Table of Contents

Collateral Value Uncertainty and Mortgage Credit Provision

Erica Xuewei Jiang and Anthony Lee Zhang*

Abstract

We document large cross-sectional differences in how effective houses are as collateral for mortgages. Mortgages collateralized by houses with higher price dispersion are more likely to be rejected, receive higher interest rates, and have lower loan-to-price ratios. We build a structural model to illustrate how under-appraisal risk and lenders' concerns about collateral recovery contribute to driving these effects. Our model shows that the impending shift from human to automated appraisals would significantly lower mortgage credit provision, particularly in high-dispersion areas with more low-income and minority households. We propose a simple policy modification to mitigate the negative effects of price dispersion.

*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

An important policy goal of housing regulators in the US is to improve homeownership rates.[1] Regulatory interventions in the market for residential mortgage are among the crucial components of these policies. The median US first-time homebuyer borrows more than 80 percent of their house's value since early 2000; seemingly small shifts in the amount of credit available to homeowners can translate to large differences in housing affordability.[2] A large body of literature has emerged studying the various frictions in credit markets that affect the ability of households to finance their home purchases.[3]

This paper analyzes a new force which affects mortgage credit availability to households: the value uncertainty of the housing stock. Older and less standardized houses have higher price dispersion; that is, their values are harder to predict based on their characteristics. Mortgages backed by these houses are more likely to be rejected, receive worse rate menus, and have lower loan-to-price ratios (LTP). We propose two main mechanisms that drive these effects. First, less standardized houses have lower expected recovery rates. Second, less standardized houses have noisier appraisal values, which makes regulatory constraints on loan-to-value ratios (LTV) linked to appraisal values more likely to bind. We build a structural model to disentangle the effects of each channel on different credit market outcomes. Through counterfactual analyses, we study the aggregate as well as the distributional implications of various ongoing policy debates.

We begin by documenting a number of stylized facts about house price dispersion. Using rich residential property transaction data from 2000 to 2020, we show that there is substantial cross-sectional variation in the predictability of house prices. Older houses, and less standardized houses – in the sense that they are much smaller or larger than the average house in a given area – have higher price dispersion, as measured by the average prediction errors from a hedonic model of house prices. Price dispersion is highly persistent over time at the zipcode level, suggesting that differences in price dispersion across zipcodes are driven by

[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.

[2] Source: CFPB report, Market Snapshot: First-time Homebuyers.

[3] See, for example, Glaeser and Shapiro (2003); Hurst et al. (2016); Agarwal et al. (2017); Adelino et al. (2020); DeFusco (2018); DeFusco and Mondragon (2020).

persistent differences in the characteristics of local housing stocks, rather than time-varying market conditions.

Price dispersion also appears to affect the house appraisals process. When a homebuyer purchases a house with a mortgage, regulators require the house to be appraised, and the value used for mortgage LTV calculation is the lower of the transaction price and the appraisal price. These appraisal values are calculated based on the transaction prices of recently sold similar houses. Thus, when overall price dispersion is higher, appraisal values should be more dispersed around transaction prices; we confirm this empirically. Finally, price dispersion tends to be higher in low-income and Black-dominant zipcodes, suggesting that households whose homeownership decisions presumably are more sensitive to credit access in fact tend to live in areas with houses that are more difficult to finance.

Since high-dispersion houses have more disperse appraisal values, these houses are more likely to under-appraise, and under-appraisal is also more costly. This should increase the rejection rates of mortgages backed by high-dispersion houses and lower the observed LTP ratios. Moreover, when house prices are more dispersed, lenders have higher expected costs if they foreclose on houses, so they should offer a worse interest rate for any given LTP.

We find empirical support for all of these hypotheses: mortgages backed by high-dispersion houses are more likely to be rejected, receive worse rate menus, and have lower LTP ratios. These results hold in the cross-section of house transactions, and the patterns remain when comparing houses transacted in the same county-year, at the same price, financed by the same lender. The effects are quantitatively large. Mortgages backed by houses in zipcodes with one standard deviation higher house price dispersion are 25% more likely to be rejected for collateral reasons and have roughly 50bps lower LTPs. The mortgage rate for these houses is about 2bps higher, conditional on borrower creditworthiness and mortgage features.

