Collateral Value Uncertainty and Mortgage Credit Provision

Erica Xuewei Jiang and Anthony Lee Zhang*

Abstract

Using property transaction and financing data, we document large cross-sectional differences in how effective houses are as collateral for mortgages. Older and less standardized houses tend to have higher price dispersion, and their appraisal values tend to deviate more from transaction prices. Mortgages collateralized by these houses have lower loan-to-price ratios, higher interest rates, and are more likely to be rejected due to collateral reasons. Using a structural model, we show that appraisal uncertainty explains most of the effects on LTP and mortgage failure, while the collateral recovery channel is the main driver of the interest rate effect. We discuss the implications of our findings for the shift from human to automated appraisals, the FHA mortgage program, and urban policy.

*Jiang is from the University of Southern California, Marshall School of Business. erica.jiang@marshall.usc.edu. Zhang is from the University of Chicago Booth School of Business. anthony.zhang@chicagobooth.edu. Pranav Garg, Willy Yu-Shiou Lin, Michael Yip, and Zichen Zhao provided outstanding research assistance. We are grateful to Kenneth Ahern, Jason Allen, Eric Budish, Sylvain Catherine, Anthony DeFusco, Jason Donaldson, Kevin Duncan, Marco Giacoletti, Adam Guren, Larry Harris, Mark Jansen, Ben Keys, Sam Kruger, David Low, Gregor Matvos, Tim McQuade, Norm Miller, Andrii Parkhomenko, Giorgia Piacentino, Tomasz Piskorski, Michael Reher, Dan Ringo, Amir Sufi, Selale Tuzel, Joe Vavra, Daniel Greenwald, Edward Glaeser, Jinyuan Zhang, Arthur Korteweg, Rodney Ramcharan, participants at Columbia Workshop in New Empirical Finance, the Guthrie Center Housing and Macroeconomics Conference, USC brownbag, UIUC, the Bank of Canada, NUS, UChicago Booth, the Craig W. Holden Memorial Conference, the FRB Philadelphia Mortgage Market Research Conference, UCLA, the FHFA EconSummit, SFS Cavalcade, PUC Chile, NBER SI Real Estate, Georgetown University, McGill University, FRB Chicago, and FRB Boston for comments. We are grateful for financial support from the Fama-Miller Center for Research in Finance at the University of Chicago Booth School of Business. TransUnion (who provided some of the data for this project) has the right to review the research before dissemination to ensure it accurately describes TransUnion data, does not disclose confidential information, and does not contain material it deems to be misleading or false regarding TransUnion, TransUnion's partners, affiliates or customer base, or the consumer lending industry. *First version: August, 2021*

1 Introduction

The residential mortgage market has been central to policies for improving homeownership and stabilizing the economy. Despite the significant amount of subsidies devoted to this market, various frictions inhibit the passthrough of these subsidies to households (Glaeser and Shapiro, 2003; Hurst et al., 2016; Agarwal et al., 2017; Adelino et al., 2020; DeFusco, 2018; DeFusco and Mondragon, 2020). Factors that prevent home buyers from borrowing against the house can significantly affect their homeownership decisions, especially for low-income families. Understanding such credit market frictions is important for improving homeownership rates, a topic which has been central in housing policy debates.[1]

In this paper, we document large cross-sectional differences in how effective houses are as collateral for mortgages. Older and less standardized houses have higher price dispersion, whose appraisal values tend to deviate more from transaction prices. Price dispersion affects financing: mortgages backed by high-dispersion houses have lower loan-to-price ratios, higher interest rates, and are more likely to be rejected. Using a structural model, we show that house price dispersion matters for two reasons: because dispersion affects lenders' expected recovery on mortgage debt, and because dispersion leads to appraisal uncertainty, which interacts with regulatory constraints that require mortgage LTVs to be tied to house appraisals. The estimated model allows us to disentangle how the collateral recovery and appraisal channels affect different mortgage outcomes, which has implications for several ongoing policy debates in the housing market.

Policymakers aiming to encourage homeownership for low-income households have considered interventions in credit markets as well as in housing markets. This paper highlights a link between these two markets: the amount of credit that mortgage lenders provide depends on the value uncertainty of the house used as collateral. Our findings have implications for the impending shift from human to automated appraisals. We also provide a rationale for interventions in the mortgage market, such as the FHA program, which allows low-income households to borrow at higher LTV ratios. More broadly, our results suggest that the ease of financing of the housing stock is quantitatively important for housing affordability.

The paper unfolds in two steps. First, we establish that collateral value uncertainty

[1]See policy reports, e.g., Herbert et al. (2005) and Boehm and Schlottmann (2008). As of 2021, homeownership rate of below-median income households is about 52 percent, compared to 79 percent of above-median income households. Source: The US Census quarterly report, Quarterly Residential Vacancies and Homeownership.

limits mortgage credit. We use rich residential property transaction data from 2000 to 2020 to document substantial cross-sectional variation in the predictability of house prices. Older and less standardized houses in terms of the number of bedrooms or square footage have more value uncertainty, as measured by the predicted errors from a hedonic model of house price. Aggregating price dispersion to the zip code level, we show that zip code price dispersion is persistent over time, suggesting that cross-sectional differences in price dispersion is mainly driven by differences in characteristics of local housing stocks.

We show that collateral value uncertainty affects three measures of mortgage lending: loan-to-price ratios, interest rates, and mortgage rejections. Comparing two houses which are transacted in the same county-year at the same transaction price, we find that houses with higher estimated price dispersion tends to have lower loan-to-price ratios (LTPs). The result holds when we further restrict to comparing houses financed by the same lender. On average, LTPs are about 50 bps lower for houses with one standard deviation higher estimated price dispersion. Moreover, we estimate the mortgage price menu (rate-LTP pairs) using loans received by homebuyers in zip-codes with different level of house price dispersion. The estimated menu shows that, conditional on borrower and loan characteristics, mortgages collateralized by high price dispersion houses tend to have higher interest rates. The risk-adjusted rate increases by about 2bps in zip codes with one standard deviation higher house price dispersion. Lastly, mortgage applications are more likely to be rejected due to collateral reasons if the underlying houses have higher price dispersion. The effect is economically significantly: given the sample average rejection rate, the effect amounts to about 25% increase in rejection likelihood.

Our baseline identification strategy exploits within county-year variation by comparing two properties that are bought in the same county-year, at the same price, and by buyers with the same credit profile and income. To address concerns that house price dispersion is associated with other unobserved characteristics that also affect mortgage credit provision, we construct instruments for price dispersion, based on the heterogeneity of a house relative to its local housing stock. The instruments help alleviate the endogeneity concerns: zip codes with more heterogenous housing stock within a county do not systematically attract less creditworthy or low-income households; nor is the instrumented price dispersion correlated with ex-post default. Our results continue to hold using the instrumental variable approach, and the ex-post default analysis suggests that our IV results are not driven by unobserved differences in borrowers' creditworthiness.

We perform two robustness tests of our results. First, our results hold even with lender-zip-year fixed effects, suggesting that the results are not driven by lender market power, or other features of lenders' behavior which affect all houses within a zip code uniformly. Second, our findings hold even restricting to a subsample of houses with sale prices below conforming loan limits, suggesting that our findings are not driven by home buyers reducing borrowing amounts to be eligible for GSE or FHA loans.

We then construct a structural model to decompose the underlying mechanisms, which helps understand the implications of our findings. The model shows two main forces through which high price dispersion limits credit provision. The first is the *collateral recovery* channel: lenders face higher losses in high-dispersion areas, so offer worse mortgage menus to homebuyers, with higher interest rates for any given LTV. The second is the *appraisal risk* channel: in high-dispersion areas, under-appraisals are more likely and larger when they occur, so buyers limit their mortgage loan size in order to lower the risk of under-appraisals.

In the model, a borrower chooses a targeted mortgage loan size to purchase a house. Competitive lenders offer menus of interest rate-LTP pairs to borrowers, such that lenders break even, given the exogenous risk of default and expected recovery rates. After borrowers choose a target mortgage loan size from the offered menu, the house undergoes an appraisal, to satisfy a regulatory constraints on mortgage LTVs. We model appraisals as upwards-biased signals of house prices, consistent with the distribution of house appraisals in practice. If the house over-appraises, the borrower proceeds with the mortgage. If the house under-appraises sufficiently much that the targeted mortgage violates the LTV constraint, the buyer must choose to either make a costly increase in her down payment, or pay a fixed cost to renege on the transaction and find a new house. Homebuyers thus face a tradeoff in choosing mortgage size: larger mortgages improve consumption smoothing, but increase the risk of under-appraisals.

We then calibrate the model to data, matching moments on how lenders' rate menus, the distribution of appraisals, and the rate of mortgage rejections depend on price dispersion. The calibrated model allows us to do two exercises. First, we do not target the dispersion-LTP relationship as a moment in our model, so we use this relationship as an out-of-sample test of model fit. The model produces a dispersion-LTP relationship of similar magnitude to what we observe empirically, indicating that the model can quantitatively rationalize the magnitude of the dispersion-LTP relationship in the data.

It is intuitive that price dispersion should lower mortgage credit provision; however, the precise manner in which this occurs in our model is subtle, combining the effects of the collateral recovery channel – the fair pricing of debt in free markets – and the appraisal risk channel, where mortgage size choice interacts with a regulatory constraint. Thus, the second exercise we perform with our model is to evaluate how each channel contributes to driving each of the three outcome measures we analyze: LTPs, interest rates, and mortgage rejections. We find that the collateral recovery channel is the primary driver of the relationship of dispersion with interest rates, whereas the appraisal risk channel is primarily responsible for the relationship of dispersion with loan size and mortgage rejections. That is, if we assume lenders offer worse rate menus, but buyers respond to this by picking similarly sized mortgages and receiving similar amounts of mortgage rejections, but bearing higher interest rates as a result. On the other hand, if we assume that appraisals become riskier while holding lenders' rate menus fixed, buyers receive more mortgage rejections, scale down borrowing in response to higher appraisal risk, but mortgage interest rates are essentially unaffected.

Our model thus has implications for how shifts in regulatory treatment of appraisals, and how lenders treat collateral recovery risk in their mortgage pricing decisions, would influence various measures of mortgage credit provision. To illustrate this point, using our model, we evaluate how much the shift to automated appraisals would influence mortgage credit provision. We estimate outcomes in the model, assuming computers generate fully fair appraisals of home values, removing the tendency of human appraisers to bias appraisals upwards. We find that automated appraisals would decrease mortgage LTPs around 2pp, and increase mortgage rejection rates by around 10pp in low-dispersion areas, and around 15pp in high-dispersion areas. Hence, shifting to automated appraisals, without somehow compensating for human appraisers' built-in upwards biases, has the potential to significantly lower mortgage credit provision.

Together, our results imply that the value uncertainty of the housing stock is a previously overlooked variable which has quantitatively large effects on mortgage credit provision in the US housing market. The effect of value uncertainty on mortgage credit does not represent a form of discrimination by lenders, or an externality which can be addressed through Pigouvian taxation. Rather, it is a structural phenomenon caused by intrinsic features of the housing stock: lenders in competitive credit markets have higher costs of lending against poor collateral, and houses tend to under-appraise by larger amounts, leading to lower credit provision for these houses.

Our results provide a rationale for interventions, such as the FHA loan insurance program, which extend credit to low-income households and first-time homebuyers at loan-to-value ratios much higher than private lenders. We have shown that low-income households face particularly high barriers to homeownership because they tend to live in high-dispersion areas, so lack access to housing with high collateral values. Thus, mortgage credit access is limited precisely for those households who are most down-payment constrained, for whom credit is most valuable. Government interventions such as the FHA loan program, can potentially alleviate this effect. Besides these programs, we also discuss implications our results have for the impending shift from human to automated housing appraisals, and for urban and zoning policies which affects the collateral value of the aggregate housing stock.

This paper relates to a number of strands of literature. Broadly, our paper fits into a literature on frictions that affect mortgage credit (Lustig and Van Nieuwerburgh, 2005; Mian and Sufi, 2011; Greenwald, 2016; Agarwal et al., 2017; Piskorski and Seru, 2018; Beraja et al., 2019; DeFusco et al., 2020; Adelino et al., 2020; Buchak et al., 2018; Jiang, 2020) and the corresponding real effects (Glaeser and Shapiro, 2003; Di Maggio and Kermani, 2017; Agarwal et al., 2022; Di Maggio et al., 2017; Gupta et al., 2021; Dokko et al., 2019; Kermani and Wong, 2021). DeFusco and Mondragon (2020) studies two counter-cyclical refinancing frictions – the need to document employment and the need to pay upfront closing costs – and show these frictions prevent borrowers who experience income shocks to refinance. DeFusco (2018) studies how changes in access to housing collateral affect homeowner borrowing behavior and estimate the marginal propensity to borrow out of housing collateral. Collier et al. (2021) shows that borrowers lower loan size to avoid collateral requirements and the impact of collateral requirements on ex-post loan performance. Lang and Nakamura (1993) theoretically argues that the precision of appraisals influences home sales through down payment requirements, leading to sub-optimal lending outcomes. Blackburn and Vermilyea (2007) empirically tests the theories of rational redlining and shows that a low volume of home sales lead to uncertainty in house appraisals, reducing mortgage lending.

Our paper also relates to the housing literature. We build on a literature on idiosyncratic price dispersion in the housing market and its consequences. Case and Shiller (1989) and Giacoletti (2021) analyze idiosyncratic risk in residential real estate markets. Sagi (2021) analyzes idiosyncratic risk in commercial real estate. Hartman-Glaser and Mann (2017) documents that lower-income zip codes have more volatile returns to housing than higher-income zip codes. They rationalize the finding with a model where shocks to the representative household's marginal rate of substitution lead to volatility in the return to housing via the collateral constraint, and lower-incomes have a more volatile marginal rate of substitution, and thus more volatile returns to housing. Sklarz and Miller (2016) propose a method to adjust loan-to-value ratios to reflect house value uncertainty.

More broadly, our paper fits into a classic literature analyzing how collateral values affect the properties of debt contracts collateralized by these assets (Titman and Wessels, 1988; Shleifer and Vishny, 1992). Literature has studied how collateral affects the cost of debt (Benmelech and Bergman, 2009; Liu, 2022) and firms' willingness to borrow (Pan et al., 2021; LaPoint, 2021), and the effect of collateral liquidation values on contract renegotiation (Benmelech and Bergman, 2008) and on ex-ante firm investments (Bian, 2021).

We contribute to the above three strands of literature by showing quantitatively that collateral value uncertainty matters in the US residential real estate market: there is substantial cross-sectional heterogeneity in housing collateral values, which has economically significant effects on mortgage credit availability. Our model also elucidates the mechanisms through which the collateral channel influences outcomes within the unique structure of the US residential mortgage market: in particular, how house price dispersion interacts with lender incentives and the housing appraisal system to influence mortgage credit access.

The paper proceeds as follows. Section 2 describes our data, measurement strategy, and stylized facts on our price dispersion measure. Section 3 studies the effect of price dispersion on mortgage provision. Section 4 constructs our model, and Section 5 calibrates the model to the data. We discuss implications of our results in section 6, and conclude in section 7.

2 Measurement and Data

2.1 Measuring Value Uncertainty

The housing market is far from a perfectly competitive, frictionless market. Houses are differentiated, buyers may have heterogeneous preferences, and sellers often only list houses for sale when they face idiosyncratic shocks forcing them to move. These forces imply that individual houses trade in thin markets: there is a relatively small set of potential buyers for each house at any given point in time. Thus, there is nontrivial randomness in house sale prices: the same house may sell for higher or lower prices, depending on whether there happens to be a high-valued buyer in the market when the owner lists the house for sale. House price dispersion induced by market thickness tends to be larger for non-standardized houses, since there are fewer interested buyers overall, and since there is likely to be larger dispersion in buyers' values for the house.

We proceed to empirically estimate house price dispersion, at the level of individual house sales, essentially by measuring what kinds of houses tend to have smaller errors when priced with a hedonic regression.[2] Our estimation has two steps. First, we regress transaction prices on house characteristics:

$$p_{it} = \eta_{kt} + f_k\left(x_i, t\right) + \epsilon_{it}, \tag{1}$$

We then regress the squared residuals, $\hat{\epsilon}^2_{it}$, from (1), on a flexible function of characteristics and time, to predict which house characteristics make them difficult to price:

$$\hat{\epsilon}^2_{it} = g_k\left(x_i, t\right) + \xi_{it} \tag{2}$$

In (1) and (2), i indexes properties, k indexes counties, and t indexes months. p_{it} is the log transaction price of house i at time t. $f_k\left(x_i, t\right)$ and $g_k\left(x_i, t\right)$ are generalized additive models in observable house characteristics x_i and time t, which we describe in Appendix A.2. $f_k\left(x_i, t\right)$ allows houses with different observable characteristics to appreciate at different rates. $g_k\left(x_i, t\right)$ allows the variance of price dispersion to vary over time. η_{kt} is county-month fixed effect. Specification (1) essentially estimates a hedonic specification for house prices, and specification (2) projects the squared residuals ϵ^2_{it} from the hedonic regression on house features and time, to predict which characteristics make houses difficult to value. We then use the square roots of the predicted values from specification (2) as our house-level measure

[2] A similar methodology is used in Buchak et al. (2020).

of idiosyncratic price dispersion:[3]

$$\hat{\sigma}_{it}^2 \equiv \hat{g}_k\left(x_i, t\right) \tag{3}$$

This measure directly captures the forces that tend to generate appraisal variance and thus the value uncertainty in the appraisal channel. Appraisers simply compare the price of a house to recent sale prices of houses with similar characteristics, which is fairly close to our procedure of taking the squared residuals from a hedonic regression. In contrast, under the fair-pricing channel, lenders should in principle care about the total price volatility of a house, which consists of the idiosyncratic volatility of a house as well as the volatility of local house price index. Our measurement strategy focuses on the idiosyncratic component, which is a large component of total volatility.[4] This methodology is justified because our empirical analyses will compare individual houses within a given region-year. To the extent that houses within a geographical region have similar exposure to local index volatility, differences in total volatility among these houses are likely to be mainly driven by differences in the idiosyncratic component.

We next discuss why our measurement strategy is robust to two potential concerns. First, we measure value uncertainty as independent shocks at each house sale rather than fluctuations over time that scale with the holding period of a house. This modeling assumption is justified by evidence in Giacoletti (2021) and Sagi (2021), which shows that idiosyncratic component of house price risk has a very flat term structure, scaling very little with the holding period of a house. Also, idiosyncratic price dispersion mainly varies in the cross-section and has relatively small time-series variation (Kotova and Zhang, 2021). Intuitively, search frictions, market thickness, and heterogeneous preference are among the main drivers of the idiosyncratic volatility; these forces tend to generate price shocks that are realized upon sales, rather than a drift term which increases in variance substantially depending on house holding periods.

Another concern is measurement errors of the hedonic approach. If we observed all characteristics of houses that market participants observed, and our functional forms for

[3] Note that it is important to use the predicted values of $\hat{\sigma}_{it}^2$ in stage 2 rather than the residuals $\hat{\epsilon}_{it}^2$ in stage 1 directly. This is because the expected value of idiosyncratic dispersion, σ_{it}^2, is the analog of σ in our model, which is relevant for loan-to-values. Each realization of $\hat{\epsilon}_{it}^2$ is a noisy measure of σ_{it}^2. If we regressed outcomes such as house-level LTP on the regression residuals $\hat{\epsilon}_{it}^2$ directly, the coefficients would be biased towards 0, relative to the first-best of regressing LTPs on σ_{it}, due to measurement error bias.

[4] According to Piazzesi and Schneider (2016), roughly half of the total volatility in a house price transaction is idiosyncratic.

house prices were fully flexible, the hedonic approach would fully filter out the effects of house characteristics, capturing only price dispersion generated by housing market frictions. In practice, our estimates are likely to be confounded by two main factors. First, our estimation cannot account for the effects of house characteristics unobserved in our data, but observed by market participants and lenders. Second, our functional forms may not be flexible enough to capture the true conditional expectation function.

To further address the concern about unobservables, we construct an alternative measure of value dispersion using a repeat-sale model in Appendix D.1. This specification absorbs all time-invariant house quality variation into house fixed effects, so the squared residuals essentially measure the extent to which a house's price fails to track local house price indices. We also purge the repeat-sales residuals of variation driven by average time-between-sales of houses, and the number of times a house is sold, to address concerns that the squared residuals are mechanically associated with house sale frequency. While the size of the residuals from the repeat-sales specification are substantially lower than in the baseline specification, the squared residuals from the two specifications are very correlated: houses that have high predicted value uncertainty under one measure also tend to have high predicted uncertainty from the other specification. Our empirical results also continue to hold using the repeat-sales residuals as a measure of value uncertainty, suggesting that our results are not purely driven by variation in unobserved house quality.

2.2 Data and Stylized Facts

2.2.1 Data Sources

Corelogic Deed & Tax Data. We obtain house transaction records in the entire US from 2000 to 2020 from the Corelogic Deed dataset, and restrict the sample to arms-length, non-foreclosure transactions in single family residences. The date set reports each house transaction attached to a specific property, and provides information on sale amount, mortgage amount, transaction date, and property location. We exclude transactions with missing sale price, date, property ID, or location information. We merge the transaction records with the Corelogic Tax records to get property characteristics, such as year built and square footage, and estimate price dispersion for each house in this merged data set. Appendix A.1 provides detailed description about data cleaning steps.

Corelogic Loan-Level Market Analytics (LLMA) Data. We obtain mortgage information from the Corelogic LLMA data, which provides detailed information on mortgage and borrower characteristics at origination – interest rates, down payments, sale prices, credit score, and debt-to-income ratio – and monthly loan performance of the loan, including delinquency status and investor type. Importantly for our analysis, the LLMA provides both appraised house value and transaction price. We use this data set to estimate the menu of LTP-interest pairs in any given market and to examine loan performance. The LLMA terms of use do not allow us to merge the data with the Deeds records; thus, we aggregate estimated idiosyncratic price dispersion to the 5-digit zip code level.

Home Mortgage Disclosure Act (HMDA). The HMDA covers the near universe of U.S. mortgage applications, including both originated and rejected applications. For rejected loans, we observe the rejection reasons. We use the HMDA for extensive margin analysis on mortgage application rejections, while we aggregate the estimated idiosyncratic price dispersion to the finest geographic regions in the HMDA (census tract).

Other Sources. We use the Booth TransUnion Consumer Credit Panel to calculate the average VantageScore credit score by county to measure the creditworthiness of the entire borrower population. We obtain zip level demographic data from the American Community Survey (ACS) 1-year and 5-year samples.

Table 1 provides summary statistics.

2.2.2 Estimated Value Uncertainty and Housing Market Frictions

We next present some stylized facts about the estimated value uncertainty of the US housing stock and discuss how the estimates reflect the housing market frictions discussed in the previous section.

We first confirm that the estimated price dispersion is very persistent over time. Figure 1 Panel A plots zip-code idiosyncratic price dispersion in 2020 against zip code dispersion in 2010. Over both time periods, zip code dispersion in the later year is lined up with the dispersion in the earlier year. This suggests that the differences in price dispersion are driven by persistent characteristics of the local housing stock, rather than time-varying local market conditions.

To explore this further, Table 2 presents the association between estimated value uncertainty and house characteristics. Panel A analyzes house features. Throughout, we control for linear and squared terms in log house prices, comparing houses with similar prices and different characteristics. Older houses have higher price dispersion (column 1). Controlling for building age, recently renovated houses within 5 years of the transaction date (column 2) have lower price dispersion.[5] Columns 3-4 present the association between property size, measured by square-footage and number of bedrooms, and price dispersion. There is a U-shaped relationship: price dispersion is low for moderately large houses, and higher for houses which are very large or very small. In terms of local housing market conditions, Panel B of Table 2 shows that houses in zip codes with larger income inequality, less population density, and more vacancies tend to have higher price dispersion. Together, Table 2 suggests that house price dispersion is essentially driven by house standardization and market thickness.[6] Lastly, Figure 1 Panel B shows the relationship between price dispersion and average zip-code incomes. Price dispersion tends to be higher in low-income zip codes.

3 Price Dispersion and Mortgage Credit

In this section, we empirically show that collateral value uncertainty affects mortgage credit at three margins: loan size, interest rate, and approval likelihood. We present baseline OLS results before discussing our instrumental variable strategies and a set of robustness tests.

3.1 Effects on Mortgage Credit

3.1.1 Loan-to-Price Ratio

We start with the effect of price dispersion on loan size as measured by loan-to-price ratio (LTP). We first visualize the relationship by plotting county average LTP against average house price dispersion in Figure 2(a).[7] Counties with higher price dispersion have lower

[5] We can partially measure house renovations, as the Corelogic tax data contains an "effective year built" variable, which tracks the last date at which a property was renovated.

[6] This finding is consistent with evidence from other papers: see, for example, Kotova and Zhang (2021) and Andersen et al. (2021).

[7] To make this plot, we first remove the average LTP differences across levels of individual house prices and then plot their county average against county average price dispersion.

average LTP. The pattern holds for all types of loans: GSE loans, FHA loans, and jumbo loans (Figure A1).

We then exploit within county-year variation by comparing two properties that are bought in the same county-year, at the same price, and by buyers with similar credit profile and income. To implement this strategy, we estimate the following property-level specification:

$$LTP_{ikt} = \alpha + \beta Dispersion_{ikt} + X_{ikt}\Gamma + \mu_{kt} + \nu_m + \epsilon_{ikt} \tag{4}$$

LTP_{ikt} is loan-to-price ratio of the mortgage that is collateralized by property i in county k in year t. $Dispersion_{ikt}$ is the estimated price dispersion of the underlying property. X_{ikt} is a set of controls, including property transaction price, mortgage type, mortgage term, and resale indicator. μ_{kt} and ν_m are county-year and transaction month fixed effects, respectively.

Table 3 presents the results. Column 2 corresponds to Specification 4, while columns 1 and 3 are less saturated specification with only transaction date fixed effects and more saturated specification further add lender-year fixed effects, respectively. For two houses in the same county that are transacted on the same date at the same price, the one with higher estimated price dispersion tends to receive a smaller sized loan. In the most saturated specification, the loan-to-price ratio is more than 20bps lower for houses with one standard deviation higher estimated price dispersion across these specifications. The estimates effects are economically significant: in Appendix C, we calibrate a lifecycle model of homeownership choice and show that price dispersion-induced changes in LTV can substantially decrease aggregate homeownership rates.

3.1.2 Interest Rates

We next turn to the effect of price dispersion on the cost of mortgage credit. Figure 2(b) visualizes residualized mortgage rates against the average house price dispersion. We residualize rates for individual mortgages on borrower and loan characteristics such as FICO, LTP, DTI, the squared terms, and their interactions with origination year. Properties in counties with higher price dispersion tend to be financed by more expensive mortgage. The positive relationship between price dispersion and interest rate suggests that the negative effect of price dispersion on LTP is not explained by borrowers substituting to smaller loans to obtain lower interest rates.

We further corroborate this result by estimating the entire menu of LTP-interest rate pairs that are available in a market using loan-level data. Figure 3 plots the mortgage price menu, separately for groups of zip codes with high and low price dispersion. We residualize interest rates by purging out the effects of borrowers' credit scores, loan type, and time fixed effects and plot a menu of average interest rates for different LTPs. As shown in Figure 3, the entire menu of interest rate-LTP pairs shifts upwards in high-dispersion zip codes: for any given LTP, borrowers in high-dispersion zip codes can expect to pay higher prices. The difference is about 3bps for loans with LTP below 80, and enlarges to 7bps for loans with LTP above 80.

Table 4 presents the above results in regression settings by estimating Specification 4 with interest rate being the outcome variable while including LTP as one of the explanatory variables.[8] Column 1 uses the full sample. We first confirm that higher loan-to-price ratios are associated with higher interest rates. The coefficient on LTP is positive and statistically significant. A one percentage point increase in LTP is associated with an 80bps increase in interest rate. Controlling for LTP, houses in zip codes with higher house price dispersion are financed with more expensive mortgages. The mortgage rate increases by 1.1bps in zip codes with one standard deviation higher average house price dispersion. Columns 2 to 3 show the results for securitized loans and portfolio loans, respectively. The results hold in all samples. For every one standard deviation increase in zip-code average house price dispersion, the mortgage rate of securitized loans increases by 1.38bps, and the mortgage rate of the portfolio loans increases by 1.98bps.

