer who levers to the maximum. However, this does not negate the importance of financial constraints; rather, it shows that constraints affect the premium more through the extensive margin of credit (q) than the intensive margin (L). Intuitively, the extensive margin, q, affects μ both directly through the seller's indifference condition, shown in Proposition 1, and indirectly through the continuation value, shown in Lemma 1. By contrast, L works only indirectly. We thank the Associate Editor for encouraging us to examine how financial constraints and leverage can explain the mortgage-cash premium.

VII. Belief Distortions as an Amplification

We use an experimental survey to evaluate whether noncanonical preferences ("belief distortions") can explain the mortgage-cash premium in states with high transaction risk. We have two goals: first, to test whether our main results summarized in Section VI obtain in an experimental setting, and, second, to rank a set of candidate belief distortions by their ability to explain the mortgage-cash premium. We consider four distortions known to affect behavior in financial markets, in particular, ambiguity-aversion, realization utility, present focus, and probability weighting. This list is not exhaustive, and so our purpose is to simply provide a ranking within this restricted set.

A. *Survey Design*

We administer the survey online to participants who have been pre-screened, compensated for their participation, and recruited through a mainstream crowdsourcing platform. We would like to simulate a lab experiment on representative homeowners as closely as possible, and, reassuringly, Casler, Bickel, and Hackett (2013) finds that online experiments can produce data of similar quality as those obtained in the lab. We recruit survey respondents through the crowdsourcing platform Prolific, a competitor to Amazon's commonly used MTurk platform. Prolific has several advantages over MTurk, including the ability to exclusively recruit U.S. homeowners and a stringent screening process that results in more honest responses. Section VII.B.1 discusses these advantages. Section IV of the Internet Appendix describes the survey in detail.

We administered the survey in three nonlongitudinal waves with a similar question structure. At a high level, the survey consists of a thought experiment in which we ask respondents to imagine that they are selling their current home. We specify a particular list price, chosen to match the typical sales price in the respondent's neighborhood. The conditions of the sale are similar to the median values in Table VII. Specifically, we tell respondents that they have an outstanding mortgage balance equal to 30% of this list price ($\ell_s = 30\%$). 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 ($\xi = 15\%$). This down payment must be made within six weeks, which also equals the mortgage closing period. We state all quantities in both percents and dollars.

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 all-cash, and the second buyer would pay using mortgage financing. We specify that the all-cash transaction has "almost no risk" of failing and will close "any time within two weeks." The mortgaged transaction would take four weeks longer than the all-cash transaction to close.

Importantly, we state in the survey that "there is a chance that the mortgaged buyer will not be able to secure money from their lender," in which case the respondent "will need to relist your home in six weeks." In the survey's second and third waves, a random subsample of participants are told that mortgaged transactions will fail q percent of the time, where q takes values of 1%, 7%, and 13%, similar to the grid points of the empirical distribution in Table VII. This randomization enables us to test for ambiguity-aversion. Finally, we specify that the respondent would need to impose a 6% price cut to attract another offer, expressed in dollars and percent. We use this price cut to approximate κ. Otherwise, describing all of the parameters to calculate Lemma 1 would overwhelm survey respondents.

The remainder of the survey consists of three 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 which offer they would prefer. We elicit the respondent's mortgage-cash premium through a sequence of pairwise comparisons (i.e., multiple price list), following standard practice. We ask the respondent which offer she would prefer at gradually increasing spreads between the mortgaged and the all-cash offer price. Each question takes the form: *"Suppose the Mortgaged Buyer offers to pay \$[*$(1 + \tilde{\mu}) \times B$*]. That is [*$100 \times \tilde{\mu}$*]% more than the Cash Buyer. Which offer would you accept now?"* The spread $\tilde{\mu}$ grows in increments of 4 pps until reaching 28%. Most respondents switch from preferring the all-cash to the mortgaged offer once $\tilde{\mu}$ exceeds some threshold strictly between 0% and 28%. 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.

In the second block, we elicit respondents' prior beliefs about mortgage transaction failure (q) for the subsample facing ambiguity. Respondents select a value for q on a sliding scale with an upper bound of 30%. The use of a sliding scale does not lead to bunching around the scale's midpoint, since 43% of respondents select a value less than 10% or greater than 20%.

Finally, we elicit the respondent's numeracy following the method of Lipkus, Samsa, and Rimer (2001), and we collect information about the respondent's annual household income, age, state of residence, education, and risk-aversion, elicited using the method in Fuster and Zafar (2021). Our resulting data set contains information on 1,019, 1,202, and 1,199 U.S. homeowners from the survey's first, second, and third waves. We administered the three waves in April 2021, November 2021, and January 2023. These periods cover three out of four seasons and include a range of conditions for home sellers, ranging from quite favorable due to strong demand and limited supply in 2021 (Gascon and Haas (2020)) to a slowdown by early 2023.

B. Experimental Mortgage-Cash Premium

Table VIII summarizes the experimental mortgage-cash premium and other characteristics of survey respondents. The top row of Panel A reports an average mortgage-cash premium of 10.5% in our pooled sample. Columns (2) to (4) show little variation across the three waves. The next row weights respondents by 2020 Census weight to address concerns of representativeness, discussed shortly. Next, following the literature's convention, we restrict the sample to single-switchers, who exhibit a single positive switch point strictly between zero and the maximum value in the multiple price list (e.g., Bernheim and Sprenger (2020), Andreoni and Sprenger (2012)). These two steps lower the average premium to 9.6%.

Panels B and C summarize variables that map to model parameters and demographic characteristics, respectively. We note that survey respondents who face ambiguity believe that mortgaged transactions fail 13% of the time, which is higher than the empirical average of 7%. Since we do not observe fine enough geographic information to cross-reference this probability against HMDA data, we refrain from characterizing respondents as "pessimistic." We next turn to demographic characteristics and the question of representativeness.

B.1. Representativeness and Response Bias

Any survey must contend with response bias, which can work through observed or unobserved characteristics. A growing body of evidence suggests that Prolific, the crowdsourcing platform we use, is the least susceptible to both margins of bias, as discussed below.

In terms of observed characteristics, Prolific allows researchers to exclusively recruit participants with a particular demographic profile. Crucially, this restriction enables us to restrict attention to individuals with knowledge of the home sale process, namely U.S. homeowners. Relative to the population of U.S. homeowners, our respondents have similar average income and somewhat higher levels of education.[18] Geographically, our sample covers all U.S. states in similar proportion to state population (see Internet Appendix Figure IA.7). Lastly, while our sample appears representative, we nevertheless weight respondents by the share of all U.S. homeowners in the same income-by-education-by-age-by-state of residence bin, using bins from the 2020 Census. Thus, our results up-weight respondents who appear more similar to the typical U.S. homeowner.

Turning to unobserved characteristics, we seek to avoid both inattentive and manipulative respondents. To motivate attentiveness, we compensate respondents for their participation at three times the federal minimum wage. Prolific is known to filter out, to the extent possible, manipulative respondents. For example, in an influential study, Peer et al. (2017) repeat a set of common experiments that test for psychological biases across MTurk, Prolific, and other mainstream crowdsourcing platforms. They find that respondents recruited by Prolific provided both reproducible and internally consistent responses, but they were less likely to manipulate the experiment and exhibited significantly higher rates of honesty.

C. Ambiguity-Aversion

Ambiguity-aversion describes a preference for risky investments with a known distribution over those with an unknown, "ambiguous" distribution (Machina and Siniscalchi (2014)). More precisely, an ambiguity-averse investor who forecasts that mortgaged transactions fail with probability $\hat{q}$ will instead make decisions according to a distorted probability $Q > \hat{q}$.

[18]Concerning income, Internet Appendix Figure IA.5 plots the share of respondents with income in bins defined by the 2020 Census, along with the share of all U.S. homeowners in those bins. Our sample resembles the overall distribution quite well within the center of the distribution. Concerning education, the 70.2% share of respondents with a bachelor's degree exceeds the 40.1% share among U.S. homeowners from the 2019 American Housing Survey.

In practice, selling a home can be a very ambiguous process. Home sales occur relatively rarely, especially for owner occupants who comprise the majority of our sample. Consequently, most home sellers have likely acquired little information about the true distribution of transaction risk. Consistent with this view, Internet Appendix Table IA.XII estimates that sellers with prior experience require a 2 pps lower mortgage-cash premium for each home they have already sold.[19]

Given the important role of heterogeneity in Figure 4, we take a similar approach and estimate the effect of ambiguity on the mortgage-cash premium across the distribution of q. Specifically, we pool our three survey waves and estimate

$$\begin{aligned} Premium_k = &\sum_p \mu_{q,p} Tercile(\hat{q},p)_k + \sum_p \Delta\mu^A_{q,p} \left(Tercile(\hat{q},p)_k \times Ambiguity_k \right) + ... \\ &... + \sum_\theta \sum_p \mu_{\theta,p} Tercile(\hat{\theta},p)_k + \epsilon_k, \end{aligned} \tag{40}$$

where k indexes survey respondent, $Ambiguity_k$ indicates whether the respondent is randomized into the 62% of the sample in which we do not provide the distribution of q, and $Tercile(\hat{q},p)_k$ indicates whether the respondent's probability of failure, $\hat{q}$, lies in the p^{th} tercile of the empirical distribution used in Figure 4. We measure $\hat{q}$ using the value of q provided to the respondent or, for those facing ambiguity, the respondent's prior probability, elicited at the end of the survey. The vector of indicators $Tercile(\hat{\theta},p)_k$ has a similar interpretation as in equation (39), with details in Section III.D.5 of the Internet Appendix. We always weight respondents using Census weights to correct for response bias.

We build on Figure 4 and use the estimates from equation (40) to trace out the mortgage-cash premium across the distribution of transaction risk, and now we do so separately for respondents facing an ambiguous versus a known distribution of q. As before, this exercise constitutes the empirical (i.e., experimental) portion of the figure. We then calculate the mortgage-cash premium according to Proposition 1 across terciles of the parameter space. This theoretical premium will vary by q, w, and κ, as implied by Panel B of Table VIII. Given practical constraints on survey size, we preserve power by presenting all respondents with same values of ℓ_s and ξ, and we use

[19] It is possible that listing agents have an incentive to keep the process ambiguous for their client. For example, doing so may shorten the time-to-close across transactions as ambiguity-averse sellers prioritize all-cash offers, which tend to close more quickly (e.g., Levitt and Syverson (2008)).

the average values of the remaining parameters in Table I to calculate the theoretical premium.

Figure 6 reports the results. Panel A plots the empirical premium across terciles of $1 - q$ conditional on ambiguity, $Ambiguity_k = 1$. As with the companion Figure 4 derived from observational data, the slope of the empirical premium with respect to q significantly exceeds the slope implied by the model. Consequently, there exists a significant 6.4 pps wedge between the empirical and theoretical premiums at low success rates (high q), while the two premiums converge at high success rates.

Essentially, Panel A of Figure 6 replicates our main finding using experimental data, which allows us to analyze the effect of ambiguity about a key parameter on the premium. Panel B evaluates ambiguity-aversion by plotting the empirical premium across terciles of q, conditional on facing a known distribution, $Ambiguity_k = 0$. Interestingly, the slope of the empirical premium with respect to q becomes much more flat relative to the case of ambiguity (Panel A). This flattening reduces the empirical premium by 4 pps at low success rates such that it becomes much closer to the theoretical premium. Internet Appendix Figure IA.9 shows that removing ambiguity also reduces the wedge between empirical and theoretical premiums across the distribution of other model parameters, w and κ, through a level shift.

Overall, removing ambiguity significantly reduces the empirical premium in states with high transaction risk and, by extension, the puzzle in those states. In reality, sellers do face substantial ambiguity. Thus, our experimental evidence suggests that a more accurate model of the sale process should allow for ambiguity-aversion of some form. We propose a brief extension of our model that does so.

C.1. Modelling Ambiguity-Aversion

Most models of ambiguity-aversion would begin with the same basic idea outlined earlier: home sellers fear that their approximation of the failure probability, $\hat{q}$, understates the true value, and so they instead act according to the distorted probability Q. The degree of ambiguity-aversion determines how far Q lies from $\hat{q}$. Models differ in how they would set up the seller's problem of choosing Q. We follow the setup in the literature on robust decision-making (e.g., Hansen and Sargent (2011)), although many of the other options in footnote 1 also seem appropriate.

An ambiguity-averse ("robust") home seller evaluates mortgaged offers according to a worst-case failure rate that solves

$$Q = \arg\min_{\tilde{q}} V^{\mathcal{S}}(M, P; \tilde{q}) \quad s.t. \quad \mathcal{R}\left(\tilde{q}; \hat{q}\right) \leq \overline{\mathcal{R}}, \tag{41}$$

where

$$\mathcal{R}\left(\tilde{q}; \hat{q}\right) = \tilde{q} \log\left(\frac{\tilde{q}}{\hat{q}}\right) + (1 - \tilde{q}) \log\left(\frac{1 - \tilde{q}}{1 - \hat{q}}\right) \tag{42}$$

is the relative entropy between $\tilde{q}$ and $\hat{q}$. The maximum permissible entropy, $\overline{\mathcal{R}}$, parameterizes the seller's ambiguity-aversion. Since $V^{\mathcal{S}}(M, P; q)$ is decreasing in q, the constraint in equation (41) binds and so Q solves

$$\overline{\mathcal{R}} = Q \log\left(\frac{Q}{\hat{q}}\right) + (1 - Q) \log\left(\frac{1 - Q}{1 - \hat{q}}\right). \tag{43}$$

We can then calculate the mortgage-cash premium as in Proposition 1 after substituting Q for q. Computationally, the methodology is the same as in Section V.

Generating an average theoretical premium of 11.1% to match the empirical premium in Table IV would require ambiguity-aversion (i.e., an entropy bound) of $\overline{\mathcal{R}} = 0.35$. It is hard to compare entropy bounds across settings, but a bound of 0.35 stands out as relatively large.[20] Barnett, Brock, and Hansen (2020), for example, report entropies between 0.01 and 0.20. Even so, reducing $\overline{\mathcal{R}}$ to 0.15 still generates a theoretical premium of 9.5%, on par with the estimates in Table V. Internet Appendix Figure IA.6 reports these results and also shows how both values of $\overline{\mathcal{R}}$ enable the model to fit the distribution of μ, not just the average. In particular, incorporating ambiguity-aversion steepens the slope of μ with respect to q, so that the model comes much

[20] It is more common for papers to report the Lagrange multiplier associated with the constraint in equation (41). The reason that multipliers are often reported is because the literature often works with a companion setup to equation (41) that does not feature an entropy constraint, but instead adds a term of the form $A\mathcal{R}\left(\tilde{q}; \bar{q}\right)$ to the objective function, where A is a constant. Hansen and Sargent (2001) show that the two setups yield the same solution, and, in particular, A may be interpreted as the Lagrange multiplier from the problem in equation (41). Hansen and Sargent (2011) stress that the values of $\overline{\mathcal{R}}$ or A should be context-specific, which makes cross-paper comparisons difficult. Further complicating cross-paper comparisons of A is the fact that multipliers are not scale-independent (Maenhout (2004)).

closer to matching the empirical premium in high-friction states.

Putting these values into context, a representative home seller who approximates $\hat{q} = 6.4\%$ as in Table I and has ambiguity-aversion of $\overline{\mathcal{R}} = 0.35$ would act according to $Q = 35\%$. This worst-case probability corresponds to the top 1.6% of the distribution of q summarized in Table VII. With ambiguity-aversion of $\overline{\mathcal{R}} = 0.15$, she acts according to $Q = 23\%$, which lies at the top 5% of the distribution. So, an empirical premium of 9% to 11% is consistent with an ambiguity-averse home seller selecting a worst-case model from the top 1% to 5% of the distribution of q, which in our sample occurred during the Great Recession. We cannot definitively say whether this precise degree of ambiguity-aversion is reasonable.[21] However, taken together with the experimental evidence, a model in which home sellers have substantial ambiguity-aversion can go a long way towards explaining the empirical premium.

D. *Additional Belief Distortions*

While ambiguity-aversion performs well in our setting, the analysis would be incomplete without considering other belief distortions known to affect behavior in financial markets: reference dependence, and, in particular, realization utility, probability weighting, and present focus.

Reference Dependence through Realization Utility. A number of frameworks have applied the basic principle of reference-dependent optimization (e.g., Tversky and Kahneman (1992)) to the case of financial markets. Realization utility is one such framework that seems natural for our setting because of its focus on the sale of an asset (e.g., Barberis and Xiong (2012), Ingersoll and Jin (2013)). At a high level, realization utility leads investors (i.e., home sellers) to organize time according to trading episodes (i.e., home sales), and they experience a burst of utility whenever an episode ends in a realized capital gain.

Section III.E of the Internet Appendix provides an analytic approximation that conveys three ideas about how the mortgage-cash premium for sellers with realization utility compares

[21] Anderson, Hansen, and Sargent (2003) propose evaluating the plausibility of a given $\overline{\mathcal{R}}$ according to whether the implied worst-case probability $Q_{\overline{\mathcal{R}}}$ would be rejected above some reasonable Type-I error threshold using data generated by $\hat{q}$. Applying this criteria with a 5% error threshold, ambiguity-aversion of $\overline{\mathcal{R}} = 0.15$ seems reasonable. We clarify that the precise Anderson, Hansen, and Sargent (2003) procedure is more involved and would require several assumptions about the size of the sample from which a home seller can draw inference.

with the baseline premium. First, sellers with realization utility may require a lower premium if they place greater value on the option of waiting for a large capital gain from a bidding war and thus have a lower cost of failure. Second, the existence of a positive mortgage-cash premium makes sellers more willing to accept mortgaged offers because they come with the potential for a higher capital gain. Third, factors that raise the overall level of the seller's potential capital gain (e.g., leverage) may have a small impact on the mortgage-cash premium unless they also affect the relative utility burst from the gain on a mortgaged versus an all-cash offer.

Since it is theoretically unclear how realization utility affects the mortgage-cash premium, on average, our experiment focuses on the specific case in which an all-cash offer results in zero capital gain, which should imply a lower mortgage-cash premium. Section III.E of the Internet Appendix describes the experiment and Internet Appendix Figure IA.8 summarizes the results. We find no significant difference in the premium among respondents for whom realization utility would predict a lower value. This finding comes with the caveat that realization utility likely influences home sellers' behavior in ways distinct from their preference for all-cash offers, as we describe in Section III.E of the Internet Appendix.

Probability Weighting. Probability weighting refers to the tendency to perceive small probabilities as larger than they actually are when making decisions. Section III.E of the Internet Appendix shows that this tendency cannot explain the puzzle because it predicts an abnormally large premium when transaction risk is low, whereas the opposite is true empirically.

Present Focus. Present focus refers to a preference for immediate gratification that leads to dynamically inconsistent decisions. Section III.E of the Internet Appendix describes why, theoretically, present focus does not affect problems like ours in which home sellers do not immediately consume their sale proceeds and payoffs arrive over the course of weeks, rather than instantaneously.

VIII. Conclusion

We find that the financing of home purchases affects transaction value to an extent that it cannot be explained by transaction frictions alone. Based on a variety of subsamples, estimators, data sets, 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 dynamic, quantitative model with a representative home seller implies a premium of only 3%. Accounting for heterogeneity in selling conditions, especially transaction risk, raises the model-implied premium and so reduces the 8 pps price puzzle by half. Our survey evidence suggests that belief distortions, in particular ambiguity-aversion, can explain the remaining half of the puzzle.

Our results have policy implications that derive from buyers' and sellers' perspectives. 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 ambiguity that amplifies them. The former route may have an outsized impact because of amplification, but this needs to be verified in a general equilibrium setting. The latter route requires more research on how ambiguity and aversion to ambiguity vary across home sellers, as well as how best to model ambiguity-aversion. We leave these questions for future research.

Initial submission: April 27, 2022; Accepted: October 2, 2023
Editors: Stefan Nagel, Philip Bond, Amit Seru, and Wei Xiong

REFERENCES

Abadie, A., and G. W. Imbens, 2006, Large sample properties of matching estimators for average treatment effects, *Econometrica* 74, 235–267.

Almeida, H., and T. Philippon, 2007, The risk-adjusted cost of financial distress, *Journal of Finance* 62, 2557–2586.

Anderson, E. W., L. P. Hansen, and T. J. Sargent, 2003, A quartet of semigroups for model specification, robustness, prices of risk and model detection, *Journal of the European Economic Association* 1, 68–123.

Andreoni, J., and C. Sprenger, 2012, Estimating time preferences from convex budgets, *American Economic Review* 102, 1593–1610.

Asabere, P. K., F. E. Huffman, and S. Mehdian, 1992, The price effects of cash versus mortgage transactions, *Real Estate Economics* 20, 141–154.

Bai, B., J. Zhu, and L. Goodman, 2015, A closer look at the data on first-time homebuyers, Working paper, Urban Institute.

Bajari, P., J. C. Fruehwirth, K. I. Kim, and C. Timmins, 2012, A rational expectations approach to hedonic price regressions with time-varying unobserved product attributes: The price of pollution, *American Economic Review* 102, 1898–1926.

Barberis, N., L. J. Jin, and B. Wang, 2021, Prospect theory and stock market anomalies, *Journal of Finance* 76, 263–298.

Barberis, N., and W. Xiong, 2012, Realization utility, *Journal of Financial Economics* 104, 251–271.

Barnett, M., W. Brock, and L. P. Hansen, 2020, Pricing uncertainty induced by climate change, *Review of Financial Studies* 33, 1024–1066.

Barnett, M., G. Buchak, and C. Yannelis, 2021, Epidemic responses under uncertainty, Working paper, NBER.

Barwick, P. J., and P. A. Pathak, 2015, The costs of free entry: An empirical study of real estate agents in greater Boston, *RAND Journal of Economics* 46, 103–145.

Bayer, P., K. Mangum, and J. W. Roberts, 2021, Speculative fever: Investor contagion in the housing bubble, *American Economic Review* 111, 567–601.

Bernheim, B. D., and C. Sprenger, 2020, On the empirical validity of cumulative prospect theory: Experimental evidence of rank-independent probability weighting, *Econometrica* 88, 301–337.

Bleichrodt, H., and H. L. Pinto, 2000, A parameter-free elicitation of the probability weighting function in medical decision analysis, *Management Science* 46, 1485–1496.

Booij, A. S., B. M. S. Van Praag, and G. Van De Kuilen, 2010, A parametric analysis of prospect theory's functionals for the general population, *Theory and Decision* 68, 115–148.

Bricker, J., L. J. Dettling, A. Henriques, J. W. Hsu, L. Jacobs, K. B. Moore, S. Pack, J. Sabelhaus, J. Thompson, and R. A. Windle, 2017, Changes in U.S. family finances from 2013 to 2016: Evidence from the Survey of consumer finances, *Federal Reserve Bulletin* .

Buchak, G., G. Matvos, T. Piskorski, and A. Seru, 2020, Why is intermediating houses so difficult? Evidence from iBuyers, Working paper, NBER.

Cagetti, M., L. P. Hansen, T. Sargent, and N. Williams, 2002, Robustness and pricing with uncertain growth, *Review of Financial Studies* 15, 977–1007.

Caskey, J. A., 2008, Information in equity markets with ambiguity-averse investors, *Review of Financial Studies* 21, 1443–1484.

Casler, K., L. Bickel, and E. Hackett, 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* 29, 2156–2160.

Chen, H., 2010, Macroeconomic conditions and the puzzles of credit spreads and capital structure, *Journal of Finance* 65, 2171–2212.

Chen, L., P. Collin-Dufresne, and R. S. Goldstein, 2009, On the relation between the credit spread puzzle and the equity premium puzzle, *Review of Financial Studies* 22, 3367–3409.

Cheng, I.-H., S. Raina, and W. Xiong, 2014, Wall street and the housing bubble, *American Economic Review* 104, 2797–2829.

Chinco, A., and C. Mayer, 2016, Misinformed speculators and mispricing in the housing market, *Review of Financial Studies* 29, 486–522.

Cororaton, S., 2022, Realtors confidence index survey: Buyer competition intensifies, Working Paper, National Association of Realtors.

D'Acunto, F., and A. Rossi, 2020, Regressive mortgage credit redistribution in the post-crisis era, *Review of Financial Studies* 33, 423–456.

Ellie Mae, 2012, Origination insight report, Technical report.

Ericson, K. M., and D. Laibson, 2019, Intertemporal choice, in B. D. Bernheim, S. DellaVigna, and D. Laibson, eds., *Handbook of behavioral economics - foudantions and applications* (North-Holland, Kidlington, Oxford U.K.).

Federal Deposit Insurance Corporation, 2019, New appraisal threshold for residential real estate loans, *Financial Institution Letter* .

Federal Reserve, 2019, Financial accounts of the United States: Mortgage debt outstanding, Technical report.

Foote, C. L., L. Loewenstein, and P. S. Willen, 2020, Cross-sectional patterns of mortgage debt during the housing boom: Evidence and implications, *Review of Economic Studies* 87, 1293–1334.

Fuster, A., and B. Zafar, 2021, The sensitivity of housing demand to financing conditions: Evidence from a survey, *American Economic Journal: Economic Policy* 13, 144–181.

Gascon, C. S., and J. Haas, 2020, The impact of covid-19 on the residential real estate market, *Federal Reserve Bank of St. Louis Regional Economist* .

Genesove, D., and C. Mayer, 2001, Loss aversion and seller behavior: Evidence from the housing market, *Quarterly Journal of Economics* 116, 1233–1260.

Gete, P., and M. Reher, 2018, Mortgage supply and housing rents, *Review of Financial Studies* 31, 2347–2386.

Gete, P., and M. Reher, 2021, Mortgage securitization and shadow bank lending, *Review of Financial Studies* 34, 546–585.

Gilboa, I., and D. Schmeidler, 1989, Maxmin expected utility with non-unique prior, *Journal of Mathematical Economics* 18, 141–153.

Gilbukh, Sonia, and Paul Goldsmith-Pinkham, 2023, Heterogeneous Real Estate Agents and the Housing Cycle, Working Paper.

Glaeser, E. L., and C. G. Nathanson, 2017, An extrapolative model of house price dynamics, *Journal of Financial Economics* 126, 147–170.

Glaeser, E. L., and T. Sinai, 2013, *Housing and the financial crisis* (The University of Chicago Press, Chicago, Illinois).

Greenwald, D., 2018, The mortgage credit channel of macroeconomic transmission, Working paper.

Guren, A. M., 2018, House price momentum and strategic complementarity, *Journal of Political Economy* 126, 1172–1218.

Han, L., and S.-H. Hong, 2011, Testing cost inefficiency under free entry in the real estate brokerage industry, *Journal of Business and Economic Statistics* 29, 564–578.

Han, L., and S.-H. Hong, 2023, Cash is king? Understanding financing risk in housing markets, Working Paper.

Han, L., and W. C. Strange, 2016, What is the role of the asking price for a house?, *Journal of Urban Economics* 93, 115–130.

Hansen, L. P., and T. J. Sargent, 2001, Robust control and model uncertainty, *American Economic Review: Papers and Proceedings* 91, 60–66.

Hansen, L. P., and T. J. Sargent, 2011, *Robustness* (Princeton University Press, Princeton, New Jersey).

Hansen, L. P., G. A. Turmuhambetova T. J. Sargent, and N. Williams, 2002, Robustness and uncertainty aversion, *Review of Financial Studies* 15, 975–1007.

Hansz, J. A., and D. K. Hayunga, 2016, Revisiting cash financing in residential transaction prices, *Real Estate Finance* 33, 55–72.

Harding, J. P., E. Rosenblatt, and V. W. Yao, 2012, The foreclosure discount: Myth or reality?, *Journal of Urban Economics* 71, 204–218.

Hsieh, C.-T., and E. Moretti, 2003, Can free entry be inefficient? Fixed commissions and social waste in the real estate industry, *Journal of Political Economy* 111, 1076–1122.

Huang, J.-Z., and M. Huang, 2012, How much of the corporate-treasury yield spread is due to credit risk?, *Review of Asset Pricing Studies* 2, 153–202.

Ingersoll, J. E., and L. J. Jin, 2013, Realization utility with reference-dependent preferences, *Review of Financial Studies* 26, 723–767.

Kaplan, G., K. Mitman, and G. L. Violante, 2020, The housing boom and bust: Model meets evidence, *Journal of Political Economy* 128, 237–271.

Kennedy, R., S. Clifford, T. Burleigh, P. D. Waggoner, R. Jewell, and N. J. G. Winter, 2021, The shape of and solutions to the mturk quality crisis, *Political Science Research and Methods* 9, 641–657.

Klibanoff, P., M. Marinacci, and S. Mukerji, 2005, A smooth model of decision making under ambiguity, *Econometrica* 73, 1849–1892.

Koszegi, B., and M. Rabin, 2006, A model of reference-dependent preferences, *Quarterly Journal of Economics* 121, 1133–1165.

Laibson, D., 1997, Golden eggs and hyperbolic discounting, *Quarterly Journal of Economics* 112, 443–477.

Lebrun, B., 1996, Existence of an equilibrium in first price auctions, *Journal of Economic Theory* 69, 157–173.

Leuven, E., and B. Sianesi, 2003, Psmatch2: Stata module to perform full mahalanobis and propensity score matching, common support graphing and covariate imbalance testing.

Levitt, S. D., and C. Syverson, 2008, Market distortions when agents are better informed: The value of information in real estate transactions, *Review of Economics and Statistics* 90, 599–611.

Lian, C., Y. Ma, and C. Wang, 2018, Low interest rates and risk taking: Evidence from individual investment decisions, *Review of Financial Studies* 31, 2103–2140.

Lipkus, I. M., G. Samsa, and B. K. Rimer, 2001, General performance on a numeracy scale among highly educated samples, *Medical Decision Making* 21, 37–44.

Liu, H., and C. Palmer, 2021, Are stated expectations actual beliefs? New evidence for the beliefs channel of investment demand, Working paper, NBER.

Loutskina, E., and P. E. Strahan, 2015, Financial integration, housing and economic volatility, *Journal of Financial Economics* 115, 25–41.

Lusht, K. M., and J. A. Hansz, 1994, Some further evidence on the price of mortgage contingency clause, *Journal of Real Estate Research* 9, 221–228.

Machina, M. J., and M. Siniscalchi, 2014, Ambiguity and ambiguity aversion, in M. J. Machina, and K. Viscusi, eds., *Handbook of the economics of risk and uncertainty* (North-Holland, Kidlington, Oxford U.K.).

Maenhout, P. J., 2004, Robust portfolio rules and asset pricing, *Review of Financial Studies* 17, 951–983.

Maskin, E., and J. Riley, 2000, Equilibrium in sealed high bid auctions, *Review of Economic Studies* 67, 575–599.

Mehra, R., and E. C. Prescott, 1985, The equity premium: A puzzle, *Journal of Monetary Economics* 15, 145–161.

Milgrom, P., and C. Shannon, 1994, Monotone comparative statics, *Econometrica* 62, 157–180.

Mills, J., R. Molloy, and R. Zarutskie, 2019, Large-scale buy-to-rent investors in the single-family housing market: The emergence of a new asset class, *Real Estate Economics* 47, 888–915.