We conduct a variety of robustness checks to argue that the results we find are indeed driven by price dispersion and its effects on mortgage credit, rather than other confounding factors. Lenders could in principle lend less against high-dispersion houses, not because the *houses* are worse, but because *buyers* of high-dispersion houses are systematically worse credit risks. In the cross-section, we find that high-dispersion areas indeed have buyers with lower incomes and FICO scores. To address this concern, we construct instruments for price dispersion, based on the heterogeneity of houses relative to their local housing stock.

Intuitively, when a zipcode has very nonstandardized houses, the market for any given house will tend to be thin, and price dispersion will tend to be high. Our instrument is correlated with price dispersion, and is not correlated with ex-ante buyer creditworthiness measures, such as FICO and income. Moreover, buyers with high predicted price dispersion are not ex-post more likely to default on mortgages. These results thus suggest that our findings are indeed driven by the fact that houses are worse collateral, rather than buyers of these houses being worse credit risks.

We then build a structural model to disentangle how two distinct mechanisms drive the effects of price dispersion on mortgage market outcomes. The first is a classic force involving *collateral recovery*: lenders face higher losses upon default in high-dispersion areas, causing them to rationally offer worse mortgage LTV-rate menus to homebuyers. The second force, unique to the housing market, is an *appraisal risk* channel: under-appraisals require households to make higher down payments, so households may rationally choose to borrow less to limit the risk of under-appraisal.

In the model, competitive lenders offer menus of interest rate-LTP pairs to a borrower, such that lenders break even, given the exogenous risk of default and expected recovery rates from the house upon foreclosure. The borrower chooses a target loan size from the menu. The house then undergoes an appraisal. We model appraisals as noisy, upward-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 as planned. If the appraisal is sufficiently low that the mortgage would violate regulatory LTV constraints, 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, which we interpret as a mortgage rejection. The tradeoff homebuyers face is that larger mortgages improve consumption smoothing, but increase the risk of under-appraisals. When price dispersion is higher, lenders offer worse menus, and under-appraisal risk is larger, leading to more mortgage rejections, higher interest rates, and lower LTPs.

We then calibrate the model to data, matching moments on how rate menus, appraisal distribution, and mortgage rejection rate depend on house price dispersion. The calibrated model can quantitatively rationalize the patterns we see in the data: in addition to our targeted moments, the model produces a dispersion-LTP relationship of similar magnitude to what we observe empirically. We use the model to evaluate how the collateral recovery and appraisal risk channels contribute to driving each of the three mortgage outcomes we observe. 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 while holding the distribution of appraisals fixed, buyers respond by picking similarly sized mortgages and receiving a similar rejection rate, but bearing higher interest rates. On the other hand, if we assume that appraisals become riskier while holding lenders' rate menus fixed, rejection rates increase and LTPs decrease, but mortgage rates are mostly unaffected.

Our calibration suggests that homebuyers care more about loan size, implied by their inelastic loan size choices to changes in interest rates. Thus, policies targeting under-appraisal risk could be particularly effective because such risk is the main driver of the extensive margin effect. Using our model, we analyze how changes in the regulatory treatment of appraisals would influence mortgage credit provision and the resulting distributional consequences.

We conduct a counterfactual evaluating the effect of a shift to automated appraisals, assuming that computers would generate fully fair appraisals of home values, removing the tendency of human appraisers to bias appraisals upwards.[4] In this counterfactual, the distribution of appraisals would shift downward, increasing under-appraisal risk overall and disproportionately so in areas with high house price dispersion. We find that shifting to automated appraisals without compensating in some way for human appraisers' upwards biases has the potential to significantly lower mortgage credit provision. Automated appraisals would increase overall mortgage rejection rates by around 10pp and decrease overall mortgage LTPs around 2pp.