3.1.3 Mortgage Rejection Rates

Besides less favorable loan terms conditional on approval, mortgage applications are more likely to be rejected in counties with higher price dispersion, according to Figure 2(c). In particular, as shown in Figure 2(d), the fraction of mortgages rejected for collateral-related reasons is higher in high-dispersion counties.

We show this negative relationship between price dispersion and loan approval likelihood in regression settings by estimating Specification 4 with mortgage rejection indicator being

[8] We use zip-code dispersion instead of property-level dispersion because our price dispersion measure is estimated using Corelogic Deeds, and the data vendor prohibited us from merging loan-level records in LLMA with property-level records in Corelogic Deeds and Tax. We therefore aggregate property-level price dispersion measures to the most granular geographic region in LLMA.

the outcome variable.[9] Panel A of Table 5 reports the results. We first confirm the effect of local house price dispersion on mortgage rejection using full sample (column 1). Zip-code house price dispersion is positively and significantly associated with mortgage rejections. This result holds for both securitized loans (column 2) and portfolio loans (column 3). The rejection rate increases by about 1.4 percentage points as house price dispersion increases by one standard deviation. The effect is economically significant: given the sample average rejection rate of about 16%, the estimate amounts to about 10% increase in rejection likelihood.

We provide more direct evidence for the collateral channel by focusing on rejections due to collateral reasons in Panel B of Table 5. A mortgage application is about 50bps more likely to be rejected due to collateral reasons in a zip code with one standard deviation higher house price dispersion, which is about 25% increase in rejection likelihood. Again, the result holds in the full sample (column 1) as well as sub-samples of securitized loans (column 2) and portfolio loans (column 3).

Mortgage Rejection Reasons. As robustness, Table 6 reports the relationship between house price dispersion and the likelihood of being rejected due to different reasons conditional on being rejected. As the sample means indicate, the most common rejection reasons in the entire sample are creditworthiness related reasons (i.e., credit score and debt-to-income ratios). However, as the house price dispersion increases, the results clearly show that the mortgage rejection is significantly more likely to due to collateral reasons and less likely to be due to creditworthiness reasons, supporting our baseline findings.

3.2 Identification

3.2.1 Identification Assumptions of Baseline Results

Our baseline specification in Section 3.1 exploits within county-year variation by comparing two properties that are bought in the same county-year, at the same price, and by buyers with similar credit profile and income. In the second stage of $Dispersion_{ikt}$ estimation (Eqn.

[9] Again, We use zip-code dispersion instead of property-level dispersion because our price dispersion measure is estimated using Corelogic Deeds. We aggregate property-level price dispersion measures to zip-code level and assign it to every loan application in HMDA based on borrowers' location recorded by lenders of the mortgage.

2), we project the squared residuals from the first stage hedonic model onto the entire set of house characteristics. In a sense, house characteristics are used as instruments for price dispersion. The identifying assumption of the baseline specification 4 is that conditional on house price and borrower and loan characteristics, characteristics of a house only affect mortgage outcomes, insofar as they affect price dispersion.

This assumption can be violated if unobservable borrower characteristics are correlated with certain house features. For example, if borrowers who are more likely to purchase smaller, and thus high price dispersion, houses are also less creditworthy after conditioning on hard information like credit score and debt-to-income ratio, then the baseline specifications are subject to omitted variable bias.

We address this concern by instrumenting price dispersion using measures of house heterogeneity relative to the local housing stock. Our baseline results are robust to the instrumental variable approach, which will be discussed in the rest of this section.

3.2.2 Instrumental Variables

We construct a set of house-level instruments, for the price dispersion of each individual house i in county c, by measuring its heterogeneity relative to the local housing stock.[10] For all houses transacted in each county c, we first calculate the average value of each key house features ($\overline{X}_c^m$), where

$$m \in \{\text{building age}(age), \text{size}(sqft), \text{bedrooms}(bed), \text{bathrooms}(bath), \text{geo-coordinates}(geo)\}.$$

For each house, there are 5 instruments, one for each characteristic m. The instrument Z_i^m is equal to the squared difference between the house's feature m, and the average value of m in county c, that is:

$$Z_i^m = (X_i^m - \overline{X}_c^m)^2, \ \ \forall m \in \{age, sqft, bed, bath, geo\}, \tag{5}$$

[10]The approach of using measures of house nonstandardization as instruments is not new to the literature: similar ideas are used in Andersen et al. (2021), and the approach can be micro-founded in a search and matching framework as done in Guren (2018).

Then, we estimate the following 2SLS specification:

$$\begin{aligned}
&\text{Stage 1: } Dispersion_{it} = \alpha + \beta_1 Z^{age}_{it} + \beta_2 Z^{sqft}_{it} + \beta_3 Z^{bed}_{it} + \beta_3 Z^{bath}_{it} + \beta_4 Z^{geo}_{it} \\
&\qquad\qquad + X_{ikt}\Gamma + \mu_{kt} + \nu_m + \epsilon_{ikt} \\
&\text{Stage 2: } Y_{ikt} = \alpha + \beta \widehat{Dispersion}_{ikt} + X_{ikt}\Gamma + \mu_{kt} + \nu_m + \epsilon_{ikt},
\end{aligned} \tag{6}$$

where $\widehat{Dispersion}_{ikt}$ is the predicted value from stage 1.

The intuition behind the instruments is that they measure how locally thin the market is for a given house i by benchmarking it to other houses within the same county. Small houses, for example, will have large Z^{sqft} in a county with mostly large houses but will have small Z^{sqft} in a county with mostly small houses. Since markets for small houses are thinner in the former than in the latter, there are likely fewer buyers at any given point in time in the former.

For our analyses that use zip-code, we take geographical averages of Z_i^m, which is a measure of how heterogeneous the housing stock in the zip-code is, along characteristic m. Zip-codes with more heterogeneous housing stocks will tend to have higher price dispersion, since they have thinner local markets for any individual house, so there are likely to be less interested buyers for any given house.

Relevance and Exclusion Restriction. Consistent with the literature, Table A1 shows that our instruments are correlated with the raw price dispersion measure in a statistically significant manner. In terms of exclusion restriction, our instruments alleviate the endogeneity concern to the extent that lenders' willingness to lend against certain house feature — e.g., driven by the clustering of unobserved borrower creditworthiness and house features — is monotone.

We believe this is a reasonable assumption because it is implausible that borrowers purchasing atypical houses in a given county could be systematically less creditworthy than other borrowers. To support this argument, we conduct a balance test in Table A2. In Panel A, we regress price dispersion on a set of credit-related borrower characteristics. column 1 reports the results using the raw price dispersion measure, while columns 2-4 use the instrumented price dispersion measures with different fixed effects that correspond to the three first-stage columns in Table A1. Unsurprisingly, households with lower credit score and lower income are more likely to live in high-dispersion zip-codes defined using our raw price dispersion (column 1). Yet, the instruments alleviate the concern about such sorting: zip-codes with more heterogeneous housing stock within a county do not systematically attract less credit-worthy or low-income households. In Panel B, we examine whether home buyers in higher price dispersion zip codes have different credit-related characteristics. We reach similar conclusions: the raw price dispersion measure is correlated with FICO, income, and household age, but the instrumented price dispersion is not statistically significantly correlated with these characteristics.

Moreover, we will show in Section 3.2.4 that, while the raw price dispersion measure is correlated with borrower default rates, our instrumented price dispersion is not, lending support to the idea that the instrument alleviates endogeneity due to borrower sorting.

3.2.3 IV Results

We estimate specification 6 for every credit outcomes in our baseline analyses in Section 3.1. We confirm our baseline results qualitatively and get reasonably stronger estimated effects.

LTP. Table 3 columns 4-6 present the LTP results, where column 5 corresponds to Specification 6, and column 4 and 6 are less and more saturated specifications, respectively. LTP decreases by 45bps for every one standard deviation increase in the estimated price dispersion in the most saturated IV specification.

Interest Rates. Table 4 columns 4-6 present the interest rate results. For every one standard deviation increase in zip-code average house price dispersion, the mortgage rate increases by 2.2bps in the full sample (column 4), increases by 2.2bps in the sample of securtized loans (column 5), and increases by 5.32bps for portfolio loans (column 6).

Approvals. Lastly, Table 5 reports the loan approval likelihood results. Zip-code house price dispersion is positively and significantly associated with mortgage rejections: the rejection rate increases by more than 2 percentage points as house price dispersion increases by one standard deviation (Panel A columns 4-6). As shown in Panel B, a mortgage application is about 80bps more likely to be rejected due to collateral reasons in a zip code with one standard deviation higher house price dispersion (Panel B column 4-6). Both results — overall rejection or rejection due to collateral reasons — hold in the full sample as well as sub-samples of securitized loans and portfolio loans.

3.2.4 Unobservable Buyer Creditworthiness

An important identification assumption of our empirical design is that home buyers of houses with high price dispersion are not more likely to default on their mortgage, after conditioning on observable borrower and loan characteristics. To address this, we assess the ex-post performance of mortgage loans, to test whether ex-post default rates are associated with house price dispersion. Table 7 Panel A estimates the specifications 4 (columns 1-3) and 6 (columns 4-6) but sets the outcome variable equal to 100 for loans that become 60 or more day-delinquent within 2 years after origination and zero otherwise. Columns 1 and 4 include the full sample. Columns 2 and 5 restricts the sample to securitized loans. Columns 3 and 6 restricts the sample to portfolio loans. All regressions include the full set of borrower and loan characteristics as in our main regression specifications.

The 2SLS results suggest that home buyers of houses with higher instrumented price dispersion are not more likely to default on their loans than home buyers of houses with lower instrumented price dispersion. This alleviates the concern that our IV results are driven by unobserved differences in buyer creditworthiness, that are associated with our house nonstandardization IV.

Note that, in our OLS specifications, the coefficient estimate on price dispersion is positive and statistically significant. This could be because certain house characteristics, which are associated with higher house price dispersion, also tend to attract homeowners who have higher default rates. If this were the case, it would upwards bias our OLS estimates of the effect of price dispersion on LTPs. This further validates the importance of using our instrument, which is associated with price dispersion, but is not associated with homeowners' default rates.

3.3 Robustness

Lender Market Power. The results are not likely to be driven by lender market power. Firstly, our empirical analysis exploits within county-year variation. Existing literature on local lender market power find local competition at county level. Therefore, it is reasonable to believe that buyers from the same county-year with similar creditworthiness are facing the same credit supply. To address further concerns about the effect of lender market power, we re-estimate specification 4 and 6 with lender-zip-year fixed effects using a sub-sample of house transactions in Corelogic Deeds records that we also observe the mortgage interest rates. Note that we cannot do this robustness check using Corelogic LLMA data as we did in Section 3.1 because we do not observe lender ID in the LLMA dataset. The inclusion of lender-zip-year fixed effects allows us to compare houses financed by the same lender-zip-year.

Panel A of Table 8 reports the results. The key variable of interest is price dispersion, which is property-level idiosyncratic price dispersion. We first confirm Table 4 results using this sub-sample in column 1. In columns 2-3, we add in more saturated lender fixed effects: lender-county-year and lender-zip-year fixed effects, respectively. The results hold in all specifications, confirming that the effect of house price dispersion on mortgage credit is not driven by lender market power.

Bunching Below Conforming Loan Limits. Lastly, we test whether the effect of price dispersion on mortgage LTP and cost menu is driven by home buyers lowering the loan-to-price ratio to be eligible for securitization with the participation of government-sponsored enterprises (GSEs). Specifically, conforming mortgages must be below the conforming loan limits, which vary across regions and time. Conforming loans are much easier to sell than non-conforming loans, also known as jumbo loans, because of the participation of GSEs. GSEs insure default risks of loans they purchase and securitize, providing subsidized credit to GSE mortgage borrowers.

We test if our main findings are robust to the sub-sample of house transactions with sale prices below local conforming loan limits. These house transactions are not subject to the concern about bunching below conforming loan limit as the transaction prices are already below the conforming loan limit.

Panel B of Table 8 reports the results. The results show that our main finding is not driven by home buyers' incentive to keep their loan amount below the conforming loan limit. Among houses with prices below the conforming loan limit, houses with higher price dispersion are financed with smaller loans given the same interest rates than houses with lower price dispersion. The result holds in both OLS and IV settings.

4 Model

4.1 Model Overview

We construct a structural model showing how price dispersion affects mortgage loan-to-value ratios (LTVs), interest rates, and application failures. We will then calibrate the model to show that the model can quantitatively rationalize the empirical results and to decompose various channels.

In the model, given a menu of pairs of interest rates and targeted LTVs, a prospective homebuyer chooses a targeted mortgage size to finance a house at an exogenous transaction price. Higher loan-to-values are riskier to lenders, so lenders require higher interest rates for higher loan-to-values. Moreover, since lenders' payoff in case of borrower default is a concave function of the foreclosure house price, the menu is uniformly worse when houses have high price dispersion.

Mortgages are also subject to a regulatory constraint on LTVs, which depends on house appraisal values. After the buyer chooses a targeted loan size, a third party generates a random appraisal value for the house. If the house over-appraises, the transaction proceeds according to its original terms. If the house under-appraises, the buyer must make an increased down payment in order for the transaction to proceed. Buyers can also choose to renege on the transaction altogether, paying a fixed cost to restart the house purchase process; we consider this to be a mortgage rejection. When idiosyncratic price dispersion is higher, appraisals are noisier, and under-appraisals are more likely and larger when they occur, increasing the probability that transactions fail.

The buyer thus faces a tradeoff. Increasing the target loan size smooths consumption more effectively if the house over-appraises. However, there are two costs of targeting larger loans: lower expected collateral recovery implies that lenders charge higher interest rates for larger loans; and larger loans increase appraisal risk, making under-appraisals more likely and more costly when they occur. When idiosyncratic price dispersion is higher, lenders offer worse interest rate menus, and under-appraisals are more likely; both forces push buyers towards choosing smaller mortgages.

4.2 Setup

4.2.1 The Buyer's Problem

A homebuyer attempts to finance a house that is sold at price P. The buyer's decision determines her consumption in two time periods: the first period is when the buyer purchases the house, and the second is when the mortgage loan is paid back. The buyer has CRRA utility, discounting consumption at rate β^T between periods

$$U(c_1, c_2) = \frac{c_1^{1-\eta} - 1}{1-\eta} + \beta^T u_2' c_2 \tag{7}$$

where u_2' is an exogeneous constant.[11] Hence, the buyer is attempting to solve a consumption smoothing problem, where utility is concave in the first period, and linear in the second. The buyer receives exogenous labor income W_1 in period 1, and W_2 in period 2.

The buyer faces a two-stage problem:

1. Lenders offer an interest rate menu $r(L, \sigma)$, determining the mortgage interest rate if the buyer targets loan size L and idiosyncratic price dispersion is σ. The buyer chooses a target loan size L, receiving interest rate $r(L, \sigma)$. We introduce how rate menu is determined in the next section.

2. The house appraisal value A is determined. The collateral value used to calculate the LTV of the mortgage takes the smaller of the appraisal value A and the transaction

[11] This functional form is a simplified version that makes a linear approximation to utility in period 2 of the following standard CRRA utility:

$$U(c_1, c_2) = \frac{c_1^{1-\eta} - 1}{1-\eta} + \beta^T \frac{c_2^{1-\eta} - 1}{1-\eta} \tag{8}$$

A similar linear approximation to utility in future periods is used in Jansen et al. (2022). In our setting, this modelling simplification is needed in order to make the appraisal problem recursive, allowing us to use tools from the search literature to model the buyer's response to under-appraisals.

price P:[12]

$$L_{final} \leq \phi \min(P, A). \tag{9}$$

If $A < P$, the final loan amount L_{final} will be below the target size L, so the buyer will need to make an additional down payment. Conditional on A, the buyer can choose to continue the transaction, or to renege, pay a fixed penalty cost, and searching for a new house, returning to period 1.

In the following, we normalize final loan size, target loan size, and appraisal values:

$$l_{final} = \frac{L_{final}}{P}, l \equiv \frac{L}{P}, a \equiv \frac{A}{P} \tag{10}$$

Hence, the target LTV is l, the final LTV is l_{final}, and the ratio of appraisals to transaction prices is a. We will write $r(l)$ to mean the interest rate if the target LTV is l. We proceed to describe the buyer's payoffs if she chooses to continue with a transaction, then if she decides to renege.

Continuation. From (9), if $a < 1$, the final loan size is capped at:

$$\phi \min(P, A) = \phi P \min(1, a) = \phi a P \tag{11}$$

Since we have restricted the target loan size to $l < \phi P$, the buyer's final loan size is $P \min(l, \phi a)$. If the buyer originally planned to borrow l, making down payment $P(1-l)$, the appraisal further constrains loan size relative to l whenever $a < \frac{l}{\phi}$. With appraisal a, the down payment is $P(1-\phi a)$, which is $P \max[0, l - \phi a]$ larger than the targeted down payment. We assume that, if the buyer faces such a down payment gap, this decreases her period-1 consumption c_1 by $\psi P \max[0, l - \phi a]$, where $\psi > 1$. That is, for every dollar in additional downpayments she must make, the buyer's period-1 consumption decreases by $\psi > 1$ dollars. This is a reduced-form modelling device, capturing the idea that an unanticipated increase in down payments, induced by an under-appraisals, is more costly than an anticipated increase. This is because the buyer can smooth consumption more effectively if she anticipates and plans to make a large down payment; if the buyer suddenly learns that

[12]This is imposed by both bank regulators and mortgage securitizers in reality.

she must make a large down payment, consumption smoothing is more difficult.[13]

Given an appraisal a, the buyer's consumption in period 1 is:

$$c_1 = \underbrace{W_1}_{\text{labor income}} - \underbrace{P(1-l)}_{\text{target down payment}} - \underbrace{\psi P \max[0, l - \phi a]}_{\text{penalty term from under-appraisal}} \tag{12}$$

That is labor income less the target down payment for the house, less the penalty term from under-appraisal. Consumption in period 2 is:

$$c_2 = \underbrace{W_2}_{\text{labor income}} - \underbrace{(1 + r(l))^T P(l - \max[0, l - \phi a])}_{\text{mortgage principal and interest}} \tag{13}$$

This is labor income, minus the principal and interest on the mortgage, which we assume is paid in a single lump sum in period 2. Since utility in period 2 is linear, the term W_2 simply increases the level of utility and does not affect any outcomes, so for notational simplicity we will set $W_2 = 0$ going forwards.

Reneging. If the appraisal is too low, the buyer can renege on the transaction, paying a cost ζ (as a fraction of house price), and then searching for a new house. For tractability, to make the problem recursive, we think of ζ as being paid in period 2 dollars. We think of this as capturing, for example, foregone deposits if there is no appraisal contingency in the sales contract, or hassle costs of searching for another house. They then revert to stage 1, to purchase another house, and have continuation value:

$$-\beta^T u_2' \zeta P + E_a[V(a,l)] \tag{14}$$

where $V(a,l)$ is the value of choosing loan size l, when the appraisal is a.

[13] Formally, consider a multi-period consumption-savings model, in which a buyer saves to purchase a house. If a buyer learns at the start of her lifecycle whether there is an increase in the required down payment for a house, this has relatively low cost, since the homebuyer can increase savings in the many years before she purchases the house. On the other hand, if the buyer only learns whether the down payment will increase in the period before she buys a house, this has higher cost, since the buyer cannot condition her savings decision on whether the down payment increases: she is left either over-saving for a low down payment, or under-saving for a high down payment. We demonstrate this quantitatively in Appendix B.5. The parameter ψ can be thought of as a reduced-form model capturing the increased utility cost of down payment shocks which cannot be anticipated.

4.2.2 Interest Rate Menu

We assume that the interest rate lenders offer depends on price dispersion and the size of the mortgage. Mortgages which are larger, and which are in higher-dispersion areas, are riskier, and lenders will thus charge higher interest rates as a result. In the main text, we assume a simple reduced-form model of the rate menu:

$$r(l, \sigma) = \bar{r} + \theta_l l + \theta_\sigma \sigma \tag{15}$$

where θ_l and θ_σ capture the dependence of the interest rate on loan size and price dispersion respectively. In Appendix B.3, we construct a more detailed microfoundation of the interest rate menu, based on competitive profit-maximizing lenders making loans with default rate and imperfect collateral recovery, and we show that the model can qualitatively and quantitatively rationalize the rate menu observed in the data.

4.2.3 The Distribution of Appraisal Values

It is known in the literature that house appraisals are systematically biased upwards, and there is substantial bunching at house transaction prices. Empirically, we observe that the distribution of appraisal prices bunches at the sale price (Figure 4a): large over-appraisals are also rare, suggesting that appraisers largely only bias appraisals upwards to the point where they are equal to sale prices.

We construct a model of the distribution of appraisals which matches these stylized facts. We assume there is an unbiased appraisal value, which is normally distributed around the house transaction price, $A_{raw} \sim N(P, \sigma)$. The appraisal value A given to the borrower is then determined by:

$$A = \begin{cases} A_{raw} + Pb & A_{raw} < P(1-b) \\ P & P(1-b) \leq A_{raw} < P \\ A_{raw} & P \leq A_{raw} \end{cases} \tag{16}$$

In words, (16) states that, when A_{raw} is above P, it is not necessary to further increase A, so appraisers simply report the raw appraisal price $A = A_{raw}$. When A_{raw} is below P but above $P(1-b)$, the appraisers biases A just enough so that it is equal to the transaction price P, generating bunching at P. When A_{raw} is below $P(1-b)$, appraisers still attempt to bias A upwards, but are only able to push it to $A_{raw} + Pb$. This is still useful to the buyer, since any upwards bias allows the buyer to receive a larger loan than if the appraisal were simply A_{raw}.[14] We will estimate b based on the distribution of appraisal-to-sale ratios in our data, as we describe in Subsection 5.1 below.

4.3 Model Outcomes

Optimal behavior in the model is described by buyers' optimal target loan size choice l and buyers' optimal decision about whether to continue or renege on the transaction for each possible value of a. The following theorem characterizes optimal buyer behavior.

Theorem 1. *For any parameter settings, and for any target loan size l, there is an optimal appraisal cutoff $\bar{a}(l)$, which is the unique value that satisfies:*

$$\omega(\bar{a}, l) = -\beta^T u_2' \zeta P + \int_0^\infty \max\big(\omega(a, l), \omega(\bar{a}, l)\big)\, dF_a(a) \tag{17}$$

where $\omega(a, l)$ is defined as:

$$\omega(a, l) \equiv u_1\big(W_1 - P(1-l) - \psi P \max(0, l - \phi a)\big) + u_2' \beta^T \big(1 + r(l)\big)^T P \max[0, l - \phi a] \tag{18}$$

The buyer optimally continues with the purchase for any $a > \bar{a}(l)$, and reneges on the transaction for any $a < \bar{a}(l)$. The buyer chooses target loan size l to solve:

$$l^* = \arg\max_l \left(-\beta^T \big(1 + r(l)\big)^T u_2' P l + \int_0^\infty \max\Big(\omega(a, l), \omega\big(\bar{a}(l), l\big)\Big)\, dF_a(a)\right) \tag{19}$$

The proof of Theorem 1, and further properties of the buyer's choice problem, are described in Appendix B.1. In words, Theorem 1 states the following. Conditional on any target loan size l, buyers will continue the transaction if the house appraises to at least

[14] In Appendix B.4, we show that (16) can be microfounded in a simple model based on Calem et al. (2021). In the model, appraisers have a convex cost of biasing appraisals upwards, and receive some linear side benefit – for example, from increased future business – to the extent that they are able to increase the amount that buyers can borrow on the loan. In this model, appraisals bunch at sale prices, because appraisers face positive costs, but no benefit, of biasing appraisals upwards past the transaction price, since the transaction price then binds in (9), and further increases in A do not affect the amount that can be borrowed.

$\bar{a}(l)$, and will renege otherwise. The cutoff $\bar{a}(l)$ is the value of the appraisal such that the consumer is just indifferent between her two options: continuing with the transaction and making a higher down payment, thus receiving the LHS of (17); and reneging, thus receiving the RHS of (17), which is negative the cost ζ multiplied by house prices and period-2 marginal utility, plus the expected value from buying a new house.

To find the optimal loan size target, (19) states that buyers simply maximize expected utility from the second-stage problem over l. In Appendix B.2, we derive a first-order condition for optimal loan choice. The buyer faces a tradeoff: larger loan sizes smooth consumption more effectively if the house over-appraises, but lead to higher interest rates, and also larger under-appraisals and thus larger consumption penalties in period 1 upon under-appraisal. At the optimal loan choice, buyers target a loan size slightly smaller than they would if the house never under-appraised: that is, buyers decrease borrowing, limiting consumption smoothing, in order to decrease interest rates, and the likelihood and size of under-appraisals.

5 Why Does Price Dispersion Affect Credit?

We next calibrate the model to show that the model can quantitatively rationalize the observed relationship between price dispersion and mortgage outcomes, and to decompose the channels through which outcomes are affected by price dispersion. While it is intuitive that higher price dispersion should lead to lower mortgage credit provision, the mechanism through which this occurs in our model is subtle, involving two distinct channels. The first one is the *collateral recovery channel*: lenders offer a more expensive rate menu to buyers of houses with larger price dispersion, which gives buyers incentives to scale down loan size to get an affordable mortgage rate. The second is the *appraisal risk channel*: houses with larger price dispersion face higher under-appraisal risk, hence buyers scale down their targeted loan sizes to minimize the cost of under-appraisal. The two channels have differential effects on the three primary measures of credit provision in the model – LTPs, interest rates, and mortgage rejections – and the calibration allows us to shed light on the mapping between channels and outcomes.

5.1 Calibration

We calibrate several parameters externally in relation to existing literature. We then estimate the remaining parameters by matching model-implied moments to the moments in the data. Table 9 summarizes the estimates.

Externally Calibrated Parameters. We set the intertemporal elasticity of substitution (η) to 2, as chosen in standard lifecycle models. We set period 1 wealth to \$60,000 and the house price to \$200,000. We set $\beta = 0.96$. We set $T = 7$, approximately equal to the duration of a 30-year mortgage.[15] The maximum LTV parameter ϕ is set to 0.8, which is the most common regulatory threshold.

Parameters Calibrated to the Data or through Moment Matching. A core parameter is the standard deviation of raw appraisal values, $\sigma_{a,i}$. We calibrate the standard deviation of appraisal prices in each of 10 quantile buckets of σ values. Under our model of appraisal in (16), the raw appraisal values a_{raw} are distorted only when they are below the transaction price, $a_{raw} < 1$. Thus, the distribution of realized appraisals, conditional on over-appraisal should be identical to the distribution of a_{raw}. Since we also assume raw appraisals have mean equal to the house price, we can thus estimate σ_i^A as:

$$\hat{\sigma}_{a,i} = \sqrt{E\left[(a_i - 1)^2 \mid a_i > 1\right]} \tag{20}$$

That is, $\hat{\sigma}_{a,i}$ is simply the conditional mean squared error of appraisals around 1. Using (20), we calculate $\hat{\sigma}_{a,i}$ for each quantile bucket of σ values.

To calibrate the interest rate menu, $r(l, \sigma)$ from expression (15), we assume:

$$r(l, \sigma) = \bar{r} + \theta_l (l - 0.8) + \theta_\sigma (\sigma - \bar{\sigma}) \tag{21}$$

That is, the interest rate $r(l)$ is equal to a constant $\bar{r}$, plus θ_l times the target LTV, plus θ_σ times idiosyncratic price dispersion. We set $\bar{r}$, the interest rate for a mortgage with $l = 0.8$, and $\sigma = \bar{\sigma}$, to $\frac{1}{\beta} - 1$, which is approximately 4.17%. We set θ_r and θ_σ to their values in

[15] Mortgages amortize and are prepayable, so their average duration is much lower than 30 years; see for example Krishnamurthy and Vissing-Jorgensen (2011).

Column 6 in Panel A of Table 4.

We choose the remaining parameters, ζ, b, ψ, u_2', through moment matching. For each σ-bucket of counties, we compute two moments. The first is the average probability of transaction failures due to under-appraisals. In the data, we calculate this as the rate of collateral-related mortgage failures.[16] In the model, we calculate mortgage failure rates as $F_a(\bar{a})$, the probability that the appraisal a falls below the boundary $\bar{a}$ below which the buyer reneges on the transaction.