Modigliani, F., and M. H. Miller, 1958, The cost of capital, corporation finance and the theory of investment, *American Economic Review* 48, 261–297.

Murray, M. P., 2006, Avoiding invalid instruments and coping with weak instruments, *Journal of Economic Perspectives* 20, 111–132.

National Association of Realtors, 2016, Realtors confidence index survey, Technical report.

National Association of Realtors, 2018, Realtors confidence index survey, Technical report.

National Association of Realtors, 2020, Realtors confidence index survey, Technical report.

O'Donoghue, T., and M. Rabin, 1999, Doing it now or later, *American Economic Review* 89, 103–124.

O'Donoghue, T., and M. Rabin, 2015, Present bias: Lessons learned and to be learned, *American Economic Review* 105, 273–279.

Peer, E., S. Samat L. Brandimarte, and A. Acquisti, 2017, Beyond the turk: Alternative platforms for crowdsourcing behavioral research, *Journal of Experimental Social Psychology* 70, 153–163.

Piazzesi, M., and M. Schneider, 2016, Housing and macroeconomics, in J. B. Taylor, and H. Uhlig, eds., *Handbook of Macroeconomics* (North-Holland, Kidlington, Oxford U.K.).

Piazzesi, M., M. Schneider, and J. Stroebel, 2020, Segmented housing search, *American Economic Review* 110, 422–452.

Prelec, D., 1998, The probability weighting function, *Econometrica* 66, 497–527.

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* 104, 687–706.

Seo, Y., C. Holmes, and M. Lee, 2021, Examining the cash-only price discount and the driving forces of cash-only transactions in the housing market, *Journal of Real Estate Finance and Economics* 53, 456–476.

Shiller, R. J., 2014, Speculative asset prices, *American Economic Review* 104, 1486–1517.

Solon, G., S. J. Haider, and J. M. Wooldridge, 2015, What are we weighting for?, *Journal of Human Resources* 50, 301–316.

Stock, J. H., and M. Yogo, 2005, Testing for weak instruments in linear iv regression, *Identification and Inference for Econometric Models* 80–108.

Strzalecki, T., 2011, Axiomatic foundations of multiplier preferences, *Econometrica* 79, 47–73.

Tauchen, G., and R. Hussey, 1991, Quadrature-based methods for obtaining approximate solutions to nonlinear asset pricing models, *Econometrica* 59, 371–396.

Townsend, R. M., 1994, Risk and insurance in village India, *Econometrica* 62, 539–591.

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 D. Kahneman, 1992, Advances in prospect theory: Cumulative representation of uncertainty, *Journal of Risk and Uncertainty* 5, 297–323.

Zillow, 2018, Consumer housing trends report, Technical report.

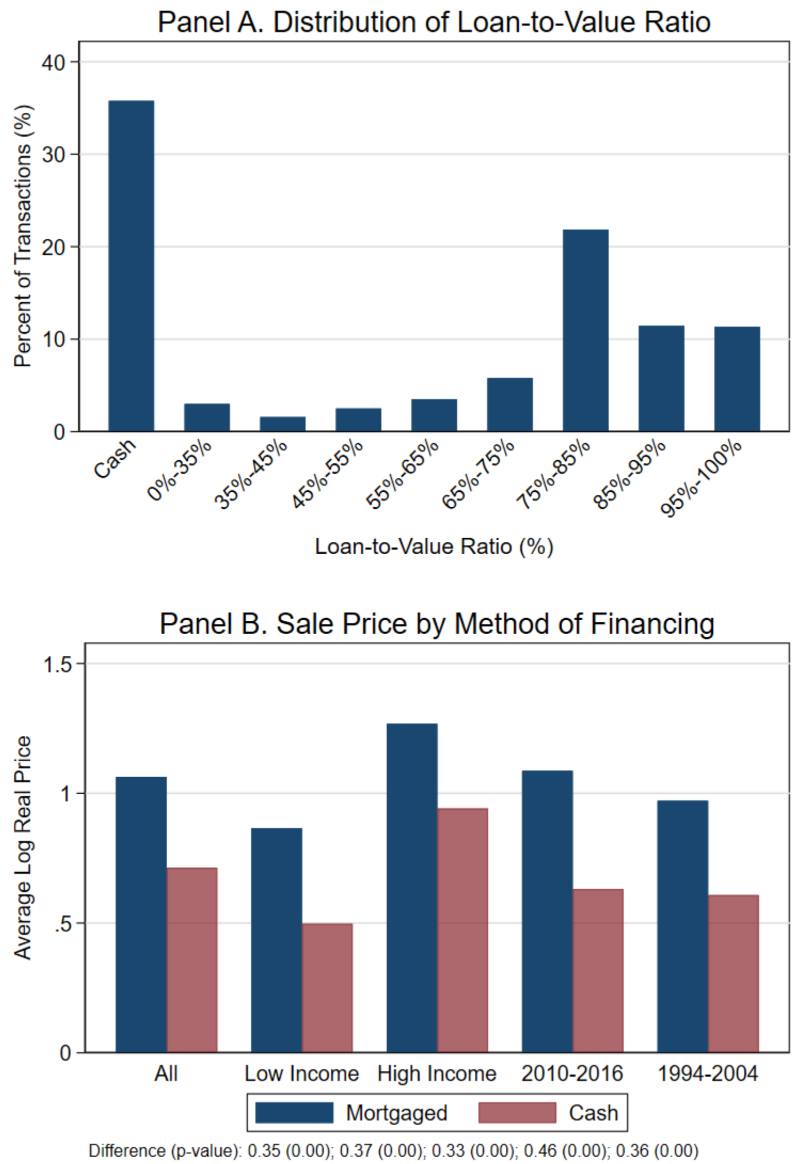


Figure 1. Facts about all-cash purchases. This figure documents two facts about cash-financed home purchases. Panel A plots the distribution of LTV ratios across purchases over 1980 to 2017. Panel B plots the average log sales price in hundreds of thousands of 2010 dollars for purchases financed by a 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 to 2016 and 1994 to 2004 denote purchases over these two periods. Data are from ZTRAX.

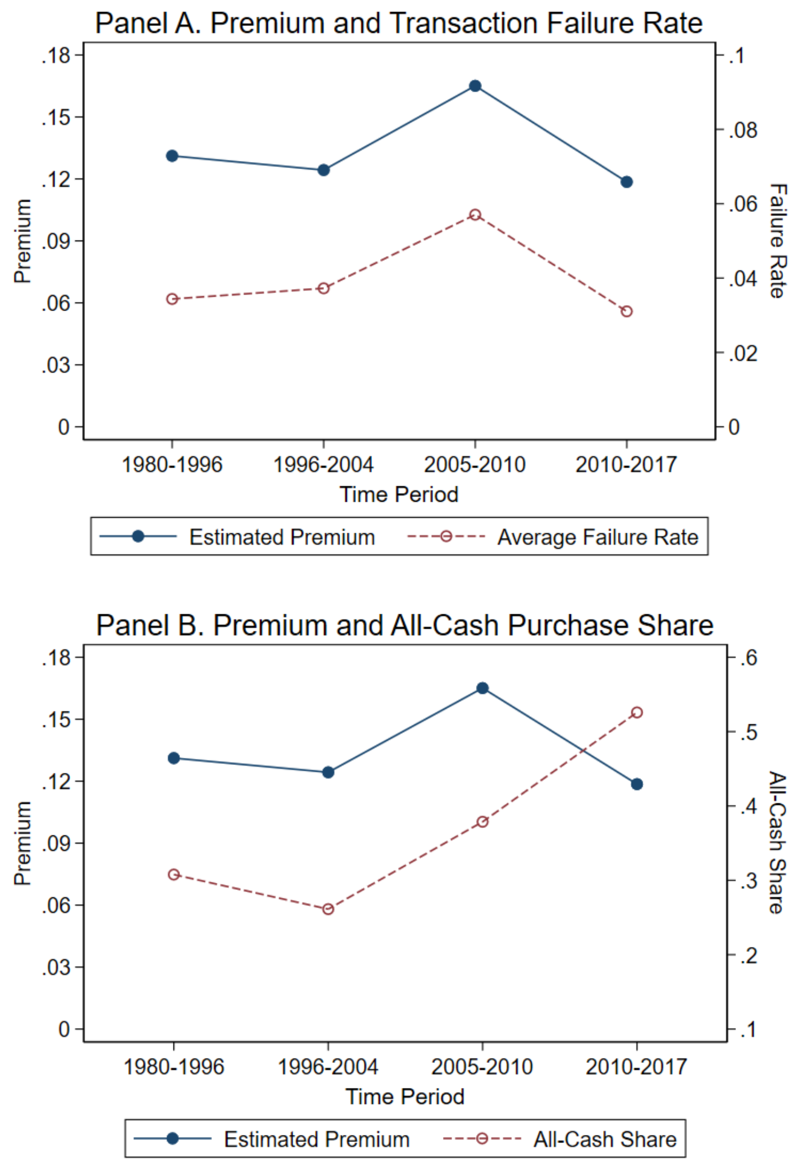


Figure 2. Empirical mortgage-cash premium over time. This figure plots the empirical mortgage cash premium over various time periods. The empirical premium is the value of μ that comes from estimating equation (28) on the subsample $\mathcal{T}$ consisting of months within the indicated time period,

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

where subscripts i and t index property and month; and the remaining terms are defined in the note to Table IV. Panel A plots the empirical premium with the mortgage application denial rate for pre-approved home purchase loans, shown on the right axis. Data on the denial rate are from HMDA and described in Section II and Section I of the Internet Appendix. Panel B plots the empirical premium with the share of purchases that are all-cash based on the ZTRAX data set. The remaining notes are the same as in Table IV.

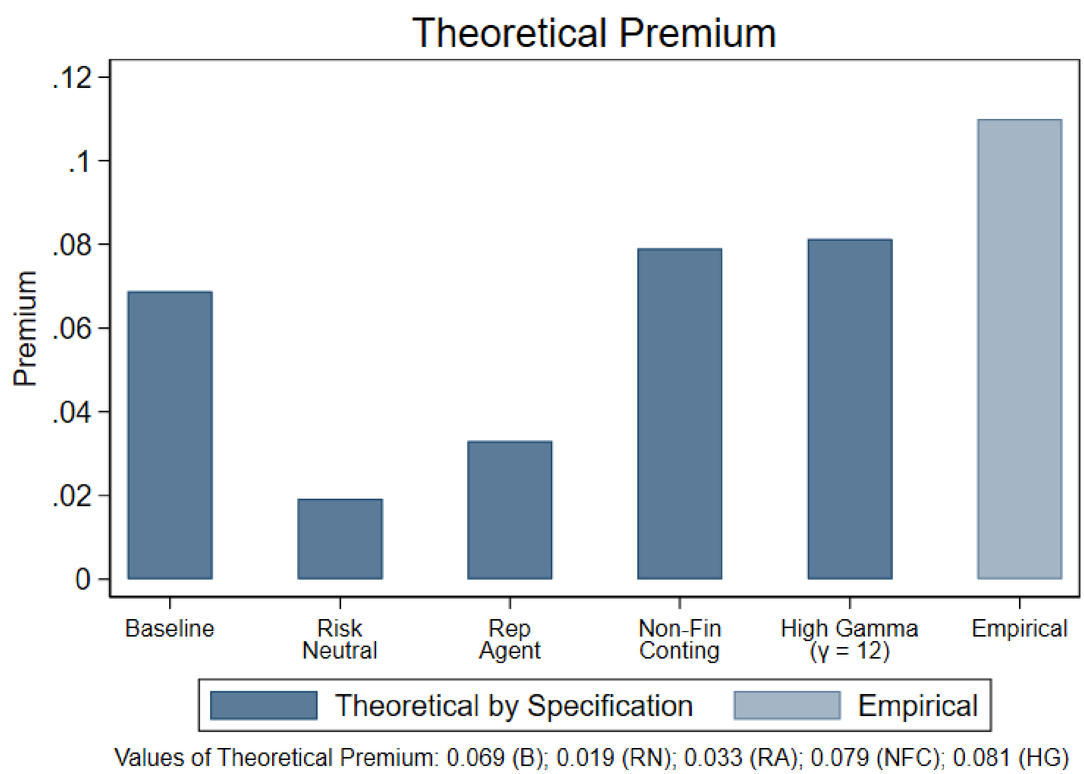


Figure 3. Summary of theoretical mortgage-cash premium. This figure summarizes calculations of the mortgage-cash premium, μ, from Proposition 1. "Baseline" denotes the average value of μ across the empirical distribution of parameters shown in Table VII and the preference parameters shown in that table. "Risk Neutral" denotes the analogous calculation when setting the coefficient of relative risk-aversion, γ, equal to zero. "Rep Agent" denotes the value of μ evaluated at the average value of each parameter, not the average of μ across the parameter distribution, and γ again equals its baseline value of 5. "Non-Fin Conting" denotes a similar calculation as in the baseline case after setting the share of mortgaged transactions with a home-sale contingency, h, equal to 7% and the probability of failure due to the home inspection contingency, q^c, equal to 1% (National Association of Realtors 2017, 2018). Equation (38) shows the combined failure rate with contingencies. "High Gamma" denotes the calculation when $\gamma = 12$ and, as in the baseline, μ is averaged across the parameter distribution. The rightmost bar marks 11%, corresponding to the empirical premium. Additional details are in Sections V.A and V.B.

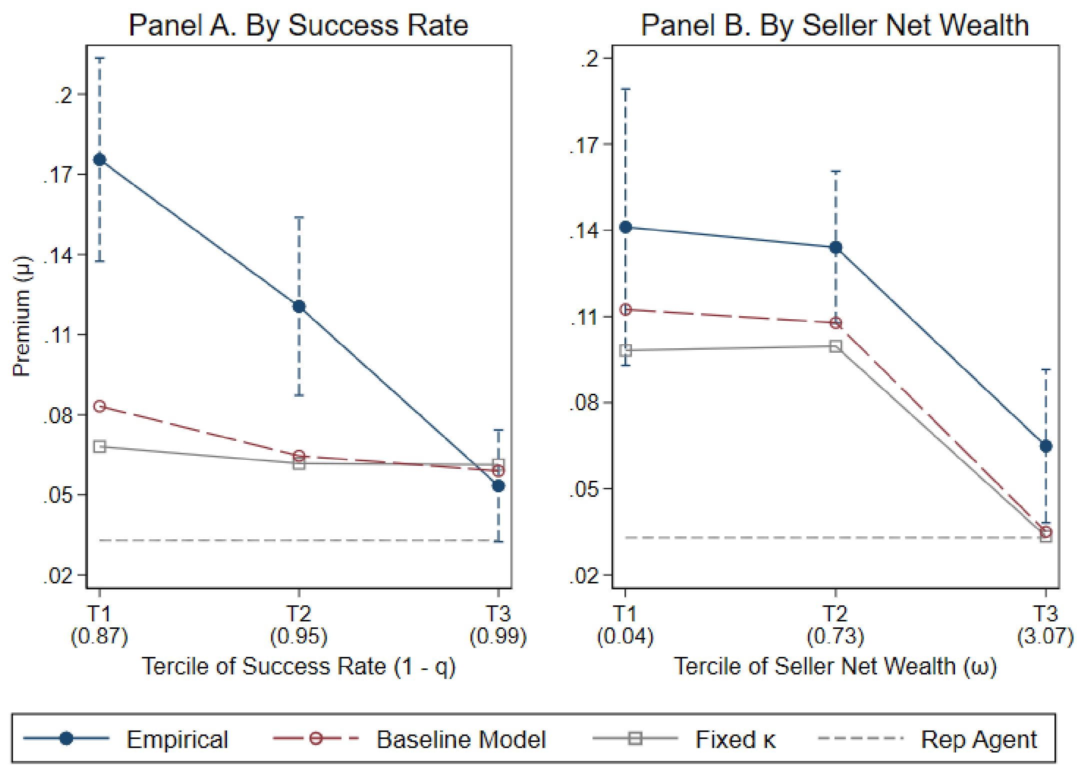


Figure 4. Heterogeneity in the empirical and theoretical mortgage-cash premium. This figure plots the empirical and theoretical premium across the model's parameters. The empirical premium is estimated as in equation (39),

$$\log\left(Price_{i,t}\right) = \hat{\mu}_0 Mortgaged_{i,t} + \sum_{\theta}\sum_{p} \hat{\mu}_{\theta,p}\left(Mortgaged_{i,t} \times Tercile(\hat{\theta}, p)_{z,t}\right) + \psi X_{i,t} + \zeta_{z(i),t} + \alpha_i + \epsilon_{i,t},$$

where $\hat{\theta}_{z,t}$ is an empirical measure for model parameter θ at the zip code-by-year level, $Tercile(\hat{\theta}, p)_{z,t}$ indicates whether $\hat{\theta}$ lies in the p^{th} tercile of the empirical distribution, and the remaining notes and notation are the same as in Table IV. The theoretical premium is calculated from the projection

$$\mu_g = \mu_0 + \sum_{\theta}\sum_{p} \mu_{\theta,p} Tercile(\theta, p)_g,$$

where g denotes grid points in the parameter distribution defined in Table VII. The coefficients $\{\mu_{\theta,p}\}$ are shown in Internet Appendix Table IA.VI. The red open circles show the theoretical premium under the baseline model (Baseline Model). The gray open squares show the premium for a naive model that fixes κ at its value from Table I (Fixed κ). The dashed lines shows the premium from the naive calibration in Table I that ignores heterogeneity (Rep Agent). The figure plots the expected value of the empirical and theoretical premium across terciles of θ_k, holding the other parameters θ_{-k} at their average. The parameters are one minus the probability of transaction failure $(1-q)$, the seller's total wealth net of current mortgage debt and a simultaneous down payment (ω), the monthly Poisson offer arrival rate, which also equals the expected number of offers (λ), and the ratio of the buyer's maximally levered liquid assets to housing value (L). Details on θ and $\hat{\theta}$ are in Sections V.A and VI. This figure shows the relationship between the premium and the parameters $1-q$ and ω. The relationship between the premium and the parameters λ and L are shown in Internet Appendix Figure IA.4. Brackets are 95% confidence intervals clustering by grid of the parameter distribution.

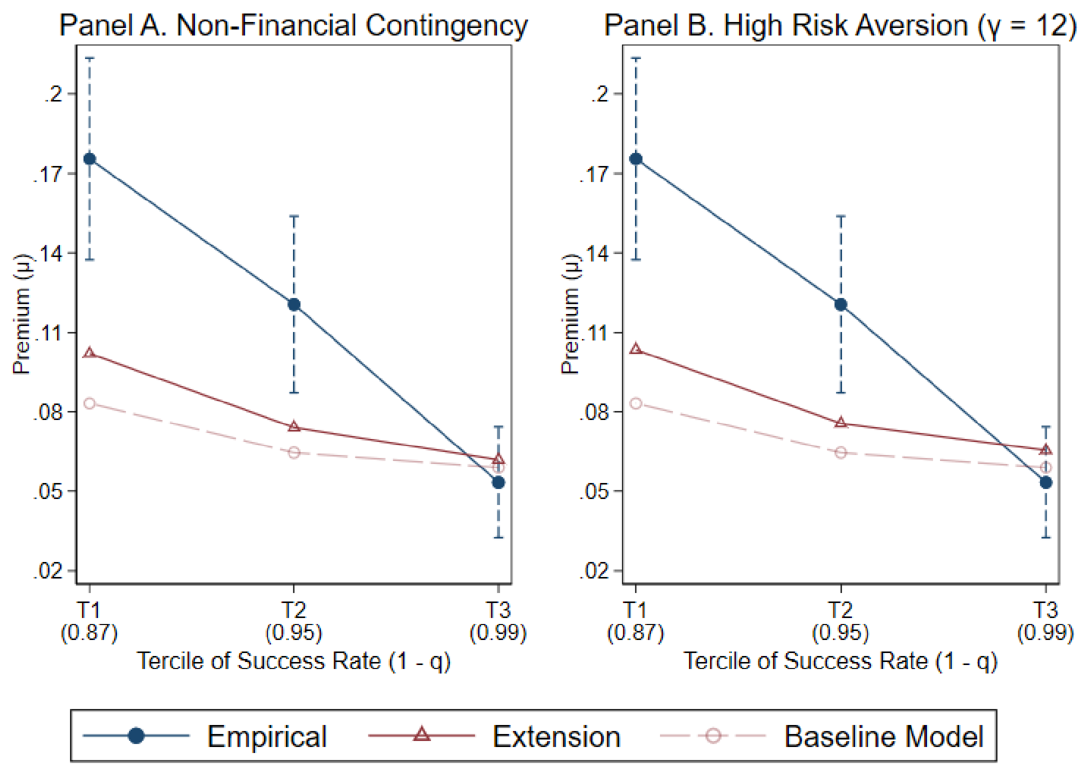


Figure 5. Heterogeneity with nonfinancial contingencies and high risk aversion. This figure is analogous to Figure 4 under two extensions of the baseline model. Panel A incorporates nonfinancial contingencies: both all-cash and mortgaged offers come with a home inspection contingency, and a share h of mortgaged offers also come with a home-sale contingency. The NAR RCI data set implies $h = 7\%$. The probability of failure due to the home inspection contingency is $q^c = 1\%$, also based on the NAR RCI data set. The probability of failure for a mortgaged offer with a home-sale contingency is shown in equation (38). Panel B uses a higher coefficient of relative risk-aversion, $\gamma = 12$. The remaining notes are the same as in Figure 4.

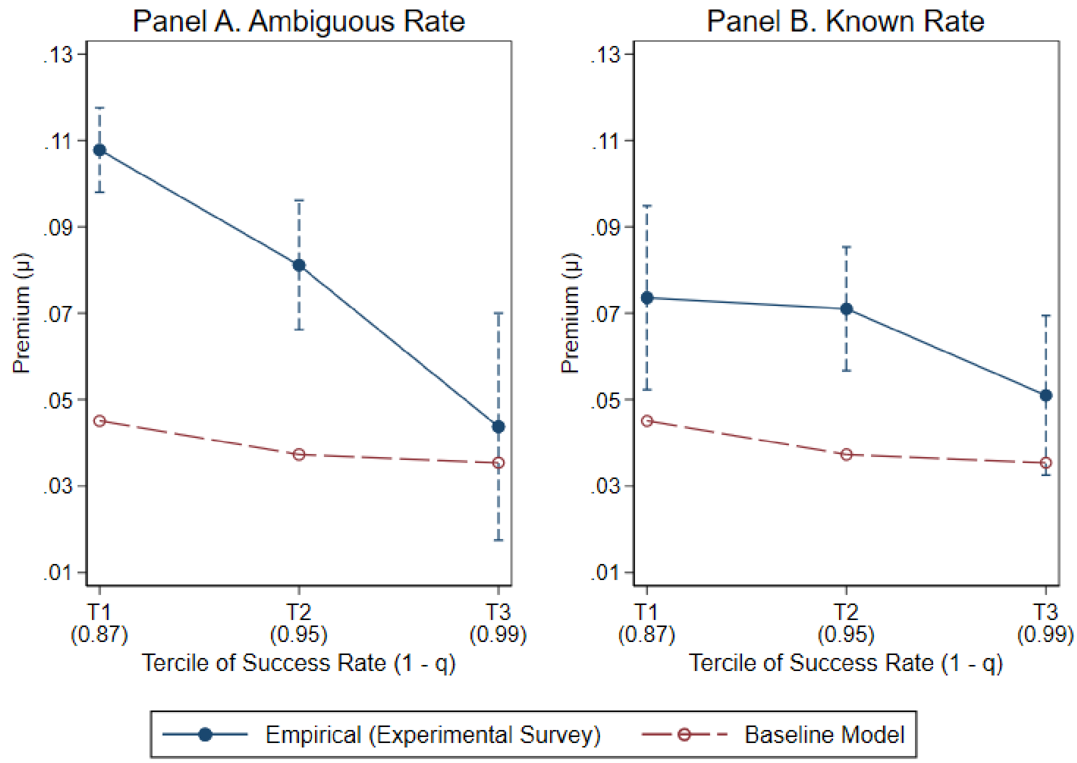


Figure 6. Heterogeneity in the experimental mortgage-cash premium. This figure plots the empirical and theoretical premium using data from the experimental survey, and it is analogous to Figure 4. The theoretical premium is calculated similarly to Figure 4. The empirical premium is estimated as in equation (40),

$$Premium_k = \sum_p \mu_{q,p} Tercile(\hat{q}, p)_k + \sum_p \Delta\mu^A_{q,p} \left(Tercile(\hat{q}, p)_k \times Ambiguity_k \right) + ...$$

$$... + \sum_\theta \sum_p \mu_{\theta,p} Tercile(\hat{\theta}, p)_k + \epsilon_k,$$

where k indexes survey respondent, $Premium_k$ denotes the respondent's mortgage-cash premium elicited from the multiple price list, $Ambiguity_k$ indicates whether the respondent is randomized into an experiment in which we do not provide the distribution of q, $Tercile(\hat{q}, p)_k$ indicates whether the respondent's probability of failure, $\hat{q}$ lies in the p^{th} tercile of the empirical distribution used in Figure 4, and we measure $\hat{q}$ using the value of q provided to the respondent or, for those facing ambiguity, the respondent's prior probability, which we elicit at the end of the survey. The vector of indicators $Tercile(\hat{\theta}, p)_k$ has a similar interpretation as in Figure 4. The left panel plots the estimated premium by q when facing ambiguity: $\mu_{q,p} + \Delta\mu^A_{q,p}$. The right panel does so when not facing ambiguity: $\mu_{q,p}$. Respondents are weighted using Census weights to correct for response bias. Standard errors are heteroskedasticity robust. The remaining notes are the same as in Figure 4 and Table VIII. Additional details are in Section VII.C.

Table I
Representative Seller Calibration of the Mortgage-Cash Premium

This table summarizes the notation of the baseline model in Section I and performs a stylized calibration of the mortgage-cash premium for a representative home seller. The endogenous variables are the mortgage-cash premium, μ, and the seller's endogenous cost of failure, κ, implied by this stylized calibration. The expressions for μ and κ are shown in Proposition 1 as a function of parameters $\Theta = \{\Theta_1; \Theta_2\}$. The elements in Θ_1 are exogenous and are used to calculate the values of the endogenous variables. These values equal the averages across the distribution of the empirical measures obtained from a more rigorous calibration, summarized in Table VII. The elements in Θ_2 are also exogenous and are externally calibrated. Under a low-friction parameterization in which $\xi = \ell_s = q^d = m = 0$, $\lambda = 20$, $\tilde{\delta} = 0.01$, and the other parameters are as in the table, we obtain $\mu = 0.009$ and $\kappa = 0.009$, as referenced in Section I.E. The model's setup and solution are in Section I.

Parameter	Notation	Value	Reference
Endogenous:			
Mortgage-Cash Premium	μ	0.033	Equation (19)
Seller Cost of Failure	κ	0.047	Equation (20)
Exogenous (Θ_1)*:*			
Probability of Transaction Failure	q	0.064	HMDA Denial
Offer Arrival Rate	λ	2.162	Offer-Level Data
Seller Financial Wealth-to-Housing	w	1.577	SCF Total Wealth
Seller Down Payment	ξ	0.052	Zillow, ZTRAX
Seller Loan-to-Value Ratio	ℓ_s	0.294	Zillow, ZTRAX
Buyer Liquid Assets-to-Housing	y	0.162	IRS Income
Loan-to-Value Constraint	$\bar{\ell}$	0.864	Zillow, ZTRAX
Probability of Transaction Delay	q^d	0.30	NAR Index
Mortgaged Offer Share	m	0.622	Offer-Level Data
Maintenance Cost-to-Value, Annualized	$\tilde{\delta}$	0.056	Zillow, ZORI
Preferences (Θ_2)*:*			
Coefficient of Relative Risk Aversion	γ	5	
Discount Rate, Annualized	ρ	0.04	
Borrower Loan Cost, Share of Par	D	1	

Table II.
Summary of Variables in the Empirical Analysis

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

Variable	Mean	Standard Deviation	Variable	Mean	Standard Deviation
Panel A: Transaction-Level Data Set					
$\textit{Real Price}_{i,t}$	\$416,813	\$776,315	$\textit{Mortgaged}_{i,t}$	0.642	0.479
$\textit{Age}_i$	28.431	6.682	$\textit{Rooms}_i$	1.409	1.492
$\textit{Bathrooms}_i$	0.237	0.715	$\textit{Stories}_i$	1.096	0.111
$\textit{Air Conditioning}_i$	0.217	0.239	$\textit{Detached}_i$	0.405	0.491
$\textit{Flip}_{i,t}$	0.115	0.319	$\textit{Foreign Buyer}_{i,t}$	0.003	0.054
$\textit{Same-County Buyer}_{i,t}$	0.009	0.093	$\textit{Institutional Buyer}_{i,t}$	0.001	0.038
$\textit{Cash Propensity}_{b(i,t)}$	0.252	0.434	$\textit{High Seller LTV}_{i,t}$	0.068	0.252
$\textit{Same-Month Seller Purchase}_{i,t}$	0.015	0.122	$\textit{Foreign Seller}_{i,t}$	0.002	0.041
$\textit{Cash Share}_{s(i,t)}$	0.179	0.327	$\textit{Number of Sales}_{s(i,t)}$	4.687	1.034
Panel B: Offer-Level Data Set					
$\textit{Real Price}_{i,j,t}$	\$512,662	\$359,480	$\textit{Mortgaged}_{i,j,t}$	0.609	0.488
$\textit{Winning}_{i,j,t}$	0.166	0.372	$\textit{Number of Offers}_{i,j,t}$	5.787	5.120
Baseline Number of Transactions: 426,256					
Baseline Number of Offers: 22,516					

Table III.
Characteristics of Mortgaged Transactions

P-values are in parentheses. This table correlates the probability that a transaction is mortgage-financed with market-level and transaction-level characteristics. Subscripts i, z, and t index property, zip code, and month. $Mortgaged_{i,t}$ indicates whether the loan amount is positive. Transaction-level variables are defined in Table II. Market-level variables are observed at the zip code-by-year level: $Mortgage\ Denial\ Rate_{z,t}$ denotes the mortgage application denial rate, using data from HMDA, $Transaction\ Volume_{z,t}$ is the total number of home purchase transactions in the zip code and year in the ZTRAX data set, $Average\ House\ Price_{z,t}$ is the FHFA All-Transactions house price index, and $Average\ Income_{z,t}$ comes from the IRS data set. Columns (3) and (4) include a zip code-by-month fixed effect and so exclude market-level variables. Transaction-level variables are included in columns (1) and (2) but not tabulated. The sample period is 1980 to 2017. Standard errors are clustered by property.