We then propose an alternative policy which could alleviate the effects of the appraisal channel: when a house under-appraises, regulators could use a weighted average of the transaction price and the appraisal price for LTV calculation. This policy would give regulators a continuous way to modulate the influence of appraisals: higher weights on transaction prices would make under-appraisals less costly. We show that this policy can substantially increase

[4] In 2021, the FHFA announced that banks and mortgage lenders could use automated appraisal software in the place of human appraisals.

mortgage credit provision disproportionately in low-income zipcodes where price dispersion is highest.

Our results imply that the degree of value uncertainty of the housing stock is a previously overlooked variable, which has significant effects on mortgage credit provision in the US housing market. This effect is not a form of discrimination by lenders or an externality which can be addressed through Pigouvian taxation. In fully competitive mortgage markets, lenders would still lend less against high-dispersion houses, since they have higher foreclosure costs for these houses, and regulatory appraisal constraints are more likely to bind. Our results provide a rationale for interventions, such as the FHA loan insurance program, which extends credit to low-income households and first-time homebuyers at loan-to-value ratios much higher than private lenders. We also discuss the 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 shows that 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 leads to uncertainty in house appraisals, thereby reducing mortgage lending.

Our paper also relates to the housing literature. We build on the 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 the commercial real estate. Hartman-Glaser and Mann (2017) documents that lower-income zipcodes have more volatile returns to housing than higher-income zipcodes. They rationalize the finding with a model in which 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 the classic literature analyzing how collateral values affect the properties of debt contracts collateralized by these assets (Titman and Wessels, 1988; Shleifer and Vishny, 1992). Research 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 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, we illuminate 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 the implications of our results in Section 6, and conclude in Section 7.

2 Measurement, Data, and Stylized Facts

2.1 Measuring Value Uncertainty

Houses trade in thin markets; houses are differentiated, buyers are heterogeneous, and sellers only list houses when they face moving shocks. Thus, there is a relatively small set of buyers for any house at any point in time; sale prices may be higher or lower depending on whether there happens to be a high-valued buyer when the house is listed. We empirically estimate house price dispersion at the level of individual house sale by measuring what kinds of houses tend to have smaller errors when priced with a hedonic regression.[5] We first 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, $\hat{\epsilon}_{it}^2$, from (1) on a flexible function of characteristics and time to predict which house characteristics make them difficult to price:

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

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 house characteristics to affect prices in a manner that varies over time. $g_k\left(x_i, t\right)$ allows the variance of price dispersion to vary with characteristics and over time. η_{kt} is a county-month fixed effect. Intuitively, specification (1) estimates a hedonic specification for house prices, and specification (2) projects the squared residuals ϵ_{it}^2 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

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

house-level measure of idiosyncratic price dispersion:[6]

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

Our main specification uses a hedonic model of house prices; one concern is that there are house-level features which affect prices, which are observed by market participants but not in our dataset. To alleviate this concern, in Appendix C.1, we repeat the analysis using a repeat-sales specification to predict prices in (1). This specification absorbs all time-invariant components of house quality, whether or not they correspond to observable characteristics in our data, into house fixed effects. The resultant estimates of price dispersion are closely correlated with those in our baseline specification, and all of our empirical results continue to hold in this specification.

Note that, when considering whether to lend against a house, lenders should care about the *total* volatility of a house, not only the idiosyncratic component. Our measurement strategy focuses on the idiosyncratic component, which Piazzesi and Schneider (2016) estimate to be approximately half of total price volatility for an individual house. Much of our empirical analysis will compare houses within a given region-year; these houses should have similar exposure to the local index volatility, so most differences in total volatility should be driven by differences in the idiosyncratic component.

2.2 Data

Corelogic Deed & Tax Data: We obtain house transaction records in the entire US from 2000 to 2020 from the Corelogic Deed dataset. The data set reports each house transaction attached to a specific property and provides information on the sale amount, mortgage amount, transaction date, and property location. We merge the transaction records with the Corelogic Tax records, which contain property characteristics such as year built and square footage. We estimate price dispersion for each house in this merged data set. Appendix A.1

[6] 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 the LTV. 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.

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 and monthly loan performance after origination.Importantly for our analysis, the LLMA provides both transaction price and the house's appraisal value. We use this data set to estimate the menu of LTP-interest pairs in any given market, to examine loan performance, and to analyze appraisal values relative to prices.