The second set of moments is the "appraisal deviation" within each σ-bucket, that is:

$$ApprDev_i = p_i^{under} E\left[\frac{a}{p} - 1 \mid \text{under-appraisal}\right] \tag{22}$$

That is the product of the under-appraisal probability, and the expectation of the the percentage deviation of appraisal prices to sale prices conditional on under-appraisal. Theoretically, this corresponds to the product of the expectation of the probability of under-appraisal, conditional on not reneging, and the conditional expectation of the gap conditional on under-appraisal without reneging.[17]

We visualize this moment in the data in Figure 4 and provide regression analysis in Appendix D.2. The figure shows that the under-appraisal probability is only weakly related to σ, but the size of under-appraisal conditional on under-appraisal increases from roughly 3% to 6% when moving from the lowest to highest buckets. The result, however, is that, $ApprDev_i$, which is the product of these two quantities, is strongly related to price dispersion in the data. In the model, we calculate (22), conditioning first on appraisal values that do not result in transaction failure, since failures are not included when we calculate $ApprDev_i$ empirically. The intuition behind the model is that the magnitude of under-appraisal pressure depends on the shape of the appraisal distribution; this is controlled largely by the appraisal bias parameter b. Buyers' preferences then determine whether under-appraisals

[16]To be precise, we calculate the mortgage failure rate as $\frac{collFailure_c}{collFailure_c + mortgage_c}$, where $collFailure_c$ is the total number of collateral-related mortgage failures in county c, from the HMDA data, and $mortgage_c$ is the total number of mortgages in county c.

[17]In principle, we could target either $ApprDev_i$, or the probability of under-appraisal, in each quantile bucket. We cannot target both, as the model has difficulty simultaneously matching both moments. This is because, as we show in panel (c) of Figure 5, the distribution of appraisal values, conditional on under-appraisal, is fairly long-tailed in the data. However, in the model, the consumer tends to renege on the transaction when appraisal values are too low, so the conditional appraisal distribution in the model is truncated from below. Thus, if we match appraisal probabilities in the model and the data, $ApprDev_i$ would tend to be much higher in the data than in the model. We choose to target the conditional appraisal deviation, because this appears to be a better measure of the downwards pressure that under-appraisals generate for sale prices, compared to the simple under-appraisal probability.

mostly result in transaction failure, or under-appraisals with larger down payments; the level of the failure-underappraisal tradeoff, and its relationship with σ, is affected by the consumption penalty parameter ψ, the fixed cost of transaction failure ζ, and consumers' utility from period-2 consumption u_2'.

5.2 Results and Model Fit

The estimated penalty for reneging (ζ) on a transaction is 17.5% of house prices, paid in period-2 dollars. Appraisers bias house prices upwards approximately 7.9% (b). The penalty for under-appraisal-induced consumption decreases ($\psi - 1$) is approximately 61.4%. We view these as roughly reasonable parameter values. While ζ is somewhat high, this may be due to our assumption that period-2 utility is linear, which increases the amount consumption must decrease to decrease utility a given amount. We show in Appendix B.5 that values of ψ in roughly this range can be attained, if consumers receive large shocks that hit suddenly in one period and cannot be saved for in advance.

Targeted Moments. The fit of the model is shown in Figure 5. The estimated appraisal standard deviations $\hat{\sigma}_{a,i}$ are shown in panel (a). In the data, the conditional standard deviation of appraisals is monotonically higher for higher σ buckets, and we feed this directly into the model.[18] Panel (b) shows the CDF of appraisals, conditional on transactions not failing, in the model and the data, for the fifth σ-percentile bucket. We are able to match the main stylized facts about the appraisal distribution: the bunching of appraisals at 1, the relatively low probabilities of under-appraisal, and the relatively large probabilities of over-appraisal. There are two main differences between the model and the data. First, the right tail of the appraisal distribution in the data deviates slightly from the normal distribution: the empirical distribution of appraisals is more likely than the model distribution to be either

[18] We note two features of the appraisal distribution. First, the implied number of houses entering appraisals is somewhat high. In the fifth percentile bucket, we estimate σ to be 22.9%, whereas we estimate σ_a to be 6.03%; this implies a number of transactions of $\left(\frac{\sigma}{\sigma_a}\right)^2 = 14.4$. This is somewhat high; anecdotally most houses use approximately 3-4 appraisals in practice. Second, the relation between σ and σ_a is somewhat weak. If σ_a were simply determined by drawing a number of independent price draws, we should have $\sigma_a = \frac{\sigma}{\sqrt{N}}$, so the implied N should be the same for each percentile bucket. Instead, σ_a scales less than proportionally with σ: in the highest percentile bucket, we get an implied N of 30.8, and in the lowest we get an implied N of 6.44. One possible explanation of these discrepancies is that, first, our estimate of σ is somewhat higher than the effective value used by appraisers, due to model misspecification, or the fact that appraisers observe somewhat more features of houses than we do; this would explain why our implied N is too high. Second, our measure of σ may contain some measurement error, leading to a weaker than proportional relationship between our measured σ and σ_a.

quite close to 1 or quite far from 1. Second, the left tail of the appraisal distribution is longer in the data than in the model. This is because, in the model, appraisal values that are too low result in transaction failure, so the distribution of a conditional on under-appraisal is truncated below.

Panels (c) and (d) show, respectively, the values of the two sets of targeted moments, mortgage failure probabilities and appraisal deviations, in the model and the data. Empirically, both moments are monotone with respect to changes in σ: counties with higher idiosyncratic price dispersion have monotonically higher collateral-related mortgage failures, and higher appraisal deviations. The fitted model matches the average level of both moments fairly well; the main difference is is that the relationship between both outcomes and σ is slightly stronger in the model and in the data.

Untargeted Moments. The main result is the optimal loan choice plot, shown in panel (e) of Figure 5, where we show the final realized loan size l_{final}, against idiosyncratic price dispersion σ. Loan-to-value ratios are systematically lower when idiosyncratic price dispersion is higher. Moreover, the magnitude of the implied relationship is quite close to the estimated empirical relationship between σ and LTP. In the model, shifting σ by one standard deviation changes the average value of l_{final} by roughly 0.2. The estimated magnitude is close to the OLS estimates in columns 2 and 3 of Table 3 but smaller than the IV estimates.

5.3 Decomposition of Channels: Collateral Recovery and Appraisal Risk

We have shown that the model can quantitatively rationalize the relationship between price dispersion and mortgage outcomes. Next, we evaluate how the two channels in the model each contribute to driving variation in LTPs, interest rates, and loan rejections. We evaluate the relative contributions of each component by solving counterfactual models in which we remove the effect of price dispersion on recovery rate and the appraisal noise, respectively. Figure 6 presents the magnitude of each channel. In short, the figure shows that the collateral recovery channel has a large effect on rates, and a smaller effect on loan size and mortgage rejections, while the appraisal risk channel is has a large effect on loan size and rejection likelihood, and a smaller effect on interest rates.

We evaluate the magnitude of the collateral recovery channel in isolation by shutting down the appraisal risk channel. To do this, we set appraisal noise constant – $A_{raw} \sim N(P, \bar{\sigma})$ for all percentile buckets – and calculate mortgage outcomes for each price dispersion bucket in the new equilibrium. This shows how mortgage outcomes would vary if lenders shifted the interest rate menu depending on price dispersion, but appraisals were distributed identically for each price dispersion bucket. Analogously, to evaluate the magnitude of the appraisal risk channel, we shut down the collateral recovery channel, setting the rate menu constant across sigma buckets:

$$r(L, \sigma) = r(L, \bar{\sigma}) = r\left(l\right) = \bar{r} + \theta_r \left(l - 0.8\right) \tag{23}$$

and then solving for mortgage outcomes in the new equilibrium. This shows how mortgage outcomes would vary, if the distribution of appraisals varied across price dispersion buckets, but lenders offered the same menu in each bucket.

Panels (a) decomposes the effect of price dispersion on interest rate. The collateral recovery channel explains most of the effect of price dispersion on mortgage rates; the appraisal risk channel plays a more minor role. Intuitively, when lenders shift the interest rate menu upwards in response to higher house price dispersion, consumers can respond either by holding interest rates fixed and decreasing loan size, or holding loan size fixed and bearing increased interest rates. In our calibrated model, loan size is relatively inelastic to changes in interest rates, so consumers mostly respond by borrowing similar amounts and bearing higher interest rates. For every unit increase in house price dispersion, lenders charge 60bps higher interest rate for any given LTV.

On the other hand, the appraisal risk channel attenuates the observed price dispersion-interest rate relationship: when appraisal noise is higher, holding the rate menu fixed, equilibrium interest rates are actually *lower*. Intuitively, when appraisal noise is higher, consumers choose lower target loan sizes to risk the risk of under-appraisal. Through the interest rate menu in (23), a side benefit of lowering target loan size is that consumers get lower interest rates. However, note that the magnitude of this effect is very small: for every unit increase in price dispersion, the observed interest rate is about 4bps lower due to the appraisal risk channel. Given that the standard deviation of price dispersion is about 0.1, the collateral recovery channel maps to a 6bps increase in interest rate for every one standard deviation of house price dispersion, while the appraisal risk channel maps to a 0.4bps decreases in interest rate. The net effect is close to the estimates in Table 4 column 6.

Panel (b) decomposes the effect of price dispersion on mortgage failures. Here, we find that appraisal risk is the main driver, with the collateral recovery channel playing a smaller role. Intuitively, when appraisals are noisier, holding fixed target loan size, under-appraisals and mortgage rejections become more likely. Consumers respond by choosing smaller targeted mortgage sizes; this allows appraisals to be lower before they bind and force the consumer to increase down payments, incurring penalty costs to consumption. However, consumers do not scale down loan size enough to counteract the direct effect of higher appraisal noise; thus, loan size is smaller and mortgage rejections are higher when appraisals are noisier. For every unit increase in price dispersion, the mortgage is 11pp more likely to fail because appraisal noise increases with house price dispersion. On the other hand, the collateral recovery channel lowers the failure likelihood because home buyers of high price dispersion houses receive higher interest rates for any given LTV, which pushes down their targeted loan size. The net effect is about 1pp increase in failure likelihood for a one standard deviation increase in house price dispersion, which is close to the estimates in Table 5 Panel B column 6.

Lastly, Panel (c) decomposes the effect of price dispersion on loan-to-price ratio. As with mortgage failures, appraisal risk is the main driver of the dispersion-LTP relationship, with the collateral recovery channel playing a smaller role. The collateral recovery channel contributes 0.1pp to the price dispersion effect on loan-to-price ratio. Intuitively, banks offer higher prices for any given loan-to-price ratio when the price dispersion is higher, which makes home buyers of high-price dispersion houses optimally borrow less. The appraisal risk channel contributes about 1.5pp to the price dispersion effect on LTP. While both the ex-ante and ex-post effects are present, Panel (c) shows that the ex-ante appraisal risk effect is dominant in the calibrated model (1.25pp for every unit increase in price dispersion).

We note that there is a third *ex-post appraisal risk* channel through which price dispersion affects LTP ratios: when appraisals are noisier, the gap between l_{final} and l will tend to be larger, putting downward pressure on l_{final}. We measure the magnitude of this effect by comparing the difference between l_{final} and l, for different price dispersion bucket. We find that ex-post appraisal risk plays a relatively small role compared to the primary ex-ante appraisal risk channel.[19]

[19]The fact that the ex-post appraisal effect must be small is not only driven by our modelling assumptions, but

To summarize, the major reason why LTVs are lower when price dispersion is higher is that buyers scale down l due to under-appraisal risk. The main reason why interest rates are higher is due to lenders offering a worse menu of interest rates. While it is qualitatively intuitive that price dispersion should lead to lower levels of mortgage credit provision, our decomposition allows us to make more fine-grained predictions about how changes to policy or institutional settings, which influence how each channel in the model functions, would affect each of the three measures of mortgage credit provision. As an example of this, in the following section, we use the model to analyze how the impending shift to automated appraisals would affect each of our measures of credit provision.

6 Discussion and Policy Implications

6.1 Implications for Desktop Appraisals

Our findings have implications for the shift from human appraisals to automated appraisals. In 2021, the FHFA announced that banks and mortgage lenders could use automated appraisal software in place of human appraisals.[20] It is known in the literature that human appraisals tend to be distorted, so that they are generally equal to or higher than transaction prices (Calem et al., 2015; Eriksen et al., 2019; Bogin and Shui, 2020; Conklin et al., 2020; Calem et al., 2021; Kruger and Maturana, 2021). Automated appraisals are likely to be less distorted, but as a result, under-appraisals will be more frequent, especially in areas with high house price dispersion. Automated appraisal thus have the potential to hurt low-income households who tend to live in areas with less predictable house prices.[21]

To evaluate the extent to which the shift from human appraisals to automated appraisals affects mortgage outcomes, we use our calibrated model to conduct counterfactual analyses. We focus on one particular aspect of such shift: the removal of biases in human appraisals. Specifically, using our calibrated model, we remove appraiser bias, setting $b = 0$, and re-

follows simply from back-of-envelope calculation using the data. From Figure 4, roughly 3.5% of mortgages under-appraise, and the magnitude of under-appraisal is approximately 4.5%. This implies that average LTVs should be roughly $0.035 * 0.045 = 0.0016$, or roughly 16 basis points higher, when moving from the lowest to highest percentile buckets. This is of a similar order of magnitude to the size of the effect in our model.

[20] https://www.americanbanker.com/news/fhfa-will-make-desktop-home-appraisals-a-permanent-option

[21] Blattner and Nelson (2021) and Fuster et al. (2020) have made similar arguments that low-income households tend to have nosier hard information, and the development of FinTech is going to increase statistical discrimination in mortgage lending.

evaluate the model, estimating loan-to-price ratios, interest rates, and mortgage rejections.

Figure 7 compares the benchmark and the counterfactual mortgage outcomes. The removal of human biases in the appraisal process will have significantly impact at the extensive margin in terms of mortgage failure and loan size, while the impact on interest rate is relatively small. The results are not surprising, given our decomposition in Section 5.3 that the appraisal risk is the main driver of the price dispersion effect on loan size and mortgage failure. The shift will lower the loan size by about 2pp of the house price (Panel a), lower the interest rate by about 5bps (Panel b), and increase the mortgage failure rate by more than 10pp (Panel c). In high-dispersion areas, the mortgage failure rate increases by more than 15pp.

Technically, this exercise demonstrates how our calibrated model can be used to predict how changes in mortgage policy would affect various measures of credit provision. From a policy perspective, our results illustrate how the biases of human appraisers in fact act to alleviate the effects of price dispersion on mortgage credit availability. Shifting to automated appraisals, without compensating for the upwards bias in appraisal prices induced by human appraisers, has the potential to significantly decrease credit provision, especially for areas where price dispersion is high.

6.2 Effects on the Homeownership Gap

A large literature has analyzed how limited access to mortgage credit influences the gap in homeownership between high- and low-income households.[22] Policymakers aiming to improve homeownership rates for low-income households have considered interventions in credit markets as well as in housing markets. Our analysis highlights a link between these two markets: the amount of credit that mortgage lenders provide depends on the value uncertainty of the house used as collateral, even in a fully competitive mortgage market.

As we show in Figure 1b, low-income households tend to live in areas with higher house price dispersion on average, and thus likely receive lower mortgage LTVs as a result. This is not a market inefficiency, or a form of credit market discrimination: it is a rational response of lenders to the fact that more volatile assets are worse as collateral for debt.

[22] See, for example, the discussion in Herbert et al. (2005).

To quantify the effect of the borrowing constraints induced by collateral value uncertainty on homeownership gap, we calibrate a standard lifecycle model of housing choices in Appendix C. Through counterfactual analyses, we show in Table A7 that the aggregate homeownership rate would increase by 1.5pp if we lower the price dispersion in the high-dispersion counties, defined as top decile counties ranked by price dispersion, to the price dispersion level in the low-dispersion counties, defined as bottom decile counties ranked by price dispersion. We then divide households into two groups, according to their initial income at age 25. The effect of price dispersion on homeownership is concentrated among low-income households: at all ages, low-income households have lower homeownership rates in the high-dispersion counterfactual than the low-dispersion counterfactual, with an average homeownership rate difference of 2.6pp. The homeownership gap is large for young households below age 30, somewhat smaller for middle-aged households from 30-40, and rises again for households above 40. In contrast, high-income households initially have higher homeownership rates, but the gap declines essentially to 0 from age 30 onwards.

The difference in collateral constraints induced by collateral value uncertainty contributes to about 6.6% of the homeownership gap between the rich and the poor in 2016, ranging from 5% to 10% across the age distribution.[23] Therefore, our results suggest that, in a standard calibrated lifecycle model of housing choice, LTV differences induced by price dispersion can have sizable effects on aggregate homeownership rates, and the homeownership gap between high- and low-income households.

Our results thus provide a rationale for interventions in the mortgage market, such as the FHA program, which promote mortgage credit access for low-income households. The FHA program allows low-income households to borrow at loan-to-value ratios up to 96%, far higher than the LTVs that private lenders and GSEs offer. This distorts mortgage credit provision. However, as our findings suggest, since low incomes tend to live in areas with older and less standardized houses, they have restricted access to mortgage credit due to their lack of access to better housing collateral. By allowing low-income households to borrow at higher LTV ratios, the FHA program effectively alleviates this structural issue in the current housing stock.

[23] The homeownership gap between above-median income households and below-median income households is about 32% in 2016 (SCF Statistics). According to the report by the U.S. Department of Housing and Urban Development, the homeownership gap between the very low-income households and high-income households is 37% in 2004. https://www.huduser.gov/Publications/pdf/HomeownershipGapsAmongLow-IncomeAndMinority.pdf

6.3 Housing Affordability

Our results highlight that the collateral value of the housing stock is an important, and previously underappreciated, determinant of housing affordability. Newer, more standardized housing is better collateral, alleviates lenders' concerns about collateral recovery risk, and presents lower under-appraisal risks to borrowers. As a result, it is easier to obtain larger mortgages at lower rates against these houses, improving the affordability of the housing stock.

This suggests that urban policymakers, who regulate the construction and renovation of residential housing, should consider the effects of policies on the collateral value of the housing stock. By encouraging rebuilding and renovation, and by zoning in a way which promotes the development of standardized housing urban policy can potentially improve affordability by increasing average collateral values. Lenders would lend more against these houses, contributing to increasing homeownership rates for low-income households, even if these policies do not decrease house prices. Interestingly, this is a channel through which housing stock renewal disproportionately benefits low-income households and first-time homebuyers, since down payment constraints tend to be most binding for these households.

7 Conclusion

In this paper, we have shown that house value uncertainty affects mortgage credit provision in the US residential real estate market. Houses differ substantially in their degree of idiosyncratic price dispersion, which affects their value as collateral and thus the availability of mortgage credit. This effect is partially due to fair pricing of collateral recovery risk, and partly through the effect of idiosyncratic price dispersion on appraisal noise. Our results have implications for policy interventions in mortgage and housing markets aimed at improving credit access and homeownership, especially among low-income individuals and first-time homebuyers.

References

Adelino, Manuel, William B McCartney, and Antoinette Schoar, 2020, The role of government and private institutions in credit cycles in the us mortgage market, Technical report, National Bureau of Economic Research.

Agarwal, Sumit, Gene Amromin, Itzhak Ben-David, Souphala Chomsisengphet, Tomasz Piskorski, and Amit Seru, 2017, Policy intervention in debt renegotiation: Evidence from the home affordable modification program, *Journal of Political Economy* 125, 654–712.

Agarwal, Sumit, Gene Amromin, Souphala Chomsisengphet, Tim Landvoigt, Tomasz Piskorski, Amit Seru, and Vincent Yao, 2022, Mortgage Refinancing, Consumer Spending, and Competition: Evidence from the Home Affordable Refinance Program, *The Review of Economic Studies* rdac039.

Andersen, Steffen, Cristian Badarinza, Lu Liu, Julie Marx, and Tarun Ramadorai, 2021, Reference dependence in the housing market, *Available at SSRN 3396506* .

Bayer, Patrick, Christopher Geissler, Kyle Mangum, and James Roberts, 2011, Speculators and Middlemen: The Strategy and Performance of Investors in the Housing Market, Technical Report w16784, National Bureau of Economic Research, Cambridge, MA.

Benmelech, Efraim, and Nittai K. Bergman, 2008, Liquidation Values and the Credibility of Financial Contract Renegotiation: Evidence from U.S. Airlines, *The Quarterly Journal of Economics* 123, 1635–1677, Publisher: Oxford University Press.

Benmelech, Efraim, and Nittai K Bergman, 2009, Collateral pricing, *Journal of financial Economics* 91, 339–360.

Beraja, Martin, Andreas Fuster, Erik Hurst, and Joseph Vavra, 2019, Regional heterogeneity and the refinancing channel of monetary policy, *The Quarterly Journal of Economics* 134, 109–183.

Berger, David, Veronica Guerrieri, Guido Lorenzoni, and Joseph Vavra, 2018, House prices and consumer spending, *The Review of Economic Studies* 85, 1502–1542.

Bian, Bo, 2021, The real effects of legal and trading frictions in collateralization, *Working Paper* .

Blackburn, McKinley, and Todd Vermilyea, 2007, The role of information externalities and scale economies in home mortgage lending decisions, *Journal of Urban Economics* 61, 71–85.

Blattner, Laura, and Scott Nelson, 2021, How costly is noise? data and disparities in consumer credit, *arXiv preprint arXiv:2105.07554* .

Boehm, Thomas P, and Alan Schlottmann, 2008, Wealth accumulation and homeownership: Evidence for low-income households, *Cityscape* 225–256.

Bogin, Alexander N., and Jessica Shui, 2020, Appraisal Accuracy and Automated Valuation Models in Rural Areas, *The Journal of Real Estate Finance and Economics* 60, 40–52.

Buchak, Greg, Gregor Matvos, Tomasz Piskorski, and Amit Seru, 2018, Beyond the balance sheet model of banking: Implications for bank regulation and monetary policy, Technical report, National Bureau of Economic Research.

Buchak, Greg, Gregor Matvos, Tomasz Piskorski, and Amit Seru, 2020, ibuyers: Liquidity in real estate markets?, *Available at SSRN 3616555* .

Calem, Paul, Jeanna Kenney, Lauren Lambie-Hanson, and Leonard Nakamura, 2021, Appraising home purchase appraisals, *Real Estate Economics* 49, 134–168.

Calem, Paul S, Lauren Lambie-Hanson, and Leonard I Nakamura, 2015, Information losses in home purchase appraisals, Working paper, Working paper.

Case, Karl E, and Robert J Shiller, 1989, The efficiency of the market for single-family homes, *The American Economic Review* 125–137.

Collier, Benjamin L, Cameron Ellis, and Benjamin J Keys, 2021, The cost of consumer collateral: Evidence from bunching, Technical report, National Bureau of Economic Research.

Conklin, James, N. Edward Coulson, Moussa Diop, and Thao Le, 2020, Competition and Appraisal Inflation, *The Journal of Real Estate Finance and Economics* 61, 1–38.

DeFusco, Anthony A, 2018, Homeowner borrowing and housing collateral: New evidence from expiring price controls, *The Journal of Finance* 73, 523–573.

DeFusco, Anthony A, Stephanie Johnson, and John Mondragon, 2020, Regulating household leverage, *The Review of Economic Studies* 87, 914–958.

DeFusco, Anthony A, and John Mondragon, 2020, No job, no money, no refi: Frictions to refinancing in a recession, *The Journal of Finance* 75, 2327–2376.

Di Maggio, Marco, and Amir Kermani, 2017, Credit-induced boom and bust, *The Review of Financial Studies* 30, 3711–3758.

Di Maggio, Marco, Amir Kermani, Benjamin J Keys, Tomasz Piskorski, Rodney Ramcharan, Amit Seru, and Vincent Yao, 2017, Interest rate pass-through: Mortgage rates, household consumption, and voluntary deleveraging, *American Economic Review* 107, 3550–88.

Dokko, Jane, Benjamin J Keys, and Lindsay Relihan, 2019, Affordability, financial innovation, and the start of the housing boom .

Druedahl, Jeppe, and Thomas Høgholm Jørgensen, 2017, A general endogenous grid method for multi-dimensional models with non-convexities and constraints, *Journal of Economic Dynamics and Control* 74, 87–107.

Eriksen, Michael D., Hamilton B. Fout, Mark Palim, and Eric Rosenblatt, 2019, The influence of contract prices and relationships on appraisal bias, *Journal of Urban Economics* 111, 132–143.

Eriksen, Michael D., Chun Kuang, and Wenyu Zhu, 2020, Attribute (Mis)Reporting and Appraisal Bias, *SSRN Electronic Journal* .

Floden, Martin, and Jesper Lindé, 2001, Idiosyncratic risk in the united states and sweden: Is there a role for government insurance?, *Review of Economic dynamics* 4, 406–437.

Fout, Hamilton, Nuno Mota, and Eric Rosenblatt, 2021, When Appraisers Go Low, Contracts Go Lower: The Impact of Expert Opinions on Transaction Prices, *The Journal of Real Estate Finance and Economics* .

Fuster, Andreas, Paul Goldsmith-Pinkham, Tarun Ramadorai, and Ansgar Walther, 2020, Predictably unequal? the effects of machine learning on credit markets, *SSRN* .

Giacoletti, Marco, 2021, Idiosyncratic risk in housing markets, *The Review of Financial Studies* 34, 3695–3741.

Giacoletti, Marco, and Victor Westrupp, 2017, The Risk-Adjusted Performance of Asset Flippers, SSRN Scholarly Paper ID 2892439, Social Science Research Network, Rochester, NY.

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

Greenwald, Daniel L., 2016, The Mortgage Credit Channel of Macroeconomic Transmission, *SSRN Electronic Journal* .

Gupta, Arpit, Christopher Hansman, and Pierre Mabille, 2021, Financial constraints and the racial housing gap, *Available at SSRN 3969433* .

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

Hartman-Glaser, Barney, and William Mann, 2017, Collateral constraints, wealth effects, and volatility: Evidence from real estate markets, *Working Paper* .

Herbert, Christopher E, Donald R Haurin, Stuart S Rosenthal, and Mark Duda, 2005, Homeownership gaps among low-income and minority borrowers and neighborhoods, *Washington, DC: US Department of Housing and Urban Development. http://www. huduser. org/publications/HOMEOWN/HGapsAmongLInMBnN. html* .

Hurst, Erik, Benjamin J Keys, Amit Seru, and Joseph Vavra, 2016, Regional redistribution through the us mortgage market, *American Economic Review* 106, 2982–3028.

Jansen, Mark, Fabian Nagel, Constantine Yannelis, and Anthony Lee Zhang, 2022, Data and welfare in credit markets, *Available at SSRN 4015958* .

Jiang, Erica Xuewei, 2020, Financing competitors: Shadow banks' funding and mortgage market competition, *Available at SSRN 3556917* .

Kaplan, Greg, Kurt Mitman, and Giovanni L Violante, 2017, The housing boom and bust: Model meets evidence, Working Paper 23694, National Bureau of Economic Research.

Kermani, Amir, and Francis Wong, 2021, Racial disparities in housing returns, Technical report, National Bureau of Economic Research.

Kotova, Nadia, and Anthony Lee Zhang, 2021, Liquidity in residential real estate markets, *Available at SSRN 3386353* .

Krishnamurthy, Arvind, and Annette Vissing-Jorgensen, 2011, The effects of quantitative easing on interest rates: channels and implications for policy, Technical report, National Bureau of Economic Research.

Kruger, Samuel, and Gonzalo Maturana, 2021, Collateral Misreporting in the Residential Mortgage-Backed Security Market, *Management Science* 67, 2729–2750.

Lang, William W, and Leonard I Nakamura, 1993, A model of redlining, *Journal of Urban Economics* 33, 223–234.

LaPoint, Cameron, 2021, You only lend twice: Corporate borrowing and land values in real estate cycles, *Available at SSRN 3633606* .

Liu, Lu, 2022, The demand for long-term mortgage contracts and the role of collateral .

Lustig, Hanno N, and Stijn G Van Nieuwerburgh, 2005, Housing collateral, consumption insurance, and risk premia: An empirical perspective, *The Journal of Finance* 60, 1167–1219.

Mian, Atif, and Amir Sufi, 2011, House prices, home equity-based borrowing, and the us household leverage crisis, *American Economic Review* 101, 2132–56.

Pan, Guangqian, Zheyao Pan, and Kairong Xiao, 2021, The shadow cost of collateral, *Available at SSRN* .

Pennington-Cross, Anthony, 2006, The value of foreclosed property, *Journal of Real Estate Research* 28, 193–214.

Piazzesi, Monika, and Martin Schneider, 2016, Housing and macroeconomics, *Handbook of macroeconomics* 2, 1547–1640.

Piskorski, Tomasz, and Amit Seru, 2018, Mortgage market design: Lessons from the great recession, *Brookings Papers on Economic Activity* 2018, 429–513.