Outcome:	$Mortgaged_{i,t}$			
	Across Markets		Within Markets	
	(1)	(2)	(3)	(4)
Market Level:				
$Mortgage\ Denial\ Rate_{z,t}$	-0.318	-0.156		
	(0.018)	(0.017)		
$\log(Transaction\ Volume_{z,t})$	0.055	0.045		
	(0.001)	(0.003)		
$\log(Average\ House\ Price_{z,t})$	0.097	0.046		
	(0.002)	(0.004)		
$\log(Average\ Income_{z,t})$	0.003	0.008		
	(0.002)	(0.004)		
Transaction Level:				
$Institutional\ Buyer_{i,t}$			-0.494	-0.656
			(0.002)	(0.003)
$Foreign\ Buyer_{i,t}$			-0.125	-0.148
			(0.021)	(0.044)
$Flip_{i,t}$			-0.080	-0.086
			(0.002)	(0.002)
$High\ Seller\ LTV_{i,t}$			0.031	-0.134
			(0.002)	(0.002)
$Foreign\ Seller_{i,t}$			-0.050	-0.138
			(0.031)	(0.066)
$Institutional\ Seller_{i,t}$			0.057	0.037
			(0.002)	(0.003)
Month FE	Yes	Yes	No	No
Zip Code FE	No	Yes	No	No
Zip Code-Month FE	No	No	Yes	Yes
Property FE	No	No	No	Yes
R^2	0.300	0.461	0.628	0.873
Number of Observations	425,389	425,389	425,386	425,370

Table IV.
Empirical Mortgage-Cash Premium

P-values are in parentheses. This table estimates equation (28), 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

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

where observations are home purchases, $Mortgaged_{i,t}$ indicates whether 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) and (1) is 1980 to 2017. Columns (2) to (5) restrict the sample to months within the indicated time periods. Column (6) restricts the sample to properties built within the previous three years (New Homes). Column (7) restricts the sample to properties that were sold at least 12 months later (No Flips). Column (8) restricts the sample to properties in which neither the buyer nor seller is an institution (Non Instit.). Column (9) restricts the sample to properties in which neither the buyer nor seller has an address outside the U.S. (Non Foreign). Column (10) drops purchases in which the seller's LTV ratio is above 100% (Positive Equity), where the numerator is imputed using a straight-line amortization according to loan term and the denominator is imputed using the average sales price in the surrounding zip code and month to avoid mechanical correlation with $\log\left(Price_{i,t}\right)$. Standard errors are clustered by property. Data are from the ZTRAX data set.

Outcome:	$\log\left(Price_{i,t}\right)$										
	(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$Mortgaged_{i,t}$	0.161	0.117	0.131	0.124	0.165	0.119	0.090	0.088	0.112	0.108	0.099
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Sample Restriction	None	None	1980-1996	1996-2004	2005-2010	2010-2017	New Homes	No Flips	Non Instit.	Non Foreign	Positive Equity
Zip Code-Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Hedonic-Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Property FE	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
R^2	0.582	0.907	0.969	0.842	0.797	0.826	0.647	0.963	0.895	0.904	0.923
Number of Observations	2,254,389	426,256	27,087	122,683	69,307	20,792	6,651	186,570	333,288	323,956	313,370

Table V.
Summary of Estimates of the Mortgage-Cash Premium

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

	Empirical Mortgage-Cash Premium ($\hat{\mu}$)	
	Estimate	Table
Repeat-Sales-Hedonic, Baseline	0.117	Table IV, Column (1)
Repeat-Sales-Hedonic, 1980-1996	0.131	Table IV, Column (2)
Repeat-Sales-Hedonic, 1996 to 2004	0.124	Table IV, Column (3)
Repeat-Sales-Hedonic, 2005 to 2010	0.165	Table IV, Column (4)
Repeat-Sales-Hedonic, 2010 to 2017	0.119	Table IV, Column (5)
Repeat-Sales-Hedonic, New Homes	0.090	Table IV, Column (6)
Repeat-Sales-Hedonic, No Flips	0.088	Table IV, Column (7)
Repeat-Sales-Hedonic, Non Institutional	0.112	Table IV, Column (8)
Repeat-Sales-Hedonic, Non Foreign	0.108	Table IV, Column (9)
Repeat-Sales-Hedonic, Positive Equity	0.099	Table IV, Column (10)
Nonaccepted Offers	0.086	Table VI, Column (2)
Homeowner Survey, Wave 1	0.106	Table VIII, Column (2)
Homeowner Survey, Wave 2	0.104	Table VIII, Column (3)
Homeowner Survey, Wave 3	0.107	Table VIII, Column (4)
Hedonic with Nonrepeat Sales	0.161	Internet Appendix Table IA.I, Column (5)
Buyer and Seller Characteristics	0.129	Internet Appendix Table IA.II, Column (3)
IV, Regulatory Appraisal Floor	0.136	Internet Appendix Table IA.III, Column (2)
IV, Propensity to Sell All-Cash	0.139	Internet Appendix Table IA.IV, Column (4)
CoreLogic Data Set	0.122	Internet Appendix Table IA.I, Column (8)
Listing Characteristics	0.143	Internet Appendix Table IA.VII, Column (2)
Semi-Structural (Bajari et al. (2012))	0.149	Internet Appendix Table IA.VIII, Column (2)
Weighting by Representativeness	0.103	Internet Appendix Table IA.IX, Column (2)
Matching	0.169	Internet Appendix Table IA.XI, Column (3)

Table VI.
Robustness to Using Data on Nonaccepted Offers

P-values are in parentheses. This table estimates equation (35), which assesses whether the baseline results are robust to using data on nonaccepted offers to estimate the mortgage-cash premium. Subscripts i, j, and t index property, offer, and month. The regression equation is of the form

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

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

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

Table VII.
Distribution of Model Parameters

This table reports the distributions of the empirical measures of the model's parameters. Panel A reports the data source for directly calibrated parameters and the midpoint of each tercile. Probability of Failure (q) is measured as the mortgage application denial rate among pre-approved, first-lien, mortgages for the purchase of an owner-occupied, single-family home, based on the HMDA data set aggregated to the zip code-by-year level. Offer Arrival Rate (λ) is measured using the relationship $\mathbb{E}[N|N>0] = \frac{\lambda}{1-e^{-\lambda}}$. The number of offers received conditional on receiving an offer, $\mathbb{E}[N|N>0]$, is measured using the Offer-Level data set aggregated to the zip code level. Seller Wealth (w) is measured as the ratio of: the sum of checking and savings accounts, certificates of deposit, cash, stocks, savings bonds, other bonds, mutual funds, annuities, trusts, IRAs, employer-provided retirement plans, and total household income, divided by the value of the home. The unit of analysis is a household in the SCF data set, excluding nonhomeowners and first-time homeowners. Seller LTV (ℓ_s) is measured as the current LTV ratio, imputed from the ratio of mortgage balance implied by a straight-line amortization to the sale price. Seller Down Payment (ξ) is measured as the ratio of next-home price minus next-home mortgage to current-home price for sellers who purchase another home within 12 months. LTV Constraint ($\bar{\ell}$) is measured as the LTV ratio conditional on lying between the GSE regulatory limit and 100%. The calibration of ℓ_s, ξ, and $\bar{\ell}$ rely on the ZTRAX data set aggregated to the zip code-by-year level by taking the average. Buyer Liquid Assets (y) is measured as the average ratio of total adjusted gross income to total tax returns across zip codes and years, based on the IRS SOI data set. Probability of Delay (q^d) is measured as one minus the share of home sales not settled on time across months, based on the NAR RCI data set. Mortgaged Offer Share (m) is measured as the average share of offers that are mortgage-financed, based on the Offer-Level data set aggregated to the MSA level. Maintenance Cost ($\tilde{\delta}$) is measured as the average ratio of Zillow's ZORI rent index to Zillow's Home Value index across zip codes and months. The table reports the annualized $\tilde{\delta}$ that comes from multiplying by 12. Panel B reports the distribution of the composite parameters ω and L implied by Panel A. Details are in Section V.A.

Parameter	Source	Percentiles		
		15^{th}	50^{th}	85^{th}
Panel A: Directly Calibrated				
Probability of Failure (q)	HMDA Denial	0.01	0.047	0.133
Offer Arrival Rate (λ)	Offer-Level Data	0.871	2.232	3.383
Seller Financial Wealth (w)	SCF Total Wealth	0.338	1.025	3.369
Down Payment (ξ)	Zillow, ZTRAX	0	0.046	0.109
Seller LTV Ratio (ℓ_s)	Zillow, ZTRAX	0.042	0.248	0.593
Buyer Liquid Assets (y)	IRS Income	0.227	0.548	1.508
LTV Constraint ($\bar{\ell}$)	Zillow, ZTRAX	0.8	0.88	0.913
Probability of Delay (q^d)	NAR Index	0.24	0.3	0.36
Mortgaged Offer Share (m)	Offer-Level Data	0.354	0.639	0.872
Maintenance Cost ($\tilde{\delta}$)	Zillow, ZORI	0.035	0.054	0.079
Panel B: Composite Parameters				
Seller Net Wealth ($\omega = w - \ell_s - \xi$)		0.044	0.731	3.075
Buyer Leverage Capacity ($L = y/[1-\bar{\ell}]$)		0.498	1.22	2.335
Panel C: Baseline Preferences				
Coefficient of Relative Risk Aversion (γ): 5				
Discount Rate, Annualized (ρ): 0.04				
Borrower Loan Cost, Share of Par (D): 1				

Table VIII.
Summary of Survey Respondents

This table summarizes the mean of key variables from the experimental survey. Panel A reports the average mortgage-cash premium, elicited through the multiple price list method. The average is calculated (i) without weighting respondents (Unweighted), (ii) weighting by the respondent's representativeness in the 2020 Census according to age, income, home tenure, education, and state of residence (Census Weighted), (iii) after applying Census weights and restricting the sample to respondents with a single switch point between zero and the maximum price in list (Single Switchers), and (iv) after partitioning the single-switch sample according to whether the respondent is told the probability of mortgage transaction failure (Known q) or faces ambiguity (Ambiguous q). All respondents in the first wave face ambiguity. Panel B reports the average model parameters used to calculate the theoretical premium. For respondents facing an ambiguous failure probability, the theoretical premium is calculated using the respondent's subjective probability, elicited at the end of the survey. Similarly, the cost of failure for such respondents equals the subjective value elicited at the end of the survey, and otherwise equals 6%. The seller's financial wealth-to-housing is imputed using the respondent's income and a projection of wealth onto log income estimated in the 2016 SCF data set. Other model parameters are as in Table I. Panel (c) reports the average of demographic variables for the unweighted sample. Section VII and Internet Appendix Section IV contain details.

	Pooled Average	Average by Wave		
		First	Second	Third
	(1)	(2)	(3)	(4)
Panel A: Mortgage-Cash Premium (μ)				
Unweighted	0.105	0.106	0.104	0.107
Census Weighted	0.106	0.109	0.106	0.104
Single Switchers	0.096	0.109	0.092	0.087
Ambiguous q	0.103	0.109	0.096	0.096
Known q	0.084	.	0.087	0.081
Panel B: Model Parameters				
Probability of Failure (q)				
Subjective	0.13	0.135	0.121	0.13
Known	0.07	.	0.07	0.07
Seller Cost of Failure (κ)	0.051	0.036	0.056	0.06
Seller Wealth (w)	2.610	2.513	2.636	2.675
Seller LTV Ratio (ℓ_s)	0.3	0.3	0.3	0.3
Seller Down Payment (ξ)	0.15	0.15	0.15	0.15
Panel C: Additional Variables				
$\textit{Household Income}_k$	\$102,078	\$101,887	\$101,259	\$103,032
$\textit{Age}_k$	42.33	40.37	40.99	45.24
$\textit{College Educated}_k$	0.702	0.723	0.703	0.684
Number of Respondents in First Wave: 1,019				
Number of Respondents in Second Wave: 1,202				
Number of Respondents in Third Wave: 1,199				

Internet Appendix for
"The Mortgage-Cash Premium Puzzle"

Michael Reher and Rossen Valkanov*

This document contains additional material referenced in the text. Section I elaborates on our data description from Section II of the main article. Section II performs additional robustness exercises mentioned in the text. Section III provides additional details related to the framework from Section V of the main article. Section IV elaborates on the description of our survey in Section VII of the main article.

*Citation format: Reher, Michael and Rossen Valkanov, Internet Appendix for "The Mortgage-Cash Premium Puzzle," *Journal of Finance* [DOI STRING]. Please note: Wiley is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the authors of the article.

I. Data Appendix

We elaborate on the data description in Section I.II of the main article by providing additional details about the ZTRAX, offer-level, and additional data sets in Sections I.A, I.B, and I.C, respectively. We conclude with a catalog of the variables used in our analysis in Section I.D.

A. ZTRAX Data Set

The ZTRAX data set is divided into a transactions data set (ZTrans) and an assessment data set (ZAsmt). The former data set contains information on deed transfers, mortgages, and other real estate transactions. The latter data set contains information on property characteristics from tax assessments. The raw data derive from public records compiled by Zillow. In Zillow's words: "The Zillow Transaction and Assessment Data Set (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 12 months. We obtained the ZTRAX data set in June 2018 as a single download. For computational convenience, we work with a 25% random sample of the raw data set that we download. We shall simply refer to this random sample as the "raw ZTRAX data set."

The raw ZTRAX data set 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:

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

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

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

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

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

(vi) The transaction involves only a single property.

(vii) There is only a single vesting type.

(viii) The property is residential, which includes one-to-four family, condominium, and multifamily 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 (i), (ii), and (iv) 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 data set as the "filtered ZTRAX data set." The filtered ZTRAX data set includes 3,528,981 transactions. In terms of coverage, the filtered ZTRAX data set spans the 1980 to 2017 period and covers 80% of U.S. counties on a population-weighted basis, or 70% on an unweighted basis. The data set does not cover the following states: Nevada, Oklahoma, Ohio, North Carolina, New Jersey, or New York. The data set is most rich over the 1994 to 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 data set that can be used to estimate equation (28). 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 have information about the hedonic characteristics described below. We refer to the resulting data set as the "baseline ZTRAX data set." The baseline ZTRAX data set includes 426,256 transactions.

We rely on three sets of variables in the ZTRAX data set. 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 Section I.D below.

The second set of variables includes information about the buyer and seller in the transaction. We attach an identifier to each seller using the seller's name, and we attach a similar identifier to each buyer. For purchase transactions with multiple sellers or buyers, we define the seller or buyer using the party whose name is listed first. Before creating this identifier, we remove nonalphabetic characters from the name. So, a "seller" is defined as a unique string, and a "buyer" is defined similarly.[1] In particular, this methodology will under-classify unique sellers and unique buyers who have common names, or who do not appear first in the deed (e.g., a spouse). It will over classify unique sellers and unique buyers whose names are spelled differently in the data. We create a separate identifier for whether the seller (buyer) thus-classified is also a buyer (seller) thus-classified within a window around the sale (purchase). We also observe the address of the buyer and indicators for whether the buyer has a foreign address or is institutional. We observe the same information for sellers, although it is less well-populated than for buyers. Lastly, Zillow defines institutions somewhat loosely, as it includes trusts in its definition.

The third set of variables are hedonic characteristics, which come from the ZAsmt component of ZTRAX. We catalog these characteristics in Section I.D. The raw data come from tax assessment records, and we assign the characteristics for the most recent assessment to each transaction. We observe hedonic characteristics for 7% of the filtered ZTRAX data set. To

[1] Our approach follows the convention in the literature (e.g., Bayer, Mangum, and Roberts (2021)), although papers may differ in precisely how they pre-process raw names. We are not aware of a technique that can systematically correct measurement error due to common names or unusual name spelling.

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.

Note on Nonbanks. The ZTRAX data set does not inflate the share of cash-financed purchases by underrepresenting mortgages originated by nonbanks. Explicitly, the share of mortgages originated by nonbanks in our baseline sample equals 40%, compared to 41% based on the HMDA data set and the Gete and Reher (2021) definition of nonbanks. The share of mortgage-financed purchases based on the CoreLogic data set equals 65%, consistent with ZTRAX.

B. *Offer-Level Data Set*

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

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

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

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

(iii) The offer comes with a contingency such that the seller cannot recoup the offer price, to match the framework from Section V of the main article.

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

We cannot directly view the microdata out of concern for client privacy. Instead, we conduct our analysis remotely by submitting a program to Redfin's economic analysis division.

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

C. *Additional Data Sets*

C.1. CoreLogic

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

C.2. National Association of Realtors (NAR)

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

C.3. Home Mortgage Disclosure Act (HMDA)

We use data on mortgage application denial rates from the Home Mortgage Disclosure Act (HMDA) to parameterize the probability of transaction failure (q) in Section V.A of the main article. The denial rate is calculated conditional on applications for the purchase (i.e., not a refinance, not a HELOC) of an owner-occupied, one-to-four family home that have been pre-approved and that are for a first-lien. To avoid double-counting, we drop all applications for which the action taken is "purchase by institution" (Gete and Reher (2018)). To preserve data quality, we drop all applications with a nonempty edit status.

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

The time frequency of the HMDA data is annual. We observe Census tracts but not zip codes in HMDA, and so we aggregate the tract-level denial rate to the zip code level using the tract-to-zip code crosswalk from the Department of Housing and Urban Development, weighting tracts by the number of applications. We access the raw HMDA data through Recursion Co.

Whenever using the denial rate in regressions (i.e., Table III, Figures 4, IA.4, and 5), we first residualize it against the log of average applicant income and log of total credit requested in the zip code and year. This residualization ensures that the denial rate captures credit supply, not credit demand.

C.4. California Association of Realtors (CAR)

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

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

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

C.5. Survey of Consumer Finances

We use the Survey of Consumer Finances (SCF) to calibrate various model parameters related to wealth. The advantage of the SCF is that it contains a rich set of balance-sheet and other variables for a representative cross-section of U.S. households. The disadvantage is that the data are anonymized and consist of repeated cross-sections, not a panel. We use the 2016 SCF to calibrate the model. Bricker et al. (2017) provide a comprehensive review of the SCF data set.

C.6. Inflation and Zip Code Income

We use data on average adjusted gross income in 2010 from the IRS SOI Tax Stats to calculate real house price growth by zip code income in Figure 1. Average income is calculated as the ratio of aggregate adjusted gross income to total number of tax returns. The IRS SOI data set is observed at the zip code by year level. The data are not observed every year, and they begin after the start of our main analysis in 1994. In the calibration from Section V.A of the main article and the semi-structural estimation from Section VI.A of the main article, we interpolate missing years and we forward-fill and back-fill average income to obtain a balanced panel across zip codes and years.

Real house prices are in 2010 dollars and are calculated using CPI excluding shelter.

D. *Catalog of Variables*

ZTRAX Data Set

- *Mortgaged*$_{i,t}$: This variable indicates whether the loan amount associated with the purchase of property i in month t is positive.
- *Price*$_{i,t}$: This variable is the sales price associated with the purchase of property i in month t.
- *Age*$_i$: This variable is the number of years from when property i was built.
- *Rooms*$_i$: This variable is the number of overall rooms in property i.
- *Bathrooms*$_i$: This variable is the number of overall bathrooms in property i.
- *Stories*$_i$: This variable is the number of overall stories in property i.
- *Air Conditioning*$_i$: This variable indicates whether property i has air conditioning.
- *Detached*$_i$: This variable indicates whether property i is a detached single-family
- *Flip*$_{i,t}$: This variable indicates whether property i is sold within 12 months of its purchase in month t. We assign a value of zero to this indicator variable when the raw variable is unobserved.
- *Foreign Buyer*$_{i,t}$: This variable indicates whether the buyer of property i in month t has a foreign address, based on the indicator variable used by Zillow to flag such transactions. We assign a value of zero to this indicator variable when the raw variable is unobserved.
- *Same-County Buyer*$_{i,t}$: This variable indicates whether the buyer of property i in month t has an address in the same county as i. We assign a value of zero to this indicator variable when the raw variable is unobserved.
- *Institutional Buyer*$_{i,t}$: This variable indicates whether the buyer of property i in month t is an institution and does not intend to occupy the property, based on the two associated indicator variables used by Zillow to flag such transactions. We assign a value of zero to this indicator variable when the raw variable is unobserved.
- *Cash Propensity*$_{b(i,t)}$: This variable indicates whether the buyer of property i in month t, denoted $b(i,t)$, makes another purchase over our sample period in which the home is purchased all-cash.

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

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

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

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

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

Offer-Level Data Set

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

- *Winning*$_{i,j,t}$: This variable indicates whether 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 data set. We do not observe the offer price directly, but we observe the list price and the percent above or below price associated with the offer price. We use these two variables to calculate *Price*$_{i,j,t}$.

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

II. Empirical Appendix

We perform additional robustness exercises referenced in Section IV of the main article.

A. Controlling for Buyer and Seller Characteristics

First, since the seller's continuation value (κ) constitutes an omitted variable, we reestimate equation (28) after controlling for observed characteristics of the seller that plausibly describe her motivation to complete the transaction smoothly and quickly.[2] The seller characteristics include indicators for whether the seller has a LTV ratio that exceeds 50%, purchases another home in the same month, and has a non-U.S. address.

Then, to account for the possibility that all-cash buyers can identify motivated sellers, we also include indicators for whether the buyer flips the home within a year, has an address within the same county as the property, has a non-U.S. address, is an institutional buyer, and purchases another home all-cash over our sample period.

The results in Table IA.II imply a mortgage-cash premium between 10.0% and 12.9%. Notably, the buyer controls function similarly to the variable $Mortgaged_{i,j,t}$ from Table VI in that they absorb price-relevant characteristics of mortgaged buyers, such as the propensity to overpay in out-of-town markets (e.g., Chinco and Mayer (2016)).

B. Instrumental Variables: Appraisal Regulation

We construct an instrumental variable using the fact that federal law requires an appraisal on all bank-originated loans with a sales price above $250,000 (U.S. Code, 12 CFR §323.3). Federal regulators introduced this requirement in the wake of the savings and loan crises of the 1980s as part of the Financial Institutions Reform, Recovery, and Enforcement Act of 1989. They raised the threshold to $400,000 in 2019 (Federal Deposit Insurance Corporation (2019)), but this change will not affect our analysis because we use the ZTRAX data set for this exercise, which ends in 2016.

For background, an appraisal is an independent evaluation of a home's value, and it plays an important role in the loan underwriting process. In particular, the appraised value is used as the denominator of the LTV ratio. An appraisal is required in almost all mortgaged offers that exceed the aforementioned threshold, but all-cash offers rarely require an appraisal. Importantly, the appraisal occurs after the seller has already accepted the buyer's offer, and the appraised value does not necessarily equal the sales price. Since a loan often cannot be originated if the LTV exceeds certain values (e.g., Figure 1), the appraisal process introduces significant risk for home sellers. In fact, "appraisal issues" are the second most commonly cited reason for transaction failure or delay, after the closely related reason of "issues obtaining financing" (e.g., National

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

Association of Realtors (2016, 2018, 2020)).

In terms of our model, therefore, transactions with a price just above \$250,000 have a higher value of q relative to transactions with a price just below \$250,000. We would like to exploit this discontinuity by constructing an instrument equal to one if the sales price lies just above the \$250,000 threshold and zero otherwise. Of course, this naive approach would lead to inconsistent results because the outcome variable, $Price_{i,t}$, mechanically defines the instrument. Therefore, we adopt a strategy akin to Loutskina and Strahan (2015), who face a related econometric problem in constructing an instrument related to bunching of house prices around a regulatory threshold. In particular, we instrument for whether the predicted price in the current transaction exceeds the \$250,000 threshold using the previous transaction price on the property and growth in the FHFA house price index over the intervening period to predict the current price.

Explicitly, define the predicted price as

$$Predicted\ Price_{i,t} = Last\ Price_{i,t} \times \frac{FHFA\ Index_{z(i),t}}{Last\ FHFA\ Index_{z(i),t}} \tag{IA1}$$

where $Last\ Price_{i,t}$ denotes the price at which the home sold for in its most recent transaction before t, $FHFA\ Index_{i,t}$ denotes the FHFA All-Transaction Price Index in the associated zip code $z(i)$ and month t, and $Last\ FHFA\ Index_{z(i),t}$ denotes the analogous index in the month of the most recent transaction. Measuring price growth using the FHFA index provides an additional layer of orthogonality relative to, say, constructing a price index using the ZTRAX data set.

Next, define the instrumental variable $Appraisal\ Required_{i,t}$ as an indicator that equals one if $Predicted\ Price_{i,t}$ is greater than \$250,000 and equals zero otherwise. Column (1) of Table IA.III verifies the first stage by regressing $Price_{i,t}$ on $Appraisal\ Required_{i,t}$. Importantly, we also control for the log of $Predicted\ Price_{i,t}$, which ensures that the instrument works through a discontinuous jump in the predicted price, as desired, versus the almost-mechanical effect wherein a higher predicted price covaries with a higher actual price.

We estimate a negative coefficient on $Appraisal\ Required_{i,t}$, consistent with its theoretical foundation: if mortgaged buyers likely require an appraisal, then sellers are less likely to accept their offer. However, the relatively high p-value of 4% requires us to adjust our approach using techniques from the literature on weak instruments. Murray (2006), for example, proposes interacting the instrument with time fixed effects or other controls to improve its strength in the first stage. Accordingly, our first-stage regression equation is

$$\begin{aligned} Mortgaged_{i,t} = \sum_t \psi_t \left(Appraisal\ Required_{i,t} \times X_{i,t}\right) + \tilde{\chi} \log\left(Predicted\ Price_{i,t}\right) + ... \\ ... + \tilde{\zeta}_{z(i)} + \tilde{\tau}_t + \tilde{\epsilon}_{i,t}, \end{aligned} \tag{IA2}$$

where $X_{i,t}$ consists of the hedonic characteristics in Table IV and a vector of quarter fixed effects.

The second-stage regression equation is

$$\log\left(Price_{i,t}\right) = \mu \widehat{Mortgaged}_{i,t} + \chi \log\left(Predicted\ Price_{i,t}\right) + \zeta_{z(i)} + \tau_t + \epsilon_{i,t}, \tag{IA3}$$

where $\widehat{Mortgaged}_{i,t}$ denotes the fitted value from equation (IA2). We estimate equation (IA3) using the subsample of home purchases with a predicted price within a fairly narrow bandwidth of \$250,000. This restriction improves the validity of the instrument set, since comparing homes with predicted prices of, say, \$100,000 and \$1,000,000 would likely confound channels apart from the desired regulatory discontinuity. The resulting drop in sample size requires us to reduce the granularity of the regression's fixed effects, which we view as a worthy tradeoff given that we now have a very transparent source of identifying variation.

Columns (2) to (4) of Table IA.III estimate a mortgage-cash premium of 13.6%, 13.7%, and 14.2% using a 5%, 10%, and 15% bandwidth, respectively. As the bandwidth and thus sample size increases, the first stage F-statistic rises comfortably above the recommendations of Stock and Yogo (2005).

Concerning internal validity, it is difficult to see how this instrument set would fail to satisfy the exclusion restriction. The results are unlikely driven by mechanical correlation between the sales price and the price index, since the FHFA index is constructed using proprietary information held by the Government Sponsored Enterprises, not using deeds records like the ZTRAX data set. Moreover, the results are unlikely to be driven by "motivated sellers" (i.e., high κ), who both list at a low price to attract more offers and also only consider all-cash offers. In particular, the use of the home's previous price to construct the instrument implies that both motivated and ordinary sellers should have similar predicted prices, and, thus the same values of $\textit{Appraisal Required}_{i,t}$. If anything, motivated sellers would bias the estimates towards zero: by listing low, they make an appraisal less likely and thus are actually more likely to sell to a mortgaged buyer. Lastly, the statistical insignificance of Hansen's J-statistic supports the internal validity of the instrument set.

These instrumental variable estimates lie close to our baseline estimate of 11.7%, which supports the internal validity of an average mortgage-cash premium around that value. In this exercise, however, we have identified the premium using a highly transparent source of variation with a clear theoretical foundation within our model (i.e., an increase in transaction risk, q). Thus, these results support not only the validity of the empirical analysis but also the relevance of the model.

C. Instrumental Variables: Seller Cash Preference

The second instrumental variable is $\textit{Cash Share}_{s(i,t)}$, defined as the share of homes sold by the seller over our sample period that are to cash buyers, excluding the sale in question. This instrument is a " leave-one-out-mean," as commonly used in the labor and development literatures (e.g., Townsend (1994)). Since $\textit{Cash Share}_{s(i,t)}$ reflects the seller's persistent preference for cash financing, we can now identify the mortgage-cash premium even if the method of financing covaries with temporary shocks that jointly affect the listing price and the seller's preference for cash. In particular, we identify the premium through structural characteristics of the seller, such as aversion to ambiguity about the probability of mortgage transaction failure (Section VII of the main article).

Columns (1) and (2) of Table IA.IV substantiate the previous paragraph. We verify the first stage in column (1), which shows that sellers with a high value of $\textit{Cash Share}_{s(i,t)}$ are less likely to sell to mortgaged buyers. The first stage is strong by the Stock and Yogo (2005) criteria. In column (2), we follow other papers in the literature (e.g., D'Acunto and Rossi (2020)) and evaluate the instrument's exclusion restriction by reestimating our baseline regression equation

(28) after including $Cash\ Share_{s(i,t)}$ as a control. We obtain a highly insignificant coefficient on $Cash\ Share_{s(i,t)}$. This finding supports the instrument's validity by showing that it has no effect on a transaction's sales price once controlling for the method of financing.

Columns (3) and (4) of Table IA.IV summarize the second stage results. We estimate a mortgage-cash premium of 13.9% in column (4). In fact, the estimate is so similar to its counterpart in the baseline Table IV that the Durbin-Wu-Hausman (DWH) test fails to reject the null hypothesis that our baseline identification assumption (29) is valid.

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

D. Property Condition

We pursue the three exercises referenced in Section IV.D of the main article.

D.1. Controlling for List Price and Time-on-Market

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

D.2. Semi-Structural Hedonic Estimator

Recall from Section III of the main article that the repeat sales and hedonic pricing 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. 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 (28), Markovian evolution of the unobserved attributes, and rational expectations of market participants.

Building on equation (28), consider a property which transacts in months t and $t{+}n$. To minimize notation and, more substantively, to avoid the incidental parameters problem, we first residualize the variables $\log(Price_{i,t})$ and $Mortgaged_{i,t}$ against the property and zip code-month fixed effects in our repeat-sales methodology.[3] This step is without loss of generality in terms of

[3] Explicitly, the residualized variables are $\overline{\log(Price_{i,t})} = \log(Price_{i,t}) - \zeta^{P}_{z(i),t} + \alpha^{P}_{i}$ and $\overline{Mortgaged_{i,t}} = Mortgaged_{i,t} - \zeta^{M}_{z(i),t} + \alpha^{M}_{i}$. In words, $\overline{\log(Price_{i,t})}$ and $\overline{Mortgaged_{i,t}}$ are the residuals from a regression of

obtaining point estimates that are comparable to those in Table IV, per the Frisch-Waugh-Lovell theorem. In particular, the associated pricing kernel is

$$\log\left(Price_{i,t}\right) = \mu Mortgaged_{i,t} + \alpha + \epsilon_{i,t}. \tag{IA4}$$

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

$$\epsilon_{i,t+n} = \varrho_t\left(n\right)\epsilon_{i,t} + \omega_{i,t+n}. \tag{IA5}$$

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 (IA5) introduces some structure in how these attributes evolve over time.