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.

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

Table 1 provides summary statistics.

2.3 Stylized Facts

Price Dispersion is Persistent Over Time Figure 1 Panel (a) plots zipcode idiosyncratic price dispersion in 2020 against zipcode dispersion in 2010. There is large cross-sectional variation in price dispersion across zipcodes, but dispersion is very persistent over time. This suggests that differences in price dispersion are not driven by time-varying factors, such as local housing market conditions; rather, dispersion appears to be driven by the persistent characteristics of the local housing stock.

Price Dispersion and House Characteristics 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, houses which were renovated within 5 years of the transaction date (column 2) have lower price dispersion.[7] 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 zipcodes 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.[8]

Price Dispersion and Appraisal Noise The appraisal values of high-dispersion houses tend to be noisier.[9] Figure 2 illustrates this point, with a slightly different measure of appraisal in each panel. Panel (a) shows the percentage absolute difference between appraisals and transaction prices, $\frac{|a_i - p_i|}{p_i}$. This difference is around 1.5 percentage points larger in high-dispersion areas. Panel (b) analyzes the average percentage size of under-appraisals, conditional on under-appraisal; that is, we take $\frac{|a_i - p_i|}{p_i}$ only for the sample of houses with $a_i < p_i$. This panel shows that under-appraised houses under-appraise by larger amounts – roughly 5pp, compared to 3pp – in high-dispersion areas.

Appraisals only constrain borrowing when appraisals are below the sale price. We define under-appraisal pressure for each loan as $\frac{|a_i - p_i|}{p_i}\mathbf{1}(a_i < p_i)$, the product of the under-appraisal percentage $\frac{|a_i - p_i|}{p_i}$ and an indicator for under-appraisal, $\mathbf{1}(a_i < p_i)$. This is a summary measure of the downwards pressure on loan size induced by appraisals, combining the probability of under-appraisal and the size of under-appraisals. Panel (c) shows that under-appraisal

[7] 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.

[8] This finding is consistent with evidence from other papers: see, for example, Kotova and Zhang (2021) and Andersen et al. (2021). In Appendix B, we discuss a number of factors and theoretical forces that may drive dispersion, such as information asymmetry (Kurlat and Stroebel, 2015; Stroebel, 2016), search frictions, and so on.

[9] Given how appraisals are constructed, appraisal prices cannot be perfectly accurate estimates of house values. Previous literature has shown that most appraisals use roughly 3-7 comparable sales (Agarwal et al., 2020; Eriksen et al., 2020a). We have estimated that an individual house's sale price has an idiosyncratic shock of roughly 26%, relative to predicted prices from a time-varying hedonic model. If we assume that all appraisals are identical to the target house in terms of characteristics, the variance of appraisal prices induced by idiosyncratic price terms in comparable sales will range from $26\%/\sqrt{7}$ to $26\%/\sqrt{3}$, or 9.83% to 15.01%. In practice, comparable houses are not identical to the target house, and prices must be adjusted for characteristic differences, which will introduce additional variance into appraisals. These estimates of predicted appraisal dispersion have similar magnitude to the estimates in the literature on the gap between appraisals and AVM prices; for example, Agarwal et al. (2020) find that appraisal prices have a standard deviation of 13.4% relative to AVM prices.

pressure is higher in high-dispersion areas.

The appraisal distribution is known to be very asymmetric, with a low probability of under-appraisal and substantial bunching at the transaction price. This presumably reflects a combination of selection on successful sales, and appraisers' incentives to bias appraisal prices upwards toward transaction prices. We will account for both effects in our model of Section 4; however, to illustrate that our results are not simply driven by appraiser bias, in panel (d), we analyze the average size of the appraisal gap conditional on *over-appraisal*, $a_i > p_i$. Appraisers should have no incentive to bias appraisals if they are above the transaction price, since increasing a_i past p_i does not change the amount a buyer can borrow. Panel (d) shows that even over-appraisals are more disperse in high-dispersion areas, lending support to the idea that price dispersion fundamentally increases the dispersion of house appraisals.