Sagi, Jacob S, 2021, Asset-level risk and return in real estate investments, *The Review of Financial Studies* 34, 3647–3694.

Shleifer, Andrei, and Robert W Vishny, 1992, Liquidation values and debt capacity: A market equilibrium approach, *The journal of finance* 47, 1343–1366.

Sklarz, Michael, and Norman Miller, 2016, Adjusting loan to value (ltv) ratios to reflect value uncertainty, Technical report, Collateral Analytics Research.

Tauchen, George, 1986, Finite state markov-chain approximations to univariate and vector autoregressions, *Economics letters* 20, 177–181.

Titman, Sheridan, and Roberto Wessels, 1988, The determinants of capital structure choice, *The Journal of finance* 43, 1–19.

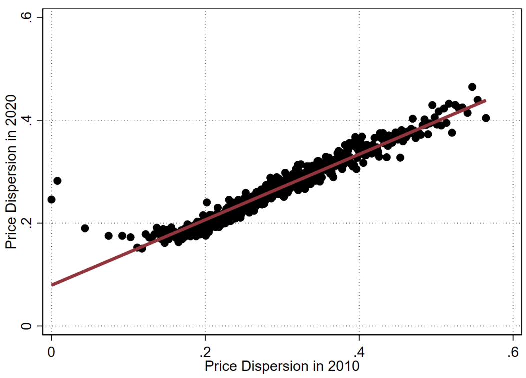
Wong, Arlene, 2019, Refinancing and the transmission of monetary policy to consumption, *Unpublished manuscript* 20.

Wood, Simon N, 2017, *Generalized additive models: an introduction with R* (CRC press).

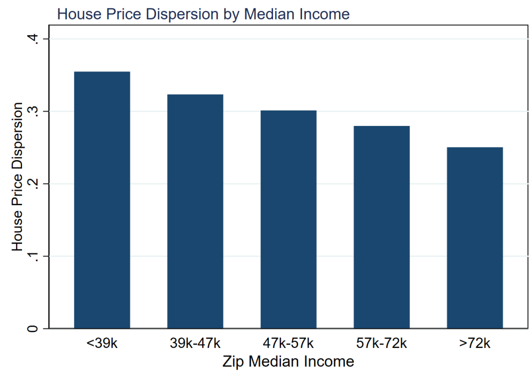
Zhou, Hanqing, Yuan Yuan, Christopher Lako, Michael Sklarz, and Charles McKinney, 2015, Foreclosure discount: definition and dynamic patterns, *Real Estate Economics* 43, 683–718.

Figures

Figure 1. Stylized Facts about Price Dispersion Estimates



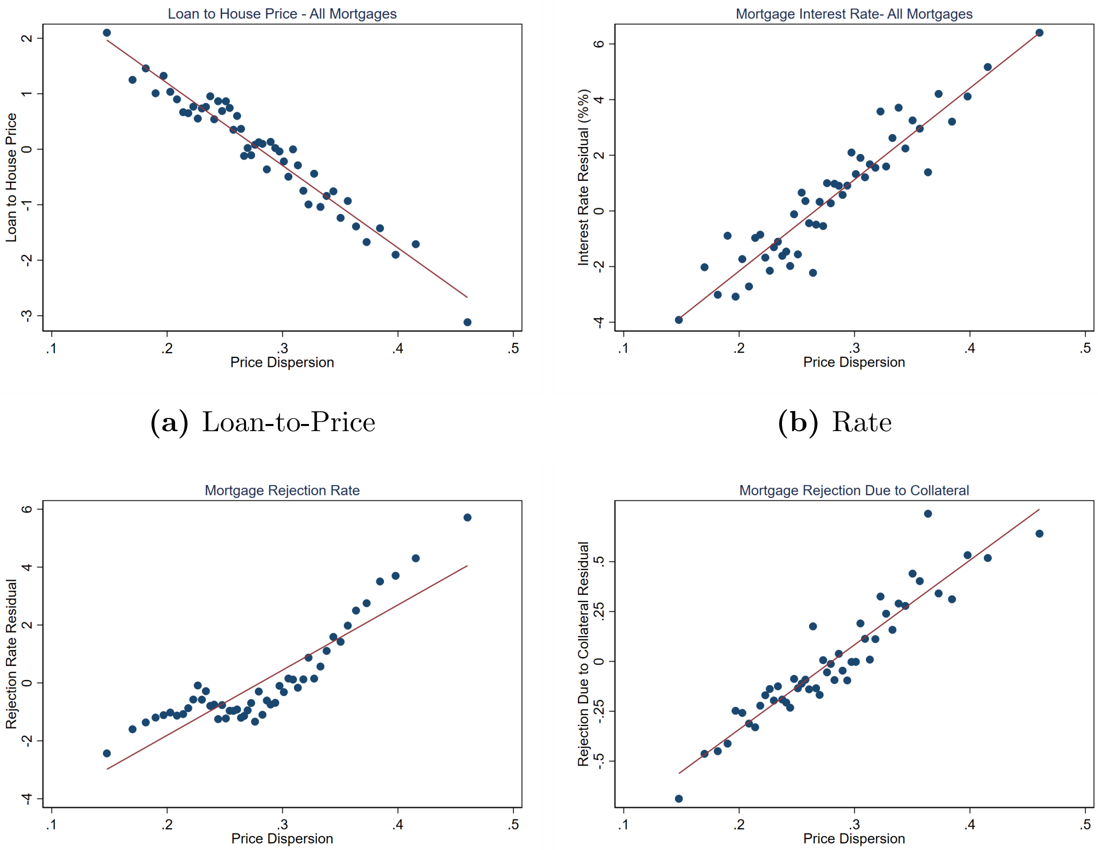
(a) Zip-Code Dispersion: 2010 vs 2020



(b) Dispersion and Income

Note: Panel (a) plots zip-code dispersion measures in 2020 against zip-code dispersion measures in 2010. Panel (b) shows the association between house price dispersion and zip-code household income prices. We divide all zip codes into five buckets based on local median household income and plot the average values in each bucket. The sample includes annual zip level observations from 2000 to 2020.*Source*: Corelogic Deeds and American Community Survey 2008-2012.

Figure 2. County Level House Price Dispersion and Credit Access

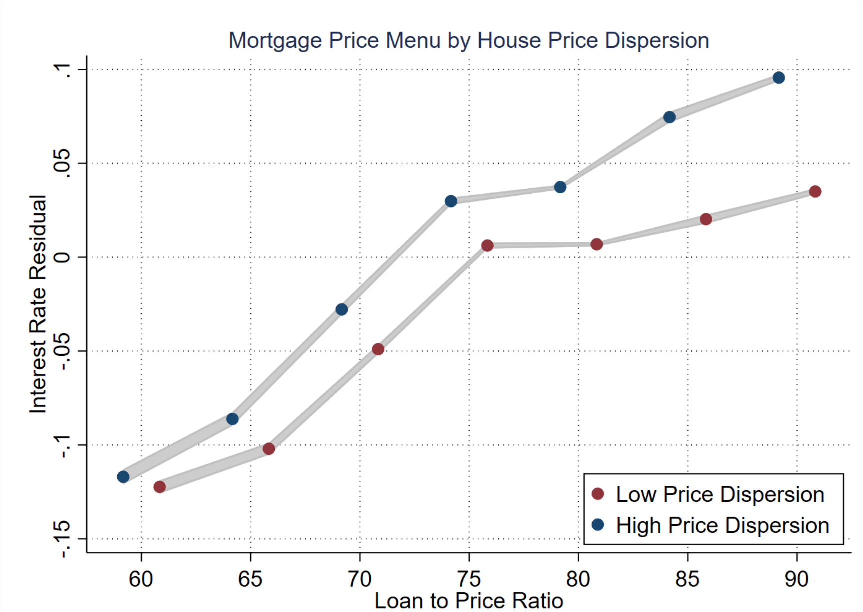


(a) Loan-to-Price **(b)** Rate

(c) Total Rejection **(d)** Rejection due to Collateral

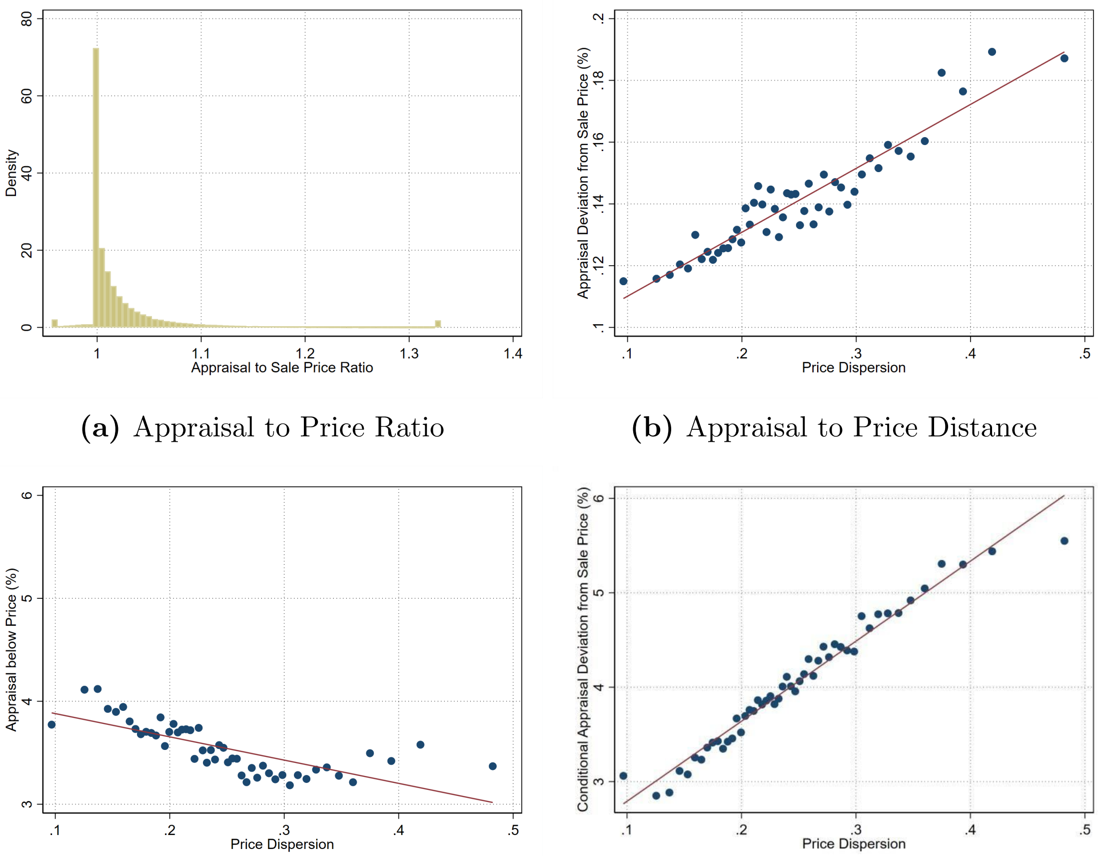
Note: This figure shows the correlation between county level house price dispersion and various credit access outcomes. Panel a plots county average LTP *after taking out the effect of underlying house prices*. The y-axis values are in percentage points. Panel b plots county average residualized mortgage interest rate (basis points). Individual mortgage interest rates are residualized using borrower and loan characteristics, such as FICO, LTP, DTI, the squared terms, and their interactions with origination year. We then take the county-average of residualized mortgage rates. Panels c and d plot mortgage rejection rate (percentage points). Panel c plots the total rejection rate. Panel d plots the rejection rate due to collateral. We residualize mortgage rejection rate by taking the residuals of regressions of mortgage rate on county average log house price, credit score, and year fixed effects. The sample includes annual county observations from 2000 to 2020 for panels (a) and (b) and from 2000 to 2017 for panels (c) and (d). *Source*: County house price dispersion is estimated using Corelogic Deeds records. Mortgage data are from Corelogic LLMA and HMDA.

Figure 3. Property Level Mortgage Menu by House Price Dispersion



Note: This figure shows mortgage price menu (rate-LTP pair) by zip-level house price dispersion. The y values are interest rate residuals from a regression of mortgage rates on borrower fico, fico-squared, DTI, DTI-squared conforming or jumbo indicator, and origination month fixed effects. The dots represent the average mortgage rate in each LTP bucket. The shaded area indicates 95% confidence interval. The sample includes loan level observations of conventional loans from 2000 to 2020. *Source*: Corelogic LLMA and Deeds.

Figure 4. Price Dispersion and Appraisals



(a) Appraisal to Price Ratio

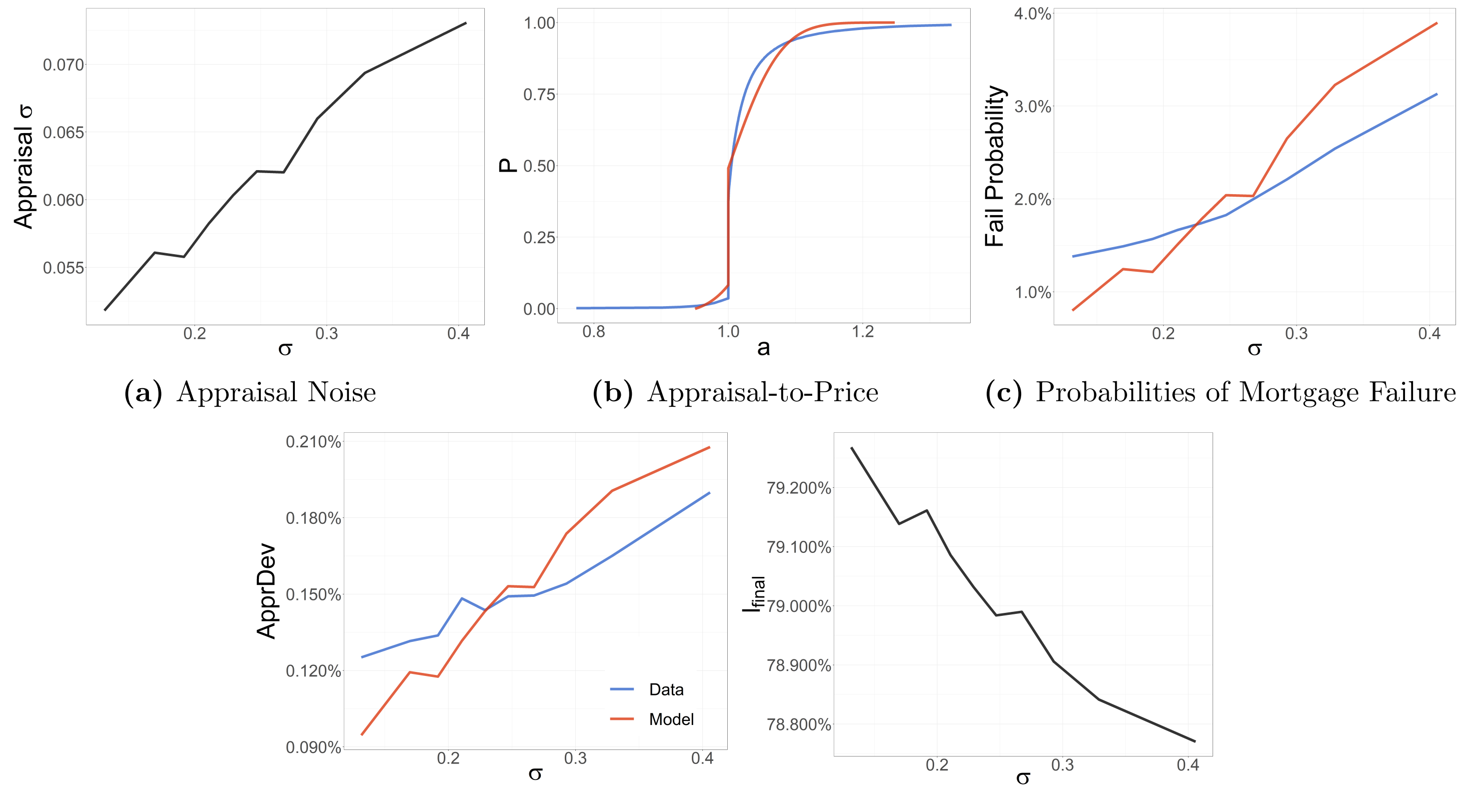
(b) Appraisal to Price Distance

(c) Appraisal below Sale

(d) Conditional Appraisal to Price Distance

Note: Panel (a) of this figure shows the histogram of appraisal-to-transaction price ratios, winsorized at 1% to remove outliers. Panel (b) shows a binned scatter plot, where the y-variable is $ApprDev_i$, the product of the percentage deviation of appraisal prices to sale prices with a dummy for a house under-appraising, defined in (A47). In panel (c), the y-variable is the probability that appraisals are below transaction prices. In panel (d), the y-variable is the average under-appraisal percentage conditional on under-appraisal, defined in (A48). In all panels, the x-variable is zip code price dispersion. We divide all loans into 50 buckets based on zip code house price dispersion. The sample includes loan level observations from 2000 to 2020. *Source*: Corelogic LLMA, Deed and Tax datasets.

Figure 5. Model Fit

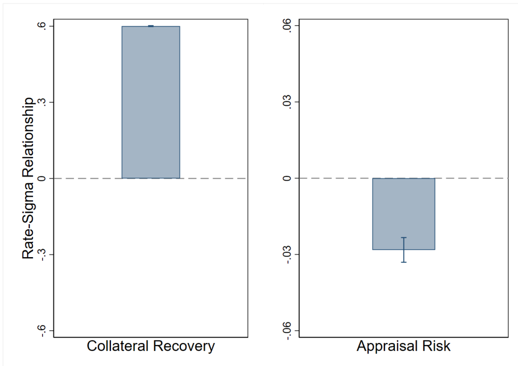


(a) Appraisal Noise **(b)** Appraisal-to-Price **(c)** Probabilities of Mortgage Failure

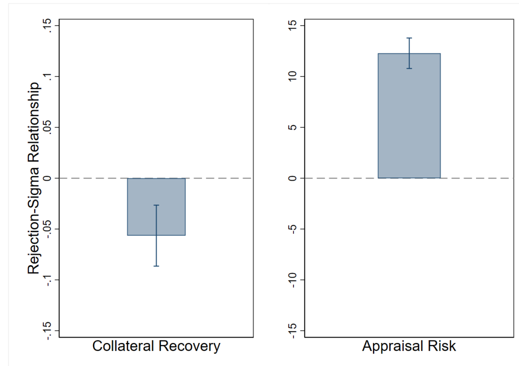
(d) Appraisal Deviation **(e)** Model Predicted LTP

Note: Panel (a) shows shows estimated appraisal standard deviations σ_a on the y-axis, and estimated idiosyncratic price dispersion σ on the x-axis. Panel (b) shows the distribution of appraisal-over-price ratios a, in the data and the fitted model, for the 5th percentile bucket (that is, counties with values of σ between the 40th and 50th percentiles). Panel (c) shows transaction failure probabilities in the data and in the fitted model. Panel (d) shows $ApprDev_i$, which is defined as $p_i^{under} E\left[\frac{a}{p} - 1 \mid under\right]$, in the data and in the fitted model. Panel (e) shows l_{final}, which is the model predicted loan-to-price ratio.

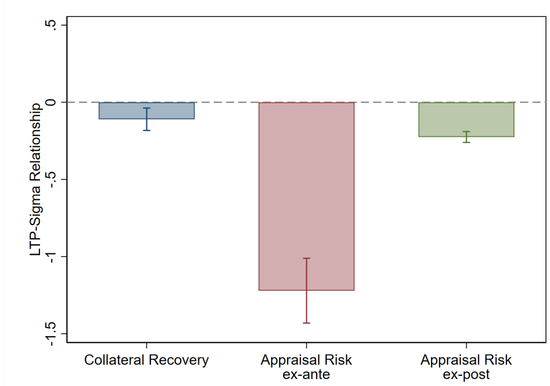
Figure 6. Decomposition of Channels



(a) Interest Rate



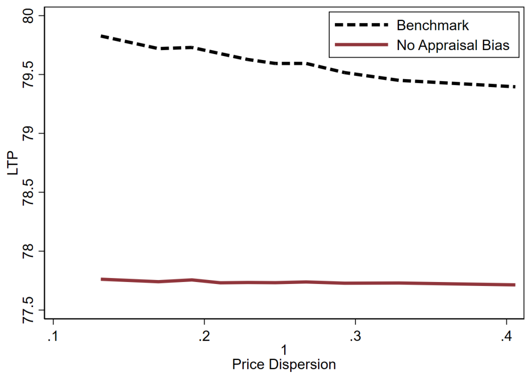
(b) Mortgage Failure



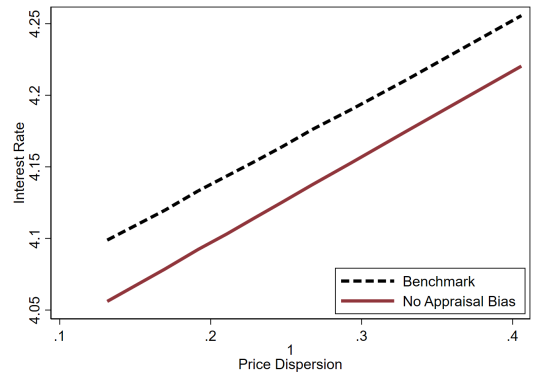
(c) Loan-to-Price

Note: Panel (a) decomposes the effect of sigma on interest rate. Panel (b) decomposes the effect of sigma on mortgage rejection. Panel (c) decomposes the effect of sigma on loan-to-price ratio. In all panels, the y-axis is the effect of sigma on the outcome variables. The rate menu channel assumes that appraisal noise does not change with sigma. The appraisal risk effect assumes that the rate menu does not change with sigma. In Panel (c), we further decompose the appraisal risk channel into ex-ante and ex-post, where ex-ante is based on the effect of sigma on targeted loan-to-price ratio, and ex-post is based on the effect of sigma on the difference between targeted and realized loan-to-price ratio.

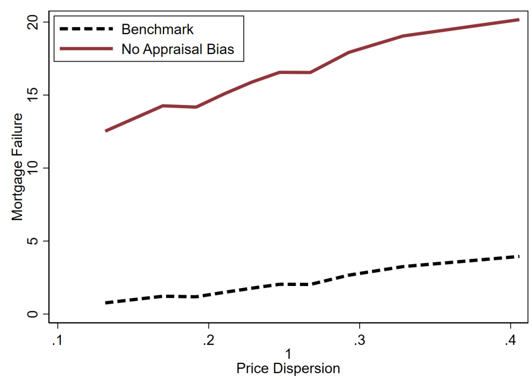
Figure 7. Counterfactual



(a) Loan-to-Price



(b) Interest Rate



(c) Mortgage Failure Rate

Note: This figure plots the counterfactual mortgage outcomes when there is no appraisal bias, that is, setting the parameter $b = 0$. Panel (a) compares the benchmark and the counterfactual loan-to-price ratio. the effect decomposes the effect of sigma on interest rate. Panel (b) compares the benchmark and the counterfactual interest rate. Panel (c) compares the benchmark and the counterfactual mortgage failures. In all panels, the y-axis is the outcome variable, and the x-axis is the house price dispersion.

Table

Table 1: Summary Statistics

This table reports summary statistics for the three main datasets: the property sample from the Corelogic Deed and Tax datasets, the loan sample from the Corelogic LLMA dataset, and the mortgage application sample from the HMDA. The Corelogic samples span the time period 2000 to 2020. The HMDA sample spans 2000 to 2017.

	N	Mean	Stdev	P25	Median	P75
Property Level Sample						
Loan to Price	29M	85.42	15.65	80.00	89.68	98.19
Price Dispersion	29M	0.24	0.11	0.17	0.23	0.30
Sale Price (Thousand)	29M	273.02	224.93	140.30	215.00	332.50
Mortgage Amount (Thousand)	29M	222.40	163.53	121.80	182.16	275.79
Building Age	29M	27.12	25.95	6.00	20.00	42.00
Square Footage	29M	1,961.57	2,982.11	1,363.00	1,774.00	2,365.00
Loan Level Sample						
Loan to Price	4.8M	85.48	14.98	80.00	90.00	98.19
Zip Price Dispersion	4.8M	0.25	0.08	0.19	0.24	0.29
Sale Price (Thousand)	4.8M	280.83	242.84	143.50	218.00	340.00
Appraised to Price Ratio	4.8M	1.03	0.19	1.00	1.00	1.02
Mortgage Amount (Thousand)	4.8M	227.66	170.25	124.00	185.18	283.00
FICO	4.8M	725.35	61.39	681.00	735.00	778.00
Debt-to-Income	4,8M	37.23	11.28	29.85	38.00	44.69
Mortgage Application Sample						
Rejection Rate	49M	15.86	36.53	0.00	0.00	0.00
Rejection due to Collateral Reasons	49M	1.95	13.83	0.00	0.00	0.00
Zip Price Dispersion	49M	0.26	0.08	0.20	0.25	0.31
Applicant Income (Thousand)	49M	102.35	193.47	47.00	72.00	114.00
Loan-to-Income	49M	242.18	6,896.19	135.83	227.78	316.51
County Credit Score	49M	667.19	22.16	650.30	666.18	684.14

Table 2: Determinants of House Price Dispersion

This table presents the association between house price dispersion and house features (Panel A) and zip code market condition (Panel B). All continuous variables are scaled by standard deviation. Recent renovation is defined as renovation in the last 5 years from the transaction year. The sample includes house transactions from 2000 to 2020. Standard errors are clustered at county level. ***, **, * represent 1%, 5%, and 10% significance, respectively.

Panel A: House Features

	Estimated Price Dispersion				
	(1)	(2)	(3)	(4)	(5)
Building Age	0.04***	0.04***	0.04***		0.04***
	(0.002)	(0.002)	(0.002)		(0.002)
Recent Renovation		-0.01***			-0.01***
		(0.003)			(0.003)
Benchmark: Square-Footage < 1281					
[1282,1601]			-0.03***	-0.03***	-0.02***
			(0.002)	(0.002)	(0.002)
[1602,1970]			-0.03***	-0.03***	-0.02***
			(0.003)	(0.003)	(0.003)
[1971,2544]			-0.02***	-0.02***	-0.00
			(0.004)	(0.004)	(0.003)
> 2544			0.01**	0.01**	0.03***
			(0.005)	(0.004)	(0.004)
Benchmark: Bedrooms < 4					
=4				-0.01***	-0.01***
				(0.001)	(0.001)
>4				0.01***	0.01***
				(0.002)	(0.001)
Log House Price	-0.48***	-0.48***	-0.51***	-0.51***	-0.37***
	(0.028)	(0.028)	(0.025)	(0.025)	(0.021)
Log House Price Squared	0.50***	0.50***	0.51***	0.51***	0.38***
	(0.029)	(0.029)	(0.025)	(0.025)	(0.022)
County-Year FE	✓	✓	✓	✓	✓
R2	0.33	0.33	0.26	0.27	0.35
Observations	29M	29M	29M	29M	29M

Panel B: Zip Code Market Condition

	Zip Code Price Dispersion			
	(1)	(2)	(3)	(4)
Gini Index	0.01***			0.01***
	(0.001)			(0.001)
Population Density		-0.01***		-0.01***
		(0.003)		(0.002)
Vacancy Share			0.03***	0.03***
			(0.002)	(0.002)
Year FE	✓	✓	✓	✓
R2	0.02	0.02	0.08	0.09
Observations	276,079	276,079	276,079	276,079

Table 3: Property-Level House Price Dispersion and LTP

This table presents property-level regression results. Columns 1-3 present OLS results. Columns 4-6 present IV results. In all columns, the outcome variable is the loan level loan-to-sale price ratio. The explanatory variable of interest in columns 1-3 is property-level house price dispersion, scaled by its standard deviation, and is the predicted price dispersion in columns 4-6. Controls include the transaction price of the property, mortgage type, mortgage term, and resale indicator. The sample includes property transaction level observations from 2000 to 2020. Standard errors are clustered at county level. ***, **, * represent 1%, 5%, and 10% significance, respectively.