Finally, suppose market participants rely on rational expectations such that

$$\mathbb{E}_t\left[\omega_{i,t+n}\right] = 0, \tag{IA6}$$

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 (IA6) 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 (IA5) and (IA6) provide a moment condition we can use to recover the mortgage-cash premium, μ. Substituting equation (IA5) into equation (IA4) in month $t{+}n$ gives

$$\begin{aligned} \log\left(Price_{i,t+n}\right) &= \mu Mortgaged_{i,t+n} + \alpha + \ldots \\ &\ldots + \varrho_t\left(n\right)\left[\log\left(Price_{i,t}\right) - \mu Mortgaged_{i,t} - \alpha\right] + \omega_{i,t+n}. \end{aligned} \tag{IA7}$$

Based on the rational expectations assumption in equation (IA6), all regressors in equation (IA7) 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

$$Mortgaged_{i,t+n} = \bar{\varphi} + \varphi_t^M\left(n\right)Mortgaged_{i,t} + \varphi_t^P\left(n\right)\log\left(Price_{i,t}\right) + \nu_{i,t+n}, \tag{IA8}$$

where, again making use of rational expectations, $\mathbb{E}_t\left[\nu_{i,t+n}\right] = 0$. In practical terms, we estimate equations (IA7) and (IA8) 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.

Table IA.VIII reports the results. We estimate a mortgage-cash premium of 14.9% using this semi-structural approach, per the result in column (2). Since the sample is necessarily restricted to properties that transact more than twice, we facilitate comparison with the repeat sales and

$\log\left(Price_{i,t}\right)$ 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.

hedonic results in Table IV by reestimating equation (28) on this subsample.[4] This results in a similar mortgage-cash premium of 10.9%, shown in column (1).

Along with the instrumental variable results and the results based on nonaccepted offers from Section IV.A of the main article, the semi-structural results from this subsection are quantitatively similar to the estimated mortgage-cash premium of 11% from Table IV. This similarity again supports the validity of the repeat sales and hedonic pricing methodology.

D.3. Nonparametric Matching Estimator

Estimating the mortgage-cash premium through propensity score matching, a nonparametric approach, avoids bias from the linear specification in equation (28). We note that this matching estimator primarily assesses bias due to a nonlinear functional form, as its focus on observed characteristics makes it ill-suited to assess bias due to property condition.

Our implementation follows Harding, Rosenblatt, and Yao (2012), who use this approach to estimate the foreclosure discount. 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 IV and the seller controls from Table IA.IV. 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 IA.XI 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 IA.XI. 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.

E. External Validity

We perform the three tests of external validity mentioned in the text. First, we estimate the mortgage-cash premium on a broader sample of properties, including those without a repeat sale. Second, we estimate the premium weighting by the inverse probability of appearing in the

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

baseline sample. Third, we cross-reference our results with contemporaneous papers studying all-cash home purchases.

E.1. Properties without a Repeat Sale

In column (1) of Table IA.I, we estimate a premium of 18.6% using the 11,367,195 transactions in the ZTRAX universe that satisfy the basic filters in Section I.A. In column (3), we estimate a premium of 12.6%, using the 3,911,805 transactions in the subset of the ZTRAX universe that occur in the zip code-by-month bins that appear in the 25% random sample of the ZTRAX universe that we study in our main analysis, which we call the "filtered ZTRAX data set" in Section I. In column (5), we estimate a premium 16.1%, using the 2,254,389 transactions in the filtered ZTRAX data set without including a property fixed effect. In column (6), we include a property fixed effect, so that we replicate the specification from column (1) of Table IV. Note that the R^2 equals 58% in column (5), versus 91% in column (6). This finding suggests both that the property fixed effects in equation (28) reduce bias and that the accompanying sample restriction does not jeopardize external validity.

E.2. CoreLogic Data Set

We estimate a premium of 12.2% in column (8) of Table IA.I after applying our repeat sales and hedonic pricing estimator to the CoreLogic data set. As described in Section I.C, the CoreLogic data set draws from public records, but it is compiled by an entirely separate vendor. Thus, the similarity of the results in Table IA.I strongly supports the external validity of the estimates obtained from our baseline ZTRAX data set.

E.3. Inverse Probability Weighting

In Table IA.IX, we reestimate equation (28) after weighting transactions by their inverse probability of appearing in the baseline sample from Table IV, based on observed characteristics (e.g., Solon, Haider, and Wooldridge (2015)). We obtain the weights through a logistic regression estimated on the filtered ZTRAX data set, shown in column (1) of Table IA.IX. The weighted point estimate equals 10.3%, shown in column (2).

E.4. Consistency with the Literature

Two contemporaneous papers have studied the price differential between mortgage-financed and cash-financed home purchases. First, using data on purchases in the Los Angeles MSA between 1999 to 2017, Han and Hong (2023) estimate a lower premium of 5%. In column (1) of Table IA.X, 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 (2023).[5] This finding is consistent with their lower estimated premium, which lends external support to our results.

[5]This exercise is based on an earlier version of Han and Hong (2023) written before January 2023, which only included microdata for Los Angeles. In the most recent version as of the time of this writing, Han and Hong (2023) separately estimate the mortgage-cash premium using microdata from the largest 100 U.S. counties,

Second, using data on purchases in the Phoenix, Las Vegas, Dallas, and Orlando MSAs and in Gwinnet County, GA between 2013 to 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.[6] Since mortgage-financed purchases account for 17% of all purchases in the sample studied by Buchak et al. (2020), as shown in Table IA.X, 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 Table IA.X, 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 (2023), this similarity supports our results.

Lastly, we corroborate the remark in footnote 6 regarding the option value of waiting in Han and Hong (2023). Preserving the notation of Han and Hong (2023) but with HH subscripts, a mortgaged transaction succeeds with probability q_{HH}. With probability ω_{HH}, a mortgaged offer arrives that the seller accepts. The value function for a seller with a mortgaged offer is

$$V^{\mathcal{S}}(M, P_M) = q_{HH} e^{-r} P_M + (1 - q_{HH}) e^{-r} \left[\omega_{HH} V^{\mathcal{S}}(M, P_M) + (1 - \omega_{HH}) V^{\mathcal{S}}(\emptyset) \right] \tag{IA9}$$

using the Han and Hong (2023) assumptions that: mortgaged buyers always offer the same price P_M, if the first offer that a seller accepts is mortgaged, she only accepts mortgaged offers thereafter, and sellers have linear preferences over the price received. For convenience, we again define κ such that $V^{\mathcal{S}}(\emptyset) = P_M e^{-\kappa}$. So, the value function becomes

$$V^{\mathcal{S}}(M, P_M) = e^{-r} P_M \frac{q_{HH} + (1 - \omega_{HH})(1 - q_{HH}) e^{-\kappa}}{1 - \omega_{HH}(1 - q_{HH}) e^{-r}}, \tag{IA10}$$

which is the same as in Han and Hong (2023) if $\kappa \to \infty$. An all-cash offer closes in one month, so that the value function for a seller with an all-cash offer is

$$V^{\mathcal{S}}(C, P_C) = P_C e^{-r}. \tag{IA11}$$

The mortgage-cash premium, δ_{HH} in the notation of Han and Hong (2023), solves

excluding Los Angeles, finding a larger premium of 12.4% in their most well-controlled specification.

[6] 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.

$V^{\mathcal{S}}(C, P_C) = V^{\mathcal{S}}(M, P_M)$. Rearranging and solving for δ_{HH} gives

$$\delta_{HH} = 1 - \frac{P_C}{P_M} = 1 - \frac{q_{HH} + (1 - \omega_{HH})(1 - q_{HH})e^{-\kappa}}{1 - \omega_{HH}(1 - q_{HH})e^{-r}}. \tag{IA12}$$

Notice that the mortgage-cash premium is decreasing in the term $e^{-\kappa}$, which reflects the seller's option value from having no offer. This term equals zero in Han and Hong (2023), corroborating the statement from footnote 6 that omitting it inflates the calibrated mortgage-cash premium. Under the Han and Hong (2023) parameterization of $q_{HH} = 15.9\%$, $\omega_{HH} = 55\%$, $r = 0.05/12$, and $e^{-\kappa} = 0$, the mortgage-cash premium from equation (IA12) equals 7.88%. When $q_{HH} = 4.7\%$, the median in Table VII, the premium equals 2.18% as in the introduction.

III. Model Appendix

Section III.A contains proofs of the paper's main analytic results. Section III.B provides details on model extensions and setup. Section III.C states and proves additional analytic results. Section III.D provides details on the calibration of the model and the semi-structural estimation of it. Section III.E contains details on the belief distortions analyzed in Sections VII.C and VII.D.

A. *Proofs of Main Analytic Results*

This section proves the main analytic results stated in the text: Proposition 1, Lemma 1, and Corollary 1.

A.1. Proof of Proposition 1 (Deriving the Mortgage-Cash Premium)

Using Definition 1, the mortgage-cash premium solves $V^{\mathcal{S}}(C, vp^C) = V^{\mathcal{S}}(M, vp^C e^{\mu})$. Explicitly

$$\frac{\left(\omega + p^C\right)^{1-\gamma}}{1-\gamma} = -\delta\left(e^{-\alpha} + q^d\right) + e^{-\alpha-\rho}\left[(1-q)\frac{\left(\omega + p^C e^{\mu}\right)^{1-\gamma}}{1-\gamma} + q\frac{\left(\omega + e^{-\kappa}\right)^{1-\gamma}}{1-\gamma}\right], \tag{IA1}$$

following equations (4) to (5) and (7). Multiplying by $\frac{(1-\gamma)e^{\alpha+\rho}}{1-q}$ and rearranging

$$\left(\omega + p^C e^{\mu}\right)^{1-\gamma} = \left[(1-\gamma)\left(\frac{(e^{\alpha+\rho}-1)u_C + \Gamma\delta + u_C - u_\kappa}{1-q} + u_\kappa\right)\right], \tag{IA2}$$

with

$$u_C \equiv \frac{\left(\omega + p^C\right)^{1-\gamma}}{1-\gamma} \tag{IA3}$$

$$u_\kappa \equiv \frac{\left(\omega + e^{-\kappa}\right)^{1-\gamma}}{1-\gamma} \tag{IA4}$$

$$\Gamma \equiv (1 + q^d e^{\alpha})e^{\rho}. \tag{IA5}$$

Solving for μ gives

$$\mu = \log\left(\left[(1-\gamma)\left(\frac{(e^{\alpha+\rho}-1)u_C + \Gamma\delta + u_C - u_\kappa}{1-q} + u_\kappa\right)\right]^{\frac{1}{1-\gamma}} - \omega\right) - \log(p^C), \tag{IA6}$$

which is what needed to be shown.

A.2. Proof of Lemma 1 (Deriving the Seller's Cost of Failure)

Use Lemma IA.3, Lemma IA.4, and equation (7) to rewrite equation (6). Consider the case $\bar{p}_b(M) \geq \overline{p_M}$ from Lemma IA.4. Equation (6) becomes

$$\begin{aligned} V^{\mathcal{S}}(\emptyset) &= -\delta + ... \\ ... & \; e^{-\rho}\varphi_1(1-m)\left[u(\omega + e^{-\kappa})\right] + ... \\ ... & \; e^{-\rho}\varphi_1 m e^{-\alpha-\rho}\left[(1-q)u(\omega + e^{\mu_\kappa - \kappa})\right] + \\ ... & \; e^{-\rho}\Phi_C\left[u(\omega + \bar{p}_b(C))\right] + ... \\ ... & \; e^{-\rho}\Phi_M(1-q)e^{-\alpha-\rho}\left[(1-q)u(\omega + \bar{p}_b(M))\right] + ... \\ ... & \; e^{-\rho}((\Phi_M + \varphi_1 m)(e^{-\alpha} + q^d))\left[-\delta\right] + ... \\ ... & \; e^{-\rho}((\Phi_M + \varphi_1 m)qe^{-\alpha-\rho} + \varphi_0)\left[V^{\mathcal{S}}(\emptyset)\right]. \end{aligned} \tag{IA7}$$

Recognize that, by definition, μ_κ solves

$$u(\omega + e^{-\kappa}) = -\delta(e^{-\alpha} + q^d) + e^{-\alpha-\rho}[(1-q)u(\omega + e^{\mu_\kappa - \kappa}) + qV^{\mathcal{S}}(\emptyset)]. \tag{IA8}$$

So, equation (IA7) becomes

$$\begin{aligned} V^{\mathcal{S}}(\emptyset) &= -\delta + ... \\ ... & \; e^{-\rho}\varphi_1\left[u(\omega + e^{-\kappa})\right] + ... \\ ... & \; e^{-\rho}\Phi_C\left[u(\omega + \bar{p}_b(C))\right] + ... \\ ... & \; e^{-\rho}\Phi_M(1-q)e^{-\alpha-\rho}\left[u(\omega + \bar{p}_b(M))\right] + ... \\ ... & \; e^{-\rho}(\Phi_M(e^{-\alpha} + q^d))\left[-\delta\right] + ... \\ ... & \; e^{-\rho}(\Phi_M qe^{-\alpha-\rho} + \varphi_0)\left[V^{\mathcal{S}}(\emptyset)\right]. \end{aligned}$$

We now iterate the previous equation forward and solve the geometric sequence. Doing so yields

$$\begin{aligned} V^{\mathcal{S}}(\emptyset) &= \left[e^{\rho} - (\Phi_M qe^{-\alpha-\rho} + \varphi_0)\right]^{-1} \times ... \\ ... & \; \left[\varphi_1\left[u(\omega + e^{-\kappa})\right] + ...\right. \\ ... & \; \Phi_C\left[u(\omega + \bar{p}_b(C))\right] + ... \\ ... & \; \Phi_M(1-q)e^{-\alpha-\rho}\left[u(\omega + \bar{p}_b(M))\right] + ... \\ ... & \; \left.(e^{\rho} + \Phi_M(e^{-\alpha} + q^d))\left[-\delta\right]\right]. \end{aligned}$$

Obtain κ by solving

$$\begin{aligned} u(\omega + ve^{-\kappa}) &= \left[e^{\rho} - (\Phi_M qe^{-\alpha-\rho} + \varphi_0)\right]^{-1} \times ... \\ ... & \; \left[\varphi_1\left[u(\omega + e^{-\kappa})\right] + ...\right. \\ ... & \; \Phi_C\left[u(\omega + \bar{p}_b(C))\right] + ... \\ ... & \; \Phi_M(1-q)e^{-\alpha-\rho}\left[u(\omega + \bar{p}_b(M))\right] + ... \\ ... & \; \left.(e^{\rho} + \Phi_M(e^{-\alpha} + q^d))\left[-\delta\right]\right]. \end{aligned}$$

so that

$$u(\omega+e^{-\kappa}) = \left[e^{\rho}-(\Phi_M q e^{-\alpha-\rho}+\varphi_0+\varphi_1)\right]^{-1}\times ... \tag{IA9}$$
$$... \left[\Phi_C\left[u(\omega+\bar{p}_b(C))\right]+...\right.$$
$$... \Phi_M(1-q)e^{-\alpha-\rho}\left[u(\omega+\bar{p}_b(M))\right]+...$$
$$\left. ... (e^{\rho}+\Phi_M(e^{-\alpha}+q^d))\left[-\delta\right]\right]$$

and, given CRRA indirect utility,

$$\kappa=-\log\left(\left[\frac{\Phi_C(\omega+\bar{p}_b(C))^{1-\gamma}+\Phi_M(1-q)e^{-\alpha-\rho}(\omega+\bar{p}_b(M))^{1-\gamma}+(\gamma-1)\delta\Xi}{e^{\rho}-(\Phi_M q e^{-\alpha-\rho}+\varphi_0+\varphi_1)}\right]^{\frac{1}{1-\gamma}}-\omega\right), \tag{IA10}$$

with

$$\Xi = e^{\rho}+\Phi_M(e^{-\alpha}+q^d) \tag{IA11}$$
$$= e^{\rho}+\Phi_M(1+q^d e^{-\rho}),$$

and, as in Lemma IA.4, $\varphi_0+\varphi_1=e^{-\lambda}(1+\lambda)$. After substituting the terms $\Pr\left[u=V^{\mathcal{S}}(\emptyset)\right]$ and $\mathbb{E}\left[u|u>V^{\mathcal{S}}(\emptyset)\right]\Pr\left[u>V^{\mathcal{S}}(\emptyset)\right]$ defined in the statement of Lemma 1, equation (IA10) matches equation (20). This is what needed to be shown for the case $\bar{p}_b(M)\geq\overline{p_M}$

Next consider the case $\bar{p}_b(M)<\overline{p_M}$. Now equation (IA7) becomes

$$V^{\mathcal{S}}(\emptyset) = -\delta+...$$
$$... e^{-\rho}\varphi_1(1-m)\left[u(\omega+e^{-\kappa})\right]+...$$
$$... e^{-\rho}\Phi_C\left[u(\omega+\bar{p}_b(C))\right]+...$$
$$... e^{-\rho}(\varphi_{MM}+\varphi_1 m)\left[-\delta\right]+...$$
$$... e^{-\rho}(\varphi_{MM}+\varphi_1 m+\varphi_0)\left[V^{\mathcal{S}}(\emptyset)\right].$$

so that

$$\kappa=-\log\left(\left[\frac{\Phi_C(\omega+\bar{p}_b(C))^{1-\gamma}+\delta e^{\rho}(\gamma-1)}{e^{\rho}-(\varphi_{MM}+\varphi_0+\varphi_1)}\right]^{\frac{1}{1-\gamma}}-\omega\right), \tag{IA12}$$

which matches equation (20) after substituting the terms Ξ, $\mathbb{E}\left[u|u>V^{\mathcal{S}}(\emptyset)\right]\Pr\left[u>V^{\mathcal{S}}(\emptyset)\right]$, and $\Pr\left[u=V^{\mathcal{S}}(\emptyset)\right]$ defined in the statement of Lemma 1. This is what needed to be shown for the case $\bar{p}_b(M)<\overline{p_M}$, which completes the proof.

A.3. Proof of Corollary 1 (Identifying the Mortgage-Cash Premium)

The proof relies on Lemma IA.3 from Section III.C. Using the notation of the corollary and the results from Lemma IA.3,

$$Price_{i,b,s} = \begin{cases} ve^{-\kappa} & \text{if } \eta = 0, Mortgaged_b = 0 \\ ve^{\mu_\kappa - \kappa} & \text{if } \eta = 0, Mortgaged_b = 1 \\ v\bar{p}_b(C) & \text{if } \eta = 1, Mortgaged_b = 0 \\ v\bar{p}_b(M) & \text{if } \eta = 1, Mortgaged_b = 1 \end{cases}, \tag{IA13}$$

where μ_κ denotes the mortgage-cash premium when $p^C = e^{-\kappa}$. Therefore, if $\eta = 0$, then

$$\log\left(Price_{i,b,s}\right) = \log(v_i) + \mu_\kappa \times Mortgaged_b - \kappa_s, \tag{IA14}$$

If $\eta = 1$, then

$$\log\left(Price_{i,b,s}\right) = \log(v_i) + \log\left(\bar{p}_b\right) + \mu_1 \times Mortgaged_b - \log\left(\tilde{v}_i/v_i\right) \tag{IA15}$$

where μ_1 denotes the mortgage-cash premium when $p^C = 1$, and $\tilde{v}$ denotes the price the buyer must offer to make the seller indifferent with respect to a certain payment of v. For an all-cash buyer, $\tilde{v} = v = v\bar{p}_b$, where the second equality uses equation (12) and the assumption $y \geq 1$ for an all-cash buyer. For a mortgaged buyer, $\tilde{v} = ve^{\mu_1}$.

Combining equations (IA14) to (IA15) gives

$$\log\left(Price_{i,b,s}\right) = \bar{\mu} \times Mortgaged_b + \log(v_i) - \kappa_s\left(1-\eta\right) + \eta\left[\log(\bar{p}_b) - \log(\tilde{v}_i/v_i)\right] \tag{IA16}$$

with $\bar{\mu} = (1-\eta)\mu_\kappa + \eta\mu_1$. Under the normalization from Section I.D of the main article we have $\tilde{v}_i = \tilde{v}_i/v_i$, which gives the expression shown in equation (27). This is what needed to be shown.

B. Model Extensions

This section provides additional detail on the model's setup and extensions. Section III.B.1 provides details on the extension with nonfinancial contingencies described in Sections I.A.3 and V.B. Section III.B.2 elaborates on the modelling of the home sale process from Section I.C.2 of the main article.

B.1. Extension with Home Sale Contingencies

Repeating the setup from Section I.A.3 of the main article, both all-cash and mortgaged offers may have contingencies unrelated to mortgage financing. The two most common nonfinancial contingencies require a satisfactory home inspection and that the buyer can sell his current home (National Association of Realtors (2018)). The home inspection contingency appears in 83% of both all-cash and mortgage-financed offers in our offer-level data set. Therefore, we make the simplifying assumption that both all-cash and mortgage-financed offers fail with probability q^c due to a failed inspection. We also assume that the home inspection occurs immediately after the offer is accepted, consistent with how real estate law in most states allow for a relatively short inspection period of around 10 days.

None of our data sources reports the distribution of the home-sale contingency across buyer types, but it seems reasonable that mortgaged buyers rely more heavily on this contingency since less-wealthy buyers are more likely to use mortgage financing. We therefore assume that the home-sale contingency accompanies a share h of mortgaged offers and never accompanies an all-cash offer.

As mentioned in Section V.B of the main article, the home-sale contingency can trigger a chain of events that amplifies the baseline probability of failure. To see this, let q_0 denote the probability that any given mortgaged offer fails to obtain financing, which is also the probability of failure for a mortgaged offer without any nonfinancial contingency ("0" for "no contingency"). Note that q_0 is simply a re-labelling of the baseline mortgage failure rate, q. Let q_1 denote the probability that a mortgaged buyer with the home-sale contingency fails to either obtain financing or fails to sell his home ("1" for "has contingency"). Lastly, let $q_{1,h}$ denote the specific probability of failure due to not selling the home, which triggers the home-sale contingency. Then,

$$\begin{aligned} q_1 &= q_0 + (1-q_0)q_{1,h} \qquad \text{(IA17)} \\ &\approx q_0 + q_{1,h}, \\ q_{1,h} &= q^c + (1-q^c)(mq_0 + m(1-q_0)hq_{1,h}) \qquad \text{(IA18)} \\ &\approx q^c + mq_0 + mhq_{1,h}, \end{aligned}$$

using the approximation that the product of two failure rates equals zero: $q_0 q_{1,h} \approx 0$, $q^c q_0 \approx 0$, and $q^c q_{1,h} \approx 0$. Rearranging terms,

$$q_{1,h} = \frac{mq_0 + q^c}{1-mh}, \qquad \text{(IA19)}$$

$$q_1 = q_0 + \frac{mq_0 + q^c}{1-mh} \qquad \text{(IA20)}$$

up to the aforementioned approximation for small probabilities. So, the combined probability of failure for a home seller with the home-sale contingency is

$$\begin{aligned} q^h &= q_0 + q^c + q_{1,h} \qquad \text{(IA21)} \\ &= q_0 + q^c + \frac{mq_0 + q^c}{1-mh}, \end{aligned}$$

as shown in equation (38).

The home-sale contingency can also influence the probability of delay. Analgously, let q_0^d denote the probability that any given mortgaged offer experiences a delay. Let q_1^d denote the probability that a mortgaged buyer with the home-sale contingency either experiences a delay himself or that his buyer experiences a delay. The derivation of q_1^d is symmetric to that for q_1,

$$q_1^d = q_0^d + \frac{mq_0^d}{1-mh}, \qquad \text{(IA22)}$$

noting that we do not have a term concerning the inspection contingency because the inspection occurs without delay.

Proposition IA.1 summarizes the resulting mortgage-cash premium and cost of failure. The expressions inherit a similar form as their analogues from the baseline case in Proposition 1 and Lemma 1. The main difference is that there are now three buyer types: all-cash, mortgaged without a home-sale contingency, and mortgaged with a home-sale contingency. By contrast, the baseline case featured only the first two of these buyer types.

The addition of another buyer type increases the dimension of the set of outcomes in the next month by one, which substantially complicates the seller's dynamic problem, and thus the expression for κ in Proposition IA.1. The basic intuition is the same as in Lemma 1, however. In addition, calculating the mortgage-cash premium μ requires us to specify whether the mortgaged buyer in question has a home-sale contingency. Figure 3 reports the average mortgage-cash premium for the case of a mortgaged buyer without a home-sale contingency, 7.7%. The mortgage-cash premium for a mortgaged buyer with a home-sale contingency equals 9.9%. Since $h = 7\%$ of mortgaged offers come with a home-sale contingency, based on the calibration in Section V.A of the main article, the average mortgage-cash premium for the case with nonfinancial contingencies equals 7.9%.

B.2. Tie-Breaking Mechanisms

As mentioned in the text, we approximate the home purchase as a sealed-bid, first-price auction. Our setting requires two modifications of the canonical setup. First, since buyers may fail to deliver on their offer, the seller evaluates bids according to their certainty equivalence, not offer price. Second, if the seller receives multiple offers during the arrival window, she invites a tie-breaking round. In this round, candidate buyers simply offer their willingness-to-pay, $\bar{p}_b$, and, for reputational reasons, buyers' real estate agents enforce truth telling. We interpret this mechanism as the activation of an escalation clause, which frequently accompany offers in competitive markets (Redfin 2021).

We require a tie-breaking round because our setting features a discrete distribution of buyer types, all-cash or mortgaged, and an equilibrium may not exist in such settings without a second-round auction to break the tie (e.g., Maskin and Riley (2000), Lebrun (1996)). Intuitively, the tie-breaking round eliminates the incentive to offer \$$\varepsilon$ more than the certainty-equivalent offer of a symmetric buyer.

Exemplifying this point, suppose that the seller simply breaks ties in the first round by selecting a bidder at random. If buyers play a symmetric strategy within their type (i.e., all-cash or mortgaged), then the probability of winning, $Win(\cdot, P)$ will exhibit discontinuities at the price that would win when there are no competitors, only competitors of the same type, and competitors of multiple types. Therefore, when ties are broken at-random, buyers can always discontinuously raise their probability of winning by offering a price ε greater than one of these jump points. Hence, equilibrium does not exist. The tie-breaker in our model avoids this issue because buyers know ex-ante that any ties will be resolved in the second round.

C. Additional Analytic Results

This section states and proves additional analytic results that are not explicitly stated in the text. Lemma IA.1 states and proves the remarks about the model's empirical implications referenced in Section I.C.1 of the main article. Lemma IA.2 derives the buyers' willingness-to-pay described in Section I.C.2 of the main article, specifically equations (12) to (13). Lemma IA.3 derives buyers' optimal offer price, also described in Section I.C.2 of the main article and specifically in equations (14) to (15). Lemma IA.4 derives various useful features of the distribution of prices within the model. Proposition IA.1 derives the mortgage-cash premium and seller cost of failure for the extension with nonfinancial contingencies.

C.1. Statement and Proof of Lemma IA.1

LEMMA IA.1 (Buyer Financing): *Consider a variant of the buyer's problem defined in equations (8) to (11) in which the only choice variable is F, so that v, P, ℓ, and Y are parameters. Define*

$$\overline{\mathcal{W}_Y} \equiv \frac{(1-q)\left(u'(Y) - u'(Y+v-P\left[1-\bar{\ell}[1-D]\right])\right) - (u'(Y) - u'(Y+v-P))}{u'(Y) - u'(Y+v-P)}, \tag{IA23}$$

$$\overline{\mathcal{W}_v} \equiv \frac{u'(Y+v-P) - (1-q)u'(Y+v-P\left[1-\bar{\ell}[1-D]\right]}{u'(Y+v-P)}, \tag{IA24}$$

where q is small enough such that $\overline{\mathcal{W}_Y}$ is positive. If

$$\frac{Win(C,P)}{Win(M,P)} < 1 + \min\left\{\overline{\mathcal{W}_Y}, \overline{\mathcal{W}_v}\right\}, \tag{IA25}$$

then the following statements hold in the variant problem:

(i) There exists a wealth threshold Y^* such that buyers optimally pay all-cash for $Y \geq Y^*$ and use mortgage financing otherwise.

(ii) There exists a housing quality threshold v^* such that buyers optimally pay all-cash for $v \leq v^*$ and use mortgage financing otherwise.

Proof. The proof involves an application of monotone comparative statics (e.g., Milgrom and Shannon (1994)). Define $\tilde{F} \equiv \mathbb{1}[F = M]$. It suffices to show that, under condition (IA25), the objective function in the variant problem exhibits: (i) decreasing differences in $\tilde{F}$ and Y, and (ii) increasing differences in $\tilde{F}$ and v. Then, by single crossing, the thresholds Y^* and v^* exist within the set of extended real numbers.

Let $\mathcal{U}(F; v, P, \bar{\ell}, Y)$ denote the objective function, where we have used the result that $\ell = \bar{\ell}$. Introduce the following simplifying notation and parameterization for this proof: $A \equiv v - P > 0$;

and $B \equiv P\bar{\ell}[1-D] > 0$. The following derivatives obtain:

$$\left.\frac{\partial \mathcal{U}}{\partial Y}\right|_{\tilde{F}=0} = -Win(C,P)\left[u'(Y) - u'(Y+A)\right] + u'(Y), \tag{IA26}$$

$$\left.\frac{\partial \mathcal{U}}{\partial Y}\right|_{\tilde{F}=1} = -Win(M,P)(1-q)\left[u'(Y) - u'(Y+A+B)\right] + u'(Y), \tag{IA27}$$

$$\left.\frac{\partial \mathcal{U}}{\partial v}\right|_{\tilde{F}=0} = Win(C,P)u'(Y+A), \tag{IA28}$$

$$\left.\frac{\partial \mathcal{U}}{\partial v}\right|_{\tilde{F}=1} = Win(M,P)(1-q)u'(Y+A+B). \tag{IA29}$$

Moreover,

$$ID(\tilde{F},Y) \equiv \left.\frac{\partial \mathcal{U}}{\partial Y}\right|_{\tilde{F}=1} - \left.\frac{\partial \mathcal{U}}{\partial Y}\right|_{\tilde{F}=0} \tag{IA30}$$

$$= \left(Win(C,P) - Win(M,P)\right)\left(u'(Y) - u'(Y+A)\right) + ... \tag{IA31}$$

$$... - Win(M,P)\left((1-q)\left[u'(Y) - u'(Y+A+B)\right] - \left[u'(Y) - u'(Y+A)\right]\right), \tag{IA32}$$

which is negative if and only if

$$\frac{Win(C,P)}{Win(M,P)} < 1 + \overline{\mathcal{W}_Y}. \tag{IA33}$$

In addition

$$ID(\tilde{F},v) \equiv \left.\frac{\partial \mathcal{U}}{\partial v}\right|_{\tilde{F}=1} - \left.\frac{\partial \mathcal{U}}{\partial v}\right|_{\tilde{F}=0} \tag{IA34}$$

$$= -\left(Win(C,P) - Win(M,P)\right)u'(Y+A) + ... \tag{IA35}$$

$$... + Win(M,P)\left(u'(Y+A) - (1-q)u'(Y+A+B)\right), \tag{IA36}$$

which is positive if and only if

$$\frac{Win(C,P)}{Win(M,P)} < 1 + \overline{\mathcal{W}_v}. \tag{IA37}$$

Collectively, $ID(\tilde{F},Y) < 0$ and $ID(\tilde{F},v) > 0$ if and only if $\frac{Win(C,P)}{Win(M,P)} < 1 + \min\left\{\overline{\mathcal{W}_Y}, \overline{\mathcal{W}_v}\right\}$, which is what needed to be shown. □

C.2. Statement and Proof of Lemma IA.2

LEMMA IA.2 (Buyer Willingness-to-Pay): *For a given v and $y \equiv \frac{Y}{v}$, a buyer who chooses to pay all-cash offers a price $\tilde{P} \leq v\bar{p}_b(C)$, with*

$$\bar{p}_b(C) \equiv \min\left\{y, 1\right\}, \tag{IA38}$$

and a buyer who choose to use mortgage financing offers a price $\tilde{P} \leq v\bar{p}_b(M)$*, with*

$$\bar{p}_b(M) \equiv \min\left\{\frac{y}{1-\bar{\ell}}, \frac{1}{1-\bar{\ell}\,[1-D]}\right\}. \tag{IA39}$$

Proof. First, suppose an all-cash buyer offers price $P_{a,C} > v\min\{y,1\}$. Substituting into equation (8), the buyer's utility equals

$$\begin{aligned} V_{a,C} &= Win(C, P_{a,C})\frac{(yv + v - P_{a,C})^{1-\gamma}}{1-\gamma} + (1 - Win(C, P_a))\frac{(yv)^{1-\gamma}}{1-\gamma} \\ &< Win(C, P_{a,C})\frac{(v\,[y + 1 - \min\{y,1\}])^{1-\gamma}}{1-\gamma} + (1 - Win(C, P_{a,C}))\frac{(yv)^{1-\gamma}}{1-\gamma} \\ &= Win(C, P_{a,C})\frac{(v\max\{y,1\})^{1-\gamma}}{1-\gamma} + (1 - Win(C, P_{a,C}))\frac{(yv)^{1-\gamma}}{1-\gamma}. \end{aligned}$$

If the buyer chooses not to make an offer, then her utility equals

$$V_b = \frac{(yv)^{1-\gamma}}{1-\gamma}.$$

When $y \geq 1$, we have $V_b > V_{a,C}$, so that the all-cash buyer strictly prefers not making an offer to an offer with $P_{a,C} > v\min\{y,1\}$. When $y < 1$, constraint (11) binds, so that $P_{a,C}$ is infeasible. Therefore, the buyer will never offer $P_{a,C}$, which implies he will offer $\tilde{P} \leq v\min\{y,1\} = v\bar{p}_b(C)$.