Regional Variation: Zipcode Income and Race Lastly, price dispersion tends to be higher in low-income and black-dominant zipcodes. Panel (b) of Figure 1 shows the relationship between price dispersion and zipcode demographics.[10] Comparing zipcodes with similar levels of median income, price dispersion in the Black-dominant zipcodes tends to be 0.03 (1/4 SD) higher than in the non-Black dominant zipcodes. Comparing zipcodes with similar racial composition, price dispersion in the low-income zipcodes tends to be 0.06 (1/2 SD) higher than in the high-income zipcodes.

3 Price Dispersion and Mortgage Credit Provision

In this section, we empirically show that collateral value uncertainty affects mortgage credit at three margins: rejection rates, interest rate, and loan size. We first show a stylized diagram of the home purchase and mortgage application process in Figure 3 to illustrate how price dispersion can affect mortgage credit.

A homebuyer first decides to purchase a home. After the homebuyer's offer is accepted,

[10] To obtain the values, we regress zipcode price dispersion on a dummy that indicates whether the zipcode is Black-dominant, i.e., at least 50% of population is Black, and a dummy that indicates whether the zipcode median income is below median among all zipcodes in a given year. The figure plots the coefficients on Black-dominant dummy and low-income dummy.

she applies for a mortgage. The lender offers a menu of interest rate-LTV pairs to the borrower. When houses have higher price dispersion, lenders should offer worse rate menus – higher interest rates for any given LTP, and vice versa – since their expected recovery rates are lower. The buyer then chooses an option from the menu: in high-dispersion areas, buyers are thus forced to choose either lower LTVs, higher interest rates, or both. We call this effect of price dispersion the "collateral recovery" channel.

After the buyer chooses her targeted mortgage LTV, the house is appraised, and the value of the house for LTV calculation is set to the smaller of the transaction price and the appraisal price. If the house over-appraises – that is, the appraisal price is at least the transaction price – the transaction proceeds as planned. If the house under-appraises, the homebuyer may need to decrease her mortgage size to meet regulatory LTV constraints and thus make higher down payments. If the buyer is unable to make the higher down payments, she may have to renege on the transaction. We showed in Figure 2 that appraisals are noisier when house price dispersion is high; thus, under-appraisal risk will be higher in high-dispersion areas. Thus, mortgages in these areas are more likely to be rejected, and buyers may decrease their targeted LTVs to lower the impact of under-appraisal risk. We call this effect the "appraisal risk" channel.

This figure thus illustrates that, when price dispersion is higher, mortgages should be more likely to be rejected, loan-to-price ratios should be lower, and mortgage interest rates should be higher. In this section, we show that these predictions hold in the data in various different empirical specifications. In Section 4, we build a structural model based on Figure 3 to quantify the effects of each channel on mortgage outcomes and to study the effects of various counterfactual policies.

3.1 Effects on Mortgage Credit

3.1.1 Mortgage Rejection Rates

Figure 4(a) shows that mortgage applications are more likely to be rejected in counties with higher price dispersion. The HMDA data also allows lenders to report the reason they reject a given mortgage, which allows us to conduct a finer test: Figure 4(b) shows that the fraction of mortgages rejected, in which the lenders indicate the mortgages are rejected for collateral-related reasons, is higher in high-dispersion counties.

We then exploit within county-year variation by estimating the following loan application-level specification:

$$Reject_{ikt} = \beta ZipDispersion_{ikt} + X_{ikt}\Gamma + \mu_{kt} + \nu_{lt} + \epsilon_{ikt} \quad (4)$$

$Reject_{ikt}$ is an indicator that equals 100 if the mortgage collateralized by property i in county k transaction in year t is rejected and 0 otherwise. $ZipDispersion_{ikt}$ is the average price dispersion of houses in property i's zipcode that are transacted in year t.[11] X_{ikt} is a set of controls, including zipcode house transaction price, credit score and its squared term, individual income, loan-to-income ratio and its squared term, and mortgage type. μ_{kt} and ν_{lt} are county-year and lender-year fixed effects, respectively.