	OLS			2SLS		
	(1)	(2)	(3)	(4)	(5)	(6)
Price Dispersion	-0.43***	-0.21***	-0.23***	-0.33***	-0.40***	-0.45***
	(0.042)	(0.036)	(0.033)	(0.087)	(0.067)	(0.063)
Log House Price	-4.30***	-4.59***	-4.31***	-4.29***	-4.59***	-4.30***
	(0.144)	(0.184)	(0.167)	(0.144)	(0.184)	(0.166)
Loan Controls	✓	✓	✓	✓	✓	✓
Transaction Date FE	✓	✓	✓	✓	✓	✓
County-Year FE		✓	✓		✓	✓
Lender-Year FE			✓			✓
R2	0.34	0.36	0.40	0.32	0.29	0.26
Observations	28M	28M	28M	28M	28M	28M
Underidentification test statistic				166.75	160.45	164.44
Underidentification test p-value				0.00	0.00	0.00
Weak identification test statistic				242.71	226.63	224.43

Table 4: Price Dispersion and Cost Menu

This table presents loan level regression results the of cost menu. The outcome variables are loan-level interest rate (bps). Columns 1-3 present OLS results, and columns 4-6 present 2SLS results. Column 1 (4) uses the full sample. Columns 2-3 (5-6) use securitized conventional loans (i.e., non-FHA loans that are securitized) and portfolio conventional loans (i.e., non-FHA loans that are held on lenders' balance sheets), respectively. The explanatory variable of interest is zip-code house price dispersion, scaled by its standard deviation. Borrower and loan controls include log house price, FICO score, FICO squared, LTP, LTP squared, DTI, DTI-squared, and loan type. The sample includes loan level observations from 2000 to 2020. Standard errors are clustered at county level. ***, **, * represent 1%, 5%, and 10% significance, respectively.

	OLS			2SLS		
	(1) Full	(2) Securitized	(3) Portfolio	(4) Full	(5) Securitized	(6) Portfolio
Zip Price Dispersion	1.10***	1.38***	1.98***	2.20***	2.20***	5.32***
	(0.139)	(0.111)	(0.464)	(0.356)	(0.261)	(1.052)
LTP	0.80***	0.59***	2.32***	0.80***	0.58***	2.31***
	(0.113)	(0.039)	(0.247)	(0.113)	(0.039)	(0.244)
Borrower and Loan Controls	✓	✓	✓	✓	✓	✓
Origination Month FE	✓	✓	✓	✓	✓	✓
County-Year FE	✓	✓	✓	✓	✓	✓
R2	0.85	0.87	0.81	0.08	0.12	0.05
Observations	4.8M	2.3M	1.1M	4.8M	2.3M	1.1M
Underidentification t-stat				85.28	88.95	70.61
Underidentification p-value				0.00	0.00	0.00
Weak identification t-stat				43.06	39.71	33.57

Table 5: Mortgage Rejections and Zip House Price Dispersion

This table presents loan level regression results of mortgage rejections. The outcome variable in Panel A is an indicator that equals 100 if a loan is rejected and 0 otherwise. The outcome variable in Panel B is an indicator that equals 100 if a loan is rejected due to collateral reasons and 0 otherwise. In both panels, columns 1-3 report OLS results, and columns 4-6 report 2SLS results. The explanatory variable of interest is zip code house price dispersion, scaled by its standard deviation. Borrower/Loan controls include zip code house price, log income, loan type, county average credit score and its square term, and loan to income ratio and its square term. The sample includes loan level observations from 2001 to 2017. Standard errors are clustered at county level. ***, **, * represent 1%, 5%, and 10% significance, respectively.

	OLS			2SLS		
	(1) Full	(2) Securitized	(3) Portfolio	(4) Full	(5) Securitized	(6) Portfolio
Panel A: Rejection						
Zip Price Dispersion	1.38***	1.37***	0.82***	2.47***	2.50***	2.60***
	(0.093)	(0.102)	(0.114)	(0.189)	(0.196)	(0.683)
Rejection Mean	15.9%	16.5%	16.0%	15.9%	16.5%	16.0%
R2	0.16	0.18	0.17	0.01	0.01	<0.005
Panel B: Rejection Due to Collateral						
Zip Price Dispersion	0.50***	0.53***	0.38***	0.78***	0.85***	0.81***
	(0.035)	(0.037)	(0.051)	(0.060)	(0.065)	(0.144)
Local Controls	✓	✓	✓	✓	✓	✓
County-Year FE	✓	✓	✓	✓	✓	✓
Lender-Year FE	✓	✓	✓	✓	✓	✓
Rejection due to Collateral Mean	1.9%	2.0%	2.2%	1.9%	2.0%	2.2%
R2	0.05	0.05	0.09	<0.005	<0.005	<0.005
Observations	49M	35M	3.7M	49M	35M	3.7M
Underidentification t-stat				87.07	81.51	25.92
Underidentification p-value				0.00	0.00	0.00
Weak identification t-stat				64.34	61.93	18.44

Table 6: Mortgage Rejection Reasons

This table presents loan level regression results of mortgage rejection reasons. The explanatory variable of interest is zip code house price dispersion, scaled by its standard deviation. Borrower/Loan controls include zip code house price, log income, loan type, county average credit score and its square term, and loan to income ratio and its square term. The sample includes loan level observations from 2001 to 2017. Standard errors are clustered at county level. ***, **, * represent 1%, 5%, and 10% significance, respectively.

	(1) Collateral	(2) Down Payment	(3) Debt-to-Income	(4) Employment	(5) Credit Score
Panel A: OLS					
Price Dispersion	1.70***	-0.10***	-0.48***	-0.13***	-0.65***
	(0.139)	(0.020)	(0.068)	(0.016)	(0.058)
R2	0.16	0.10	0.18	0.05	0.25
Panel B: IV					
Price Dispersion	2.65***	-0.10**	-0.66***	-0.14***	-1.28***
	(0.163)	(0.039)	(0.107)	(0.027)	(0.106)
Underidentification t-stat	67.86	67.86	67.86	67.86	67.86
Underidentification p-value	0.00	0.00	0.00	0.00	0.00
Weak identification t-stat	71.51	71.51	71.51	71.51	71.51
Sample mean	12.23	4.90	17.08	2.76	21.12
Observations	8M	8M	8M	8M	8M
Local Controls	✓	✓	✓	✓	✓
County-Year FE	✓	✓	✓	✓	✓
Lender-Year FE	✓	✓	✓	✓	✓

Table 7: Ex-Post Performance

This table analyzes ex-post performance of mortgage loans. Columns 1 and 4 use full sample. Columns 2 and 4 use securitized conventional loans (i.e., non-FHA loans that are securitized). Columns 3 and 6 use portfolio conventional loans (i.e., non-FHA loans that are held on lenders' balance sheets). Outcome variable is 100 if the loan defaults in two years since origination and 0 otherwise. The explanatory variable of interest is zip-code house price dispersion, scaled by its standard deviation. Other controls include house price and loan type. The sample includes all loans originated from 2000 to 2018. Since we need at least two-year performance to define *default*, we remove loans originated after 2018 from the full sample for this analysis. Standard errors are clustered at county level. ***, **, * represent 1%, 5%, and 10% significance, respectively.

	OLS			2SLS		
	(1) Full	(2) Securitized	(3) Portfolio	(4) Full	(5) Securitized	(6) Portfolio
Zip Price Dispersion	0.16***	0.10***	0.17***	-0.10	-0.10	-0.01
	(0.056)	(0.038)	(0.064)	(0.096)	(0.099)	(0.118)
Interest Rate	2.25***	2.61***	1.55***	2.25***	2.61***	1.55***
	(0.123)	(0.185)	(0.127)	(0.125)	(0.188)	(0.128)
FICO	-20.89***	-19.68***	-18.23***	-20.89***	-19.68***	-18.23***
	(0.238)	(0.345)	(0.532)	(0.237)	(0.344)	(0.534)
DTI	0.06***	0.03***	0.05***	0.06***	0.02***	0.05***
	(0.005)	(0.006)	(0.005)	(0.005)	(0.006)	(0.005)
Origination Month FE	✓	✓	✓	✓	✓	✓
County-Year FE	✓	✓	✓	✓	✓	✓
Property & Loan Controls	✓	✓	✓	✓	✓	✓
R2	0.15	0.13	0.19	0.09	0.06	0.09
Observations	4.3M	2.1M	0.9M	4.3M	2.1M	0.9M
Underidentification test statistic				82.86	86.90	68.21
Underidentification test p-value				0.00	0.00	0.00
Weak identification test statistic				41.98	38.03	34.89

Table 8: Robustness Tests

This table presents robustness tests. Panel A is for lender market power. We use a subsample of loans from Corelogic Deeds that we observe mortgage interest rate to estimate the effect of property-level price dispersion on LTP for any given interest rate. Panel B presents robustness test for bunching below conforming limit. We use the sample to house transactions with non-missing mortgage interest rates from Corelogic Deeds and further restrict the sample to houses whose transaction price is smaller than the local conforming loan limit. Standard errors are clustered at county level.

Panel A: Not about Lender Market Power

	OLS			2SLS		
	(1)	(2)	(3)	(4)	(5)	(6)
Price Dispersion	-0.34***	-0.31***	-0.25***	-1.01***	-0.97***	-1.05***
	(0.031)	(0.029)	(0.028)	(0.070)	(0.062)	(0.094)
Interest Rate	1.01***	0.87***	0.92***	1.02***	0.88***	0.92***
	(0.067)	(0.044)	(0.055)	(0.068)	(0.044)	(0.055)
Loan Controls	✓	✓	✓	✓	✓	✓
Origination Month FE	✓	✓	✓	✓	✓	✓
County-Year FE	✓			✓		
Lender-County-Year FE		✓			✓	
Lender-Zip-Year FE			✓			✓
R2	0.47	0.59	0.67	0.27	0.21	0.18
Observations	5M	5M	4M	5M	5M	4M
Underidentification test statistic				119.15	104.27	131.69
Underidentification test p-value				0.00	0.00	0.00
Weak identification test statistic				182.03	171.90	169.75

Panel B: Not about Bunching ($\frac{SalePrice}{ConformingLimit} < 1$)

	OLS		2SLS	
	(1)	(2)	(3)	(4)
Price Dispersion	-0.21***	-0.18***	-0.52***	-0.43***
	(0.025)	(0.022)	(0.046)	(0.039)
Interest Rate	1.01***	0.84***	1.02***	0.84***
	(0.080)	(0.059)	(0.081)	(0.059)
Loan Controls	✓	✓	✓	✓
Origination Month FE	✓	✓	✓	✓
County-Year FE	✓		✓	
Lender-Year FE		✓		✓
R2	0.43	0.50	0.27	0.23
Observations	4M	4M	4M	4M
Underidentification test statistic			123.24	121.76
Underidentification test p-value			0.00	0.00
Weak identification test statistic			181.03	171.42

Table 9: Model Estimates

This table presents the model estimates. Panel A reports the externally calibrated parameters. Panel B reports the estimated parameters.

Panel A: Externally Calibrated Parameters

Description	Parameter	Value
Intertemporal elasticity of substitution	η	2
Wealth at time of home purchase	W_1	\$60,000
House price	P	\$200,000
Discount factor	β	0.96
	T	7
Maximum LTV parameter	ϕ	0.8

Panel B: Parameters Calibrated to the Data or through Moment Matching

Description	Parameter	Value
Appraisal Standard Deviation	$\sigma_1, ..., \sigma_{10}$	See Figure A10
Search cost	ζ	0.175
Appraisal Bias	b	0.0792
Penalty rate on consumption	ψ	1.614
Marginal utility of next period consumption	u_2'	0.0021

Online Appendix for "Collateral Value Uncertainty and Mortgage Credit Provision"

Table of Contents

A Data Cleaning and Measure Estimation

A.1 Data Cleaning

Corelogic tax & deed data. We clean the datasets using a number of steps. First, we use only arms-length new construction sales or resales of single-family residences, which are not foreclosures, which have non-missing sale price, date, APN, and county FIPS code in the Corelogic deed data, and which have non-missing year built and square footage in the Corelogic tax data. We use only data from 2000 onwards, as we find that Corelogic's data quality is low prior to this date. Even after throwing out pre-2000 data, we find that some counties have very low total sales for early years, suggesting that some data is missing. To address this, we manually filter out some early county-years for which the total number of sales is low.

We also filter out "house flips", as well as instances where reported sale price seems anomalous. If a house is ever sold twice within a year, we drop all observations of the house. Most of these kinds of transactions appear to be either flips, which are known to be a peculiar segment of the real estate market (Bayer et al., 2011; Giacoletti and Westrupp, 2017), or duplication bugs in the data, where a single transaction is recorded twice or more. To filter for potentially anomalous prices, if we ever observe a property whose annualized appreciation or depreciation is above 50% for any given pair of sales, we drop all observations of the property. Finally, if a house is ever sold at a price which is more than 5 times higher or lower than the median house price in the same county-year, we drop all observations of the house from our dataset.

Our model of prices involves a fairly large number of parameters, so we filter to counties with a fairly large number of house sales in order to precisely estimate the model. Thus, we filter to counties with at least 1,000 house sales remaining, and with at least 10 sales per month on average, after applying the filtering steps described above.

Corelogic LLMA data. We filter to only purchase loans, excluding refinancing loans. As in the Corelogic Deed data, we calculate the loan-to-price ratio as the mortgage loan amount, divided by the house transaction price. We dropped observations with empty property zipcode, FICO score, initial interest rate, mortgage amount, origination date, sale price, and back-end ratio. We divide the market into conforming and non-conforming loans, using a flag provided by corelogic. We dropped all observations with balloon loans, and with loan to price ratio > 100. We kept observations with full documentation and fixed interest rates. We dropped observations with outliers. Specifically, we dropped all observations lower than 1 percentile and higher than 99 percentile with respect to loan to price and initial interest rate.

HMDA data. We filter to approved purchase or refinancing loans, omit FHA loans, filtering to one-to-four family homes, and filtering to loan amounts greater than 0. We drop observations with missing state or county codes, and with LTV higher than 130, and we Winsorize loan amounts, rate spreads, and LTVs.

A.2 Measurement: f_c and g_c Functions

In order to estimate price dispersion, we need to model prices as a flexible function of characteristics. We do this using generalized additive models, which are a class of flexible nonparametric models; Wood (2017) describes the theory of GAMs. We use the mgcv package in R to implement the GAMs. We use this class of functions because, in our simulations, they provide a better fit to house prices than standard high-order polynomials.

We implement a two-stage regression using general additive model (GAM) on a county level. Instead of a high order polynomial, GAM implements cubic spline basis (or tensor product for multivariates) to fit the regressors. Therefore, to avoid overfitting, we first throw out counties with less than 400 observations. In order to estimate the GAM, there needs to be sufficient variation in characteristics; thus, we only keep counties with at least 10 unique values of each of the following characteristics: geographic information (latitude and longitude), year built, square footage, and transaction date. We also normalize the months, latitude, and longitude, building square feet, and year built. Furthermore, we winsorize geographic information, year built and building square feet.

We then estimate the following generalized additive model:

$$\begin{aligned} f_c(x_i,t) = {} & h_c^{f,latlong}(t, lat_i, long_i) + h_c^{f,sqft}(t, sqft_i) + \\ & h_c^{f,yrbuilt}(t, yrbuilt_i) + h_c^{f,bedrooms}(t, bedrooms_i) + h_c^{f,bathrooms}(t, bathrooms_i) \end{aligned}$$

The functions $h_c^{f,latlong}$, $h_c^{f,sqft}$, and $h_c^{f,yrbuilt}$ are tensor products of 5-dimensional cubic splines in their constituent components: hence, for example, the $h_c^{f,latlong}(t, lat_i, long_i)$ is a three-dimensional spline tensor product, with a total of $5^3 = 125$ degrees of freedom. To combat overfitting, the spline terms also includes a shrinkage penalty term on the second derivative of the spline functions, with the smoothing penalty determined through generalized cross-validation. The functions $h_c^{f,bedrooms}$ and $h_c^{f,bathrooms}$ interact dummies for a given house having 1, 2, 3 or more bedrooms and 1, 2, 3 or more bathrooms respectively with cubic spline basis in time.

The functional form for $g_c(x_i, t)$ in (2) is exactly analogous to $f_c(x_i, t)$:

$$g_c(x_i, t) = h_c^{g,latlong}(t, lat_i, long_i) + h_c^{g,sqft}(t, sqft_i) + h_c^{g,yrbuilt}(t, yrbuilt_i) + \\ h_c^{g,bedrooms}(t, bedrooms_i) + h_c^{g,bathrooms}(t, bathrooms_i)$$

B Proofs and Supplementary Material for Section 4

B.1 Proof of Theorem 1

Conditional on the appraisal value a, the buyer can choose to proceed with the loan and purchase the property (continue), or renege on the offer and search for a new house and loan (renege). Let the value of each option, with loan size l and appraisal a, be respectively $V(a, l, continue)$ and $V(a, l, renege)$. The maximized value at any a and l is:

$$V(a, l) \equiv \max\left[V(a, l, continue), V(a, l, renege)\right] \tag{A1}$$

We proceed to characterize $V(a, l, continue)$ and $V(a, l, renege)$.

B.1.1 Characterizing $V(a, l, continue)$

If the buyer proceeds with appraisal a, her utility is:

$$V(a, l, continue) = \frac{c_1^{1-\eta} - 1}{1-\eta} + \beta^T u_2' c_2 \tag{A2}$$

Where, from (12) and (13) in the main text, we have:

$$c_1 = \underbrace{W_1 - P(1-l)}_{Targeted\ consumption} - \underbrace{\psi P \max(0, l - \phi a)}_{Appraisal\ shortfall} \tag{A3}$$

$$c_2 = -\left(1 + r(l)\right)^T P\left(l - \max\left[0, l - \phi a\right]\right) \tag{A4}$$

where, as we discussed in the main text, we have set $W_2 = 0$, since second-period wealth only linearly shifts buyers' utility and does not interact with any of the buyer's decisions. In words, (A3) states that the buyer's consumption in period 1 is equal to her targeted consumption $W_1 - P(1-l)$, minus an "appraisal shortfall" term $\max(0, l - \phi a)$. If $a < \frac{l}{\phi}$, then the buyer must decrease her borrowing from l to ϕa; this decreases her period-1 consumption by $l - \phi a$, multiplied by the price, and the penalty term $\psi > 1$. Since the final loan size l_{final} is smaller, this also decreases the amount that the buyer must pay back in period 2 by

$\left(1+r\left(l\right)\right)^{T}P\max\left[0,l-\phi a\right]$. Substituting (A3) and (A4) into (A2), we have:

$$\begin{aligned}V\left(a,l,continue\right)=&u_{1}\left(W_{1}-P\left(1-l\right)-\psi P\max\left(0,l-\phi a\right)\right)\\&-u_{2}'\beta^{T}\left(1+r\left(l\right)\right)^{T}Pl\\&+u_{2}'\beta^{T}\left(1+r\left(l\right)\right)^{T}P\max\left[0,l-\phi a\right]\end{aligned}\tag{A5}$$

where, $u_{1}\left(c\right)\equiv\frac{c^{1-\eta}-1}{1-\eta}$. Recall that, in (18), we defined:

$$\omega\left(a,l\right)\equiv u_{1}\left(W_{1}-P\left(1-l\right)-\psi P\max\left(0,l-\phi a\right)\right)+u_{2}'\beta^{T}\left(1+r\left(l\right)\right)^{T}P\max\left[0,l-\phi a\right]$$

Using this definition, we have:

$$V\left(a,l,continue\right)=-u_{2}'\beta^{T}\left(1+r\left(l\right)\right)^{T}Pl+\omega\left(a,l\right)\tag{A6}$$

B.1.2 Characterizing $V\left(a,l,renege\right)$

If the buyer reneges, she receives:

$$V\left(a,l,renege\right)=-\beta^{T}u_{2}'\zeta P+E_{a}\left(V\left(a,l\right)\right)\tag{A7}$$

In words, she pays a cost ζP in period 2 consumption, which costs $-\beta^{T}u_{2}'\zeta P$ in utility terms. She then returns to the beginning of the game, and thus receives the expectation of $V\left(a,l\right)$ over uncertainty in a. Expanding $E_{a}\left(V\left(a,l\right)\right)$, we have:

$$E_{a}\left(V\left(a,l\right)\right)=\int_{0}^{\infty}\max\left(V\left(a,l,continue\right),V\left(a,l,renege\right)\right)dF_{a}\left(a\right)\tag{A8}$$

Now, note that $V\left(a,l,renege\right)$, is independent of a, whereas from (A5), $V\left(a,l,continue\right)$ is increasing in a. Thus, there is some cutoff value $\bar{a}\left(l\right)$, such that continuing is optimal for all $a>\bar{a}\left(l\right)$. At the boundary $\bar{a}\left(l\right)$, continuing and reneging have equal value:

$$V\left(\bar{a},l,renege\right)=V\left(\bar{a},l,continue\right)\tag{A9}$$

Substituting for $V(\bar{a}, l, continue)$ using (A6), we have:

$$V(\bar{a}, l, renege) = -\beta^T \left(1 + r(l)\right)^T u_2' Pl + \omega(\bar{a}, l)$$

Substituting into (A8), we have:

$$\begin{aligned} &E_a\left(V(a,l)\right) = \\ &\quad \int_0^\infty \max\left(-\beta^T \left(1+r(l)\right)^T u_2' Pl + \omega(a,l), -\beta^T \left(1+r(l)\right)^T u_2' Pl + \omega(\bar{a},l)\right) dF_a(a) \end{aligned}$$

$$E_a\left(V(a,l)\right) = -\beta^T \left(1+r(l)\right)^T u_2' Pl + \int_0^\infty \max\left(\omega(a,l), \omega(\bar{a},l)\right) dF_a(a) \tag{A10}$$

Substituting into (A7), we have:

$$\begin{aligned} &V(a,l,renege) = \\ &\quad -\beta^T u_2' \zeta P - \beta^T \left(1+r(l)\right)^T u_2' Pl + \int_0^\infty \max\left(\omega(a,l), \omega(\bar{a},l)\right) dF_a(a) \end{aligned} \tag{A11}$$

B.1.3 Solving For $\bar{a}$

Having characterized $V(a, l, renege)$ and $V(a, l, continue)$, we now solve for $\bar{a}$. Plugging in expressions for $V(\bar{a}, l, renege)$ and $V(\bar{a}, l, continue)$ into (A9), we have:

$$\begin{aligned} &-\beta^T \left(1+r(l)\right)^T u_2' Pl + \omega(\bar{a},l) = \\ &\quad -\beta^T u_2' \zeta P - \beta^T \left(1+r(l)\right)^T u_2' Pl + \int_0^\infty \max\left(\omega(a,l), \omega(\bar{a},l)\right) dF_a(a) \end{aligned}$$

Rearranging, and deleting the shared term $\beta^T \left(1+r(l)\right)^T u_2' Pl$, we have:

$$\omega(\bar{a},l) = -\beta^T u_2' \zeta P + \int_0^\infty \max\left(\omega(a,l), \omega(\bar{a},l)\right) dF_a(a) \tag{A12}$$

This is (17) of Theorem 1. Equation (A12) characterizes $\bar{a}(l)$. In words, the LHS of (A12) is the period-1 utility from continuing with the appraisal $\bar{a}$, suffering the cost from under-appraising. The RHS is the expected value from reneging, which is the utility cost $-\beta^T u_2' \zeta P$, plus the expected period-1 utility from drawing a new appraisal. At $\bar{a}$, these must be equal.

We can rearrange (A12) to:

$$\int_{a>\bar{a}} \big(\omega(a,l) - \omega(\bar{a},l)\big)\, dF_a(a) = \beta^T u_2' \zeta P \tag{A13}$$

Since ω is increasing in a, the LHS of (A13) is strictly decreasing in $\bar{a}$, hence for any parameters, there is at most one value of $\bar{a}$ which solves (A13). Note also that (A13) shows that the optimal $\bar{a}$ must satisfy:

$$\bar{a} < \frac{l}{\phi}$$

that is, the optimal cutoff $\bar{a}$ must be low enough that it constrains the amount that can be borrowed. To see this, note that from (18), we have:

$$\omega(a,l) = u_1\big(W_1 - P(1-l)\big) \;\; \forall a > \frac{l}{\phi}$$

That is, when $a > \frac{l}{\phi}$, so the appraisal is high enough that it does not constrain borrowing, then $\omega(a,l)$ is constant in a. As a result,

$$\int_{a>\bar{a}} \big(\omega(a,l) - \omega(\bar{a},l)\big)\, dF_a(a) = 0 \; \forall \bar{a} \geq \frac{l}{\phi}$$

Hence, the LHS of (A13) is 0 for all $\bar{a} > \frac{l}{\phi}$; the RHS is positive, so it can never be optimal to set $\bar{a} > \frac{l}{\phi}$.

B.1.4 Optimal Loan Choice

Repeating (A10), we have that, given the optimal appraisal cutoff $\bar{a}(l)$, the expected value attained by the buyer, in expectation over uncertainty in a, is:

$$E\Big(V\left(\bar{a}(l), l\right)\Big) = -\beta^T \left(1 + r(l)\right)^T u_2' P l + \int_0^\infty \max\Big(\omega(a,l), \omega\left(\bar{a}(l), l\right)\Big)\, dF_a(a) \tag{A14}$$

The buyer picks l to maximize (A14); this is (19).

B.2 Comparative Statics: Optimal Loan Choice

To do comparative statics, we will apply the envelope theorem to the optimization framing of the buyer's choice problem. Define:

$$\Gamma(l) \equiv E\Big(V\left(\bar{a}(l), l\right)\Big)$$

We can write Γ as:

$$\Gamma(l) = \max_{\bar{a}} \left[\int_{\bar{a}}^{\infty} \left[-\beta^T \left(1 + r(l)\right)^T u_2' P l + \omega(a, l)\right] dF_a(a) + F_a(\bar{a}) \left[\Gamma(l) - P\beta^T u_2' \zeta\right]\right] \quad \text{(A15)}$$

In words, the buyer receives $-\beta^T \left(1 + r(l)\right)^T u_2' P l + \omega(a, l)$ in the range $[\bar{a}, \infty]$ where the buyer continues, and $\Gamma(l) - P\beta^T u_2' \zeta$ in the range $[0, \bar{a}]$ where she reneges. In this framing, since $\bar{a}$ is chosen optimally given any l, we have:

$$\frac{\partial}{\partial \bar{a}} \max_{\bar{a}} \left[\int_{\bar{a}}^{\infty} -\beta^T \left(1 + r(l)\right)^T u_2' P l + \omega(a, l)\, dF_a(a) + F_a(\bar{a}) \left[\Gamma(l) - P\beta^T u_2' \zeta\right]\right] = 0$$

Hence, the envelope theorem applies; we have:

$$\frac{d\Gamma}{dl} = \frac{\partial}{\partial l} \int_{\bar{a}^*}^{\infty} -\beta^T \left(1 + r(l)\right)^T u_2' P l + \omega(a, l)\, dF_a(a) + F_a(\bar{a}^*) \left[\Gamma(l) - P\beta^T u_2' \zeta\right]$$

Now, we can write $\Gamma(l)$ substituting for $\omega(a, l)$ using (18), to get:

$$\begin{aligned} \Gamma(l) = \max_{\bar{a}} \int_{\bar{a}}^{\infty} & \left[-\beta^T \left(1 + r(l)\right)^T u_2' P l \right. \\ & + \beta^T \left(1 + r(l)\right)^T u_2' P \max\left[0, l - \phi a\right] \\ & \left. + u_1 \Big(W_1 - P(1 - l) - \psi P \max\left(0, l - \phi a\right)\Big)\right] dF_a(a) \\ & + F_a(\bar{a}) \left[\Gamma(l) - P\beta^T u_2' \zeta\right] \end{aligned} \quad \text{(A16)}$$

Now, note that:

$$l - \max\left[0, l - \phi a\right] = \min\left[l, \phi a\right]$$

Hence, we can write (A16) as:

$$\begin{aligned}\Gamma(l) = &\max_{\bar{a}} \int_{\bar{a}}^{\infty} \left[-\beta^T \left(1 + r(l)\right)^T u_2' P \min(l, \phi a)\right.\\ &\left.+ u_1 \left(W_1 - P(1-l) - \psi P \max(0, l - \phi a)\right)\right] dF_a(a) \\ &+ F_a(\bar{a}) \left[\Gamma(l) - P \beta^T u_2' \zeta\right]\end{aligned} \tag{A17}$$

Differentiating with respect to l, we have:

$$\begin{aligned}\frac{d\Gamma}{dl} = &\frac{\partial}{\partial l}\left[\int_{\bar{a}}^{\infty} -\beta^T \left(1 + r(l)\right)^T u_2' P \min(l, \phi a)\right.\\ &\left.+ u_1 \left(W_1 - P(1-l) - \psi P \max(0, l - \phi a)\right) dF_a(a)\right] + F_a(\bar{a}^*) \frac{d\Gamma}{dl}\end{aligned} \tag{A18}$$

$$\begin{aligned}\frac{d\Gamma}{dl}\left(1 - F_a(\bar{a}^*)\right) = &\frac{\partial}{\partial l}\left[\int_{\bar{a}}^{\infty} -\beta^T \left(1 + r(l)\right)^T u_2' P \min(l, \phi a)\right.\\ &\left.+ u_1 \left(W_1 - P(1-l) - \psi P \max(0, l - \phi a)\right) dF_a(a)\right]\end{aligned} \tag{A19}$$

Now, we can separately analyze the RHS, in the under-appraisal region $a \in \left[\bar{a}, \frac{l}{\phi}\right]$ and the over-appraisal region $a \in \left[\frac{l}{\phi}, \infty\right]$. In the over-appraisal region, we have $\min(l, \phi a) = l$ and $\max(0, l - \phi a) = 0$, hence:

$$\frac{\partial}{\partial l}\int_{\frac{l}{\phi}}^{\infty} -\beta^{T}\left(1+r\left(l\right)\right)^{T}u_2' P\min\left(l,\phi a\right)+u_1\left(W_1-P\left(1-l\right)-\psi P\max\left(0,l-\phi a\right)\right)dF_a\left(a\right)=$$

$$\frac{\partial}{\partial l}\int_{\frac{l}{\phi}}^{\infty} -\beta^{T}\left(1+r\left(l\right)\right)^{T}u_2' Pl+u_1\left(W_1-P\left(1-l\right)\right)dF_a\left(a\right)=$$

$$\left(1-F_a\left(\frac{l}{\phi}\right)\right)\left(\underbrace{-T\beta^{T}\left(1+r\left(l\right)\right)^{T-1}r'\left(l\right)u_2' Pl}_{Rate\ Change}+\underbrace{Pu_1'\left(W_1-P\left(1-l\right)\right)-P\beta^{T}\left(1+r\left(l\right)\right)^{T}u_2'}_{Consumption\ Smoothing}\right)$$
$$-\underbrace{\left[-\beta^{T}\left(1+r\left(l\right)\right)^{T}u_2' Pl+u_1\left(W_1-P\left(1-l\right)\right)\right]f\left(\frac{l}{\phi}\right)}_{Nuisance\ Term} \quad \text{(A20)}$$

The "rate change" term in (A20) represents the increase in interest payments in period 2 from increasing $r\left(l\right)$. The "consumption smoothing" term represents gains from more effectively smoothing consumption over the two periods. The intuition is that, if the house over-appraises, targeting a larger loan allows the buyer to borrow more, smoothing consumption, and gaining on the margin the gap between period-1 and period-2 marginal utilities. The "nuisance term" will cancel once we consider the under-appraisal region.