The logic for a mortgaged buyer is symmetric. Suppose a mortgaged buyer offers price $P_{a,M} > v\min\left\{\frac{y}{1-\bar{\ell}}, \frac{1}{1-\bar{\ell}[1-D]}\right\}$. When $\frac{y}{1-\bar{\ell}} < \frac{1}{1-\bar{\ell}[1-D]}$, constraint (11) binds, so that $P_{a,M}$ is infeasible. In the opposite case, the buyer's utility equals

$$\begin{aligned} V_{a,M} &= Win(M, P_{a,M})\frac{\left(yv + v - P_{a,M}\left[1 - \bar{\ell}\,[1-D]\right]\right)^{1-\gamma}}{1-\gamma} + (1 - Win(M, P_{a,M}))\frac{(yv)^{1-\gamma}}{1-\gamma} \\ &< Win(M, P_{a,M})\frac{(yv)^{1-\gamma}}{1-\gamma} + (1 - Win(M, P_{a,M}))\frac{(yv)^{1-\gamma}}{1-\gamma} \\ &= V_b. \end{aligned}$$

So, the buyer strictly prefers not making an offer to an offer with $P_{a,M} > v\frac{1}{1-\bar{\ell}[1-D]}$. Combining cases, the buyer will never offer $P_{a,M}$, which implies that he will offer $P \leq v\min\left\{\frac{y}{1-\bar{\ell}}, \frac{1}{1-\bar{\ell}[1-D]}\right\} = v\bar{p}_b(M)$. This is what needed to be shown. □

C.3. Statement and Proof of Lemma IA.3

LEMMA IA.3 (Offer Price): *Consider a sealed-bid auction with a discrete number of offers, which the seller observes but buyers do not. The auction has the following features:*

(i) The seller evaluates offers according to the value functions shown in (4) to (6),

(ii) Buyers submit an offer price $\tilde{P}$ and their willingness-to-pay $v\bar{p}_b$, which real estate agents truthfully verify,

(iii) If the total number of offers is $N = 1$, then the buyer pays his offer price,

(iv) If $N > 1$, then the buyer's offer price is escalated to his willingness-to-pay, and

(v) The seller can decline all offers.

There exists a Nash equilibrium in which

$$\tilde{P}^C = \begin{cases} \min\left\{\tilde{v}(C)e^{-\kappa}, v\bar{p}_b(C)\right\}, & \text{if} \quad N = 1 \\ v\bar{p}_b(C), & \text{if} \quad N > 1 \end{cases}, \tag{IA40}$$

$$\tilde{P}^M = \begin{cases} \min\left\{\tilde{v}(M)e^{-\kappa}, v\bar{p}_b(M)\right\}, & \text{if} \quad N = 1 \\ v\bar{p}_b(M), & \text{if} \quad N > 1 \end{cases}, \tag{IA41}$$

where $\tilde{v}$ denotes the price of an offer with certainty equivalence v. In particular, $\tilde{v}(C) = v$ and $\tilde{v}(M) = ve^{\mu_\kappa}$, where μ_κ is the mortgage-cash premium when $P^C = ve^{-\kappa}$.

Proof. The proof requires us to show that strategy (IA40) is a best response by all-cash buyer b_c to the same strategy used by an unknown number of competing all-cash buyers $-b_c$ and to strategy (IA41) used by an unknown number of competing mortgaged buyers b_m, and, similarly, strategy (IA41) is a best response by mortgaged buyer b_m to the same strategy used by an unknown number of competing mortgaged buyers $-b_m$ and to strategy (IA40) used by an unknown number of all-cash buyers b_c.

We focus on the case $N = 1$. The case $N > 1$ is trivial due to feature (iv) of the auction. We establish the result for all-cash buyers b_c, and then we do so for mortgaged buyers. Suppose $v\bar{p}_b(C) < \tilde{v}(C)e^{-\kappa} = ve^{-\kappa}$. Lemma IA.2 implies that b_c will offer at most $\tilde{P}^C = v\bar{p}_b(C)$. The definition of κ in equation (7) implies that the seller will decline any offer $\tilde{P}^C < ve^{-\kappa}$. Therefore, b_c is indifferent to any offer price $\tilde{P}^C \leq v\bar{p}_b(C)$, since the seller will decline all such offer prices. So, b_c has no incentive to deviate from strategy (IA40) when $v\bar{p}_b(C) < \tilde{v}(C)e^{-\kappa}$.

Suppose, instead, that $v\bar{p}_b(C) \geq \tilde{v}(C)e^{-\kappa}$. Since $N = 1$, the probability of winning is

$$Win(C, \tilde{P}^C) = \begin{cases} 0, & \text{if} \quad \tilde{P}^C < ve^{-\kappa} \\ 1, & \text{if} \quad \tilde{P}^C \geq ve^{-\kappa} \end{cases} \tag{IA42}$$

$$\tag{IA43}$$

Using equation (8), the buyer's utility equals

$$U(\tilde{P}^C) = \begin{cases} \frac{(yv)^{1-\gamma}}{1-\gamma}, & \text{if} \quad \tilde{P}^C < ve^{-\kappa} \\ \frac{\left(yv+v-\tilde{P}^C\right)^{1-\gamma}}{1-\gamma}, & \text{if} \quad \tilde{P}^C \geq ve^{-\kappa} \end{cases}. \tag{IA44}$$

By supposition, $e^{-\kappa} \leq \min\{y, 1\}$, which implies that equation (IA44) is maximized at $\tilde{P}^C = ve^{-\kappa}$. In particular, if $y \geq 1$, then $yv + v - ve^{-\kappa} \geq yv$, and, if $y < 1$, then $yv + v - ve^{-\kappa} \geq v > yv$. So, b_c has no incentive to deviate from strategy (IA40) when $v\bar{p}_b(C) \geq \tilde{v}(C)e^{-\kappa}$.

We turn to the mortgaged buyer next. The argument is symmetric. Suppose $v\bar{p}_b(M) < \tilde{v}(M)e^{-\kappa} = ve^{\mu_\kappa-\kappa}$. From Lemma IA.2, b_m will offer at most $\tilde{P}^M = v\bar{p}_b(M)$ Definition 1 implies that such an offer is equivalent, from the perspective of the seller's expected utility ("certainty equivalence") to an all-cash offer at price $ve^{\mu_\kappa-\kappa}$. Equation (7), therefore, implies that the seller declines any offer $\tilde{P}^M < ve^{\mu_\kappa-\kappa}$. Consequently, b_m is indifferent with respect to any such offer. So, b_m has no incentive to deviate from strategy (IA41) when $v\bar{p}_b(M) < \tilde{v}(M)e^{-\kappa}$.

Suppose, instead, that $v\bar{p}_b(M) \geq \tilde{v}(M)e^{-\kappa}$. The probability of winning takes a similar form as equation (IA42)

$$Win(M, \tilde{P}^M) = \begin{cases} 0, & \text{if } \tilde{P}^M < ve^{\mu_\kappa-\kappa} \\ 1, & \text{if } \tilde{P}^M \geq ve^{\mu_\kappa-\kappa} \end{cases}, \tag{IA45}$$

(IA46)

and the buyer's utility equals

$$U(\tilde{P}^C) = \begin{cases} \frac{(yv)^{1-\gamma}}{1-\gamma}, & \text{if } \tilde{P}^M < ve^{\mu_\kappa-\kappa} \\ \frac{\left(yv+v-\tilde{P}^M\left[1-\bar{\ell}[1-D]\right]\right)^{1-\gamma}}{1-\gamma}, & \text{if } \tilde{P}^M \geq ve^{\mu_\kappa-\kappa} \end{cases}. \tag{IA47}$$

(IA48)

By supposition, $e^{\mu_\kappa-\kappa} \leq \min\left\{\frac{y}{1-\bar{\ell}}, \frac{1}{1-\bar{\ell}[1-D]}\right\}$, which implies that equation (IA47) is maximized at $\tilde{P}^C = ve^{\mu_\kappa-\kappa}$. So, b_m has no incentive to deviate from strategy (IA41) when $v\bar{p}_b(M) \geq \tilde{v}(M)e^{-\kappa}$.

Since neither b_c nor b_m have an incentive to deviate from strategies (IA40) to (IA41), these strategies constitute a best-response for each buyer. This is what needed to be shown. □

C.4. Statement and Proof of Lemma IA.4

LEMMA IA.4 (Distribution of Offers): *The probabilities that the seller receives no offers, exactly one offer, multiple offers consisting only of mortgaged offers, and multiple offers consisting only of all-cash offers are, respectively:*

$$\begin{aligned} \varphi_0 &= e^{-\lambda}, & \text{(IA49)} \\ \varphi_1 &= \lambda e^{-\lambda}, & \text{(IA50)} \\ \varphi_{MM} &= e^{-\lambda(1-m)}[1 - e^{-\lambda m}(1 + \lambda m)], & \text{(IA51)} \\ \varphi_{CC} &= e^{-\lambda m}[1 - e^{-\lambda(1-m)}(1 + \lambda(1 - m))]. & \text{(IA52)} \end{aligned}$$

The probability that the seller receives multiple offers and accepts a mortgaged offer is:

$$\Phi_M = \begin{cases} e^{-\lambda(1-m)}[1 - e^{-\lambda m}(1 + \lambda m)] & \text{if } \bar{p}_b(M) \leq \overline{p_C} \\ 1 - e^{-\lambda}(1 + \lambda) - e^{-\lambda m}[1 - e^{-\lambda(1-m)}(1 + \lambda(1 - m))] & \text{if } \bar{p}_b(M) > \overline{p_C}, \end{cases} \tag{IA53}$$

where $\log(\overline{p_C})$ *is the mortgage-cash premium from Proposition 1 evaluated at* $p^C = \bar{p}_b(C)$*. The probability that the seller receives multiple offers and accepts an all-cash offer is:*

$$\Phi_C \quad = \begin{cases} 1 - e^{-\lambda}(1+\lambda) - e^{-\lambda(1-m)}[1 - e^{-\lambda m}(1+\lambda m)] & \text{if } \bar{p}_b(M) \leq \overline{p_C} \\ e^{-\lambda m}[1 - e^{-\lambda(1-m)}(1+\lambda(1-m))] & \text{if } \bar{p}_b(M) > \overline{p_C}. \end{cases} \tag{IA54}$$

Lastly, there exists $\overline{p_M}$ *such that if* $\bar{p}_b(M) \geq \overline{p_M}$*, the seller accepts mortgaged offers with positive probability, and otherwise she always declines mortgaged offers.*

Proof. First, we derive $\{\varphi_j\}$. Since $N \sim Poisson(\lambda)$, we have $\Pr[N = x] = e^{-\lambda}\frac{\lambda^x}{x!}$. Therefore,

$$\varphi_0 = \Pr[N = 0] = e^{-\lambda} \tag{IA55}$$

$$\varphi_1 = \Pr[N = 1] = \lambda e^{-\lambda} \tag{IA56}$$

Moreover,

$$\varphi_{MM} = \Pr\left[N^C = 0, N^M > 1\right] \tag{IA57}$$

$$= \sum_{k=2}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} \frac{k!}{0!k!} m^k \tag{IA58}$$

$$= e^{-\lambda(1-m)} \sum_{k=2}^{\infty} \frac{e^{-\lambda m}(\lambda m)^k}{k!} \tag{IA59}$$

$$= e^{-\lambda(1-m)}[1 - e^{-\lambda m}(1+\lambda m)] \tag{IA60}$$

$$\varphi_{CC} = \Pr\left[N^M = 0, N^C > 1\right] \tag{IA61}$$

$$= \sum_{k=2}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} \frac{k!}{0!k!} (1-m)^k \tag{IA62}$$

$$= e^{-\lambda m} \sum_{k=2}^{\infty} \frac{e^{-\lambda(1-m)}(\lambda(1-m))^k}{k!} \tag{IA63}$$

$$= e^{-\lambda m}[1 - e^{-\lambda(1-m)}(1+\lambda(1-m))] \tag{IA64}$$

Next, we derive $\{\Phi_j\}$. Suppose $\bar{p}_b(M) > \overline{p_C}$. Then a seller prefers an offer from a mortgaged buyer at his willingness-to-pay, $\bar{p}_b(M)$, over an offer from an all-cash buyer at the willingness-to-pay of the latter, $\bar{p}_b(C)$. Therefore, the probability that the seller receives multiple offers and accepts a mortgaged offer is the complement probability of receiving: zero offers, exactly one offer, or multiple offers consisting only of all-cash offers. Explicitly, this probability equals

$$1 - e^{-\lambda}(1+\lambda) - e^{-\lambda m}[1 - e^{-\lambda(1-m)}(1+\lambda(1-m))]. \tag{IA65}$$

The probability that the seller receives multiple offers and accepts an all-cash offer is φ_{CC}.

Conversely, suppose $\bar{p}_b(M) \leq \overline{p_C}$, so that a seller prefers an offer from an all-cash buyer at his willingness-to-pay, $\bar{p}_b(C)$, over an offer from a mortgaged buyer at the latter's willingness-to-pay, $\bar{p}_b(M)$. The probability that the seller receives multiple offers and accepts an all-cash offer is the complement probability of receiving: zero offers, exactly one offer, or multiple offers consisting only of mortgaged offers. Explicitly, this probability equals

$$1 - e^{-\lambda}(1+\lambda) - e^{-\lambda(1-m)}[1 - e^{-\lambda m}(1+\lambda m)]. \tag{IA66}$$

The probability that the seller receives multiple offers and accepts a mortgaged offer is φ_{MM}

Collecting the previous two paragraphs gives the expressions for Φ_M and Φ_C shown in equations (IA53) to (IA54).

Lastly, define $\overline{p_M}$ such that $\log(\overline{p_M})$ equals the mortgage-cash premium when $p^C = e^{-\kappa}$. In the notation from Lemma IA.3, $\overline{p_M} = e^{\mu_\kappa - \kappa}$. Then if $\bar{p}_b(M) \geq \overline{p_M}$, the seller accepts a mortgaged offer with positive probability $m\varphi_1 + \Phi_M$. If $\bar{p}_b(M) < \overline{p_M}$, the seller always declines mortgaged offers.

This completes what needed to be shown.

□

C.5. Statement and Proof of Proposition IA.1

PROPOSITION IA.1 (Deriving the Mortgage-Cash Premium with Contingencies): *Consider the extension described in Section III.B.1 with nonfinancial contingencies. The mortgage-cash premium equals*

$$\mu = \log\left(\left[\frac{e^{\alpha_k+\rho}(1-q^c)\left(\omega+p^C\right)^{1-\gamma} - q_k\left(\omega+e^{-\kappa}\right)^{1-\gamma} + (1-\gamma)\Gamma_k\delta}{1-q^c-q_k}\right]^{\frac{1}{1-\gamma}} - \omega\right) - \log(p^C), \tag{IA67}$$

where q_k equals the combined probability that a mortgaged buyer fails to obtain financing or sell his home, based on whether the buyer has a home-sale contingency ($k=1$) or not ($k=0$), and analogously for the probability of transaction delay q_k^d. In particular,

$$q_0 = q, \tag{IA68}$$

$$q_0^d = q^d, \tag{IA69}$$

$$q_1 = q_0\left(\frac{1+m(1-h)}{1-mh}\right) + \frac{q^c}{1-mh}, \tag{IA70}$$

$$q_1^d = q_0^d\left(\frac{1+m(1-h)}{1-mh}\right), \tag{IA71}$$

up to the approximation that the product of two failure rates equals zero. The continuation value, κ, equals

$$\kappa = -\log\left(\left[\frac{(1-\gamma)\left(\mathbb{E}\left[u|u>V^{\mathcal{S}}(\emptyset)\right] Pr\left[u>V^{\mathcal{S}}(\emptyset)\right] - \delta\Xi\right)}{e^\rho - Pr[u=V^{\mathcal{S}}(\emptyset)]}\right]^{\frac{1}{1-\gamma}} - \omega\right). \tag{IA72}$$

The magnitude of $\bar{p}_b(M)$ determines the expressions for the terms in equation (IA72).

The first case obtains when $\bar{p}_b(M) \geq \overline{p_{M,h}}$. In this case,

$$\mathbb{E}\left[u|u > V^{\mathcal{S}}(\emptyset)\right] Pr\left[u > V^{\mathcal{S}}(\emptyset)\right] = \Phi_C[1-q^c]\frac{(\omega+\bar{p}_b(C))^{1-\gamma}}{1-\gamma} + ... \tag{IA73}$$

$$... + \Phi_{M_0}[1-q_0-q^c]e^{-\alpha_0-\rho}\frac{(\omega+\bar{p}_b(M))^{1-\gamma}}{1-\gamma} + ... \tag{IA74}$$

$$... + \Phi_{M_1}[1-q_1-q^c]e^{-\alpha_1-\rho}\frac{(\omega+\bar{p}_b(M))^{1-\gamma}}{1-\gamma}, \tag{IA75}$$

$$Pr\left[u = V^{\mathcal{S}}(\emptyset)\right] = e^{-\alpha_0-\rho}\Phi_{M_0}q_0 + ... \tag{IA76}$$

$$... + e^{-\alpha_1-\rho}\Phi_{M_1}q_1 + ... \tag{IA77}$$

$$... + \varphi_0 + \varphi_1 + q^c(1-\varphi_0-\varphi_1), \tag{IA78}$$

$$\Xi = e^{\rho} + \Phi_{M_0}(1+q_0^d e^{-\rho}) + \Phi_{M_1}(1+q_1^d e^{-\rho}), \tag{IA79}$$

The terms $\{\varphi\}$ equal the probabilities of receiving no offers (φ_0), one offer (φ_1), multiple offers consisting only of mortgaged offers with a home-sale contingency (φ_{M_1}), multiple offers consisting only of all-cash offers (φ_C), multiple offers with zero all-cash offers ($\varphi_{\text{-}C}$), or multiple offers, none of which are mortgaged offers without a home-sale contingency ($\varphi_{\text{-}M_0}$).

The second case obtains when $\overline{p_{M,h}} > \bar{p}_b(M) \geq \overline{p_{M,-h}}$. In this case

$$\mathbb{E}\left[u|u > V^{\mathcal{S}}(\emptyset)\right] Pr\left[u > V^{\mathcal{S}}(\emptyset)\right] = \Phi_C[1-q^c]\frac{(\omega+\bar{p}_b(C))^{1-\gamma}}{1-\gamma} + ... \tag{IA80}$$

$$... + \Phi_{M_0}[1-q_0-q^c]e^{-\alpha_0-\rho}\frac{(\omega+\bar{p}_b(M))^{1-\gamma}}{1-\gamma} \tag{IA81}$$

$$Pr\left[u = V^{\mathcal{S}}(\emptyset)\right] = e^{-\alpha_0-\rho}\Phi_{M_0}q_0 + ... \tag{IA82}$$

$$... + \varphi_{M_1} + \varphi_0 + \varphi_1 + q^c(1-\varphi_{M_1}-\varphi_0-\varphi_1), \tag{IA83}$$

$$\Xi = e^{\rho} + \Phi_{M_0}(1+q_0^d e^{-\rho}). \tag{IA84}$$

The third case obtains when $\overline{p_{M,-h}} > \bar{p}_b(M)$. In this case

$$\mathbb{E}\left[u|u > V^{\mathcal{S}}(\emptyset)\right] Pr\left[u > V^{\mathcal{S}}(\emptyset)\right] = \Phi_C[1-q^c]\frac{(\omega+\bar{p}_b(C))^{1-\gamma}}{1-\gamma}, \tag{IA85}$$

$$Pr\left[u = V^{\mathcal{S}}(\emptyset)\right] = \varphi_{\text{-}C} + \varphi_0 + \varphi_1 + q^C(1-\varphi_{\text{-}C}-\varphi_0-\varphi_1), \tag{IA86}$$

$$\Xi = e^{\rho}. \tag{IA87}$$

Proof. The proof has a similar structure as that for Proposition 1 and Lemma 1. The difference is that there are three buyer types, as opposed to two in the main results. This leads to only minor changes in the expression for the premium, μ, but to more material changes in the expression for the cost of failure, κ.

Deriving the premium again requires solving $V^{\mathcal{S}}(C, vp^C) = V^{\mathcal{S}}(M, vp^C e^{\mu})$. The value functions change slightly after incorporating nonfinancial contingencies,

$$V^{\mathcal{S}}(C, vp^C) = (1-q_c)\frac{\left(\omega + p^C\right)^{1-\gamma}}{1-\gamma} + q_c\frac{\left(\omega + e^{-\kappa}\right)^{1-\gamma}}{1-\gamma}, \tag{IA88}$$

$$V^{\mathcal{S}}(M, vp^C e^{\mu}) = -\delta\left(e^{-\alpha_k} + q^d\right) + ... \tag{IA89}$$

$$... + e^{-\alpha_k - \rho}\left[(1-q_c-q_k)\frac{\left(\omega + p^C e^{\mu}\right)^{1-\gamma}}{1-\gamma} + q_k\frac{\left(\omega + e^{-\kappa}\right)^{1-\gamma}}{1-\gamma}\right] + ... \tag{IA90}$$

$$... + q_c\frac{\left(\omega + e^{-\kappa}\right)^{1-\gamma}}{1-\gamma} \tag{IA91}$$

Solving and rearranging gives

$$\mu = \log\left(\left[\frac{e^{\alpha_k+\rho}(1-q^c)\left(\omega + p^C\right)^{1-\gamma} - q_k\left(\omega + e^{-\kappa}\right)^{1-\gamma} + (1-\gamma)\Gamma_k\delta}{1-q^c-q_k}\right]^{\frac{1}{1-\gamma}} - \omega\right) - \log(p^C), \tag{IA92}$$

using the same logic as in Proposition 1. The expressions for $\{q_k, q_k^d\}$ are in Section III.B.1.

Before deriving the cost of failure, we follow the same logic as in Lemma IA.4 to derive

$$\varphi_0 = e^{-\lambda}, \tag{IA93}$$

$$\varphi_1 = \lambda e^{-\lambda}, \tag{IA94}$$

$$\varphi_C = e^{-\lambda m}[1 - e^{-\lambda(1-m)}(1 + \lambda(1-m))], \tag{IA95}$$

$$\varphi_{\text{-}C} = e^{-\lambda(1-m)}[1 - e^{-\lambda m}(1 + \lambda m)], \tag{IA96}$$

$$\varphi_{\text{-}M_0} = e^{-\lambda m(1-h)}[1 - e^{-\lambda(1-m(1-h))}(1 + \lambda(1-m(1-h)))], \tag{IA97}$$

$$\varphi_{M_1} = e^{-\lambda(1-mh)}[1 - e^{-\lambda mh}(1 + \lambda mh)]. \tag{IA98}$$

and, similarly,

$$\Phi_C = \begin{cases} 1 - \varphi_0 - \varphi_1 - \varphi_{\text{-}C} & \text{if } \bar{p}_b(M) \le e^{\bar{\mu}_0}\bar{p}_b(C) \\ \varphi_{\text{-}M_0} - \varphi_{M_1} & \text{if } e^{\bar{\mu}_0}\bar{p}_b(C) < \bar{p}_b(M) \le e^{\bar{\mu}_1}\bar{p}_b(C) \\ \varphi_C & \text{if } \bar{p}_b(M) > e^{\bar{\mu}_1}\bar{p}_b(C) \end{cases}, \tag{IA99}$$

$$\Phi_{M_0} = \begin{cases} \varphi_{\text{-}C} - \varphi_{M_1} & \text{if } \bar{p}_b(M) \le e^{\bar{\mu}_0}\bar{p}_b(C) \\ 1 - \varphi_0 - \varphi_1 - \varphi_{\text{-}M_0} & \text{if } e^{\bar{\mu}_0}\bar{p}_b(C) < \bar{p}_b(M) \le e^{\bar{\mu}_1}\bar{p}_b(C) \\ 1 - \varphi_0 - \varphi_1 - \varphi_{\text{-}M_0} & \text{if } \bar{p}_b(M) > e^{\bar{\mu}_1}\bar{p}_b(C) \end{cases}, \tag{IA100}$$

$$\Phi_{M_1} = \begin{cases} \varphi_{M_1} & \text{if } \bar{p}_b(M) \le e^{\bar{\mu}_0}\bar{p}_b(C) \\ \varphi_{M_1} & \text{if } e^{\bar{\mu}_0}\bar{p}_b(C) < \bar{p}_b(M) \le e^{\bar{\mu}_1}\bar{p}_b(C) \\ \varphi_{\text{-}M_0} - \varphi_C & \text{if } \bar{p}_b(M) > e^{\bar{\mu}_1}\bar{p}_b(C) \end{cases} \tag{IA101}$$

with

$$\bar{\mu}_k = \log\left(\left[\frac{e^{\alpha_k+\rho}(1-q^c)\left(\omega+\bar{p}_b(C)\right)^{1-\gamma} - q_k\left(\omega+e^{-\kappa}\right)^{1-\gamma} + (1-\gamma)\Gamma_k\delta}{1-q^c-q_k}\right]^{\frac{1}{1-\gamma}} - \omega\right). \quad \text{(IA102)}$$

for $k \in \{0,1\}$. The three cases that define the $\{\Phi_F\}$ have a similar interpretation as in Lemma IA.4: (i) $\bar{p}_b(M) > e^{\bar{\mu}_1}\bar{p}_b(C)$ implies that the seller prefers an offer by any mortgaged buyer at his willingness-to-pay over an offer from an all-cash buyer at the latter's willingness-to-pay, (ii) $e^{\bar{\mu}_0}\bar{p}_b(C) < \bar{p}_b(M) \leq e^{\bar{\mu}_1}\bar{p}_b(C)$ implies that the seller prefers an offer by a mortgaged buyer without home-sale contingency at his willingness-to-pay over an offer from an all-cash buyer at the latter's willingness-to-pay, but the seller still prefers the all-cash offer over an offer from a mortgaged buyer with home-sale contingency at his willingness-to-pay, and (iii) $\bar{p}_b(M) \leq e^{\bar{\mu}_0}\bar{p}_b(C)$ implies that the seller prefers an offer by an all-cash buyer at his willingness-to-pay over an offer from any mortgaged buyer at the latter's willingness-to-pay.

The derivation of the cost of failure follows a similar logic as in Lemma 1. We must account for three cases. These cases depend on the thresholds $\overline{p_{M,h}}$ and $\overline{p_{M,-h}}$, where: $\overline{p_{M,h}}$ is defined such that $\log\left(\overline{p_{M,h}}\right)$ equals the mortgage-cash premium for a buyer with home-sale contingencies when $p^C = e^{-\kappa}$, denoted $\underline{\mu}_1$, and $\overline{p_{M,-h}}$ is defined such that $\log\left(\overline{p_{M,-h}}\right)$ is the mortgage-cash premium for a buyer without home-sale contingencies when $p^C = e^{-\kappa}$, denoted $\underline{\mu}_0$.