Panel A of Table 3 reports the results. We first confirm the effect of local house price dispersion on mortgage rejection using the full sample (column 1). Zipcode 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 for every standard deviation increase in house price dispersion. The effect is economically significant; given the sample average rejection rate of about 16%, the estimate amounts to about 10% increase in rejection likelihood.

In Panel B of Table 3, we focus on collateral-related rejections. A mortgage application is about 50bps more likely to be rejected due to collateral reasons in a zipcode with one standard deviation higher house price dispersion, which is about 25% increase in the likelihood of collateral-related rejections. 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 a robustness check, we examine the relationship between house price dispersion and different rejection reasons among rejected loans. We restrict the sample to only rejected loans and estimate Specification 4 using various rejection reason indicators as the outcome variables. Intuitively, this specification estimates, conditional on

[11] We aggregate property-level price dispersion measures estimated using Corelogic Deeds to zipcode level and assign it to every loan application in HMDA based on borrowers' location recorded by lenders of the mortgage.

a mortgage being rejected, whether rejections are more likely to be attributed to collateral-related reasons in high-dispersion areas.

Table 4 reports the results. 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 house price dispersion increases, the results show that the mortgage rejections are significantly more likely due to collateral reasons, and less likely to be due to creditworthiness reasons, thereby supporting our baseline findings.

3.1.2 Interest Rates

In high price dispersion areas, lenders offer worse rate menus, so mortgage interest rates are higher for any given LTP ratio, and LTPs are lower for any given interest rate. To visually demonstrate this, we estimate the entire menu of LTP-interest rate pairs available to borrowers in high- and low-dispersion areas. We first residualize interest rates using borrowers' credit scores, loan type, and time fixed effects; in Figure 5, we plot the residuals against LTP separately for zipcodes with above-median and below-median dispersion. Figure 5 shows that the entire menu of interest rate-LTP pairs shifts upwards in high-dispersion zipcodes: for any given LTP, borrowers in high-dispersion zipcodes 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 5 presents the above results in a regression setting by estimating Specification 4 using interest rates as the outcome variable, including LTP as a control.[12] Column 1 uses the full sample. Higher loan-to-price ratios are associated with higher interest rates: a one percentage point increase in LTP is associated with an 80bp increase in interest rate. Controlling for LTP, houses in zipcodes with higher house price dispersion have higher interest rates. The mortgage rate increases by 1.1bps in zipcodes with one standard deviation higher average house price dispersion. Columns 2 to 3 show the results for securitized loans and portfolio loans; the results hold in all samples. For every 1SD increase in zipcode average house price dispersion, the mortgage rate of securitized loans increases by 1.38bps, and the

[12] We use zipcode 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.

rates on portfolio loans increases by 1.98bps.

3.1.3 Loan-to-Price Ratio

Lastly, we show that price dispersion is associated with smaller loan sizes, as measured by loan-to-price ratios (LTP). Figure 4(c) illustrates the relationship by plotting county average LTP against average house price dispersion.[13] Counties with higher price dispersion have lower average LTPs. The pattern holds for all types of loans: GSE loans, FHA loans, and jumbo loans (Figure A3).

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 profiles and incomes. To implement this strategy, we estimate the following property-level specification:

$$LTP_{ikt} = \beta Dispersion_{ikt} + X_{ikt}\Gamma + \mu_{kt} + \nu_d + \epsilon_{ikt} \tag{5}$$

LTP_{ikt} is the loan-to-price ratio of a mortgage, 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.[14] μ_{kt} and ν_m are county-year and transaction date fixed effects, respectively.

Table 6 presents the results. Column 2 corresponds to Specification 5. Column 1 includes only transaction date fixed effects, and column 3 adds lender-year fixed effects. 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 estimated effects are economically significant: in Appendix E, we calibrate a lifecycle model of homeownership choice and show that price dispersion-induced changes in LTV can substantially decrease aggregate homeownership rates.

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

[14] Our results are not sensitive to the inclusion of transaction price, though we believe including price is the right specification. We discuss this in detail in Appendix C.2.