In the underappraisal region, we have $\min\left(l,\phi a\right)=\phi a$ and $\max\left(0,l-\phi a\right)=l-\phi a$, hence:

$$\frac{\partial}{\partial l}\int_{\bar{a}}^{\frac{l}{\phi}} -\beta^{T}\left(1+r\left(l\right)\right)^{T}u_2' P\min\left(l,\phi a\right)+u_1\left(W_1-P\left(1-l\right)-\psi P\max\left(0,l-\phi a\right)\right)dF_a\left(a\right)=$$

$$\frac{\partial}{\partial l}\int_{\bar{a}}^{\frac{l}{\phi}} -\beta^{T}\left(1+r\left(l\right)\right)^{T}u_2' P\phi a+u_1\left(W_1-P\left(1-l\right)-\psi P\left(l-\phi a\right)\right)dF_a\left(a\right)=$$

$$\int_{\bar{a}}^{\frac{l}{\phi}}\underbrace{-T\beta^{T}\left(1+r\left(l\right)\right)^{T-1}r'\left(l\right)u_2' P\phi a}_{Rate\ Change}+\underbrace{\left(1-\psi\right)Pu_1'\left(W_1-P\left(1-l+\psi\left(l-\phi a\right)\right)\right)}_{Under-Appraisal\ Penalty}dF_a\left(a\right)$$
$$+\underbrace{\left[-\beta^{T}\left(1+r\left(l\right)\right)^{T}u_2' P\phi\left(\frac{l}{\phi}\right)+u_1\left(W_1-P\left(1-l\right)-\psi P\left(l-\phi\left(\frac{l}{\phi}\right)\right)\right)\right]f\left(\frac{l}{\phi}\right)}_{Nuisance\ Term}$$
$$\text{(A21)}$$

The "rate increase" term is analogous to (A20). The intuition behind the "under-appraisal penalty" term is that, if the house eventually under-appraises, targeting a larger loan does not increase the eventual borrowing amount, but increases the size of any under-appraisal, causing the buyer to have to pay a penalty $\psi - 1 > 0$ of the incremental loan amount. The "nuisance term" simply cancels with the corresponding term from (A20) once we add the two components.

Combining (A20) and (A21), we have:

$$\begin{aligned}\frac{\partial}{\partial l}\int_{\bar{a}}^{\infty} u_1\left(W_1 - P\left(1-l\right) - \psi P \max\left(0, l - \phi a\right)\right) dF_a\left(a\right) = \\ \left(1 - F_a\left(\bar{a}\right)\right)\left(-T\beta^T\left(1 + r\left(l\right)\right)^{T-1} r'\left(l\right) u_2' P l\right) + \\ \left(1 - F_a\left(\frac{l}{\phi}\right)\right) P\left(u_1'\left(W_1 - P\left(1-l\right)\right) - \beta^T\left(1 + r\left(l\right)\right)^T u_2'\right) - \\ \int_{\bar{a}}^{\frac{l}{\phi}} \left(1-\psi\right) P u_1'\left(W_1 - P\left(1 - l + \psi\left(l - \phi a\right)\right)\right) dF_a\left(a\right)\end{aligned} \tag{A22}$$

Finally, combining (A22) with (A19), we have:

$$\begin{aligned}\frac{d\Gamma}{dl} = \left(-T\beta^T\left(1 + r\left(l\right)\right)^{T-1} r'\left(l\right) u_2' P l\right) + \\ \frac{1}{\left(1 - F_a\left(\bar{a}^*\right)\right)}\Bigg[\left(1 - F_a\left(\frac{l}{\phi}\right)\right) P\left(u_1'\left(W_1 - P\left(1-l\right)\right) - \beta^T\left(1 + r\left(l\right)\right)^T u_2'\right) - \\ \int_{\bar{a}}^{\frac{l}{\phi}} \left(1-\psi\right) P u_1'\left(W_1 - P\left(1 - l + \psi\left(l - \phi a\right)\right)\right) dF_a\left(a\right)\Bigg]\end{aligned}$$

Setting $\frac{d\Gamma}{dl}$ to 0 and rearranging, we can write the FOC for optimal loan choice as:

$$\underbrace{\left(1-F_a\left(\frac{l}{\phi}\right)\right)P\left(u_1'\left(W_1-P\left(1-l\right)\right)-\beta^T\left(1+r\left(l\right)\right)^T u_2'\right)}_{Consumption\ \ Smoothing}=$$
$$\underbrace{\left(1-F_a\left(\bar{a}^*\right)\right)\left(T\beta^T\left(1+r\left(l\right)\right)^{T-1}r'\left(l\right)u_2'Pl\right)}_{Rate\ \ Change}+$$
$$\underbrace{\int_{\bar{a}}^{\frac{l}{\phi}}\left(1-\psi\right)Pu_1'\left(W_1-P\left(1-l+\psi\left(l-\phi a\right)\right)\right)dF_a\left(a\right)}_{Under-Appraisal\ \ Penalty} \quad \text{(A23)}$$

The LHS of (A23) captures the effect of increasing loan size on consumption smoothing. If the house eventually appraises successfully, increasing targeted loan size by a dollar moves consumption from period 2, where marginal utility is lower, to period 1, where it is higher. The RHS captures the two costs of increasing l: first, the interest rate paid increases; second, conditional on under-appraisal, increasing l does not change the final loan size, but increases the consumption penalty from under-appraisal, since under-appraisals are larger. Hence, at the optimal choice of l, the LHS is positive: the buyer would prefer to increase loan size slightly, to shift consumption from period 2 to period 1, but is deterred from doing so by the rate change and under-appraisal penalty effects.

B.3 Microfounding the Mortgage Rate Menu

In this appendix, we construct a microfounded model showing how mortgage interest rates depend on targeted loan size and price dispersion. We assume mortgage rates arise from competition between profit-maximizing lenders. Suppose that, once a homebuyer has purchased the house with a mortgage, the buyer will default on the mortgage at rate δ. If the buyer defaults, the lender incurs a proportional cost Pc to foreclose the house, reflecting foreclosure discounts and other hassle costs of foreclosing. The foreclosure price is a function of the initial transaction price and a random component, ϵ_F, which has standard deviation σ_F that depends on idiosyncratic price dispersion. Thus, the final recovery value is as follows:

$$F=P\left(1-c+\epsilon_F\right) \quad \text{(A24)}$$

Thus, for a non-recourse mortgage, lender's expected loss conditional on default is:[24]

$$Loss = E\left[P\left(l - \max\left[l, 1-c+\epsilon_F\right]\right)\right] = PE\left[\max\left[0, l-(1-c+\epsilon_F)\right]\right] \tag{A25}$$

Lender's expected loss is increasing in σ_F because the lender can recover at most l and bears the cost when the foreclosure price is less than l.[25] Thus, when the variance of the foreclosure price is larger, the lender's expected losses on loans is higher.

Now, suppose lenders have cost of funds ρ, and let r represent the mortgage interest rate. Lenders' profit if buyers do not default is $Pl\,(r-\rho)$. In a competitive equilibrium, the menu of interest rates and loan size must be set such that the lender will break even on any mortgage-rate pair:

$$Pl\,(1-\delta)\,(r-\rho) = \delta PE\left[\max\left[0, l-(1-c+\epsilon_F)\right]\right] \tag{A26}$$

The LHS of (A26) is lenders' expected profit, which is the product of mortgage size l, the repayment probability $(1-\delta)$, and the mortgage spread $(r-\rho)$. The RHS is lenders' expected losses conditional on default, multiplied by the default probability δ.

Expression (A26) defines a menu of (l, r) pairs available to buyers. As we increase idiosyncratic price variance, thus increasing the variance of prices upon foreclosure σ_F, the menu of (l, r) pairs shifts to be worse for the borrower. Formally, when ϵ_F is normally distributed, the RHS of (A26) is always increasing in σ_F.[26] Thus, holding l fixed, increasing σ_F must cause r to increase. This rationalizes our observations in Figure 3 and Table 4. Expression (15) in the main text can be thought of as a linear approximation to this menu.

B.3.1 Mortgage Rate Menu Calibration

We next do a simple calibration, to show that this microfoundation can also quantitatively rationalize the relationships between interest rates, loan size, and price dispersion observed

[24] Mortgages are recourse in some states, but wage garnishment and other methods for collecting debt from buyers after the house has been sold are expensive, and buyers cannot be collected from if they file Chapter 7 bankruptcy.

[25] We assume that if the borrower defaults, it happens before Period 2. This assumption is reasonable because buyers are more likely to default in early stage when they have less equity in the house. If we relax this assumption, the loss function will be as follows, which will result in similar results: $Loss = E\left[P\left(l(1+\rho) - \max\left[l(1+r), 1-c+\epsilon_F\right]\right)\right]$

[26] Note that the RHS of (A26) is equal to δ times the value of a European call option on $l-(1-c+\epsilon_F)$ with strike 0; the value of such a call option is always increasing in volatility.

in the data. Essentially, in the calibration, we will group the data into buckets with different default rates δ. We will estimate σ_F based on price dispersion in the data, and we will choose a foreclosure discount c to minimize the distance between the model and data interest rate menus. We will then show that the fitted model, optimizing over a single parameter, can fit the empirical relationships between loan size l, price dispersion σ_F, and interest rates r, simultaneously for many levels of default rates.

We restricting the sample to all portfolio loans. We first group the data into four FICO score bins, Excellent (800-850), Very Good (740-799), Good (670-739), and Fair (580-669), indexed by f. We split each FICO score bin into high- and low-dispersion counties, indexed by d, and also split loans into LTP bins, from 60-65, 65-70, up to 80. For each FICO score bucket f, dispersion case d, and LTP bin l, we estimate average residualized interest rates r_{fld} in our sample of loans. Since the level of r_{fld} is meaningless after residualization, we normalize by subtracting the mean rate $\bar{r}_f$ within each FICO bucket f:

$$\tilde{r}_{fld} = r_{fld} - \bar{r}_f \tag{A27}$$

Since we normalize within FICO buckets, we preserve the relationships between $\tilde{r}_{fld}$, loan size l, and price dispersion d within each FICO bucket. The residuals $\tilde{r}_{fld}$ are essentially the points in the interest rate menu of Figure 3, separate for each of the four FICO buckets.

Next, we describe how we simulate value of model-predicted interest rate menu points $\tilde{r}^{model}_{fld}(c)$, given the foreclosure discount c. We assume that ϵ_F is normally distributed, with mean 0 and variance σ_F. In each FICO score bin, we calculate a homogeneous value of δ as the average delinquency rate across all loans. To determine σ_F in the high- and low-dispersion areas, we calculate the average repeat-sales residual, as described in Appendix D.1, separately for high-dispersion and low-dispersion counties.[27] We find $\sigma_F = 0.0941$ for low-dispersion areas, and $\sigma_F = 0.131$ for high-dispersion counties. Given δ, σ_F, and loan size l, for any value of the foreclosure discount c, we can calculate the interest rate spread

[27] We use repeat-sales residuals to estimate σ_F, rather than the hedonic model residuals in the main text, because repeat-sales are closer to the thought experiment in the collateral recovery model. We are interested in, when a house forecloses, how variable its price is relative to its purchase price, which is captured in a repeat-sales specification. If a house has large errors in the hedonic model, but not the repeat-sales model – that is, a house has persistently high values relative to its characteristics – this does not affect the variability of the house price relative to loan value upon foreclosure, so this should not be included in ϵ_F.

$r_{fld}^{model} - \rho$ using (A26):

$$r_{fld}^{model} - \rho = \frac{\delta E\left[\max\left[0, l - (1 - c + \epsilon_F)\right]\right]}{l\,(1-\delta)} \tag{A28}$$

where the expectation on the RHS of (A28) can be analytically calculated, since we assumed ϵ_F is normally distributed. We can then calculate the model counterpart of the interest rate residuals (A27), by subtracting the mean interest rate in each FICO bucket f:

$$\begin{aligned}\tilde{r}_{fld}^{model}(c) = r\left(l, \delta, c, \sigma_F\right) - \frac{\sum_l \sum_f r\left(l, \delta, c, \sigma_F\right)}{\sum_l \sum_f 1} = \\ \left(r\left(l, \delta, c, \sigma_F\right) - \rho\right) - \frac{\sum_l \sum_f r\left(l, \delta, c, \sigma_F\right) - \rho}{\sum_l \sum_f 1}\end{aligned} \tag{A29}$$

Note that (A29) implies that $\tilde{r}_{fld}^{model}$ does not depend on the choice of ρ, so we set an arbitrary value of ρ in calculating $\tilde{r}_{fld}^{model}(c)$. We then choose a value of the foreclosure discount c through generalized method of moments, to minimizes the squared distance between the data residuals $\tilde{r}_{fld}$, and the model residuals $\tilde{r}_{fld}^{model}$:

$$c^* = \arg\min_c \sum_l \sum_f \sum_d w_{fd}\left(\tilde{r}_{fld} - \tilde{r}_{fld}^{model}\right)^2$$

where, we set the weights w_{fld} equal to the inverse of the standard deviation of residuals $\tilde{r}_{fld}$ within each FICO and dispersion bucket; this is useful since, without weights, the errors in the low-FICO buckets would dominate the GMM objective function, since rates are higher and more variable when FICO scores are lower.

Our GMM estimate of the foreclosure discount c^* is 0.2018. This is within the range of foreclosure discounts estimate in the literature; for example, Pennington-Cross (2006) estimate a foreclosure discount of 22%, and Zhou et al. (2015) estimate discounts ranging from 11% to 26%.

Figure A9 illustrates the fit of the model. In the top two panels, we show the data and model rate residuals, $\tilde{r}_{fld}$ and $\tilde{r}_{fld}^{model}$, on the y-axis, against the LTP on the x-axis, separately for low-dispersion (top left) and high-dispersion (top right) areas. Different colors represent different credit score bins. In the data, the interest rate menu is steeper when FICO scores are lower; the model is able to quantitatively match this feature of the data, with some errors from the model-predicted interest rate menus being slightly too flat for low FICO bins. This shows that the collateral recovery model is able to quantitatively explain the relationship between interest rates and loan size.

To focus on the effect of price dispersion of credit, in the bottom panel of Figure A9, we show the difference in interest rates between high- and low-dispersion cases, for each FICO bucket and LTP; that is, each point is in the bottom panel shows:

$$r_{lf,d=H} - r_{lf,d=L} \tag{A30}$$

This is the difference between interest rates in high-dispersion and low-dispersion areas. In other words, the solid green line in the bottom panel is equal to the difference between the solid green line in the top right panel (rates for high-dispersion areas in FICO bin 4) and the solid green line in the top left panel (rates for low-dispersion areas in FICO bin 4). In the data, (A30) is larger when FICO scores are lower: dispersion affects mortgage credit more when default rates are higher. We showed a related pattern, using LTP as the dependent variable, in Figure A4. The model lines are very close to the data lines in Figure A9, implying that the model produces a surprisingly good fit of the relationship between default rates, and the relationship of price dispersion with mortgage interest rates: we are able to match the average level of each of the lines, as well as the slope for the green line, representing the lowest FICO scores.

Thus, we have shown that the interrelationships between interest rate residuals, LTP, default rates, and price dispersion in the portfolio segment of our data are quantitatively consistent with a simple collateral recovery model, under realistic parameter settings. The simple model fits the data surprisingly well, given that we only optimize a single parameter, the foreclosure discount c, in the model fitting.

B.4 Appraiser Incentives

This appendix constructs a microfounded model of appraiser behavior, which rationalizes our assumptions on how appraisers bias appraisal prices in (16) of Subsection 4.2. Our model is essentially a special case of Calem et al. (2021). The model also shares some similarities with Conklin et al. (2020), but does not model competition between appraisers. Our model is simplified and disregards some stylized facts shown in the literature: for example, we rule out the possibility that house prices are renegotiated downwards when appraisals fall below sale prices, a phenomenon which is analyzed in Fout et al. (2021).

From (9), the max loan the borrower can take out is:

$$L_{max} = \phi \max(P, A)$$

Suppose that the house appraiser receives utility χL_{max} if the loan size is L_{max}; that is, the appraiser receives some side benefit χ, for every unit they can increase the borrower's max loan size by. This could capture, for example, possible repeat business incentives to produce high appraisals, relationships with lenders (Eriksen et al., 2019), and other such forces.

We also assume that appraisers have some convex cost of biasing appraisals. If the "true" raw appraisal price is A_{raw}, and the appraiser generates appraisal A, then the appraiser incurs a cost:

$$c(A, A_{raw}) = \gamma (A - A_{raw})^2 \tag{A31}$$

This cost is a reduced-form way to capture the fact that it is more costly for appraisers to generate larger distortions in appraisal prices. The literature has documented that appraisers have a number of methods to shift appraisal prices, such as misreporting certain house attributes (Eriksen et al., 2020) and changing the weights on comparable sales used to calculate appraisals (Eriksen et al., 2019). Appraisers would have to misreport attributes or shift weights more to bias appraisals by larger amounts, which may be more costly to the appraiser in terms of legal and reputational risk, or psychological costs.

Appraisers thus solve:

$$\max_A U_{appr}(A) = \chi L_{max}(A) - \gamma (A - A_{raw})^2 \tag{A32}$$

The optimization problem in (A32) has three distinct regions. First, if $A_{raw} > P$, then the appraiser cannot increase L_{max}; it is thus optimal to set $A = A_{raw}$.

Second, suppose A_{raw} is very low. Conjecture that the optimal A is below P, so that the first-order condition for optimality holds:

$$\chi \frac{\partial L_{max}}{\partial A} = 2\gamma \left(A - A_{raw}\right)$$

This gives $A - A_{raw} = \frac{\chi\phi}{2\gamma}$. Define $b \equiv \frac{\chi\phi}{2\gamma P}$. We then have:

$$A - A_{raw} = bP$$

Third, suppose that:

$$P\left(1-b\right) \leq A_{raw} \leq P$$

In this range, we have that:

$$\frac{\partial U_{appr}}{\partial A} > 0 \ \forall A < P$$

Hence, it is optimal for the appraiser to set A=P.

We have thus shown that the appraiser's optimal appraisal A^* satisfies:

$$A^* = \begin{cases} A_{raw} + bP & A_{raw} \leq \left(1-b\right)P \\ P & \left(1-b\right)P < A_{raw} \leq P \\ A_{raw} & P < A_{raw} \end{cases}$$

which is exactly (16) in the main text.

B.5 Microfounding the Penalty Cost Parameter ψ

This appendix constructs a microfoundation for the "penalty cost" parameter ψ, which implies that increases in down payments caused by under-appraisals decrease consumption more than one-for-one. We do a simple calculation to illustrate that the penalty cost can be fairly large in reasonable models. Suppose an agent lives for T periods, and maximizes discounted CRRA utility over consumption:

$$\sum_{t=1}^{T} \beta^t \frac{c_t^{1-\eta} - 1}{1-\eta}$$

$$s.t.\ a_{t+1} + c_t = y_t + a_t\,(1+r)$$

Income y_t is exogeneous and nonrandom. As is standard in the lifecycle literature, we set $\eta = 2$. We set $\beta = 0.95, r = \frac{1}{\beta} - 1$, so that the optimal solution without uncertainty involves consuming equal amounts in every time period. We set $T = 10$, so a time period can be thought of as representing a year, and consumers can be thought of as have 10 years to save for a home purchase at time T. We set $y_t = 10$ for each period.

We compare two cases. The first is an *anticipated* shock to income in period T, whose realization is known in period 1. The anticipated shock can be thought of as the homebuyer choosing a lower target loan size: since she plans to make a larger down payment, she can consumption-smooth for this in advance. The second is an *unanticipated* shock, whose realization is only known in period T. This can be thought of as the homebuyer targeting a large loan size and anticipating that under-appraisals may force her to borrow less than the target loan size. This kind of shock is more costly because the consumer can consumption-smooth the first kind of shock in expectation, but cannot condition her consumption on the under-appraisal. We will show that the second kind of shock decreases total utility more than the former.

For both cases, we suppose that $y_T = 10$ and $y_T = 0$ with equal probability, and $y_t = 10$ for all periods $t \neq T$. In the anticipated case, we assume y_T is known when the buyer makes consumption decisions in earlier periods. Thus, to solve this problem, we simply solve a zero-uncertainty finite-horizon dynamic program for the consumer for each value of y_T, and then take the average lifetime value at $t = 0$ from each case. In the unanticipated case, the consumer's value function in period $T - 1$ is the average of her value if $y_t = 10$ and if $y_t = 0$. The rest of the consumer's problem can be solved with standard backwards induction. We solve both cases using the standard endogeneous gridpoint method for solving lifecycle problems.

We compare the consumer's lifetime value in both the anticipated and unanticipated income decrease cases to the baseline case where $y_t = 10$ for all time periods. In the anticipated case, lifetime value drops by 0.0361, whereas in the unanticipated case lifetime value drops by 0.050. Hence, under these parameter settings, an unanticipated shock is roughly 40% more costly, in utility terms, than an anticipated shock of the same magnitude, due to the inability to condition early-period consumption on the realization of the shock. Thus, unanticipated shocks to consumption can have much larger effects on utility than equally sized anticipated shocks. Our consumption penalty parameter ψ is a reduced-form way to capture this effect.

C Implications for Homeownership: A Quantitative Model

In the main text, we showed that house price dispersion is substantially correlated with loan-to-price ratios. To measure how large these effects are economically, in this section, we build a life-cycle model of housing choices. We show that the LTP changes associated with price dispersion are economically significant: if LTPs are decreased to their levels in high-dispersion areas, aggregate homeownership rates drop by 1.5pp, and low-income homeownership rates drop by 2.6pp.

C.1 Model

We consider a partial-equilibrium model of housing choice, in which households live for a finite number of periods, receive stochastic income, and purchase housing using mortgages. Our main departure from the standard model is that we will allow the loan-to-value constraint to vary according to house quality, in a way that is informed by our empirical results; we will then vary this relationship in the counterfactuals.

Income. A household lives for $T = 65$ periods, from age 25 to age 80. The household works for the first $T_{ret} - 1$ periods, then retires at age 60. At age t, the household receives exogeneous after-tax labor income $(1 - \tau)y_t$, where τ is the income tax rate, and:

$$\log (y_t) = \chi_t + \zeta_t \tag{A33}$$

χ_t is an age-specific constant which matches the lifecycle pattern of income. ζ_t is a transitory shock, which follows an AR(1) process:

$$\zeta_t = \rho_\zeta \zeta_{t-1} + \varepsilon_t$$

Households retire at 60, and receive social security benefits thereafter. ζ_t is the only source of uncertainty in the model. We also allow households to begin life with different initial incomes, a_0. Agents can save using riskless bonds, and also buy houses and borrow using mortgages against the house.

Housing. There is a discrete grid of house qualities $h_i \in S = \{s_1, s_2 \ldots s_H\}$, ordered in increasing order. There is a cutoff s_R, for $R < H$. All house qualities below $s_1 \ldots s_R$ are available for rent only, and all house qualities $s_{R+1} \ldots s_H$ are available to purchase only. Thus, the household can only rent low-quality houses, and must purchase a house to receive housing services above s_R. Rental housing has a flow cost of $p^r h_i$, that is, p^r per unit of housing services rented. The price of an owned house of quality h_i is $p^h h_i$. Homeowners pay a depreciation cost of δ^h times the value of the house, or $\delta^h p^h h_t$, each period they own the house. This can be thought of as a maintainence cost. Buying a new house also costs some fixed cost of F^{pur} of the value of the house, or $F^{pur} p^h h_t$; this can be thought as representing realtor fees and other costs of buying a house.

Households can borrow up to a fraction $\phi(h_t)$ of the house's value, that is, at mortgage rate $r^h > r^b$. $\phi(h_t)$ can depend on h_t, so lower quality houses can have different LTV requirements, in a way disciplined by data; we describe in detail how we calibrate $\phi(h_t)$ in Subsection C.2 below, and Appendix C.4.2. Let a_t represent cash-on-hand; homeowners' borrowing constraint is thus:

$$a_t \geq -\phi(h_t) p^h h_t \tag{A34}$$

The household faces a mortgage rate $r^m > r^b$. Thus, the household will never want to hold cash and mortgages together.

Utility. Households have CRRA preferences, and maximize expected utility:

$$V_0 = E\left[\sum_{t=1}^{T} \beta^j U(c_t, h_t) + \beta^T U_B(w_{T+1})\right]$$

discounting at rate β. Per-period utility is:

$$U(c,h) = \frac{\left(c^\alpha h^{1-\alpha}\right)^{1-\sigma} - 1}{1-\sigma}$$

Households also receive utility from bequests, U_B:

$$U_B(w_{T+1}) = K_B \frac{w_{T+1}^{1-\sigma} - 1}{1-\sigma}$$

where w_{T+1} is final-period wealth from housing and cash-on-hand:

$$w_{T+1} = a_{T+1} + p^h h_{T+1}$$

and K_B is parameter which determines the importance of bequests to the household.