In the first case, $\bar{p}_b(M) \geq \overline{p_{M,h}}$. The continuation value equals

$$\begin{aligned}
V^{\mathcal{S}}(\emptyset) = -\delta + ...\\
&...e^{-\rho}\varphi_1(1-m)(1-q^c)\left[u(\omega+e^{-\kappa})\right] + ...\\
&...e^{-\rho}\varphi_1 m(1-h)e^{-\alpha_0-\rho}\left[(1-q_0-q^c)u(\omega+e^{\underline{\mu}_0-\kappa})\right] + ...\\
&...e^{-\rho}\varphi_1 mhe^{-\alpha_1-\rho}\left[(1-q_1-q^c)u(\omega+e^{\underline{\mu}_1-\kappa})\right] + ...\\
&...e^{-\rho}\Phi_C(1-q^c)\left[u(\omega+\bar{p}_b(C))\right] + ...\\
&...e^{-\rho}\Phi_{M_0}(1-q_0-q^c)e^{-\alpha_0-\rho}\left[u(\omega+\bar{p}_b(M))\right] + ...\\
&...e^{-\rho}\Phi_{M_1}(1-q_1-q^c)e^{-\alpha_1-\rho}\left[u(\omega+\bar{p}_b(M))\right] + ...\\
&...e^{-\rho}((\Phi_{M_0}+\Phi_{M_1}+\varphi_1 m)(1+q_0^d e^{-\rho}) + [\Phi_{M_1}+\varphi_1 mh](q_1^d-q_0^d)e^{-\rho})\left[-\delta\right] + ...\\
&...e^{-\rho}(e^{-\alpha_0-\rho}(\Phi_{M_0}+\varphi_1 m(1-h))q_0 + e^{-\alpha_1-\rho}(\Phi_{M_1}+\varphi_1 mh)q_1) + \varphi_0(1-q^c)+q^c)\left[V^{\mathcal{S}}(\emptyset)\right].
\end{aligned}$$

Recognize that, by definition, μ solves $V^{\mathcal{S}}(C, ve^{-\kappa}) = V^{\mathcal{S}}(M_k, e^{\underline{\mu}_k-\kappa})$ for $k \in \{0,1\}$. Therefore,

$$\begin{aligned}
V^{\mathcal{S}}(\emptyset) = -\delta + ...\\
&...e^{-\rho}\varphi_1\left[(1-q^c)u(\omega+e^{-\kappa})\right] + ...\\
&...e^{-\rho}\Phi_C(1-q^c)\left[u(\omega+\bar{p}_b(C))\right] + ...\\
&...e^{-\rho}\Phi_{M_0}(1-q_0-q^c)e^{-\alpha_0-\rho}\left[(1-q_0-q^c)u(\omega+\bar{p}_b(M))\right] + ...\\
&...e^{-\rho}\Phi_{M_1}(1-q_1-q^c)e^{-\alpha_1-\rho}\left[(1-q_1-q^c)u(\omega+\bar{p}_b(M))\right] + ...\\
&...e^{-\rho}((\Phi_{M_0}+\Phi_{M_1})(1+q_0^d e^{-\rho}) + \Phi_{M_1}(q_1^d-q_0^d)e^{-\rho})\left[-\delta\right] + ...\\
&...e^{-\rho}(e^{-\alpha_0-\rho}\Phi_M q_0 + e^{-\alpha_1-\rho}\Phi_{M_1}q_1) + \varphi_0(1-q^c)+q^c)\left[V^{\mathcal{S}}(\emptyset)\right].
\end{aligned}$$

Rearranging terms gives

$$e^{-\kappa} = -\omega + \left[\frac{\Phi_C(1-q^c)(\omega+\bar{p}_b(C))^{1-\gamma} + (\gamma-1)\delta\Xi}{e^{\rho} - (e^{-\alpha_0-\rho}\Phi_{M_0}q_0 + e^{-\alpha_1-\rho}\Phi_{M_1}q_1 + (\varphi_0+\varphi_1)(1-q^c)+q^c)} \cdots\right.$$

$$\left. \cdots + \frac{[(\Phi_{M_0}+\Phi_{M_1}e^{\alpha_0-\alpha_1})(1-q_0-q^c) - \Phi_{M_1}e^{\alpha_0-\alpha_1}(q_1-q_0)]e^{-\alpha_0-\rho}(\omega+\bar{p}_b(M))^{1-\gamma}}{e^{\rho} - (e^{-\alpha_0-\rho}\Phi_{M_0}q_0 + e^{-\alpha_1-\rho}\Phi_{M_1}q_1) + (\varphi_0+\varphi_1)(1-q^c)+q^c)}\right]^{\frac{1}{1-\gamma}}$$

with $\Xi = (e^{\rho} + (\Phi_{M_0}+\Phi_{M_1})(1+q_0^d e^{-\rho}) + \Phi_{M_1}[q_1^d - q_0^d]e^{-\rho})$. Taking the log transform and substituting $\mathbb{E}\left[u|u > V^{\mathcal{S}}(\emptyset)\right]\Pr\left[u > V^{\mathcal{S}}(\emptyset)\right]$ and $\Pr\left[u = V^{\mathcal{S}}(\emptyset)\right]$ defined in the statement of Proposition IA.1 gives the result, completing the proof for the case $\bar{p}_b(M) \geq \overline{p_{M,h}}$.

In the second case, $\overline{p_{M,-h}} \leq \bar{p}_b(M) < \overline{p_{M,h}}$. The logic is similar to the first case, except that the seller optimally declines all mortgaged offers with a home-sale contingency. Accordingly, the probability that the value function in the next period is $V^{\mathcal{S}}(\emptyset)$ equals

$$\varphi_{M_1} + \varphi_0 + \varphi_1 + q^c[1-(\varphi_{M_1}+\varphi_0+\varphi_1)].$$

Therefore,

$$e^{-\kappa} = -\omega + \left[\frac{\Phi_C(1-q^c)(\omega+\bar{p}_b(C))^{1-\gamma} + \delta(\gamma-1)(e^{\rho}+\Phi_{M_0}(1+q_0^d e^{-\rho}))}{e^{\rho} - (e^{-\alpha_0-\rho}\Phi_{M_0}q_0 + (\varphi_{M_1}+\varphi_0+\varphi_1)(1-q^c)+q^c)} \cdots\right.$$

$$\left. \cdots + \frac{\Phi_{M_0}(1-q_0-q^c)e^{-\alpha_0-\rho}(\omega+\bar{p}_b(M))^{1-\gamma}}{e^{\rho} - (e^{-\alpha_0-\rho}\Phi_{M_0}q_0 + (\varphi_{M_1}+\varphi_0+\varphi_1)(1-q^c)+q^c)}\right]^{\frac{1}{1-\gamma}}$$

Taking the log transform and substituting Ξ, $\mathbb{E}\left[u|u > V^{\mathcal{S}}(\emptyset)\right]\Pr\left[u > V^{\mathcal{S}}(\emptyset)\right]$, and $\Pr\left[u = V^{\mathcal{S}}(\emptyset)\right]$ defined in the statement of Proposition IA.1 gives the result, completing the proof for the case $\overline{p_{M,-h}} \leq \bar{p}_b(M) < \overline{p_{M,h}}$.

In the third case, $\overline{p_{M,-h}} > \bar{p}_b(M)$. The logic is similar to the first and second cases, except that the seller optimally declines all mortgaged offers, regardless of whether they come with a home-sale contingency. Accordingly, the probability of that the value function in the next period is $V^{\mathcal{S}}(\emptyset)$ equals

$$\varphi_{\text{-}C} + \varphi_0 + \varphi_1 + q^c[1-(\varphi_{\text{-}C}+\varphi_0+\varphi_1)].$$

Therefore,

$$e^{-\kappa} = -\omega + \left[\frac{\Phi_C(1-q^c)(\omega+\bar{p}_b(C))^{1-\gamma} + \delta(\gamma-1)e^{\rho}}{e^{\rho} - ((\varphi_{\text{-}C}+\varphi_0+\varphi_1)(1-q^c)+q^c)}\right]^{\frac{1}{1-\gamma}}$$

Taking the log transform and substituting Ξ, $\mathbb{E}\left[u|u > V^{\mathcal{S}}(\emptyset)\right]\Pr\left[u > V^{\mathcal{S}}(\emptyset)\right]$, and $\Pr\left[u = V^{\mathcal{S}}(\emptyset)\right]$ defined in the statement of Proposition IA.1 gives the result, completing the proof for the case $\overline{p_{M,-h}} > \bar{p}_b(M)$. Since this is the final case, the proof is now complete.

□

D. Calibration and Numerical Methods

We describe the numerical methods used to calculate the theoretical mortgage-cash premium, details on measuring the distribution of the model's parameters, details on calculating heterogeneity in the theoretical premium, and details on estimating heterogeneity in the empirical premium.

D.1. Numerical Methods

Since Proposition 1 delivers μ in closed form, there are relatively few numerical details of note. The two relevant sets of details are accounting for various cases in the calculation of κ, and accounting for regions of the parameter space in which μ does not exist. First, in calculating κ, we first evaluate (20) as if the condition $\bar{p}_b(M) \geq \overline{p_M}$ holds. We perform this calculation at each point in the parameter space. Then, we calculate the corresponding μ using equation (19), again at each point in the parameter space. Since $\overline{p_M}$ depends on μ, as shown explicitly in the proof of Lemma 1, we then check whether the condition $\bar{p}_b(M) \geq \overline{p_M}$ indeed holds. If so, we conclude the calculation at the given grid point. Otherwise, we recalculate μ and κ using the values corresponding to the case $\bar{p}_b(M) < \overline{p_M}$. A similar procedure applies to the extension with nonfinancial contingencies, except that there are now three cases (Appendix Proposition IA.1).

The second detail of note concerns regions of the parameter space in which a positive, real-valued μ or κ does not exist. We drop such regions in our main analyses of Sections V.B and VI.A. This reduces the size of the parameter space from 177,147 to 118,986.

D.2. Additional Calibration Details

The text references several extensions and details about our calibration that we describe in this appendix.

Details on Probability of Transaction Failure

As described in the text, we measure q using the mortgage application denial rate among pre-approved, first-lien, mortgages for the purchase of an owner-occupied, single-family home. We observe this denial rate at the zip code-by-year level, and we calculate the distribution weighting zip code-years by number of applications.

This parameterization constitutes an upper bound because not all denials as recorded in HMDA necessitate that the buyer terminate the purchase. There are two reasons why a HMDA-recorded denial does not necessitate termination. First, based on discussions with practitioners, originators typically record a mortgage application as "denied" in HMDA if doing so violates their underwriting protocol, but they will subsequently counsel a borrower on how to modify their application so as to be approved in the second round of underwriting. Since the revised application is recorded separately in HMDA, a mortgage-financed transaction that is ultimately successful would nevertheless be associated with a denied application. Second, if borrowers cannot negotiate a second-round application with their initial originator, they can apply for credit from one of its competitors.

Accordingly, we also parameterize q using the share of transactions that were terminated over 2015 to 2021, based on the National Association of Realtors RCI data set described in Section I.C. This measure is reasonable because the most common reasons reported for delayed or terminated settlements in the RCI data set are "issues related to obtaining financing" (25%), "appraisal issues" (15%), and "home inspection" (14%), all of which pertain more commonly to mortgage-financed purchases. However, we only observe this measure at the monthly level, and so the distribution only reflects time-series variation.

The corresponding support of q based on the RCI data set is $\{0.04, 0.06, 0.07\}$, based on the discretization described in Section V.A of the main article applied to all other parameters. The corresponding average mortgage-cash premium equals 6.6%.

Details on Offer Arrival Rate

The baseline calibration of λ involves calculating the average number of offers received conditional on receiving an offer, $\mathbb{E}[N|N > 0]$. We calculate the expectation at the zip code level using the offer-level data set. The expectation filters out listings with more than the maximum number of offers from Zillow's 2018 CHTR report (Zillow (2018)), which we use as a cross-reference. We then discretize the distribution across zip codes, weighting zip codes by the total number of listings. Finally, we transform this distribution using the relationship: $\mathbb{E}[N|N > 0] = \frac{\lambda}{1-e^{-\lambda}}$.

In the text, we referenced other methods of calibrating λ. One method calculates $\mathbb{E}[N|N > 0]$ using the number of mortgage applications per origination in HMDA. This method interprets an application as "an offer" and an origination as "a sale," and it implicitly assumes that the corresponding measure of offers-per-sale applies to both all-cash and mortgaged offers. Thus, a disadvantage of this measure is that, while still reasonable, the conceptual connection to λ is not as strong as with the measure based on the offer-level data set. However, this measure has the advantage of being observed at the zip code-by-year level, as it comes from the HMDA data set. Accordingly, we can rely on much broader variation, which will be critical when turning to our semi-structural estimation of the empirical premium. We discretize the distribution of $\mathbb{E}[N|N > 0]$ in the same manner as we discretize the distribution of q. Finally, we apply the Poisson transformation to recover λ. The corresponding support of λ is $\{0.42, 0.74, 1.45\}$. The corresponding average mortgage-cash premium equals 5.4%.

A second method uses Zillow's 2018 CHTR report (Zillow (2018)), which reports the distribution of total number of offers within a listing spell and includes zero in the support. Therefore, we can directly calculate the distribution $\mathbb{E}[N] = \lambda$, which gives λ the support $\{1, 2, 4\}$. The average mortgage-cash premium equals 5.4%.

A third method uses the NAR's RCI data set, which includes a monthly measure of the average number offers received conditional on receiving an offer, $\mathbb{E}[N|N > 0]$. So, we use time-series variation to discretize the distribution of $\mathbb{E}[N|N > 0]$. The standard transformation gives λ the support $\{1.85, 2.47, 4.02\}$. The average mortgage-cash premium equals 7.0%.

Collectively, we robustly find an average mortgage-cash premium between 5.4% and 7.0% based on four unique calibrations of λ.

Details on Seller Net Total Wealth

As described in the text, we measure w using the SCF data set. The unit of observation in the SCF data set is the household. For each household, we calculate w as the ratio of the sum of checking and savings accounts, certificates of deposit, cash, stocks, savings bonds, other bonds, mutual funds, annuities, trusts, IRAs, employer-provided retirement plans, and total household income, divided by the value of the home. We exclude nonhomeowners and first-time homeowners, and we weight households using SCF sample weights.

We measure ℓ_s and ξ using the baseline ZTRAX data set as described in the text. Summarizing, we obtain the grid points for ℓ_s by first calculating the ratio of mortgage balance outstanding, imputed by a straight-line amortization using the loan's initial term, to the sale price. We obtain the grid points for ξ by first calculating the difference in price between the home to which the seller moves next in the ZTRAX data set and her mortgage on that home. Then, we divide by the price on her current home sale, conditional on the two transactions lying within 12 months. For both measures, we do not divide by the price of the transaction in question, but by the average sale price within the same zip code and month, after excluding the transaction in question ("leave-one-out-mean"). We divide in this manner to maintain consistency with our regressions, which will also rely on these measures but require division by a leave-one-out-mean for econometric reasons. In particular, using the actual sales price induces mechanical correlation that leads to biased estimates.

We aggregate these transaction-level measures of ℓ_s and ξ to the zip code by year level by taking the average, excluding implausibly large values of 10 for the LTV ratio and 1 for the down payment-to-sale price ratio. We discretize the distribution across zip codes and years, weighting by the number of transactions in a zip code and year.

As in the text, the seller's net total wealth-to-housing ratio is defined as $\omega \equiv w - \ell_s - \xi$. Thus, while the supports of w, ℓ_s, and ξ each have 3 unique values, the support of ω has $27 = 3^3$ unique values.

Details on Buyer Liquid Assets-to-Housing

Since buyers typically cannot tap retirement wealth to finance their home purchase, we do not use the SCF data set to measure y. Instead, we measure y using average income in a given zip code and year based on the IRS SOI data set. Consequently, we can also incorporate this measure into our semi-structural variation. Accordingly, we divide average income by the leave-one-out-mean sales price, for econometric reasons just described. Details on the IRS SOI data set are in Section I.C. We discretize the distribution of y across zip codes and years like we do with the other parameters. We weight zip code-years by the number of transactions in the ZTRAX data set.

Details on Loan-to-Value Constraint

We measure $\bar{\ell}$ using the average LTV ratio conditional on lying between the GSE regulatory limit and 100%. We calculate the regulatory limit as the minimum of 80% and the ratio of the conforming loan limit that applies to a given zip code and year to the leave-one-out-mean sales price. We calculate the LTV ratio using the leave-one-out-mean sales price, for the same econometric reasons as with the other parameters, and we round the transaction-level LTV ratio to the nearest 5% to reduce idiosyncratic measurement error. Then, we aggregate to the zip code-by-year level. We obtain the discretized distribution of $\bar{\ell}$ as with the other parameters, weighting zip code-years by the number of transactions in the ZTRAX data set.

In an alternative calibration, we measure $\bar{\ell}$ as the modal LTV ratio within each zip code and year. This method assumes that the modal mortgaged homebuyer faces a binding collateral constraint. We again use the leave-one-out-mean sales price and round the transaction-level LTV ratio to the nearest 5%, which effectively discretizes the transaction-level distribution of LTV ratios, which, in addition to reducing idiosyncratic variation, is necessary to calculate the mode. Additionally, we exclude LTV ratios greater than 1 when calculating the mode. As with the calibration based on regulatory limits, we weight zip code-years by the number of transactions in the ZTRAX data set. This method gives an average mortgage-cash premium of 6.7%.

Details on Additional Parameters

Section III.D describes the calibration of the probability of transaction delay (q^d), the mortgage offer share (m), and the maintenance cost ($\tilde{\delta}$). Our baseline calibration does not feature nonfinancial contingencies: $q^c = h = 0$. We only observe these parameters as short monthly time series of less than three years in the NAR RCI data set. So, in the associated extensions, we simply use the average values. Accordingly, $q^c = 1\%$. We set $h = 7\%$ to match the share of all offers come with a home-sale contingency. Dividing by the share of offers that are mortgaged would raise h to 11.3%, and it would result in quantitative results within 0.2 pps of those under our main calibration of h.

The probability of delay (q^d) is measured as one minus the share of home sales not settled on time across months, based on the NAR RCI data set. The data are at the monthly frequency, and so the distribution used to obtain the grid points is purely time-series. In the text, we mentioned that the average time to issue an approval conditional on delay equals two months. Substantiating this value: in the NAR RCI data set, delays occur a quarter of the time, and, in Zillow's Time-to-Close data set, the average times to close conditional on lying below or above the 75^{th} percentile equal 37 days (around one month) and 64 days (around two months), respectively. Ellie Mae (2012) reports a similar unconditional average time to close for a successful mortgaged transaction of 47 days.

The mortgaged offer share (m) is measured as the average share of offers that are mortgage-financed, based on the Offer-Level data set aggregated to the MSA level. Since m is treated as a slow-moving parameter in our model, aggregating to the MSA level reduces the overall variance in the discretized distribution. In addition, this avoids issues of bunching around zero when aggregating to the zip code level, given the sparsity of data in some zip codes due to the data provider's low market share there. We weight MSAs by the number of offers when calculating the discretized distribution.

We measure the maintenance cost-to-value ratio ($\tilde{\delta}$) using the ratio of Zillow's ZORI rent index to Zillow's Home Value index, both of which are observed at the zip code by month level. This statistic approximates the user cost of housing, which, conceptually, is what we would like to observe. We transform the observed maintenance cost-to-value ratio, $\tilde{\delta}$, to the utility loss from home maintenance, δ, using the relationship

$$\delta = \frac{w^{1-\gamma}}{1-\gamma} - \frac{\left(w - \tilde{\delta}\right)^{1-\gamma}}{1-\gamma}.$$

Since utility is measured in units of dollars for a risk-neutral seller, $\tilde{\delta} = \delta$ when $\gamma = 0$. Table VII reports the annualized $\tilde{\delta}$ that comes from multiplying by 12.

Lastly, we parameterize the share of offers with a home-sale contingency, h, using the RCI data set over September 2017 to May 2020, which is the longest period over which we observe this statistic. So, the variation used to construct the distribution is purely time-series. As mentioned in the text, we set $h = 7\%$ to match the share of all offers come with a home-sale contingency. If, instead, we divide by the share of offers that are mortgaged to match our model's assumption that only mortgaged buyers rely on this contingency, then h rises to 11.3%. This has a minor 0.2 pps effect on the theoretical mortgage-cash premium.

Similarly, we parameterize q^c using the product of share of contracts terminated times conditional probability of inspection issue in terminated contract. We observe this statistic in the RCI data set for a short monthly time series. Given the sparsity of data on h and q^c and their limited variation in the data that we do have, we treat both parameters as fixed and do not attempt to calibrate across their distributions.

D.3. Heterogeneity in Theoretical Premium

The theoretical premium is calculated from the projection

$$\mu_g = \mu_0 + \sum_{\theta} \sum_{p} \mu_{\theta,p} \, \mathit{Tercile}(\theta, p)_g, \tag{IA103}$$

where g denotes grid points in the parameter distribution defined in Table VII, and the remaining notation is the same as in Sections V and VI. Figures 4, IA.4, and 5 plot the expected value of the empirical and theoretical premium across terciles of θ_k, holding the other parameters θ_{-k} at their average. The parameters are one minus the probability of transaction failure $(1 - q)$, the seller's total wealth net of current mortgage debt and a simultaneous down payment (ω), the monthly Poisson offer arrival rate, which also equals the expected number of offers (λ), and the ratio of the buyer's maximally levered liquid assets to housing value (L).

In terms of implementation, we calculate the projection in equation (IA103) across all feasible grid points in the parameter distribution, where, as described earlier in this section, feasibility is defined as a grid point with a positive and real-valued μ and κ. We define the indicators $\mathit{Tercile}(\theta, p)_g$ according to whether the grid point lies in one of the three ranges defined by the 35^{th} and 65^{th} of the parameter's discretized distribution.

We follow a similar methodology within the survey when calculating the theoretical premium across the distribution of q, shown in Figure 6. Note that each point in the theoretical premium in Figure 6 is an average across the values of all other parameters, holding q fixed at the value shown in the figure. We discretize the distribution of q using the terciles of the empirical distribution, based on the HMDA data set as described in Section V.B of the main article.

The other two model parameters that vary when calculating the theoretical premium in Figure 6 are w and κ. We do not observe w directly, and so we impute it using the respondent's log household income, which we do observe, and a projection of w onto log household income in the SCF data set. Accordingly, we impute $\hat{w}_k = 1.45 \log(Income_k) - 14.25$ for each survey respondent k. We calculate terciles of $\hat{w}$ and κ within the survey, weighting respondents by their Census weight for consistency. For simplicity, κ is treated as an exogenous parameter in the survey, with some respondents being told κ explicitly and others relying on their forecast that we elicit at the end of the survey, as described in Section III.E. Figure IA.9 traces out the distribution of the experimental mortgage-cash premium across $\hat{\mu}$ and κ.

Lastly, since the focus of the survey is to evaluate belief distortions, we also calculate terciles of the respondent's risk-aversion and numeracy score, following the same method as with $\hat{w}$ and κ. These are treated as parameters for the purposes of calculating the projection in equation (IA103).

D.4. Heterogeneity in Empirical Premium

Section VI of the main article describes how we semi-structurally estimate the empirical mortgage-cash premium across the distribution of four key model parameters: the probability of transaction failure (q), the seller's total net wealth (ω), the monthly Poisson offer arrival rate, which also equals the expected number of offers (λ), and the ratio of the buyer's maximally levered liquid assets to housing value (L). The text mentions how, sometimes, the data sets used in the model's calibration summarized in Table VII cannot be merged to our ZTRAX data set, as is the case for w, which relies on the anonymous SCF data set. Or, the intersection is so small that estimating equation (39) is infeasible, as is the case for λ. In these cases, we choose from alternative data sets placing a priority on data sets that vary by zip code and year with a large match rate with ZTRAX. As we clarify in the text, it suffices for the alternative measure to preserve the ranking of the parameter in question across zip codes and years, even if the absolute magnitude is inaccurate.

Since the SCF data set consists of an anonymized cross-section, we must construct an empirical proxy for ω and, in particular, w. We do so using the ratio of average household income to the leave-one-out-mean sales price, which is exactly how we measure the buyer's liquid assets, y. Section III.D.2 of this appendix provides full detail on how we construct this measure. Since the measures of ℓ_s and ξ used to calibrate the model are already at the zip code-by-year level, we can incorporate them almost as-is. We do not drop outliers because we are more interested in ranking zip code-years by a parameter's value than obtaining an accurate magnitude of that value. With these zip code-by-year measures of w, ℓ_s, and ξ, we can calculate ω at the zip code-by-year level using the relationship $\omega = w - \ell_s - \xi$.

We cannot directly incorporate the measure of λ used in our model calibration, since it derives from the offer-level data set and thus has a time dimension that is too limited relative to our core ZTRAX data set. Instead, we use the measure based on HMDA, described in Section III.D.2. This measure relies on the number of mortgage applications per origination in a zip code and year. Thus, repeating from Section III.D.2, it interprets an application as "an offer" and an origination as "a sale," and it implicitly assumes that the corresponding measure of offers-per-sale applies to both all-cash and mortgaged offers. The key advantage of this measure over other measures of λ is that we observe it at the zip code-by-year level over a cross-section that includes almost the entire U.S. and a time-series extending back to 1990.

Fortunately, we can incorporate the main measure of q from our model calibration into our regression with minimal modification. In particular, this measure is the zip code-by-year mortgage application denial rate from the HMDA data set, which can easily be merged to our ZTRAX data set. We must make two modifications. First, we cannot condition on pre-approval or first-lien status when calculating the zip code-by-year denial rate, as these variables are only collected by the administrators of HMDA for an inadequately small subset of years. The remaining conditions still apply (e.g., for-purchase loans). Second, since q governs the extensive margin of credit supply in our model, we must residualize the empirical measure against two measures of credit demand: the log of the applicant's income, and the log of the requested loan size. This residualized measure maps more directly to the model analogue than the raw mortgage denial rate, and so it is the correct measure to use in our regressions.

Lastly, the buyer's leverage capacity, L, depends on y and $\bar{\ell}$. Since the measures of these two variables used in our model calibration are already at the zip code-by-year level, we can incorporate them as-is.

For all empirical measures $\hat{\theta}_{z,t}$ described in this section, we construct the tercile indicators in equation (39) using terciles across zip code-years in our baseline ZTRAX data set.

D.5. Heterogeneity in Experimental Premium

We study heterogeneity in the experimental premium in Section VII.C of the main article following a similar approach as when using observational data. The text describes the most important parts of our methodology. We elaborate on details here. In particular, the parameters $\hat{\theta}$ used in equation (40) differ, by necessity, from those in the analogous equation (39) estimated using observational data. This difference reflects how we simply observe different parameters in the experimental data. The parameters are: (i) the financial wealth-to-housing ratio (w), which we impute using the respondent's income and a projection of wealth onto log income estimated in the 2016 SCF data set (see Section III.D.3), (ii) the cost of failure (κ), which equals 6% for respondents without ambiguity and equals the respondent's prior value elicited at the end of the survey for those facing ambiguity, (iii) risk-aversion, measured on a 1-to-4 scale following Fuster and Zafar (2021), and (iv) numeracy, based on the average difference between the respondent's answer and the correct answer on the survey's numeracy quiz described in Section IV. The terciles of these parameters are defined using their distribution within the experimental data.

E. Details on Belief Distortions

We provide details on the analytic and numerical results on belief distortions in Section VII.D of the main article.

Realization Utility

We focus on how realization utility affects the marginal decision to accept an all-cash versus mortgaged offer. This specific focus matches our paper's research question, but we imagine that realization utility would also make interesting predictions about when investors choose to list their home in the first place. To begin, consider a sale at price p relative to fundamental value. Let g denote the ratio of fundamental value to initial purchase price. The realized capital gain is

$$G(p) = pg\frac{1-\ell_s}{1-\ell_0}, \tag{IA104}$$

where ℓ_0 denotes the seller's initial LTV ratio. The term pg in equation (IA104) equals the return for a seller who initially bought all-cash, and the term $\frac{1-\ell_s}{1-\ell_0}$ equals a leverage multiple for mortgaged sellers. Since $\ell_s \leq \ell_0$ in most cases, leverage amplifies a seller's capital gain.

The seller experiences realization utility $R(G)$ upon sale, where $R'(G) > 0$ and R enters additively into the standard utility function. Such additivity is a common convention in models of reference dependence (e.g., Koszegi and Rabin (2006)). The value functions (4) to (5) become

$$V_r^{\mathcal{S}}(C) = u\left(\omega + p^C\right) + R(G^C), \tag{IA105}$$

$$V_r^{\mathcal{S}}(M) = -\delta\left(e^{-\alpha} + q^d\right) + e^{-\alpha-\rho}\left[(1-q)\left[u(\omega+p^M) + R(G^M)\right] + qV_r^{\mathcal{S}}(\emptyset)\right], \tag{IA106}$$

where the r subscript emphasizes that the value functions are those a seller with realization utility and G^F denotes the capital gain under a successful sale under method of financing F. The continuation value, $V_r^{\mathcal{S}}(\emptyset)$, has the same general form as equation (6), after substituting $V_r^{\mathcal{S}}(C)$ and $V_r^{\mathcal{S}}(M)$. Like in Proposition 1, the mortgage-cash premium, μ_r, solves $V_r^{\mathcal{S}}(C) = V_r^{\mathcal{S}}(M)$.

To highlight the intuition, let μ_r and κ_r denote the premium and cost of failure for a seller with realization utility. Let μ and κ (i.e., with no r subscript) denote the values for a seller with our baseline preferences. Consider the indifference conditions that define the mortgage-cash premium for the two types of sellers,

$$V_r^{\mathcal{S}}(M) - V_r^{\mathcal{S}}(C) = 0 = V^{\mathcal{S}}(M) - V^{\mathcal{S}}(C), \tag{IA107}$$

which reduces to

$$u(\omega + p^C e^{\mu_r}) = u(\omega + p^C e^{\mu}) - \left[R(G^M) - \frac{e^{\alpha+\rho}}{1-q}R(G^C)\right] + ... \tag{IA108}$$
$$... + \frac{q}{1-q}\left[u(\omega + e^{-\kappa}) - u(\omega + e^{-\kappa_r})\right].$$

Taking a first-order approximation of equation (IA108) around $\mu_r = \mu = 0$ and $G^M = G^C = 0$ and $\kappa_r = \kappa = 0$ yields

$$\mu_r\chi_0 = \mu\chi_0 + \frac{q\chi_1}{1-q}\left[\kappa_r - \kappa\right] - R'(0)\left[G^M - \frac{e^{\alpha+\rho}}{1-q}G^C\right] \tag{IA109}$$

$$\chi_0 = (\omega + p^C)^{-\gamma}p^C \tag{IA110}$$

$$\chi_1 = (\omega + 1)^{-\gamma} \tag{IA111}$$

or

$$\mu_r = \mu + \frac{q\bar{\chi}_0}{(1-q)}\left[\kappa_r - \kappa\right] - \bar{\chi}_1\left[G^M - \frac{e^{\alpha+\rho}}{1-q}G^C\right], \tag{IA112}$$

$$\bar{\chi}_0 = \frac{1}{p^C}\left(\frac{\omega+1}{\omega+p^C}\right)^{-\gamma} \tag{IA113}$$

$$\bar{\chi}_1 = \frac{R'(0)}{(\omega+p^C)^{-\gamma}p^C} \tag{IA114}$$

We clarify that equation (IA112) does not identify μ_r in closed form, since it also enters through κ_r and G^M.

Equation (IA112) conveys three ideas about how incorporating realization utility can affect the mortgage-cash premium. First, realization utility can lower the mortgage-cash premium if it raises sellers' value of having no offer, so that $\kappa > \kappa_r$. Barberis and Xiong (2012) and Ingersoll and Jin (2013) show in a more general dynamic setting that realization utility introduces option value from of waiting for a potentially large capital gain. Applying this insight to our setting, it seems plausible that a seller with realization utility finds it less painful to have no offer because she retains the option value of a large capital gain from a bidding war: $\kappa > \kappa_r$.

Second, realization utility can also lower the mortgage-cash premium if the expected capital gain from a mortgaged offer exceeds that of an all-cash offer: $(1-q)e^{-(\alpha+\rho)}G^M > G^C$. By definition, a positive mortgage-cash premium implies $G^M > G^C$. Therefore, a seller with realization utility finds mortgaged offers attractive simply because they have the potential for a higher expected capital gain, which, in equilibrium would actually reduce the mortgage-cash premium relative to a seller with our baseline preferences. Of course, what matters is the expected capital gain, and, so, if $q > 1 - \frac{e^{\alpha+\rho}G^C}{G^M}$, then a seller with realization utility requires a higher mortgage-cash premium.