3.2 Identification

3.2.1 Identification Assumptions of Baseline Results

Mortgage outcomes could be worse for high-dispersion houses, not because the *houses* are worse, but because *borrowers* purchasing these houses are systematically worse. Lenders would then rationally lend less against these houses; however, the channel of effect would be due to borrower quality rather than house quality. One indication that this does not explain our results is that, in Subsection 3.1.1, we showed that lenders in high-dispersion areas systematically indicate that they are rejecting more mortgages because the collateral is low quality. To further alleviate this concern, we develop an instrument for house value dispersion, show that the instrument is not correlated with ex-ante or ex-post buyer credit characteristics, and show that all our baseline results hold using the instrument.

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.[15] For all home purchases 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:

$$Z_i^m = (X_i^m - \overline{X}_c^m)^2, \;\; \forall m \in \{age, sqft, bed, bath, geo\}, \tag{6}$$

[15]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:

$$\begin{aligned}
\text{Stage 1: } Dispersion_{it} &= \alpha + \beta_1 Z_{it}^{age} + \beta_2 Z_{it}^{sqft} + \beta_3 Z_{it}^{bed} + \beta_3 Z_{it}^{bath} + \beta_4 Z_{it}^{geo} \\
&\quad + X_{ikt}\Gamma + \mu_{kt} + \nu_m + \epsilon_{ikt} \\
\text{Stage 2: } Y_{ikt} &= \alpha + \beta \widehat{Dispersion}_{ikt} + X_{ikt}\Gamma + \mu_{kt} + \nu_d + \epsilon_{ikt},
\end{aligned} \tag{7}$$

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

The instruments 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 zipcodes, we take geographical averages of Z_i^m. The aggregated instrument is essentially a measure of the heterogeneity of the zipcode's housing stock along characteristic m. Zipcodes with more heterogeneous housing stocks tend to have higher price dispersion, since they have thinner local markets for any individual house.

In order for the instruments to be valid, we must argue that they are relevant in the sense that they are associated with house price dispersion and excluded in the sense that they are not associated with other factors which may affect mortgage market outcomes, in particular, buyers' observable and unobservable creditworthiness. For relevance, consistent with the literature, Table A2 shows that our instruments are correlated with the raw price dispersion measure in a statistically significant manner.

For exclusion, we show that the instrument is uncorrelated with ex-ante measures of buyer creditworthiness and also that borrowers in areas with higher values of the instrument are not ex-post more likely to default. We first conduct a balance test on ex-ante characteristics in Table A3. 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. The results imply that sorting is potentially a concern with our raw price dispersion measure: households with lower credit score and lower income are more likely to live in high-dispersion zipcodes. Columns 2-4 then report the correlations of predicted price dispersion, using our instrument, with these buyer characteristics. All coefficients are insignificant and economically very small. Zipcodes with high values of our instrument – that is, more heterogeneous housing stocks – do not systematically attract less creditworthy or low-income households. Panel B conducts a similar analysis using creditworthiness measures as the Y-variable and raw or instrumented price dispersion as the X-variable. 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.

While these results suggest that the instrument is not associated with ex-ante *observable* measures of borrowers' creditworthiness, a further concern is that borrowers in heterogeneous zipcodes may be *unobservably* worse credit risks. To address this concern, we can further do an *ex post* test, measuring whether borrowers in more heterogeneous zipcodes are more likely to default on mortgages. Table 7 Panel A estimates specifications 5 (columns 1-3) and 7 (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 restrict the sample to securitized loans. Columns 3 and 6 restrict the sample to portfolio loans. All regressions include the full set of borrower and loan characteristics as in our main regression specifications.

In these specifications, interest rates, FICO scores, DTI, and LTV are generally associated with ex-post default rates in the expected directions. Consistent with the results from Table A3, in the OLS results, price dispersion is positively associated with default rates: buyers in high-dispersion zipcodes are more likely to default, even after controlling for observable mortgage and buyer features. However, instrumented price dispersion is not associated with default rates: buyers in high heterogeneity zipcodes are not more likely to default on their mortgages. This lends support for our exclusion restriction, that zipcode heterogeneity shifts house price dispersion without shifting buyer creditworthiness.