Value functions. There are three state variables for the household's problem: house quality h_t, start-of-period cash-on-hand a_t, and the persistent income shock ζ_t. The household's value function is:

$$V_t\left(h_t, a_t, \zeta_t\right) = \max\left\{V_t^{renter}\left(h_t, a_t, \zeta_t\right), V^{purchase}\left(h_t, a_t, \zeta_t\right)\right\}$$

If the household decides to rent in period t, it solves:

$$V^{renter}\left(h_t, a_t, \zeta_t\right) = \max_{c_t, a_{t+1}, h_{t+1}} u\left(c_t, h_{t+1}\right) + \beta E\left[V_{t+1}\left(h_{t+1}, a_{t+1}, \zeta_{t+1}\right) \mid \zeta_t\right] \tag{A35}$$

$$s.t.\ c_t + \frac{a_{t+1}}{1+r_t} = a_t + y_t + \underbrace{p^h h_t 1\left(h_t > s_R\right)}_{Selling\ old\ house} - p_r h_{t+1} \tag{A36}$$

$$r_t = \begin{cases} r^m & a_{t+1} < 0 \\ r^b & a_{t+1} \geq 0 \end{cases}$$

$$a_{t+1} \geq 0, h_{t+1} < s_R$$

That is, consumption plus cash-on-hand at the end of the period is equal to cash-on-hand a_t, plus labor income y_t, minus rent. If the household decides to own in period t, it solves:

$$V^{purchase} = \max_{c_t, a_{t+1}, h_{t+1}} u\left(c_t, h_{t+1}\right) + \beta E\left[V\left(h_{t+1}, a_{t+1}, \zeta_{t+1}\right) \mid \zeta_t\right] \tag{A37}$$

$$s.t.\ c_t + \frac{a_{t+1}}{1+r_t} = a_t + y_t + \underbrace{p^h h_t 1\left(h_t > s_R\right)}_{Selling\ old\ house} - \underbrace{\left(1 + \delta^h + F^{pur} 1\left(h_{t+1} \neq h_t\right)\right) p^h h_{t+1}}_{Buying\ new\ house} \tag{A38}$$

$$r_t = \begin{cases} r^m & a_{t+1} < 0 \\ r^b & a_{t+1} \geq 0 \end{cases}$$

$$a_{t+1} \geq -\phi\left(h_{t+1}\right) p^h h_{t+1}, \ h_{t+1} \geq s_R$$

C.2 Calibration

The model period is annual. Most of our choices for parameter calibrations are standard, and we discuss them in Appendix C.4.1. The core way in which we deviates from the standard lifecycle model calibration is in the $\phi(h)$ function, which determines the relationship between house qualities and average LTV. We calibrate three different versions of $\phi(h)$, to represent the loan-to-price ratios available to households in counties with high (top decile), medium (median decile), and low (bottom decile) average price dispersion. We plot these functions in Appendix Figure A8, and describe details of how we construct these functions in Appendix C.4.2. We essentially estimate the relationship between prices and average price dispersion σ in each group of counties, and then calculate LTVs by multiplying the differences in σ by the coefficient from specification 1 in Table A3, which is the reduced-form relationship between price dispersion and LTVs, controlling for other observable features that may affect LTV. The average difference in σ between high- and low-dispersion counties is roughly 2.7SD. From Table A3 column 1, a 1SD change in σ is associated with around a -0.8% change in LTV for households with fair credit score, so we set the average difference in LTVs to roughly 2.2%.

Additional details on how we numerically solve the model are in Appendix C.4.3. Table A6 shows values of parameters we use. To simulate model outcomes, we simulate the lives of 1,000,000 households, and calculate averages of model quantities for households at any given age. Appendix Figure A10 evaluates the fit of the model, comparing homeownership rates and debt-to-assets in the model to data from the 2016 SCF. We are able to match the path of homeownership rates very well, and the path of debt-to-assets over the lifecycle fairly well.

C.3 Results

Our core counterfactual is to compare homeownership rates between the high-dispersion and the low-dispersion versions of our calibration. The baseline medium-dispersion case is calibrated to match aggregate homeownership rates, so the high-dispersion calibration represents how homeownership rates would shift in counties where mortgage LTVs available to homebuyers were lower because house price dispersion is high. The change in homeownership rates, moving from the high-dispersion to low-dispersion cases, can be thought of as modelling how much homeownership rates would increase if the housing stock in high-dispersion areas were renewed and rebuilt sufficiently that dispersion dropped to the level of low-dispersion areas, while holding the level of house prices fixed. Average LTVs would then increase, making housing more affordable and causing homeownership rates to increase.[28]

Table A7 shows homeownership rate differences between the high-dispersion and low-dispersion cases. The aggregate homeownership rate difference is roughly 1.5pp. We then divide households into two groups, according to their initial income at age 25.[29]

The effect of price dispersion on homeownership is concentrated among low-income households: at all ages, low-income households have lower homeownership rates in the high-dispersion counterfactual than the low-dispersion counterfactual, with an average homeownership rate difference of 2.6pp. The homeownership gap is large for young households below age 30, somewhat smaller for middle-aged households from 30-40, and rises again for households above 40. In contrast, high-income households initially have higher homeownership rates, but the gap declines essentially to 0 from age 30 onwards.

The difference in collateral constraints induced by collateral value uncertainty contributes to about 6.6% of the homeownership gap between the rich and the poor in 2016, ranging from 5% to 10% across the age distribution.[30] Therefore, our results suggest that, in a standard calibrated lifecycle model of housing choice, LTV differences induced by price dispersion can have sizable effects on aggregate homeownership rates, and the homeownership gap between high- and low-income households.

[28]Note that we showed in Subsection 2.2.2 that price dispersion is lower for houses that are newer. It is important also that house prices are held fixed: in practice, rebuilding houses would likely change the level of average house prices, and this would also affect homeownership rates. We disregard this effect in the calibration, though it may be important in practice.

[29]Since incomes are fairly persistent in lifecycle models, initial incomes have persistent effects on wealth and income at later ages.

[30]The homeownership gap between above-median income households and below-median income households is about 32% in 2016 (SCF Statistics). According to the report by the U.S. Department of Housing and Urban Development, the homeownership gap between the very low-income households and high-income households is 37% in 2004. https://www.huduser.gov/Publications/pdf/HomeownershipGapsAmongLow-IncomeAndMinority.pdf

C.4 Additional Calibration Details

C.4.1 Parameter choices for calibration

Average log earnings over the lifecycle, χ_t, are from the 2016 SCF. The income tax rate τ is set to 0.25. For retired households, χ_t is set to \$15,000 annually, which is approximately the average social security payout in the US.[31] We use standard values of β, σ, α in the literature. Housing transaction costs F^{pur} are set to 0.05, which is the typical fee charged by real estate brokers in the US. This value is also used in Berger et al. (2018) and Wong (2019), among other papers. We set the depreciation rate to 0.01, approximately matching the depreciation rate in BEA data. We set house prices p^h to:

$$p^h = K^H \frac{p^r}{1 - \beta + \delta^h}$$

that is, p^h is rent adjusted for discount rates β and depreciation rates δ^h, multiplied by an adjustment parameter K^H which influences how attractive homeownership is compared to rental. We set the initial distribution of ζ_t, the idiosyncratic income shock, for 25-year-olds such that probabilities are log-linear in the level of ζ_t, that is:

$$P_{25}(\zeta) \propto \exp\left(K_\zeta \zeta\right)$$

where k_ζ controls whether probability weights are higher for high or low values of ζ.[32] We calibrate the persistence of idiosyncratic income shocks ρ_ζ to 0.91, and the standard deviation of shocks σ_ε to 0.21, following Floden and Lindé (2001).

We choose the set of house qualities, the bequest parameter K_B, the housing attractiveness parameter K^H, and the initial income shock distribution slope parameter K^B to match the level and path of homeownership and debt-to-assets from the 2016 SCF, as well as the ratio of of median net worth at age 75 to net worth at age 50 of 1.51, as in Kaplan et al. (2017). While all parameters affect both moments, intuitively, the homeownership rate helps to pin down the level of house prices, and the net worth ratio pins down the bequest

[31] See Table A in the Social Security Program Fact Sheet.

[32] Without adjusting the initial distribution of ζ, we found that homeownership rates rose too quickly in the model relative to the data

parameter. The set of house qualities we use is:

$$\{0.1, 0.3, 0.7, 0.9, 1.1, 1.3, 1.7\}$$

Where all qualities from 0.7 upwards correspond to owned housing.

C.4.2 Calibrating the $\phi(h)$ Functions

We calibrate $\phi(h)$ based on the average price dispersion for each level of house prices and the relationship between price dispersion and LTV that we empirically identified. Our goal for calibrating $\phi(h)$ is to match the relationship between house prices and σ within three segments of the housing market, with high, medium, and low price dispersion. Since we will calibrate $\phi(h)$ based on house prices, with slight abuse of notation, we will write $\phi(p)$ to refer to ϕ as a function of house prices rather than qualities.

We first select a set of counties with comparable house price: average house prices must lie between \$140,000 and \$160,000. We do this filtering because our goal in the model counterfactual is to vary price dispersion holding average prices fixed. We then split these counties into five quintile buckets, by average price dispersion in the county. Within the top, middle, and bottom quintiles, we then calculate conditional expectations of price dispersion as a function of house prices. For the middle quintile, call this conditional expectation:

$$\sigma_{med}(p) \equiv E\left[\sigma_{ict} \mid p_{ict} = p, c \in \mathcal{C}_{mid}\right] \tag{A39}$$

where we used c to index counties, and $c \in \mathcal{C}_{mid}$ means that county c is in the middle quantile of counties by price dispersion. We define $\sigma_{high}(p)$ and $\sigma_{low}(p)$ analogously to (A39), for the high- and low-dispersion set of counties. The three curves $\sigma(p)$ curves are shown in the left panel of Figure A8. We normalized σ by its standard deviation across houses, so the units are identical to those of Table A3. High-dispersion counties have roughly a standard deviation higher values of σ than low-dispersion counties.

To calculate LTVs, let:

$$p_{min\_\sigma} \equiv \arg\min \sigma_{med}(p)$$

represent the house price level with the lowest value of σ, within the medium-dispersion group of counties. We then set $\phi_{med}(p_{min\_\sigma})$ to 80%: that is, the maximal LTV in the medium version of the calibration is set to 80%. To calculate $\phi_{med}(p)$ for other price levels, we set:

$$\phi_{med}(p) = 0.8 + \beta_{LTV\_\sigma}\left(\sigma_{med}(p) - \sigma_{med}(p_{min\_\sigma})\right) \tag{A40}$$

Where β_{LTV_σ} is the coefficient from regressing LTV on price dispersion, from column 1 of Table A3. In words, (A40) states that we adjust LTVs depending on the difference in $\sigma(p)$ values. Formally, the LTV at price p is equal to 0.8, the LTV at p_{min_σ}, plus an adjustment which is the difference between price dispersion at p, and price dispersion at p_{min_σ}, multiplied by β_{LTV_σ}, the effect of price dispersion on LTVs identified in our reduced-form results. Note that we adjust using β_{LTV_σ}, instead of simply taking the empirical relationship between house prices and LTVs, because the price-LTV relationship can be contaminated by many other factors, such as credit demand, which we account for in the specifications we use to identify β_{LTV_σ}.

Similarly, to calculate $\phi_{high}(p)$ for high-dispersion counties, we set:

$$\phi_{high}(p) = 0.8 + \beta_{LTV\_\sigma}\left(\sigma_{high}(p) - \sigma_{med}(p_{min\_\sigma})\right) \tag{A41}$$

That is, analogous to (A40), $\phi_{high}(p)$ is set so that, for any price p, the difference $\phi_{high}(p) - \phi_{med} p_{min\_\sigma}$ is equal to the dispersion difference, $\sigma_{high}(p) - \sigma_{med}(p_{min\_\sigma})$, multiplied by β_{LTV_σ}.

Analogously, for $\phi_{low}(h)$, we set:

$$\phi_{low}(p) = 0.8 + \beta_{LTV\_\sigma}\left(\sigma_{low}(p) - \sigma_{med}(p_{min\_\sigma})\right) \tag{A42}$$

Figure A8 shows the resultant $\phi_{low}(p), \phi_{med}, \phi_{high}$ functions. The left panel shows that high and low-dispersion groups differ by around 1SD of σ; multiplying by the β_{LTV_σ} coefficient, we get an average difference in LTVs of approximately 1.1% between $\phi_{low}(p)$ and $\phi_{high}(p)$ in the right panel. Moreover, the U-shape of the $\sigma(p)$ function, relating house prices to price dispersion, implies that the $\phi(p)$ function has an inverse U-shape: LTVs are highest for moderately-priced houses, and lower for cheap or expensive houses. Thus, a simple way to think of our exercise is that we vary LTVs by around 1.1% around a calibrated model, and measure the effect on resultant homeownership rates.

C.4.3 Numerically Solving the Model

To rectangularize the household problem, we change variables to keep track of agents' total wealth, instead of cash-on-hand:

$$w_t = a_t + 1\left(h_t > s_R\right) p^h h_t$$

From (A34), the leverage constraint then becomes:

$$w_t \geq \left(1 - \phi\left(h_t\right)\right) p^h h_t$$

That is, the household must always have total wealth at least $\left(1 - \phi\left(h_t\right)\right)$ times the price of the house $p^h h_t$.

Combining the owner and renter budget constraints, (A36) and (A38), and rewriting expressions in terms of wealth, we can write the budget constraint equation as:

$$\begin{aligned} w_{t+1} = & \\ (1+r_t) & \left(w_t + y_t - c_t - \underbrace{1\left(h_{t+1} > s_R\right)\left(1 + \delta^h + F^{pur} 1\left(h_{t+1} \neq h_t\right)\right) p^h h_t}_{\textit{Buying new house}} - \underbrace{1\left(h_t < s_R\right) p^r}_{\textit{Rent}} \right) + \\ & 1\left(h_{t+1} > s_R\right) p^h h_t \end{aligned} \quad \text{(A43)}$$

Using (A43), we eliminate consumption c_t from the household's optimization problem, (A35) and (A35). The household thus chooses end-of-period wealth w_{t+1} and house quality h_{t+1} each period, where the state variables are w_t, h_t, ζ_t.

To solve the problem, we discretize ζ_t into 8 states using the Tauchen (1986) method. We use a 150-point approximately exponential grid for w_t, and a 7-point grid for house qualities.

We solve the model using backwards induction, using the generalized endogeneous grid-point method of Druedahl and Jørgensen (2017), which allows for the consumer's problem to be nonconvex. In short, the method involves solving for candidate optimal consumption choices on an endogeneous grid by using inverting the consumer's consumption FOC on the final-period wealth grid, interpolating the results onto an exogeneous grid, and then taking the maximum value attained across candidate optima on the exogeneous grid. This method is thus robust to nonconvexities in the household's problem induced by discrete home purchase decisions and leverage constraints.

To simulate the model, we initialize households with wealth uniformly distributed on from 0 to 20 thousand USD. We initialize ζ_t at its stationary distribution. We then simulate 1,000,000 households over their lifespan, and take average quantities over all households.

D Additional Empirical Results

D.1 Repeat-Sales Estimation and Results

One possible concern regarding our analysis is that our measure of value uncertainty relies heavily on our hedonic model (1) for house prices. To alleviate this concern, in this appendix, we construct an alternative measure of value dispersion using a repeat-sales model. We estimate the following regression specification:

$$p_{it} = \eta_{kt} + \mu_i + \epsilon_{it} \tag{A44}$$

where i indexes properties, k indexes counties, and t indexes months. Equation (A44) is a repeat-sales model for house prices: log prices p_{it} are determined by county-month fixed effects η_{kt}, time-invariant house fixed effects μ_i, and a mean-zero error term ϵ_{it}. Specification (A44) thus models log house prices as following parallel trends, plus error terms: if house A sells for twice the price of house B in June of 2011, house A should sell for twice as much as house B in June of 2017, and any deviation from this is attributed to the error term ϵ_{it}.

There are two additional concerns with measuring idiosyncratic dispersion using a repeat-sales specification. First, the number of data points used to estimate each house fixed effect is very low; thus, the estimated residuals $\hat{\epsilon}_{it}^2$ will tend to be larger for houses which are sold more times, because the house fixed effect γ_i is estimated more precisely. Second, (A44) implicitly assumes that idiosyncratic price dispersion does not depend on the house holding period; a concern is that there is a idiosyncratic price dispersion behaves partially like a random walk, so the error terms may be systematically larger for houses that are sold less frequently.[33] To alleviate the concern that our estimates of $\hat{\epsilon}_{it}^2$ are mechanically driven by sale frequency and time-between-sales, we purge $\hat{\epsilon}_{it}^2$ of any variation which can be explained by tbs_i and $sales_i$. First, we filter to houses sold at most four times over the whole sample period, with estimated values of $\hat{\epsilon}_{it}^2$ below 0.25. We then run the following regression, separately for each county:

$$\hat{\epsilon}_{it}^2 = h_k\left(sales_i, tbs_i\right) + \zeta_{it} \tag{A45}$$

Where, $h_k\left(sales_i, tbs_i\right)$ interacts a vector of $sales_i$ dummies with a fifth-order polynomial in

[33]Note that Giacoletti (2021) and Sagi (2021) show that a large component of idiosyncratic dispersion does not scale with holding period, for both residential and commercial real estate transactions.

tbs_i. The residual $\hat{\zeta}_{it}$ from this regression can be interpreted as the component of the house's price variance which is not explainable by $sales_i$ and tbs_i. We then add back the mean of $\hat{\epsilon}^2_{it}$ within county k:

$$\hat{\epsilon}^2_{TBSadj,it} = \hat{\zeta}_{it} + E_k\left[\hat{\epsilon}^2_{it}\right] \tag{A46}$$

$\hat{\epsilon}^2_{TBSadj,it}$ can be interpreted as the baseline estimates, $\hat{\epsilon}^2_{it}$, nonparametrically purged of all variation which is explainable by a smooth function of $sales_i$ and tbs_i. We then project $\hat{\epsilon}^2_{TBSadj,it}$ onto house characteristics and time, as in (2) in the main text, and take the predicted values as our house-level measure of idiosyncratic price dispersion, which we will call $\hat{\sigma}^2_{RS,it}$.

In comparison to the hedonic model, the repeat-sales model in (A44) is able to capture observable and unobservable features of houses that have time-invariant effects on house prices. Moreover, house fixed effects allow us to capture time-invariant house quality components in a fully nonparametric way, alleviating concerns that the specific functional form we use in (1) is driving our results. A weakness of specification (A44) are that it is unable to capture any features of houses which have time-varying effects on house prices.

Figure A7 shows a binscatter of $\hat{\sigma}^2_{RS,it}$ against our baseline estimates $\hat{\sigma}^2_{it}$. There is a very strong positive relationship. The repeat-sales and hedonic methodologies for measuring house value uncertainty are econometrically quite different; the fact that they produce very correlated results at the house level suggests that both measurement strategies are picking up fundamental value uncertainty among properties, rather than simply reflecting misspecification in the model we use for house prices.

Next, we repeat our regression specifications utilizing $\hat{\sigma}^2_{RS,it}$ as our measure of house price dispersion. Table A5 shows the results; all of our baseline results continue to hold, using $\hat{\sigma}^2_{RS,it}$ as our measure of house price dispersion.

D.2 Price Dispersion and Appraisals

Figure 4 shows binned scatter plots illustrating how under-appraisal is associated with price dispersion across zip codes. For mortgage i, let a_i be the appraisal price, and p_i be the transaction price of the house. The dependent variable in panel (a) of Figure 4, which we call the appraisal deviation from sales price, is defined as:

$$ApprDev_i \equiv \frac{a_i - p_i}{p_i}\mathbf{1}\left(a_i < p_i\right) \tag{A47}$$

That is, the percent deviation of appraisal prices from transaction prices, multiplied by an indicator for the house under-appraising (that is, the appraisal price a_i being below the sales price p_i). This variable captures the downwards pressure that appraisals produce on mortgage limits, combining the probability of under-appraisal with the average magnitude of under-appraisals. Panel (a) of Figure 4 shows that the appraisal deviation from sales prices is much higher in high-dispersion zip codes, suggesting that the extent to which under-appraisals put downwards pressure on LTVs is larger in high-dispersion zip codes.

We then decompose the appraisal deviation into two components. Panel (b) shows the probability that the house under-appraises, $P\left(a_i < p_i\right)$. Panel (c) shows the average deviation of the appraisal price from the sales price from the sales price conditional on under-appraisal, that is,

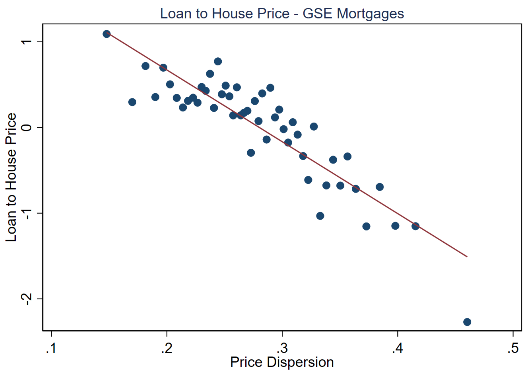
$$E\left[\frac{a_i - p_i}{p_i} \mid a_i < p_i\right] \tag{A48}$$

Panel (b) shows that the probability that a house under-appraises is similar in high- and low-dispersion zip codes; in fact, underappraisals are slightly less likely in high-dispersion zip codes, though this difference is not statistically significant in regression form. However, conditional on under-appraisal, the difference between appraisal and sale prices is much larger in high-dispersion areas. The average magnitude of under-appraisal is around 3% in low-dispersion zip codes, compared to around 5% in high-dispersion zip codes.

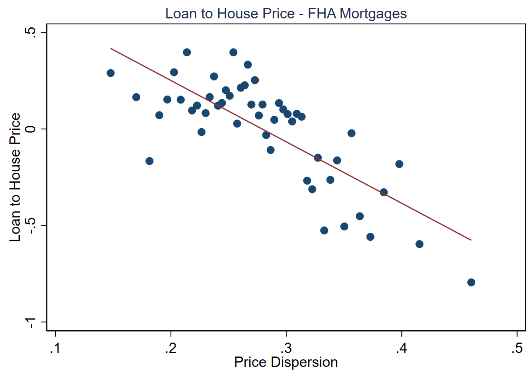
Table A4 confirms Figure 4 findings in regression settings with origination month fixed effects, county-year fixed effects, and borrower and loan controls. In high-dispersion zip codes, appraisal deviations tend to be larger: a 1SD increase in dispersion is associated with a 2bp change in the appraisal deviation (column 1). This is mostly because, conditional on under-appraisal, houses under-appraise by larger amounts: a 1SD increase in dispersion is associated with a 53bp increase in the conditional appraisal deviation (column 3). The probability of under-appraisal is statistically insignificant (column 2). The results are robust in the IV setting as in columns 4-6.

E Appendix Figures and Tables

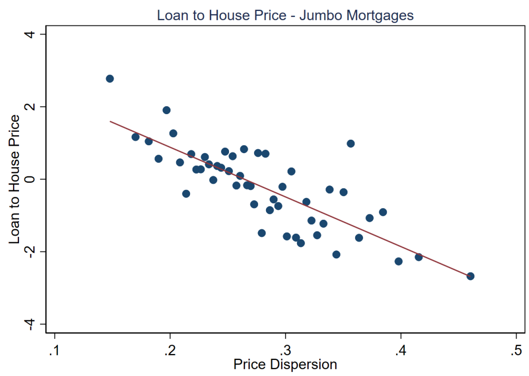
Figure A1. County Level House Price Dispersion and LTP



(a) GSE



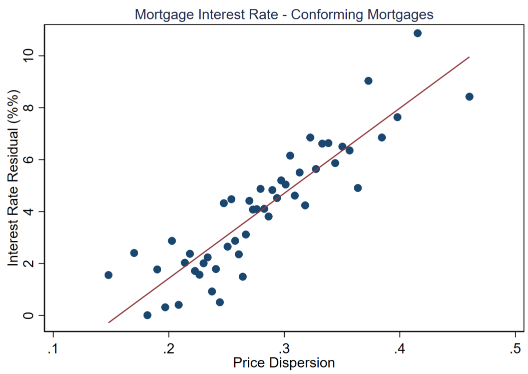
(b) FHA



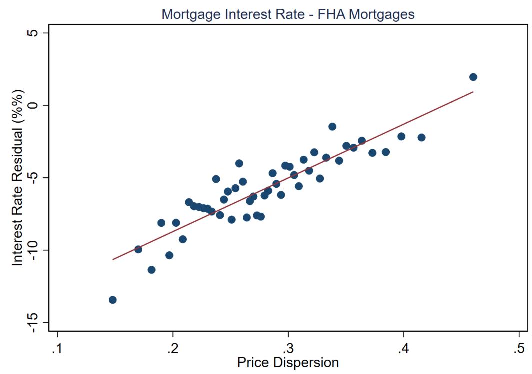
(c) Jumbo

Note: This figure shows the correlation between county level house price dispersion and residualized county average LTP. Panels a-c plot GSE loans, FHA loans, and jumbo loans, respectively. The sample includes annual county observations from 2000 to 2020. *Source*: County house price dispersion is estimated using Corelogic Deeds records. Mortgage data are from Corelogic LLMA.

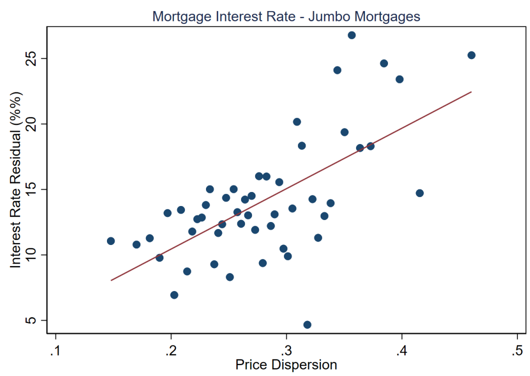
Figure A2. County Level House Price Dispersion and Mortgage Rate



(a) GSE



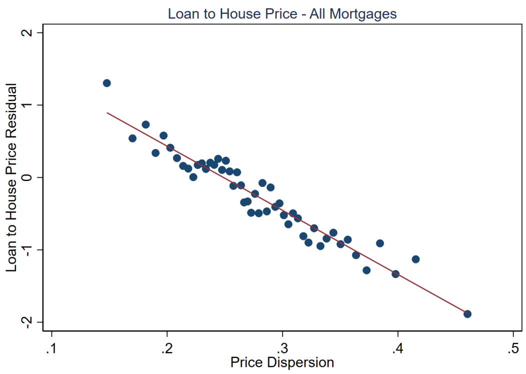
(b) FHA



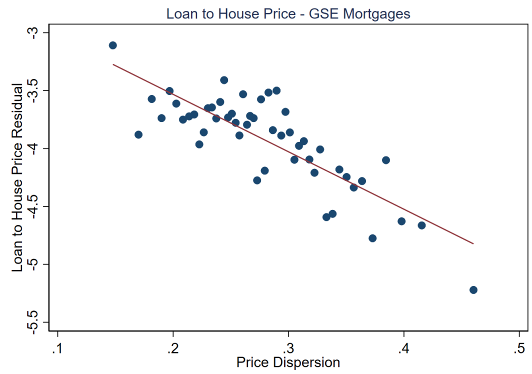
(c) Jumbo

Note: This figure shows the correlation between county level house price dispersion and residualized county average mortgage interest rate. Panels a-c c plot GSE loans, FHA loans, and jumbo loans, respectively. Individual mortgage interest rates are residualized using borrower and loan characteristics, such as FICO, LTP, DTI, the squared terms, and their interactions with origination year. We then take the county-average of residualized mortgage rates. The sample includes annual county observations from 2000 to 2020. *Source*: County house price dispersion is estimated using Corelogic Deeds records. Mortgage data are from Corelogic LLMA.