A final remark concerns the effect of leverage on the seller's realized capital gain. The amplification term $\frac{1-\ell_s}{1-\ell_0}$ in equation (IA104) implies that a home seller who levered her initial equity contribution may experience a large capital gain under any method of sale, regardless of whether the buyer is mortgaged or all-cash. Note from equation (IA108), however, that realization utility affects the mortgage-cash premium through the relative utility burst from a capital gain on a mortgaged versus an all-cash offer, not the overall level of this burst. So, while leverage can indeed amplify the level of the seller's capital gain, and thus the level of the utility burst, it may not affect the mortgage-cash premium insofar as it raises the burst from a mortgaged versus all-cash capital gain by a similar amount. This would obtain if, for example, both G^M and G^C are large (e.g., due to the seller's leverage) and R is concave in the gain region, as in Ingersoll and Jin (2013).

Putting the previous three remarks together, it is theoretically unclear how accounting for realization utility would affect the magnitude of the mortgage-cash premium puzzle. In particular, a model with realization utility could feature a lower theoretical premium and thus a larger puzzle.

In the specific case where $G^C = 0$ and $R(G^C) = 0$, however, we would expect that any mortgaged offer that would yield positive capital gain, $G^M > G^C = 0$, should require a lower mortgage-cash premium when sellers have realization utility than when sellers have the baseline preferences. In particular, equation (IA108) implies that sellers with realization utility like how such a mortgaged offer gives them at least some chance of realizing a capital gain. We evaluate this hypothesis experimentally by randomizing respondents to different thought experiments in which the realized capital gain from an all-cash offer varies from 0% to 100%, on a levered basis. Figure IA.8 tests for a difference in the mortgage-cash premium between respondents for whom the all-cash offer results in zero capital gain versus a positive gain. The results in Figure IA.8 show no difference in the premium between the two subsamples, which is inconsistent with the hypothesis predicted by realization utility.

We conclude this section by reiterating that realization utility may very well affect home sellers' behavior in ways distinct from their preference for mortgaged versus all-cash offers. For example, realization utility discourages sellers from listing their home at a capital loss, which may explain the well-known disposition effect in real estate (e.g., Genesove and Mayer (2001)). This channel would significantly impact the timing of sale, even though it may have a small impact on the price premium of a mortgaged versus all-cash transaction.

Probability Weighting

Probability weighting refers to the tendency to perceive small probabilities as larger than they actually are when making decisions and, conversely, to perceive large probabilities as smaller. This mechanism compresses both very small and very large probabilities toward some central value. Thus, it predicts that home sellers will require an exceptionally large premium when transaction risk is low. Empirically, however, we find the largest puzzle at high levels of risk (e.g., Figures 4 and 6).

A number of economics papers have proposed and experimentally verified specific functional forms that map the probability an individual has in mind onto the probability they use in optimization. We recalculate μ from Proposition 1 using several well-known weighting functions. First, Tversky and Kahneman (1992) propose the following functional form for probability weighting,

$$q_s^W = \frac{q_s^\alpha}{\left[q_s^\alpha + (1-q_s)^\alpha\right]^{\frac{1}{\alpha}}}. \tag{IA115}$$

The corresponding mortgage-cash premium equals

$$\mu_s^W \approx \xi + \frac{1}{\gamma-1}\left[q_s^W\left(\left(\frac{1-\ell-\delta}{e^{-\kappa-\sigma}-\ell-\delta}\right)^{\gamma-1}-1\right)-e^{r(1-\gamma)}+1\right]. \tag{IA116}$$

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

We also consider the two-parameter function proposed by Prelec (1998),

$$q_s^{W'} = e^{-\bar{\alpha}_0[-\log(q_s)]^{\bar{\alpha}_1}}. \tag{IA117}$$

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

We parameterize $a_{TK} = 0.65$ to match the estimate in Tversky and Kahneman (1992). This value lies close to the midpoint of subsequent estimates summarized by Booij, Praag, and Kuilen (2010), and it also equals the value used by Barberis, Jin, and Wang (2021). More recently, Bernheim and Sprenger (2020) estimate $\alpha = 0.71$, and so we view the interval [0.6, 0.75] as a reasonable range. Our results are robust to all values of a_{TK} within this range. We parameterize $a_{PR} = 1.08$ and $b_{PR} = 0.53$ based on the most aggressive estimates summarized by Booij, Praag, and Kuilen (2010), which come from Bleichrodt and Pinto (2000).

Figure IA.10 shows the result, plotting the theoretical premium under probability weighting across the distribution of q with the experimental premium. While probability weighting helps reduce the fairly small puzzle that exists low levels of q, it does little to resolve the much larger puzzle that obtains when q is high.

Present Focus

Repeating from the text, present focus refers to a preference for immediate gratification that leads to dynamically inconsistent decisions (e.g., Ericson and Laibson (2019), O'Donoghue and Rabin (2015)). It is commonly modelled by incorporating a quasi-hyperbolic time discount factor (e.g., Laibson (1997), O'Donoghue and Rabin (1999)). At first glance, present focus would seem like a plausible explanation for the large mortgage-cash premium. For example, when asked to describe their preference for an all-cash buyer, most survey respondents select: "Even if the Mortgaged Buyer would never back out, the Cash Buyer would close more quickly and end the stressful process of selling my home" (Table IA.XIII). The desire to "end the stressful process" may relate to the stress induced by uncertainty, consistent with how respondents facing ambiguity select this reason at an 11 pps higher rate. The attractiveness of "closing more quickly" would loosely suggest a role for present focus.

However, two features of our setting make present focus an inappropriate explanation. First, present focus concerns consumption, as distinct from wealth. Since sellers in our model primarily maximize indirect utility over wealth, we cannot properly incorporate present focus without assuming, for example, extreme liquidity constraints. Second, present focus concerns a distinction between the immediate present ("now") and the future ("later"). Realistically, an all-cash offer can close quickly but not immediately. Nevertheless, the "now" period may last longer in the home-sale context, and so we test for present focus in our survey by randomly presenting a subset of survey respondents with a thought experiment in which all-cash offers close "in four weeks," whereas they close "any time within two weeks" for the remaining respondents. If the "now" period indeed lasts as long as two weeks, then we should find a significantly lower mortgage-cash premium for the "four week" cohort. Figure IA.8 finds no such difference. We conclude that present focus cannot explain the mortgage-cash premium.

IV. Survey Appendix

In this appendix, we elaborate on the description of the experimental survey in Section A of the main article. We describe how we administer the survey in Section IV.A. In Section IV.B, 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_eu4yrLgIgbPYhBc.

A. *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 in the first two waves are paid \$2.00 for completing the survey and take an average of six minutes to do so. Survey respondents in the third wave are paid \$2.25 to account for inflation and take the same average time. The text referenced a \$24 average hourly wage. This value equals the average of the average hourly wage for the first (\$21.88), second, (\$22.78), and third (\$27.75) waves.

We administered the survey's first wave in April 2021, the second wave in November 2021, and the third wave in January 2023. All respondents in both surveys reside in the U.S. and own their home. We drop a small number of participants who are recruited despite not satisfying these two conditions. The three waves of the survey comprise three repeated cross-sections, not a panel. We retain respondents who spend at least two minutes on the survey and are at least twenty years old, which reduces the sample sizes to 1,019 and 1,202 and 1,199 in the three waves, respectively, as shown in Table VIII.

B. *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 nonproportional thinking, and a similar technique is used in the Federal Reserve Bank of New York's Survey of Consumer Expectations (Liu and Palmer (2021)). For reference, the share of respondents who fall into these five bins are, respectively, 1.2%, 38.0%, 41.3%, 16.5% and 2.9%. This frequency distribution rather closely matches the analogous distribution of real house prices in the ZTRAX data set, in which the corresponding shares of transactions are 1.7%, 43.6%, 32.4%, 15.7%, and 6.6%. This similarity supports the relevance of our survey results for the rest of the paper.

Description of Thought Experiment

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

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.[7] For respondents in the second and third waves, the share of the list price at which the home was bought ten years ago randomly equals either 0.5 or 0.95, with 50% probability for each value. The quantity \$[$0.3 \times B$] corresponds to $\ell_s v$ in the model.

The third screen introduces the time deadline. We tell respondents: *"To make the down payment on the new home, you will need to finalize the sale of your current home, pay off the balance, and then use the money that's left over. The down payment is \$[$0.15 \times B$]. That is 15% of the price at which you have listed your current home. You must make this down payment within [T] weeks."* The parameter T randomly equals either 6 or 8, with 50% probability for each value. To match our theoretical framework, T will also equal the time it will take to close a mortgaged transaction. Accordingly, the values of T are chosen such that the difference between the all-cash and mortgage closing periods equals one month (Ellie Mae (2012)), as we describe shortly. The quantity \$[$0.15 \times B$] corresponds to ξv in the model.

[7]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."*

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

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

The sixth screen describes a mortgaged offer: *"The second buyer has borrowed money from a mortgage lender, whom we'll call the 'Mortgaged Buyer'. If you accept the offer from the Mortgaged Buyer, it will take [T] weeks to close the transaction. There is a chance that the Mortgaged Buyer will not be able to secure money from their lender by the end of the [T]-week period. If that happens, then you will need to relist your home in [T] weeks."* This statement pertains to respondents in the first wave, a 50% random sample of the second wave, and a 40% random sample in the third wave. The remaining respondents are given a distribution of outcomes for the mortgage transaction: *"The second buyer has borrowed money from a mortgage lender, whom we'll call the 'Mortgaged Buyer'. If you accept the offer from the Mortgaged Buyer, it will take [T] weeks to close the transaction. There is a [$100 \times q$] percent chance that the Mortgaged Buyer will not be able to secure money from their lender by the end of the [T]-week period. If that happens, then you will need to relist your home in [T] weeks. To attract another offer, you would also need to cut your list price by 6%, that is, reduce your list price to$[$0.94 \times B$] from $[B]."* In the second wave, $q = 7\%$. In the third wave, q equals 1%, 7% and 13% with equal probability, which approximately maps to the empirical distribution of q based on the HMDA data set.

We avoid overwhelming survey respondents by either providing a fixed value of $\kappa = 6\%$, or not specifying the value of κ at all. Indeed, Lemma 1 shows how κ requires a complicated, dynamic calculation. Fixing κ and interpreting it as an actual price cut enables our respondents to focus their attention on q, which, per our analysis of observational data, is an important driver of the mortgage-cash premium. We do not provide a value of κ to respondents in the first wave and those who face ambiguity in the second wave are not given a price cut. Interestingly, when asked to report the price cut (i.e., κ) they would impose, the average value of 4.6% lies close to 6%.

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, and her beliefs about the probability of transaction failure (i.e., q) and the price cut after failure (i.e., κ).

The first core question asks respondents: *"Suppose that both the Mortgaged Buyer and the Cash Buyer offer to pay $[B]. Which offer would you accept today?"* The Cash Buyer's offer dominates the Mortgaged Buyer's offer in our experiment, and so we expect that respondents will answer that they prefer the Cash Buyer. Reassuringly, 97% of respondents do so. For a random 50% of participants in the survey's second wave, both buyers offer to pay 95% of the list price. We include this randomization to test for loss aversion relative to the list price.

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

1. *"If the Mortgaged Buyer backs out and I relist at a lower price, I may not have enough money to meet my mortgage and moving expenses."*

2. *"If the Mortgaged Buyer backs out and I relist at a lower price, I will sell the home at a loss relative to my target price."*

3. *"Even if the Mortgaged Buyer would never back out, the Cash Buyer would close more quickly and end the stressful process of selling my home."*

4. *"Other (please describe)"*

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

1. *"The price at which I bought my home."*

2. *"The price at which my home is currently listed."*

3. *"Other (please describe)"*

The order of the first two of these options is again randomized. Table IA.XIII summarizes the responses.

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

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

$$\begin{aligned} Premium_k = \frac{1}{2}\left[\min\left\{\tilde{\mu}\middle|\, \textit{Mortgaged Offer} \succ \textit{All-Cash Offer}\right\} + \ldots\right. \\ \left.\ldots + \max\left\{\tilde{\mu}\middle|\, \textit{All-Cash Offer} \succ \textit{Mortgaged Offer}\right\}\right]. \end{aligned} \tag{IA1}$$

We assign a missing value to the 2% of respondents who exhibit multiple switch points, following convention (e.g., Bernheim and Sprenger (2020)). Moreover, we assign a missing value to respondents who never switch to the mortgaged offer. On the one hand, such respondents could indeed require a mortgage-cash premium that exceeds 28% in the second and third waves or 20% in the first wave. For such respondents, we would like to top-code the mortgage-cash premium. On the other hand, such respondents may not understand the experiment, in which case we would like to drop them. Therefore, our decision to assign a missing value plausibly leads to conservative estimates of the mortgage-cash premium.

The remaining screens in the survey's core elicit the respondent's beliefs about transaction risk and asks them to perform basic arithmetic calculations. First, we ask respondents: *"What percent of the time do you think the Mortgaged Buyer will back out of the transaction because they fail to secure money from their lender?"* We bound the answers to lie within a range by allowing respondents to select a value between 0% and 30% on a sliding scale. As mentioned in the text, 41% of respondents select a value less than 10% or greater than 20%, suggesting that the scale does not lead to bunching around the midpoint. Let $\textit{Failure Probability}_k$ denote the respondent's prior probability, or, for respondents facing a given distribution, let $\textit{Failure Probability}_k$ equal the value to which the respondent is assigned. We then assess numeracy and attentiveness following the commonly used approach of Lipkus, Samsa, and Rimer (2001), asking: *"If indeed the Mortgaged Buyer backs out of the transaction [$\textit{Failure Probability}_k$]% of the time, then how many times out of 1,000 will the Mortgaged Buyer back out?"*. Respondents answer this question by supplying a number.

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

We construct a numeracy score − or, alternatively, an attentiveness score − for respondent k as follows. First, we calculate the absolute error for both the question based on q and the question based on κ. Then, we standardize the error in each question to have a mean of zero and variance of one. We do so separately for respondents facing a given distribution and those facing ambiguity, since the question based on κ differs slightly between the two groups. Then, we take the average of the standardized error across the two questions for each respondent. This average defines the numeracy score. Lower values correspond to greater accuracy, which may either be interpreted as greater numeracy or sharper attention.

Background Questions

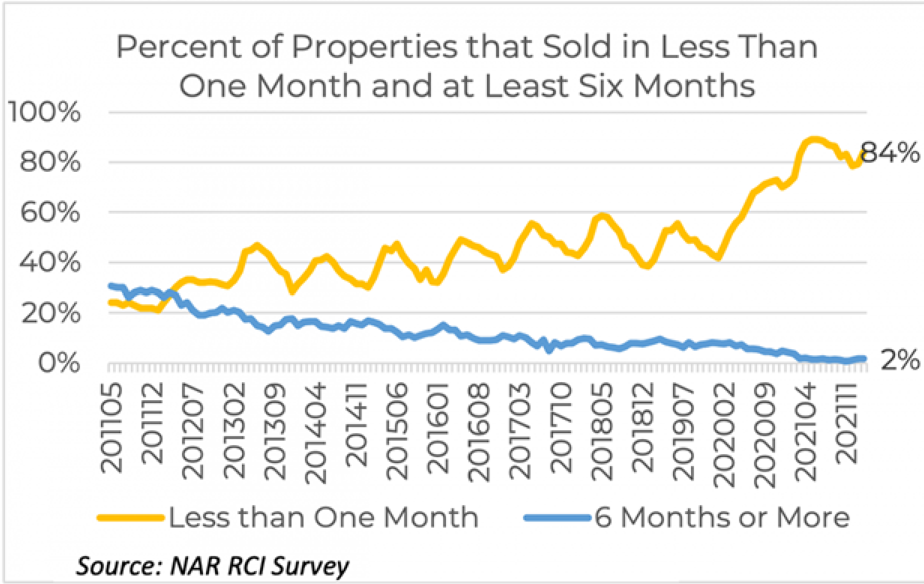
The remainder of the survey asks respondents background questions related to their demographic characteristics, risk attitude, and experience in real estate markets. The first four questions ask the respondent to provide her age, approximate annual household income, state of residence, and highest level 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:

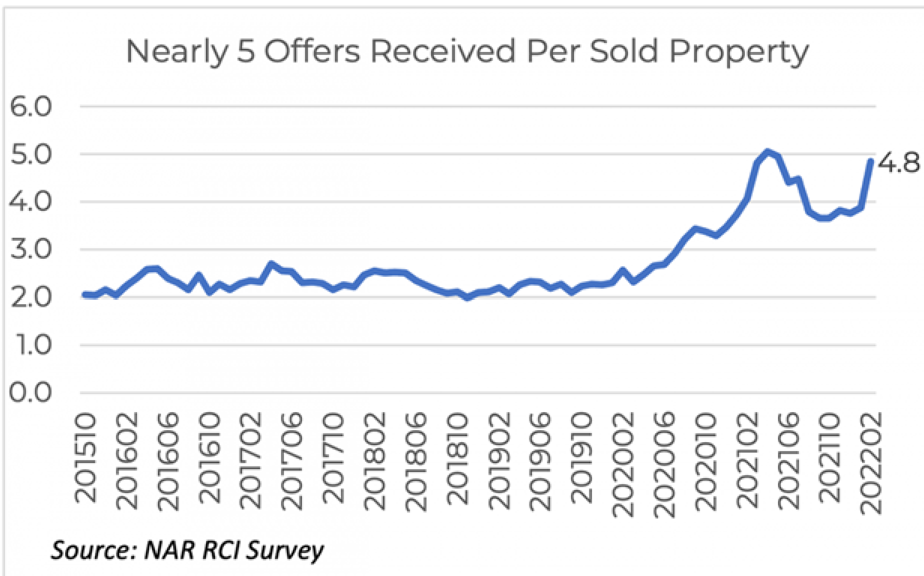
1. *"I am always willing to take risks."*
2. *"I am usually willing to take risks, but I am sometimes reluctant to do so."*
3. *"I am usually reluctant to take risks, but I am sometimes willing to do so."*
4. *"I am never willing to take risks."*

We code a respondent's risk-aversion on a scale of one to four based on her chosen response. Pooling the three waves, the distribution across these four responses is 5.5%, 32.6%, 56.0%, and 5.8%, respectively.

As described in Section III.D.3 we use the respondent's household income to impute financial wealth. We do not elicit wealth directly, since doing so would likely result in substantial idiosyncratic variance from measurement error, as calculating one's wealth is typically more complicated than doing so for income. Repeating from Section III, we use the SCF data set to first project the financial wealth-to-housing ratio (w) onto log household income. The estimated mapping is $\hat{w}_k = 1.45 \log(Income_k) - 14.25$, which we calculate for each survey respondent k. Table VIII reports average seller (i.e., survey respondent) wealth based on this mapping.

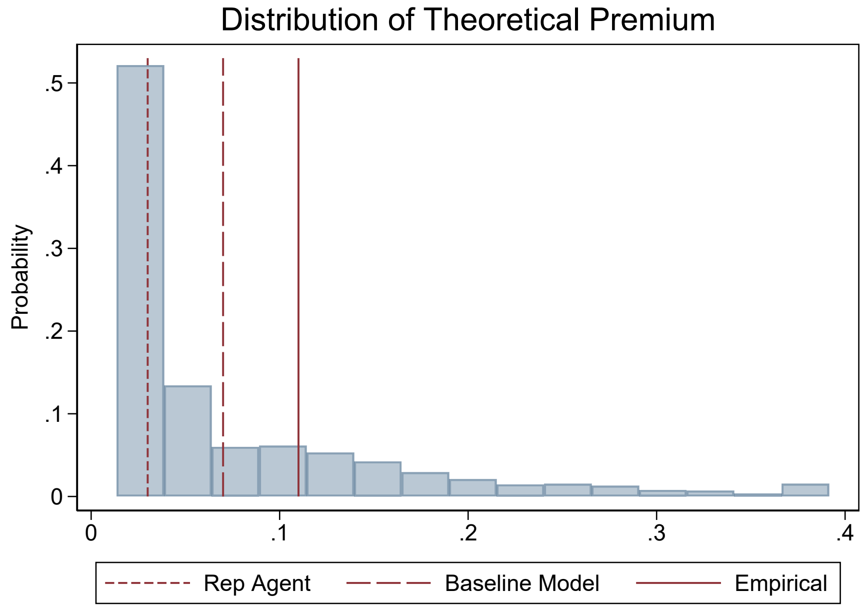


Panel A. Time to Sell



Panel B. Offers per Sold Home

Figure IA.1. Time series of time to sell and number of offers. Data are from the National Association of Realtors (Cororaton (2022)).



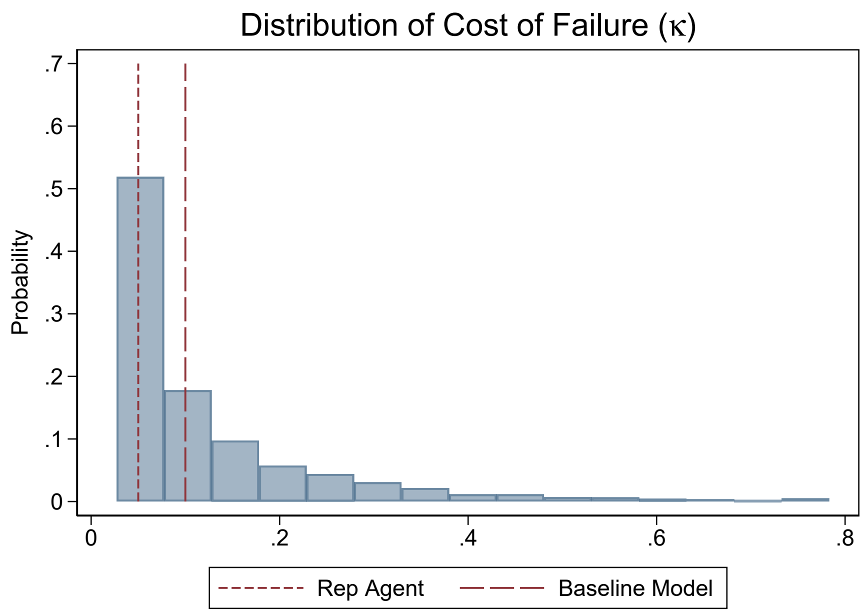


Figure IA.2. Distribution of the theoretical mortgage-cash premium and cost of failure. This figure plots the distribution of the theoretical mortgage-cash premium, μ, and seller's endogenous cost of failure, κ, across the distribution of model parameters. The expressions for μ and κ are shown in Proposition 1. The distribution of parameters is reported in Table VII. The red vertical lines in Panel A mark: the theoretical mortgage-cash premium obtained when all parameters equal their average values, as reported in Table I (Rep Agent), the average theoretical premium across the parameter distribution, as reported in Figure 3 (Baseline Model), and the baseline estimated premium, as reported in Table IV (Empirical). The red vertical lines in Panel B have a similar interpretation in terms of the seller's cost of failure.

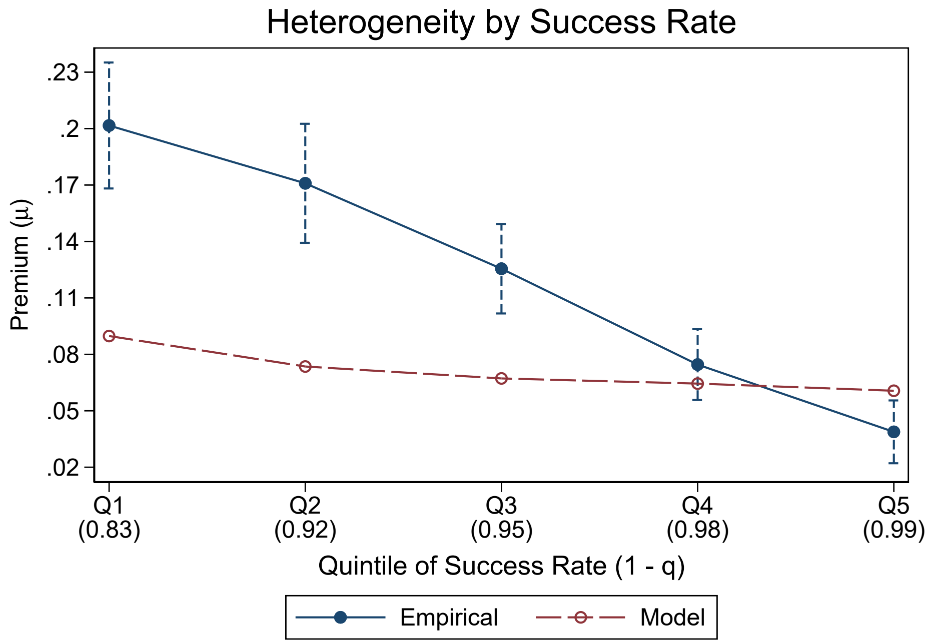


Figure IA.3. Heterogeneity with quintile discretization. This figure is analogous to Figure 4 when discretizing the distribution of each parameter into quintiles, as opposed to terciles. The remaining notes are the same as in Figure 4.

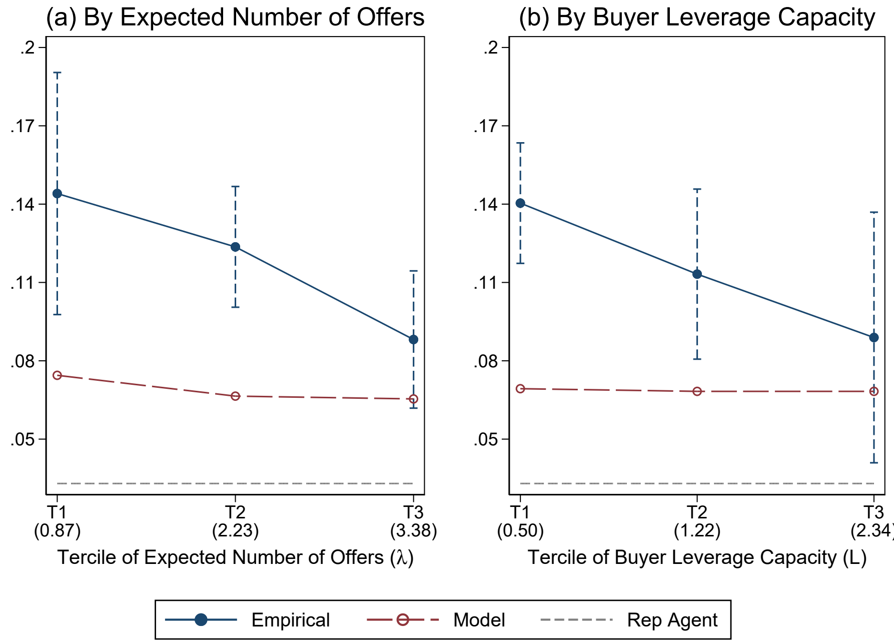


Figure IA.4. Heterogeneity by parameters that work through continuation value. This figure builds on Figure 4 by plotting the empirical and theoretical premium across the distribution of parameters that affect the premium through the seller's continuation value, κ, but do not affect it directly. The parameters are the monthly Poisson offer arrival rate, which also equals the expected number of offers (λ), and the ratio of the buyer's maximally levered liquid assets to housing value (L). The slopes of the theoretical premium with respect to λ and L are negative. In particular, the values of the theoretical premium across the three terciles of λ are: 7.44%, 6.65%, and 6.54%, respectively. The values of the theoretical premium across the three terciles of L are: 6.93%, 6.83%, and 6.83%, respectively. The remaining notes are the same as in Figure 4.

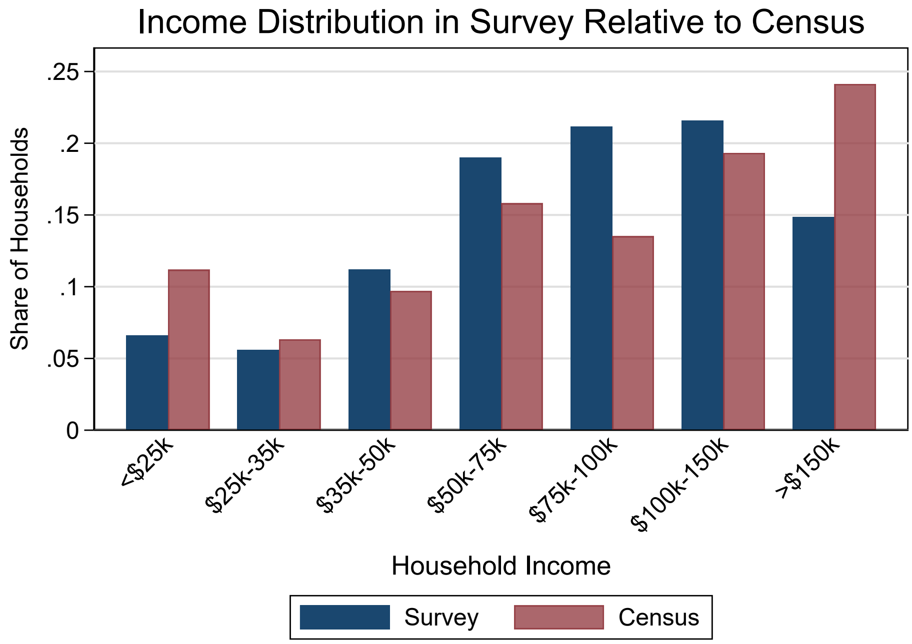


Figure IA.5. Representativeness of survey respondents by income. This figure plots the distribution of survey respondents studied in our main analysis across income bins specified by the 2020 Census. The dark blue bars plot the share of respondents in each bin. The light red bars plot the share of all U.S. homeowners in each bin. Data are from the survey described in Section VII of the main article.

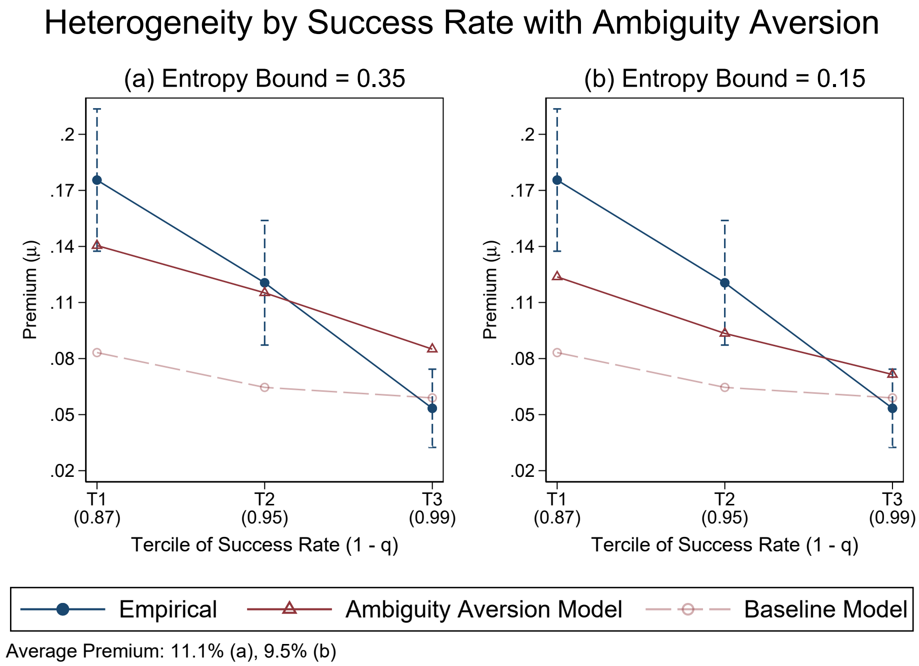


Figure IA.6. Heterogeneity under model with ambiguity-aversion. This figure is analogous to Figure 4 under an extension of the baseline model that features ambiguity-aversion (i.e., robust decision-making). As described in Section VII.C.1 of the main article, an ambiguity-averse home seller who approximates the mortgage transaction failure rate as $\hat{q}$ acts according to the worst-case failure rate $Q > \hat{q}$, where Q solves

$$\overline{\mathcal{R}} = Q \log\left(\frac{Q}{\hat{q}}\right) + (1 - Q) \log\left(\frac{1 - Q}{1 - \hat{q}}\right),$$

and $\overline{\mathcal{R}}$ has the interpretation of the seller's ambiguity-aversion. Specifically, $\overline{\mathcal{R}}$ is the largest permissible entropy between $\hat{q}$ and Q. Panels A and B plot the theoretical premium for entropy bounds of $\overline{\mathcal{R}} = 0.35$ and $\overline{\mathcal{R}} = 0.15$, respectively. The bottom of the figure notes the average theoretical premium under each value of $\overline{\mathcal{R}}$. The remaining notes are the same as in Figure 4.