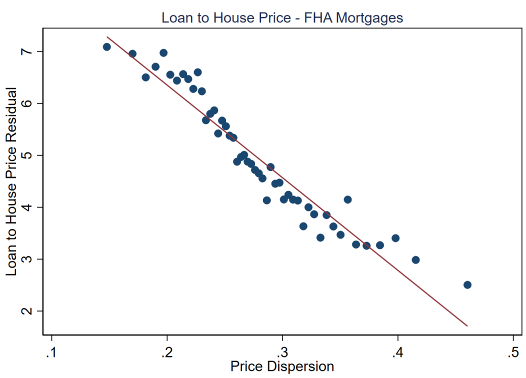
Figure A3. County Level House Price Dispersion and LTP Residuals



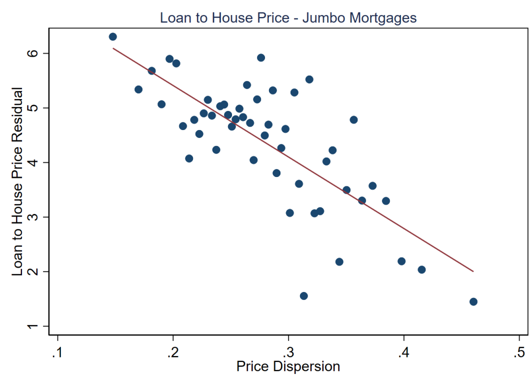
(a) All



(b) GSE



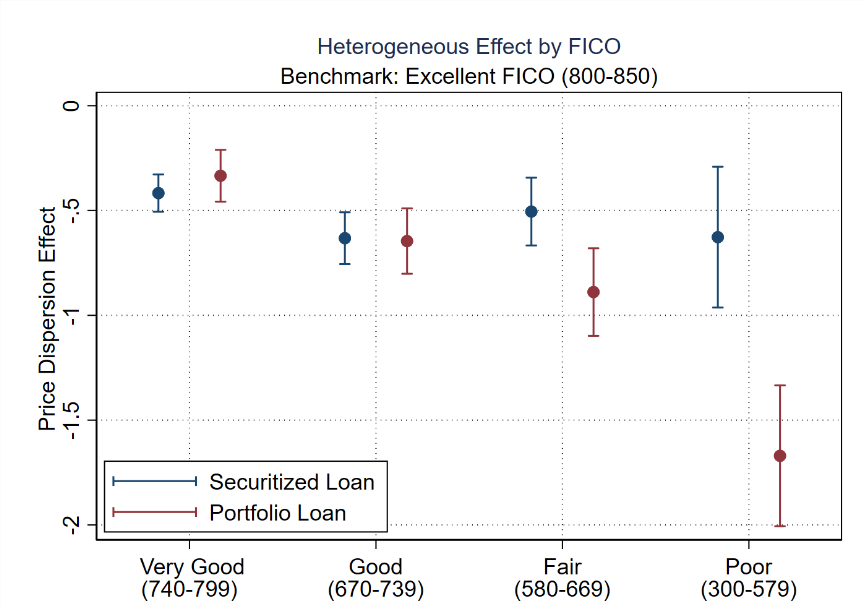
(c) FHA



(d) Jumbo

Note: This figure shows the correlation between county level house price dispersion and residualized county average LTP. Panel a plots the full sample. Panels b-d plot GSE loans, FHA loans, and jumbo loans, respectively. We residualize LTP values by taking the residuals of regressions of LTP on mortgage interest rate, debt-to-income ratio (DTI), DTI-square, FICO, FICO-square, log house price, and their interactions with origination years, and origination year fixed effects. We then take the county-average of residualized LTP. The sample includes annual county observations from 2000 to 2020. *Source*: County house price dispersion is estimated using Corelogic Deeds records. Mortgage data are from Corelogic LLMA.

Figure A4. Heterogeneous Effect of Price Dispersion by FICO

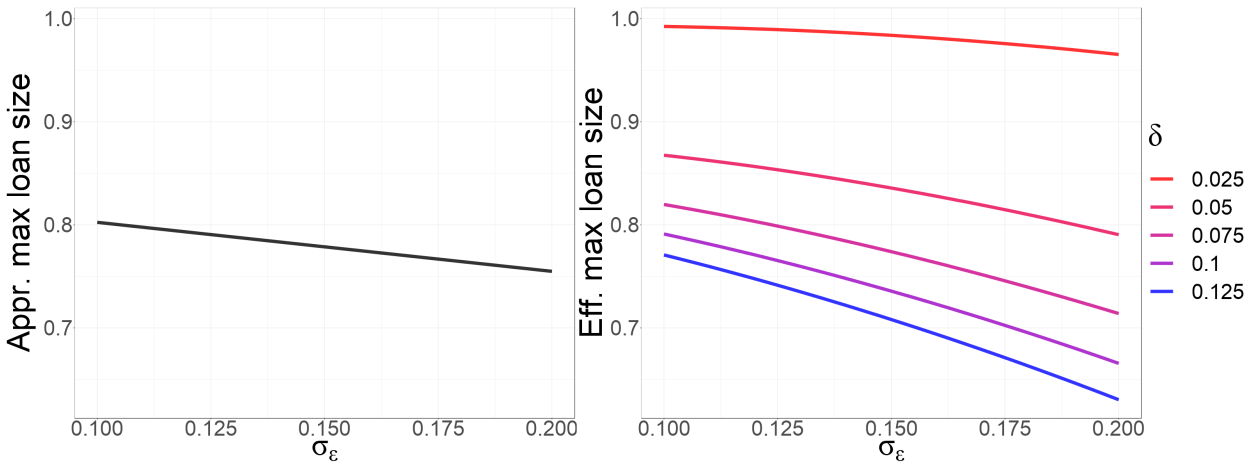


Note: This figure shows heterogeneous effect of price dispersion by FICO score. We estimate the following specification:

$$LTP_{ikt} = \alpha + \beta rate_{ikt} + \gamma ZipDispersion_{ikt} \times CreditScore_{ikt} + X_{ikt}\Gamma + \mu_{kt} + \nu_m + \epsilon_{ikt}$$

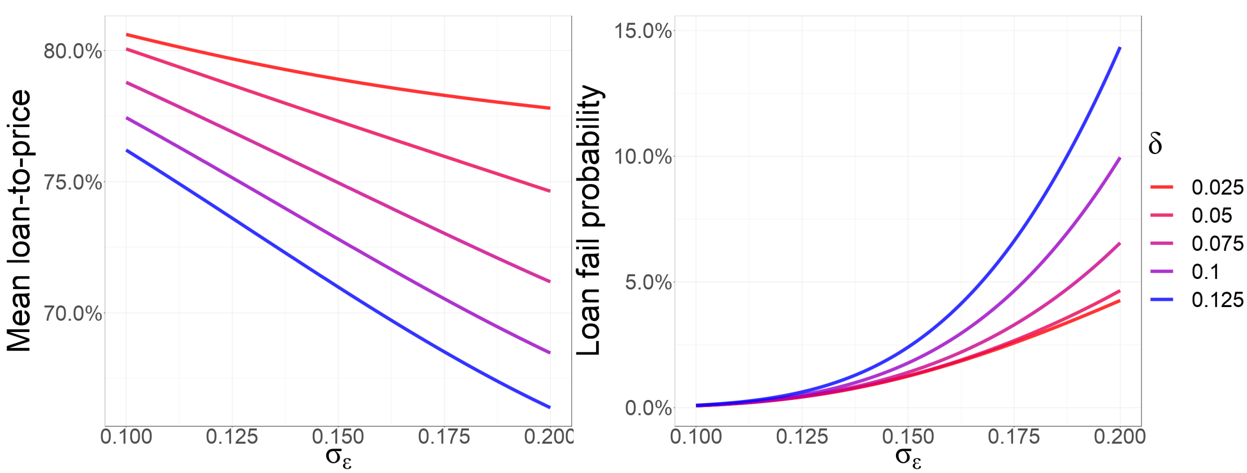
where $ZipDispersion_{ikt} \times CreditScore_{ikt}$ is zip code price dispersion interacted with home buyer's credit score, which is divided into five groups based on lenders' common practice: Excellent (800-850), Very Good (740-799), Good (670-739), Fair (580-669), and Poor (300-579). X_{ikt} includes zip code price dispersion, credit score, and other controls in Table 4. We plot γ estimated using the securitized loan sample and the portfolio loan sample, respectively. Blue nodes represent securitized loans. Red nodes represent portfolio loans. The bars indicates 95% confidence intervals. The sample includes loan level observations of conventional loans from 2000 to 2020. *Source*: Corelogic LLMA and Deeds.

Figure A5. Behavior of $\bar{L}_{appr}$ and $\bar{L}_{fair}$



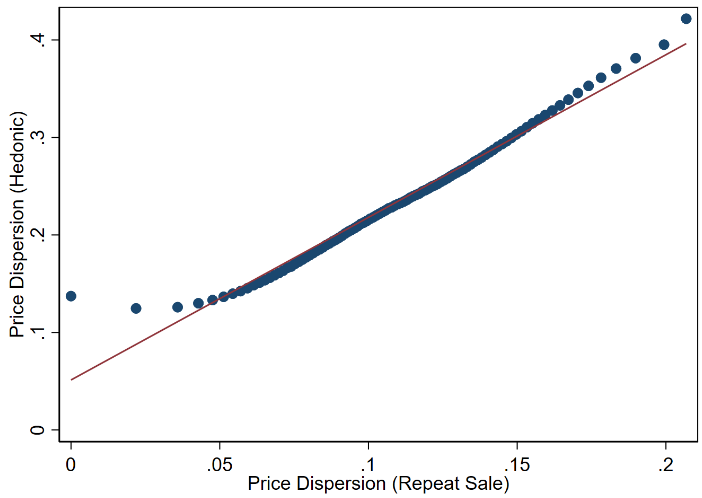
Note: In the above figure, the left panel shows the behavior of the average value of $\bar{L}_{appr}$ for successful loans (which does not depend on δ), and the right panel shows the average value of $\bar{L}_{fair}$, as σ_ϵ varies, for different values of δ. Throughout, we set $\phi = 0.85$, $c = 0.2$, $r - \rho = 0.005$.

Figure A6. LTP and fail probabilities



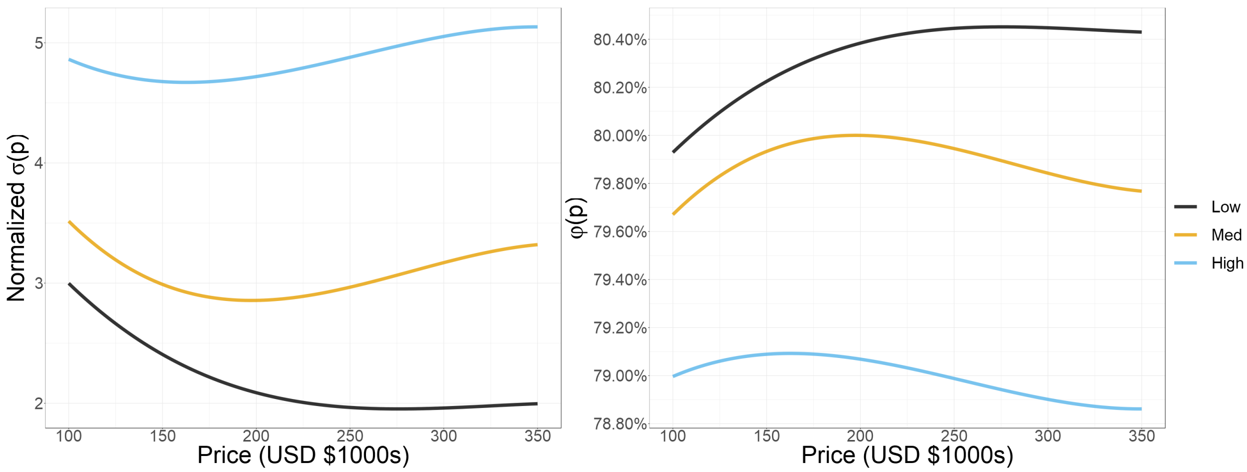
Note: In the above figure, the left panel shows the mean loan-to-price ratio. The right panel shows the probability of loans failing. Colored lines represent different values of δ. Throughout, we set $\phi = 0.85$, $c = 0.2$, $r - \rho = 0.005$.

Figure A7. Repeat-Sales Estimates and Hedonic Estimates



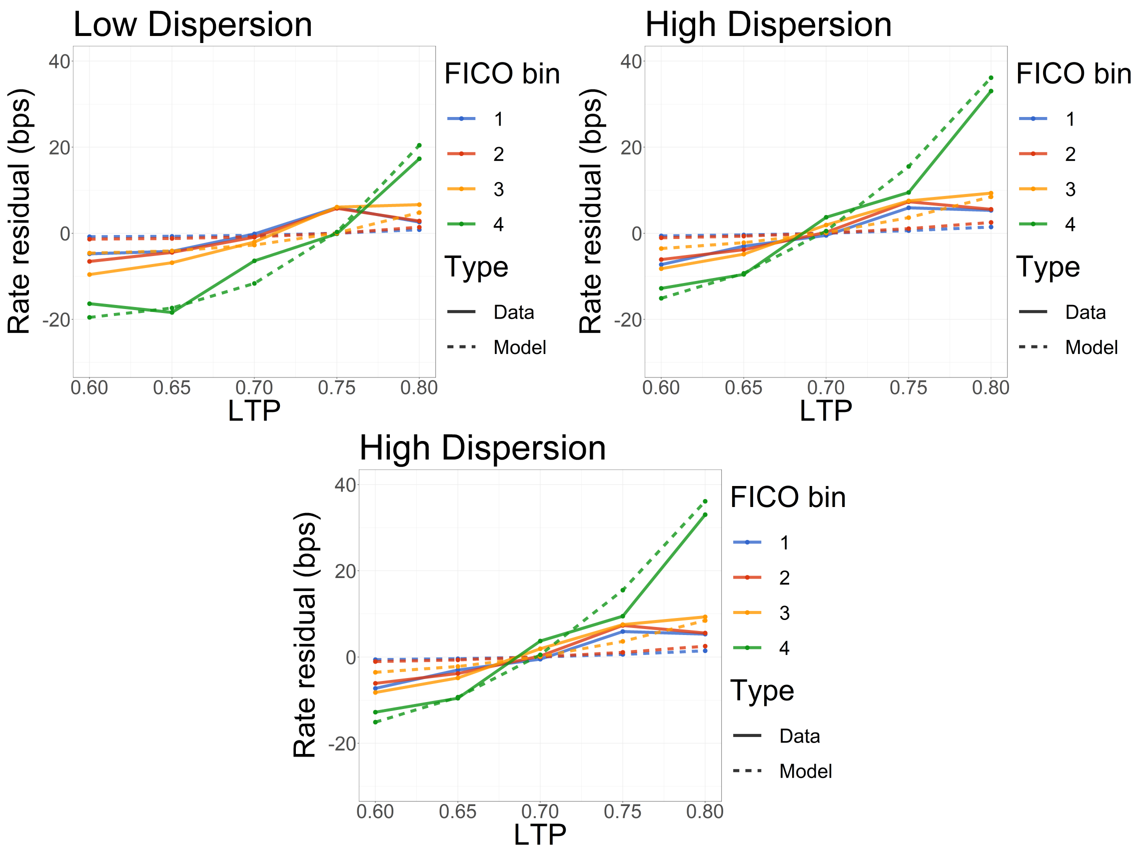
Note: This figure compares the repeat-sale estimates and the hedonic estimates by making the binned scatterplot. The x-axis is the repeat-sale estimates, and the y-axis is the hedonic estimates used in the main analysis. The sample includes property-level observations from 2000 to 2020. *Source*: Corelogic Deeds.

Figure A8. $\sigma(p)$ and $\phi(p)$ functions



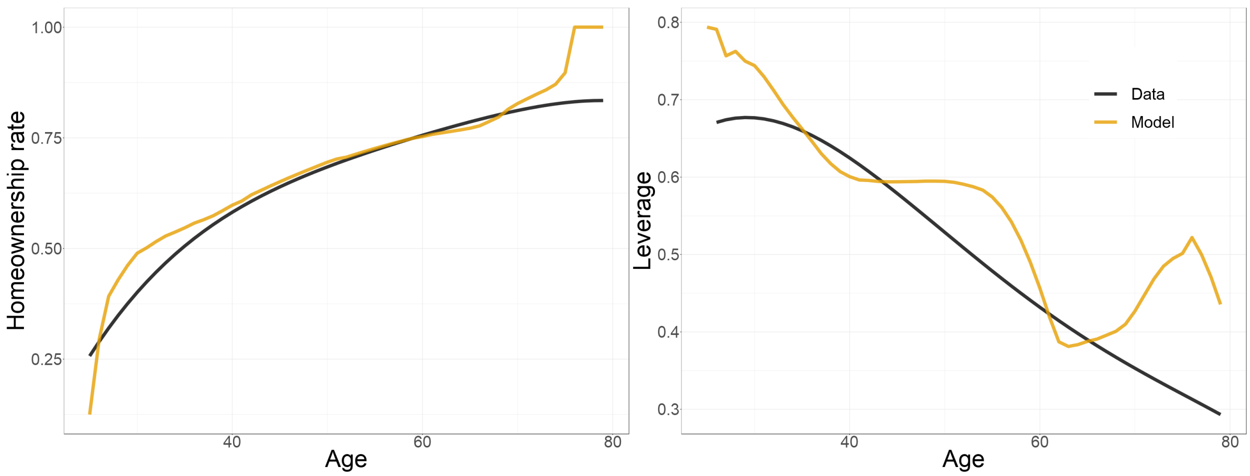
Note: The left panel shows $\sigma(p)$, the average of price dispersion σ conditional on house prices, for the low, medium, and high dispersion versions of our calibration. We normalize $\sigma(p)$ by its standard deviation across houses, the same units used in Table A3. The right panel shows the resultant $\phi(p)$ functions which we use for the three versions of our calibration. The y-axis shows LTVs available at each house price.

Figure A9. Model Fit - Rate Menu



Note: This figure shows our model fit of the rate menu. The top two panel shows empirical interest rate residuals $\tilde{r}_{fld}$ (solid lines), from (A27), and model-predicted rate residuals $\tilde{r}^{model}_{fld}(c)$ (dashed lines), from (A29), in the fitted model. LTP ratios are shown on the x-axis, and different FICO buckets are shown as different colors. The top left plot shows results for low-dispersion areas, and the top right plot shows results for high-dispersion areas. The bottom plot shows the differences $r_{lf,d=H} - r_{lf,d=L}$ in the data (solid) and in the model (dashed). In other words, each line in the bottom panel is the difference between the corresponding line in the top right panel (the high-dispersion menu) and the line in the top left panel (the low dispersion menu).

Figure A10. Model Fit



Note: The left plot shows homeownership rates in the model and in the data. The right plot shows debt-to-assets in the model and in the data. The data is from the 2016 SCF. For both SCF data series, we smooth the input series by projecting values on a fourth-degree polynomial in age and taking the predicted values.

Table A1: IV Relevance Condition

This table presents the relevance condition of our instruments. The outcome variable is house price dispersion, scaled by its standard deviation. The explanatory variables are the five instruments, introduced in Section 3.2. The sample includes property transaction level observations from 2000 to 2020. Standard errors are clustered at county level. ***, **, * represent 1%, 5%, and 10% significance, respectively.

	Price Dispersion		
	(1)	(2)	(3)
IV: Geo-coordinates	0.0419***	0.0579***	0.0553***
	(0.011)	(0.010)	(0.010)
IV: Square Footage	0.1809***	0.1799***	0.1758***
	(0.007)	(0.007)	(0.007)
IV: Number of Bedrooms	0.0341***	0.0554***	0.0557***
	(0.005)	(0.004)	(0.004)
IV: Number of Bathrooms	0.0485***	0.0495***	0.0480***
	(0.007)	(0.005)	(0.004)
IV: Building Age	0.2043***	0.1959***	0.1935***
	(0.014)	(0.013)	(0.013)
Transaction Date FE	✓	✓	✓
County-Year FE		✓	✓
Lender-Year FE			✓
R2	0.1175	0.3068	0.3199
Observations	28M	28M	28M

Table A2: IV Balance Test

This table presents the balance test results. In Panel A, the outcome variable in columns 1 is price dispersion. The outcome variables in columns 2-4 are the predicted price dispersion in the first stage as reported in Table A1. The outcome variable in column 2 corresponds to column 1 in Table A1, in column 3 corresponds to column 2 in Table A1, and in column 4 corresponds to column 3 in Table A1. The explanatory variables are borrower and property characteristics. In Panel B, the outcome variables in columns 1-4 are FICO score, in columns 5-8 are median income, in columns 9-12 are household age, and in columns 13-16 are minority population share. In both panels, the underlying sample contains zip-code level observations. Standard errors are clustered at county level. ***, **, * represent 1%, 5%, and 10% significance, respectively.

Panel A

	Price Dispersion	Predicted Price Dispersion		
	(1)	(2)	(3)	(4)
FICO	-0.01**	0.00	0.00	0.00
	(0.006)	(0.001)	(0.001)	(0.001)
Population Median Age	0.15***	0.00	0.00	0.00
	(0.012)	(0.003)	(0.003)	(0.003)
Median Income	-0.19***	-0.01	-0.01	-0.01
	(0.018)	(0.004)	(0.005)	(0.005)
Minority Population Share	0.05	-0.00	-0.00	-0.00
	(0.029)	(0.006)	(0.007)	(0.007)
House Characteristics Controls	Ln(Square Footage), Building Age, House Price per Square Footage			
County-Year FE	✓	✓	✓	✓
R2	0.51	0.46	0.43	0.43
Observation	186,164	186,164	186,164	186,164

Panel B

	FICO				Income			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Price Dispersion	-0.02***				-0.17***			
	(0.009)				(0.013)			
Predicted Dispersion		0.04	0.04	0.04		-0.05	-0.05	-0.05
		(0.023)	(0.022)	(0.022)		(0.047)	(0.044)	(0.045)
House Characteristics Controls	Ln(Square Footage), Building Age, House Price per Square Footage							
County-Year FE	✓	✓	✓	✓	✓	✓	✓	✓
R2	0.43	0.43	0.43	0.43	0.66	0.65	0.65	0.65
Observation	186343	186343	186343	186343	186167	186167	186167	186167

	Median Age				Minority Population Share			
	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Price Dispersion	0.16***				0.04			
	(0.027)				(0.028)			
Predicted Dispersion		0.03	0.10	0.09		-0.02	-0.05	-0.05
		(0.064)	(0.067)	(0.068)		(0.064)	(0.066)	(0.067)
House Characteristics Controls	Ln(Square Footage), Building Age, House Price per Square Footage							
County-Year FE	✓	✓	✓	✓	✓	✓	✓	✓
R2	0.43	0.42	0.42	0.42	0.67	0.67	0.67	0.67
Observation	186250	186250	186250	186250	186277	186277	186277	186277

Table A3: Heterogeneous Effect by FICO

This table presents heterogeneous effects of price dispersion on LTPs by FICO scores. Columns 1-3 present OLS results. Columns 4-6 present 2SLS results. In all columns, the outcome variable is the loan to price ratio. The explanatory variable of interest is the interaction between zip-code house price dispersion, scaled by its standard deviation, and FICO score buckets. The omitted benchmark credit score bucket is Excellent, including FICO score of 800 or above. Borrower/Loan controls include zip price dispersion, FICO score, FICO-squared, mortgage interest rate, and loan type. Columns 1 and 4 use the full sample. Columns 2 and 5 use securitized conventional loans. Columns 3 and 6 use portfolio conventional loans. The sample includes loan level observations from 2000 to 2020. Standard errors are clustered at county level. ***, **, * represent 1%, 5%, and 10% significance, respectively.

	OLS			2SLS		
	(1) Full	(2) Securitized	(3) Portfolio	(4) Full	(5) Securitized	(6) Portfolio
Zip Price Dispersion	0.06	0.02	0.18*	0.45**	0.34	0.84***
	(0.081)	(0.083)	(0.107)	(0.189)	(0.219)	(0.269)
Baseline: Excellent FICO						
Zip Price Dispersion × Very Good	-0.41***	-0.42***	-0.33***	-0.61***	-0.46**	-1.02***
	(0.040)	(0.045)	(0.063)	(0.167)	(0.193)	(0.244)
Zip Price Dispersion × Good	-0.70***	-0.63***	-0.65***	-1.02***	-0.85***	-1.97***
	(0.054)	(0.063)	(0.080)	(0.202)	(0.260)	(0.299)
Zip Price Dispersion × Fair	-0.86***	-0.51***	-0.89***	-1.54***	-1.25***	-1.99***
	(0.069)	(0.082)	(0.107)	(0.237)	(0.307)	(0.321)
Zip Price Dispersion × Poor	-1.05***	-0.63***	-1.67***	-2.23***	-1.44***	-2.83***
	(0.108)	(0.171)	(0.171)	(0.403)	(0.554)	(0.496)
Origination Month FE	✓	✓	✓	✓	✓	✓
County-Year FE	✓	✓	✓	✓	✓	✓
Borrower/Loan Controls	✓	✓	✓	✓	✓	✓
R2	0.40	0.27	0.30	0.32	0.19	0.18
Observations	6M	28M	1.3M	5M	2.3M	1.1M
Underidentification test statistic				140.04	145.01	82.24
Underidentification test p-value				0.00	0.00	0.00
Weak identification test statistic				91.78	59.46	81.15

Table A4: Price Dispersion and Appraisals

This table presents evidence showing that price dispersion is associated with the magnitude of under-appraisals. Columns 1-3 report OLS results. Columns 4-6 report 2SLS results. The outcome variable in columns 1 and 4 is the appraisal deviation $ApprDev_i$, which is the product of the percentage deviation of appraisal prices to sale prices with an under-appraisal dummy, defined in (A47). The outcome variable in columns 2 and 5 is a dummy for appraisals being below transaction prices. The outcome variable in columns 3 and 6 is the percentage difference between appraisal prices and sale prices, conditional on under-appraisal. The explanatory variable is zip code price dispersion scaled by its sample standard deviation. Borrowers and loan controls include mortgage rate, log house price, FICO, FICO-squared, DTI, DTI-squared, LTV, LTV-squared, GSE indicator, and loan type. The sample includes all loans originated from 2000 to 2020. Standard errors are clustered at county level. ***, **, * represent 1%, 5%, and 10% significance, respectively.

	OLS			2SLS		
	(1) Appraisal Deviation	(2) I(Appraisal <Price)	(3) Conditional Appraisal Deviation	(4) Appraisal Deviation	(5) I(Appraisal <Price)	(6) Conditional Appraisal Deviation
Zip Price Dispersion	0.02***	0.00	0.53***	0.01**	-0.00***	0.53***
	(0.002)	(0.000)	(0.033)	(0.005)	(0.001)	(0.068)
Origination Month FE	✓	✓	✓	✓	✓	✓
County-Year FE		✓		✓		✓
Property & Loan Controls	✓	✓	✓	✓	✓	✓
R2	0.02	0.03	0.16	0.00	0.00	0.02
Observations	5M	5M	0.2M	5M	5M	0.2M
Underidentification test statistic				92.24	92.26	60.65
Underidentification test p-value				0.00	0.00	0.00
Weak identification test statistic				15971.93	15974.44	1350.66

Table A5: Property-Level House Price Dispersion and LTP - Repeat Sale

This table presents the results of property-level regressions with repeat sale sigma estimates. The outcome variable is loan-to-sale price ratio. The explanatory variable of interest is property-level house price dispersion estimated using repeat sales, scaled by its standard deviation. Controls include the mortgage rate, transaction price of the property, mortgage type, mortgage term, and resale indicator. The sample includes property transaction level observations from 2000 to 2020. Standard errors are clustered at county level. ***, **, * represent 1%, 5%, and 10% significance, respectively.

	OLS			2SLS		
	(1)	(2)	(3)	(4)	(5)	(6)
Price Dispersion	-0.77***	-0.36***	-0.36***	-1.01***	-1.06***	-0.97***
	(0.090)	(0.038)	(0.036)	(0.106)	(0.055)	(0.052)
Interest Rate	0.83***	0.91***	0.69***	0.84***	0.93***	0.70***
	(0.073)	(0.058)	(0.042)	(0.076)	(0.058)	(0.042)
Log House Price	-3.29***	-3.41***	-3.07***	-3.32***	-3.48***	-3.14***
	(0.112)	(0.119)	(0.119)	(0.116)	(0.113)	(0.111)
Loan Controls	✓	✓	✓	✓	✓	✓
Transaction Date FE	✓	✓	✓	✓	✓	✓
County-Year FE		✓	✓		✓	✓
Lender-Year FE			✓			✓
R2	0.44	0.47	0.54	0.35	0.28	0.24
Observations	3M	3M	3M	3M	3M	
Underidentification test statistic				77.81	84.67	83.45
Underidentification test p-value				0.00	0.00	0.00
Weak identification test statistic				47.31	73.13	72.76

Table A6: Calibration parameters

This table shows parameter values used in our calibration. All price units, such as p^r and p^h, are in USD thousands.

Parameter	Symbol	Value
Discount factor	β	0.96
Intertemporal elasticity of substitution parameter	σ	2
Housing budget share	α	0.4
Bequest parameter	K_B	300
Earning persistence	ρ_ζ	0.91
Standard deviation of earnings shocks	σ_ε	0.21
Income tax rate	τ	0.25
Saving rate	r^B	0.02
Mortgage rate	r^M	0.04
House transaction cost	F^{pur}	0.05
House depreciation rate	δ^h	0.01
Rent price	p^r	12
House price	p^h	192

Table A7: Counterfactual Homeownership Rate

This table presents the counterfactual change of homeownership rate if we reduce the dispersion of the current housing stock. Each row shows the difference in homeownership rates between the high-dispersion and low-dispersion versions of our calibration, for a certain income and age group. High- and low-income households are defined using households' initial income at age 25. High income is defined as above median income households, and low income is defined as below-median-income households.

Age	Total	Low Income	High Income
<30	3.5	2.4	4.6
30-40	0.8	1.6	0
40-50	1.1	2.3	0
50-60	1.4	2.8	0
60-70	1.6	3.2	0
Overall	1.5	2.6	0.5