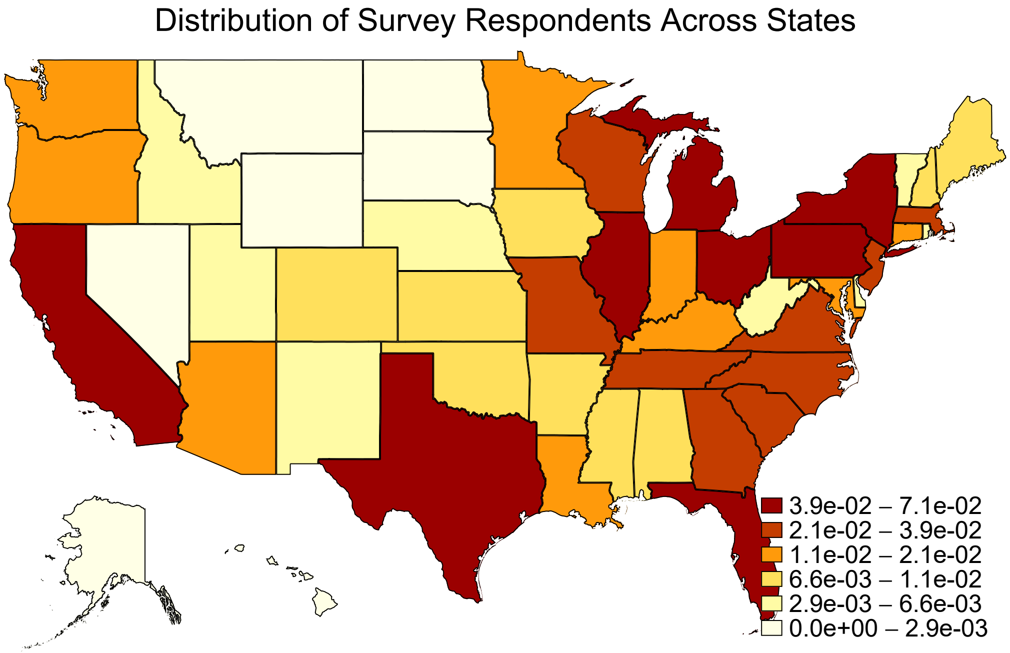


Figure IA.7. Geographic representativeness of survey respondents. This figure plots the share of survey respondents studied in our main analysis from each U.S. state. Warmer colors correspond to a larger share. Data are from the survey described in Section VII of the main article.

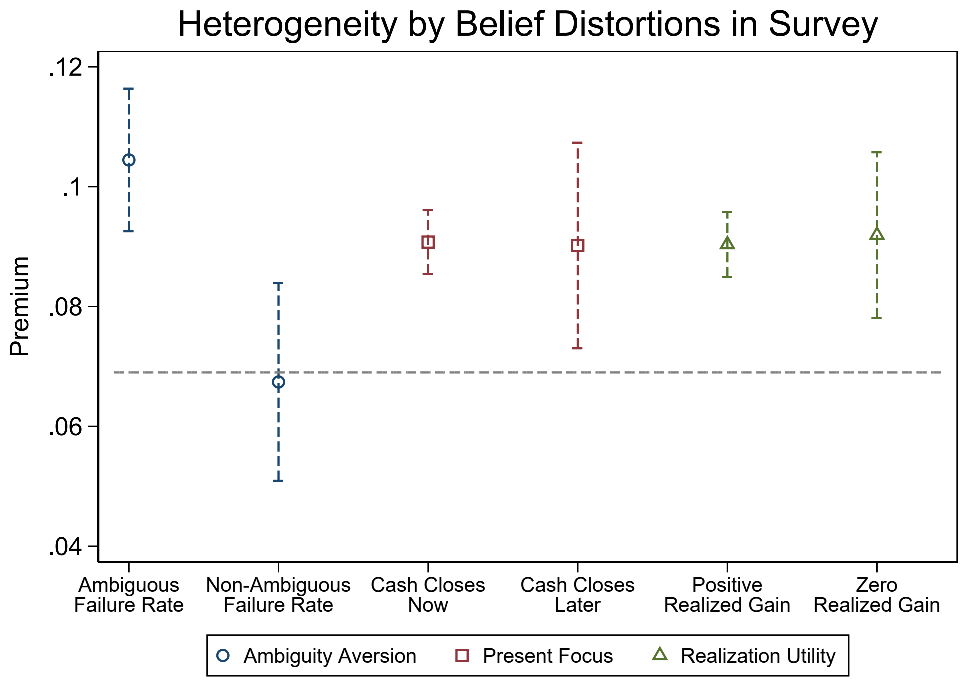


Figure IA.8. Additional belief distortions and the experimental premium. This figures summarizes tests of how various belief distortions affect the experimental mortgage-cash premium. There are three tests, which correspond to three separate regressions of a similar form as equation (40)

$$Premium_k = \mu_0 + \Delta\mu^B B_k + \sum_\theta \sum_p \mu_{\theta,p} Tercile(\hat{\theta}, p)_k + \epsilon_k,$$

where B_k indicates whether a given belief distortion is relatively strong. The remaining notation is the same as in Figure 6, except that q is now included in the index of parameters, θ, instead of writing it explicitly. The left region of the figure tests ambiguity-aversion, and B_k indicates whether k is told the distribution of q ($B_k = 1$) versus facing ambiguity ($B_k = 0$). The central region of the figure tests present bias, and B_k indicates whether k is told that the all-cash offer will close "in four weeks" ($B_k = 1$) versus "any time within two weeks" ($B_k = 0$). The right region of the figure tests realization utility, and B_k indicates whether k is given a thought experiment in which the all-cash offer would result in zero capital gain ($B_k = 1$) versus a positive gain ($B_k = 0$). Capital gains are calculated on a levered basis. The definition of B_k is such that the respective theories predict a lower experimental mortgage-cash premium when $B_k = 1$. The gray dashed line marks the average theoretical premium from Figure 3. The remaining notes are the same as in Figure 6.

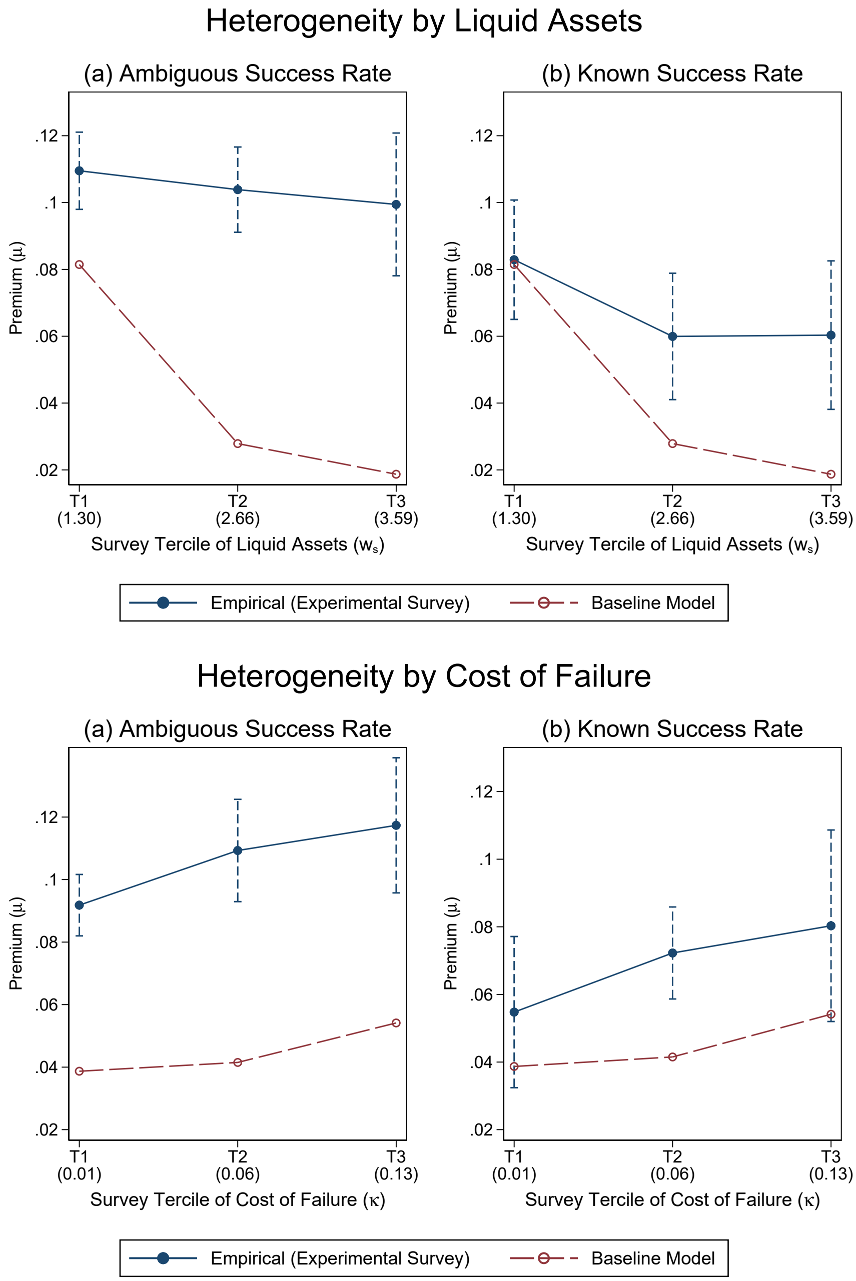


Figure IA.9. Additional heterogeneity in the experimental premium. This figure is analogous to Figure 6 in terms of the seller's cost of failure κ and the seller's financial wealth w. Additional Details on calculation of the theoretical premium are in the note to Table VIII.

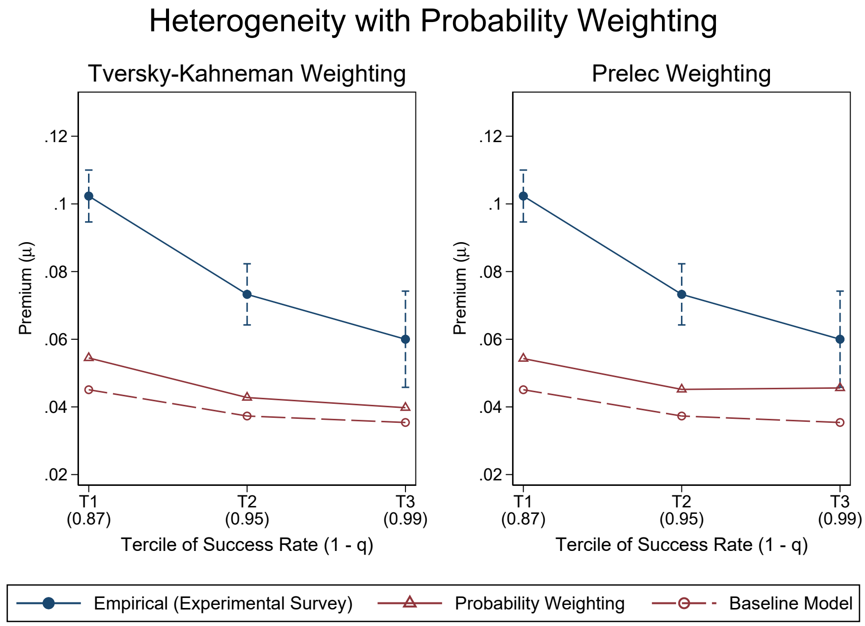


Figure IA.10. Theoretical premium with probability weighting. This figure plots the empirical and theoretical premiums across the distribution of the transaction success rate, using data from the experimental survey and a probability weighting function, $Q(q)$ to calculate the theoretical premium. The left panel uses the Tversky and Kahneman (1992) weighting function,

$$Q(q) = \frac{q^{a_{TK}}}{\left[q^{a_{TK}} + (1-q)^{a_{TK}}\right]^{\frac{1}{a_{TK}}}}.$$

The right panel uses the Prelec (1998) function,

$$Q(q) = e^{-a_{PR}[-\log(q)]^{b_{PR}}}.$$

We parameterize $a_{TK} = 0.65$, $a_{PR} = 1.08$, and $b_{PR} = 0.53$ following the literature. Section III.E provides detail. The blue solid lines plot the estimated coefficients $\{\mu_{q,p}\}$ from the regression equation

$$Premium_k = \sum_p \mu_{q,p} Tercile(\hat{q}, p)_k + \sum_\theta \sum_p \mu_{\theta,p} Tercile(\hat{\theta}, p)_k + \epsilon_k,$$

where the notation is the same as in Figure 6. The remaining notes are the same as in Figure 6.

Table IA.I.
Robustness to Data Set

P-values are in parentheses. This table estimates equation (28) using various data sets, which assesses the external validity of the baseline results. Subscripts i and t index property and month. All data sets are property-by-month panels that derive from the ZTRAX database. The data set used in columns (1) and (2) consists of all purchases in the ZTRAX database, after the imposing the basic filters described in Section I.A (ZTRAX Universe). The data set used in columns (5) and (6) consists of all purchases in a 25% random sample of the ZTRAX Universe, which is the paper's core data set (Core Data Set). The data set used in columns (3) and (4) consists of all purchases in the ZTRAX Universe that are also in a zip code-by-month bin that lies in the Core Data Set. The data set used in columns (7) and (8) consists of all purchases in the CoreLogic database, after the imposing the basic filters described in Section I.A (CoreLogic). The CoreLogic database relies on the same underlying public records as the ZTrans component of the ZTRAX database. The hedonic characteristics used in column (8) is an indicator for whether the property is a detached single-family home, as we describe in Section I.C. Columns (2), (4), (6), and (8) include a property fixed effect, while the remaining columns do not. The remaining notes are the same as in Table IV.

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

Table IA.II.
Robustness to Controlling for Buyer and Seller Characteristics

P-values are in parentheses. This table estimates a variant of equation (28) that controls for characteristics of the seller and buyer, which accounts for the possibility that parties in all-cash transactions hold lower private valuations than parties in mortgaged transactions. Subscripts i and t index property and month. Column (1) controls for seller characteristics: $\textit{High Seller LTV}_{i,t}$ indicates whether the seller's LTV 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, $\textit{Same-Month Purchase}_{i,t}$ indicates whether the seller purchases another home in the same month, and $\textit{Foreign Seller}_{i,t}$ indicates whether the seller has a foreign address. Column (2) controls for buyer characteristics: $\textit{Flip}_{i,t}$ indicates whether the property is subsequently sold within 12 months, $\textit{Foreign Buyer}_{i,t}$ indicates whether the buyer has a foreign address, $\textit{Same-County Buyer}_{i,t}$ indicates whether the buyer's address is in the same county as the property, and $\textit{Institutional Buyer}_{i,t}$ indicates whether the buyer is an institution and the property is not to be owner-occupied. Column (3) for $\textit{Cash Propensity}_{b(i,t)}$, defined as an indicator for whether the buyer of property i in month t, denoted $b(i,t)$, buys another home all-cash over our sample period. With the exception of $\textit{Mortgaged}_{i,t}$, all indicator variables are assigned a value of zero when the raw variable is unobserved. The remaining notes are the same as in Table IV.

Outcome:	$\log(\textit{Price}_{i,t})$		
	(1)	(2)	(3)
$\textit{Mortgaged}_{i,t}$	0.113	0.100	0.129
	(0.000)	(0.000)	(0.000)
Seller Characteristics:			
$\textit{High Seller LTV}_{i,t}$	-0.036	-0.049	-0.047
	(0.000)	(0.000)	(0.000)
$\textit{Same-Month Purchase}_{i,t}$	0.048	0.048	0.056
	(0.000)	(0.000)	(0.000)
$\textit{Foreign Seller}_{i,t}$	-0.115	-0.125	-0.118
	(0.255)	(0.213)	(0.236)
Buyer Characteristics:			
$\textit{Flip}_{i,t}$		-0.049	-0.062
		(0.000)	(0.000)
$\textit{Foreign Buyer}_{i,t}$		0.027	0.043
		(0.635)	(0.442)
$\textit{Same-County Buyer}_{i,t}$		-0.047	-0.039
		(0.002)	(0.009)
$\textit{Institutional Buyer}_{i,t}$		0.204	0.188
		(0.096)	(0.117)
$\textit{Cash Propensity}_{b(i,t)}$			0.101
			(0.000)
Zip Code-Month FE	Yes	Yes	Yes
Hedonic-Month FE	Yes	Yes	Yes
Property FE	Yes	Yes	Yes
R^2	0.907	0.907	0.908
Number of Observations	426,256	426,256	426,256

Table IA.III.
Robustness to Regulatory Appraisal Discontinuity as an Instrumental Variable

P-values are in parentheses. This table estimates equation (28) after instrumenting for $Mortgaged_{i,t}$ using an indicator for whether the home is predicted to be sold above the \$250,000 floor at which federal regulators require bank-originated loans to come with an appraisal. The predicted price equals

$$Predicted\ Price_{i,t} = Last\ Price_{i,t} \times \frac{FHFA\ Index_{z(i),t}}{Last\ FHFA\ Index_{z(i),t}},$$

where $Last\ Price_{i,t}$ denotes the price at which the home sold for in its most recent transaction before t, $FHFA\ Index_{i,t}$ denotes the FHFA All-Transaction Price Index in the associated zip code $z(i)$ and month t, and $Last\ FHFA\ Index_{z(i),t}$ denotes the analogous index in the month of the most recent transaction. The main instrument is $Appraisal\ Required_{i,t}$, which is an indicator for whether $Predicted\ Price_{i,t}$ is greater than or equal to \$250,000. Column (1) validates the existence of a first stage by regressing $Mortgaged_{i,t}$ on $Appraisal\ Required_{i,t}$. Columns (2) to (4) report the second-stage results. To ensure the strength of the instrument set, the first-stage regression equation is

$$Mortgaged_{i,t} = \sum_t \psi_t \left(Appraisal\ Required_{i,t} \times X_{i,t}\right) + \tilde{\chi} \log\left(Predicted\ Price_{i,t}\right) + \tilde{\zeta}_{z(i)} + \tilde{\tau}_t + \tilde{\epsilon}_{i,t},$$

where $X_{i,t}$ consists of the hedonic characteristics in Table IV and a vector of quarter fixed effects, which are included as interactions to improve the strength of the first stage (e.g., Murray (2006)). The second-stage regression equation is

$$\log\left(Price_{i,t}\right) = \mu \widehat{Mortgaged}_{i,t} + \chi \log\left(Predicted\ Price_{i,t}\right) + \zeta_{z(i)} + \tau_t + \epsilon_{i,t}.$$

Observations in the second stage are restricted to those with a predicted price that lies within a ±5%, ±10%, or ±15% bandwidth of the \$250,000 cutoff. The table drops observations for which the predicted price differs from the actual price by more than the longest bandwidth length (35%). The remaining notes are the same as in Table IV.

Outcome:	$Mortgaged_{i,t}$	$\log\left(Price_{i,t}\right)$	$\log\left(Price_{i,t}\right)$	$\log\left(Price_{i,t}\right)$
	(1)	(2)	(3)	(4)
$Mortgaged_{i,t}$		0.136	0.137	0.142
		(0.000)	(0.000)	(0.000)
$Appraisal\ Required_{i,t}$	-0.005			
	(0.043)			
$\log\left(Predicted\ Price_{i,t}\right)$	0.009	0.872	0.855	0.858
	(0.000)	(0.000)	(0.000)	(0.000)
Estimator	OLS	2SLS	2SLS	2SLS
Bandwidth around Cutoff		5%	10%	15%
Zip Code FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
First Stage F-Statistic		7.702	15.090	22.632
J-Statistic (p-value)		0.146	0.065	0.187
Number of Observations	250,924	7,207	14,575	22,021

Table IA.IV.
Robustness to Using Seller Cash Preference as an Instrumental Variable

P-values are in parentheses. This table estimates equation (28) 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 IV 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. The regression equation in columns (3) and (4) is of the same form as in column (1) of Table IV, but it is estimated through 2SLS using $Cash\ Share_{s(i,t)}$ as an instrument for $Mortgaged_{i,t}$. Seller controls are the same as in column (1) of Table IA.II. Standard errors are clustered by seller. The remaining notes are the same as in Table IV.

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	2SLS
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
Number of Observations	425,398	425,395	425,395	425,395

Table IA.V.
Additional Calculations of the Theoretical Premium

This table summarizes the theoretical mortgage-cash premium under model extensions. The first row reports the baseline value, for reference. The other extensions are as in Figure 3. The extensions by the borrower's loan cost adjust the borrower's loan cost as a share of par, D, and the remaining aspects of the calculation are the same as in the first row. The feasible parameter space is held fixed to account for the possibility that changes in D expand the set of feasible parameters. Dropping this restriction results in a calculated premium between 7.0% and 8.1%. The remaining notes are the same as in Figure 3.

Specification	Mortgage-Cash Premium (μ)
Baseline:	0.069
By Nonfinancial Contingency:	
Home Sale ($h = 0.07, q^c = 0$)	0.078
Home Sale and Inspection ($h = 0.07, q^c = 0.01$)	0.079
By Risk Aversion:	
$\gamma = 0$	0.019
$\gamma = 5$	0.069
$\gamma = 12$	0.081
Panel D. By Borrower Loan Cost:	
$D = 0.2$	0.066
$D = 0.4$	0.066
$D = 0.6$	0.066
$D = 0.8$	0.067
$D = 1.0$	0.069

Table IA.VI.
Heterogeneity in the Premium. Coefficient Estimates.

P-values are in parentheses. This table reports the coefficient estimates for the regressions in Figure 4. For reference, column (1) re-estimates the baseline regression equation (28) on the subsample with enough information to test for heterogeneity. Column (2) estimates equation (39), which performs the main test of heterogeneity. Column (4) contains the analogous projection across the model's parameter space, described in the note to Figure 4. Column (3) reports the baseline mortgage-cash premium already shown in Figure 3, and it is analogous to column (1). The remaining notes are the same as in Figure 4.

Outcome:	$\log(Price_{i,t})$			
	Data		Model	
	(1)	(2)	(3)	(4)
$Mortgaged_{i,t}$	0.107	0.191	0.069	0.109
	(0.000)	(0.000)		
$Mortgaged_{i,t} \times Tercile(\hat{q},2)_{z,t}$		0.067		0.006
		(0.001)		
$Mortgaged_{i,t} \times Tercile(\hat{q},3)_{z,t}$		0.122		0.024
		(0.000)		
$Mortgaged_{i,t} \times Tercile(\hat{\omega},2)_{z,t}$		-0.007		-0.005
		(0.774)		
$Mortgaged_{i,t} \times Tercile(\hat{\omega},3)_{z,t}$		-0.076		-0.077
		(0.024)		
$Mortgaged_{i,t} \times Tercile(\hat{\lambda},2)_{z,t}$		-0.020		-0.008
		(0.382)		
$Mortgaged_{i,t} \times Tercile(\hat{\lambda},3)_{z,t}$		-0.056		-0.009
		(0.042)		
$Mortgaged_{i,t} \times Tercile(\hat{L},2)_{z,t}$		-0.027		-0.001
		(0.191)		
$Mortgaged_{i,t} \times Tercile(\hat{L},3)_{z,t}$		-0.051		-0.001
		(0.084)		
Zip Code-Month FE	Yes	Yes		
Hedonic-Month FE	Yes	Yes		
Property FE	Yes	Yes		
R^2	0.880	0.880		
Number of Observations	267,583	267,583		

Table IA.VII.
Robustness to Controlling for Listing Characteristics

P-values are in parentheses. This table estimates a variant of equation (28) using the California Association of Realtors (CAR) data set, which assesses the robustness of the baseline results to controlling for characteristics of the listing. Subscripts h and t index home sale and month. The main variables are defined as follows: $\textit{Sale Price}_{h,t}$ is the sale price, $\textit{Mortgaged}_{h,t}$ indicates whether the home was sold to a buyer with an "all cash offer," the "ability to close fastest," or an "offer without contingencies", $\textit{List Price}_{h,t}$ is the price at which the home was initially listed, $\textit{Days on Market}_{h,t}$ is the number of days from initial listing to sale, and $\textit{First Sale}_{s(h,t)}$ indicates whether the seller, denoted $s(h,t)$, is selling a home for the first time. All specifications include county fixed effects, include month fixed effects, control for $\textit{First Sale}_{s(h,t)}$, and control the following additional seller and hedonic variables: the seller's age, an indicator for whether the seller is female, an indicator for whether the seller is black or Hispanic, an indicator for whether the seller is moving to another home at the same time, an indicator for whether the property is a detached single-family home, and the property's age, log square feet, and number of bedrooms. Columns (3) and (4) are analogous to Table IA.XII in that they interact the method of financing with a measure of the seller's experience, a proxy for less uncertainty. Standard errors are heteroskedasticity robust. Data are from the CAR data set. Details on this data set are in Section I.

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

Table IA.VIII.
Robustness to Semi-Structural Estimator from Bajari et al. (2012)

P-values are in parentheses. This table estimates equation (28) 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 (28) using OLS. Column (2) estimates equation (28) using the Bajari et al. (2012) estimator. Explicitly, column (2) shows the estimated value of μ from the following second-stage regression equation,

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

where the first-stage regression equation is

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

and where n denotes holding period, and $\varrho_t\left(n\right), \varphi_t^M\left(n\right)$, and $\varphi_t^P\left(n\right)$ are vectors of holding period-year fixed effects. Holding periods are rounded to the nearest year. The sample includes all properties that transacted at least twice over the sample period, and the smaller sample size relative to Table IV 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 IV.

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

Table IA.IX.
Robustness to Weighting by Representativeness

P-values are in parentheses. This table estimates equation (28) after weighting repeat sales by their representativeness, which assesses the external validity of the baseline results. Subscripts i and t index property and month. Column (1) estimates a logistic regression in which the outcome variable is an indicator for whether the transaction is in the baseline ZTRAX data set, denoted $Baseline_{i,t}$, and the independent variables are observed characteristics of the transaction: Age_i through $Cash\ Propensity_{b(i,t)}$ are described in the note to Table II, $Sales\text{-}per\text{-}Property_{z(i)}$ is the total number of home sales in the raw ZTRAX data set over 1980 to 2017 in property i's zip code, denoted $z(i)$, divided by the total number of properties in $z(i)$ in the raw ZTRAX data set, and $Period\ 1996\ to\ 2004_t$ through $Period\ 2010\ to\ 2017_t$ indicate whether month t lies in the associated time period from Table IV, where 1980-1996 constitutes the reference period. The values shown in column (1) are the coefficient estimates, not the marginal effects, and the p-values associated with each coefficient are not shown to conserve space. Column (2) estimates equation (28) through WLS, where transactions are weighted by the reciprocal probability of appearing in the baseline sample, based on the predictions from column (1). The remaining notes are the same as in Table IV.

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

Table IA.X.
Consistency with Contemporaneous Estimates in the Literature

P-values are in parentheses. This table estimates a variant of (28), 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 to 2017, corresponding to the sample studied in Han and Hong (2023), BMPS refers to the subsample of purchases in the Phoenix, Las Vegas, Dallas, and Orlando MSAs and in Gwinnet County, GA between 2013 to 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 (2023) 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 IV.

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^2	0.714	0.714
Number of Observations	426,256	426,256

Table IA.XI.
Robustness to Nonparametric Matching Estimator

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 IV and seller characteristics in Table IA.IV 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 IV.

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

Table IA.XII.
Sale Experience as a Proxy for Less Uncertainty

P-values are in parentheses. This table assesses the relationship between the mortgage-cash premium and the seller's experience, a proxy for less uncertainty about mortgaged transactions. Subscripts i, and t index property and month. The data set is the baseline ZTRAX data set. The table estimates a variant of the repeat sales and hedonic pricing equation (28) that interacts $Mortgaged_{i,t}$ with the number of sales made by the seller of property i in month t, denoted $s(i,t)$, as of month t in the baseline ZTRAX data set. The remaining notes are the same as in Table IV.

Outcome:	$\log(Price_{i,t})$	
	(1)	(2)
$Mortgaged_{i,t} \times Number\ of\ Sales_{s(i,t)}$	-0.005	-0.004
	(0.001)	(0.005)
Other Variables:		
$Mortgaged_{i,t}$	0.149	0.143
	(0.000)	(0.000)
$Number\ of\ Sales_{s(i,t)}$	-0.020	-0.020
	(0.000)	(0.000)
Data Set	ZTRAX	
Seller Controls	No	Yes
Zip Code-Month FE	Yes	Yes
Hedonic-Month FE	Yes	Yes
Property FE	Yes	Yes
R^2	0.907	0.907
Number of Observations	426,256	426,256

Table IA.XIII
Motivation for Preferring All-Cash Offers for Survey Respondents

This table summarizes the motivation for preferring all-cash offers from an experimental survey of U.S. homeowners. Subscript k indexes survey respondent. The upper panel summarizes responses to a question of why a respondent would prefer an offer from an all-cash buyer relative to a mortgage-financed buyer offering the same price. Respondents are restricted to choosing their most important motivation, and so the shares sum to one across motivations. Column (1) summarizes respondents who are told the probability of transaction failure and the subsequent price cut (Known Failure Rate). Column (2) summarizes respondents who are not told these parameters (Ambiguous Failure Rate). The sample consists of respondents in the survey's second and third waves, in which these questions are asked. Respondents are shown each possible answer in a random order, and so the ordering in this table does not correspond to the actual ordering. The remaining notes are the same as in Table VIII.

	Share of Respondents	
	Known Failure Rate	Ambiguous Failure Rate
	(1)	(2)
"Why would you prefer to sell your home to the Cash Buyer?":		
"If the Mortgaged Buyer backs out and I relist at a lower price, I may not have enough money to meet my mortgage and moving expenses."	0.118	0.260
"Even if the Mortgaged Buyer would never back out, the Cash Buyer would close more quickly and end the stressful process of selling my home."	0.618	0.732
"If the Mortgaged Buyer backs out and I relist at a lower price, I will sell the home at a loss relative to my target price."	0.223	0.129
"Other (please describe)"	0.039	0.048