3.2.3 IV Results

We then estimate specification 7 for every credit outcomes in our baseline analyses in Section 3.1. We confirm our baseline results qualitatively and get reasonably stronger estimated effects.

Approvals. Table 3 reports the loan approval likelihood results. Zipcode 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 zipcode 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.

Interest Rates. Table 5 columns 4-6 present the interest rate results. For every one standard deviation increase in zipcode 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).

LTP. Lastly, Table 6 columns 4-6 present the LTP results, where column 5 corresponds to Specification 7, 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.

The IV coefficient estimates are mostly larger than the OLS estimates. This is potentially driven by the fact that the independent variable, price dispersion, is measured imperfectly by our first-stage regression, causing the coefficients in the OLS specifications to be biased toward 0. When instrumental variables alleviate measurement error in the independent variable, they tend to lead to larger coefficient estimates; the pattern that IV estimates tend to be larger than OLS estimates is common in empirical studies across many areas (Pancost and Schaller, 2021).

3.3 Additional Robustness Checks

We conduct two additional robustness checks. In Appendix C.3, we show that the results survive controls for lender-zip-year fixed effects, suggesting that the results are not driven by market power or markups at the lender-zipcode level. In Appendix C.4, we show that the results hold in a subsample of transactions with sale prices below conforming loan limits, suggesting that the results are not driven by homebuyers' incentives to keep prices below conforming loan limits.

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, through the appraisal risk and collateral recovery channels. The model follows the structure of Figure 3. A prospective homebuyer chooses a targeted mortgage size to finance a house at an exogenous transaction price. By choosing a larger mortgage, the buyer smoothens consumption more effectively, but also face higher interest rates and a greater risk of under-appraisal and mortgage rejection. When idiosyncratic price dispersion is higher, lenders offer borrowers 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, by choosing a target loan size L.[16] The choice of L determines the buyer's 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:

$$U\left(c_1, c_2\right) = \frac{c_1^{1-\eta} - 1}{1 - \eta} + \beta^T u_2' c_2 \tag{8}$$

[16] We ignore heterogeneity in houses in the model. If houses in a given zipcode A are on average higher-dispersion than houses in zipcode B, then all homebuyers in A will on average purchase higher-dispersion houses; sorting cannot change the average dispersion within a zipcode. Assuming houses are identical is a reduced-form way to analyze the average outcomes of homebuyers in zipcode A compared to B. We also assume the buyer cannot respond by delaying her house purchase. The main purpose of the model is to analyze how much buyers precautionarily scale down borrowing in response to appraisal risk; appraisals are ordered and performed after buyers have decided on a target mortgage and home to buy, so buyers cannot reduce appraisal risk by delaying home purchases. There is a separate question of how much reductions in LTV induced by price dispersion affect home purchase decisions; we construct a lifecycle home purchase model to analyze this effect in Appendix E.

where u_2' is an exogeneous constant. Hence, the buyer solves a consumption smoothing problem, where utility is concave in the first period, and linear in the second.[17] The buyer receives exogenous labor income W_1 in period 1, and W_2 in period 2.

The mortgage application process has two stages:

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 price P:[18]

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

   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.

[17] Expression (8) can be thought of as a reduced-form of a richer model in which a consumer smoothes consumption between a single period, in which the house purchase is made, and a large number of future periods in which the mortgage is paid down. The term $\beta^T u_2' c_2$ in (8) can be thought of as a linear approximation to the consumer's value function over wealth in future periods after the house purchase. A similar linear approximation to utility in future periods is used in Jansen et al. (2022). The concavity of utility over consumption in the single house purchase period is high relative to the concavity over the value function of wealth in future periods, since there are many future periods to smooth consumption over, so assuming post-purchase utility is linear in consumption is likely a reasonable approximation. 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.

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

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 constrains loan size further whenever $a < \frac{l}{\phi}$. With appraisal a, the required 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, because it is harder to smooth consumption in response to unanticipated shocks.[19]

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:

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

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.

[19] We demonstrate this point quantitatively in Appendix D.5.

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:

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

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

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:

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

where θ_l and θ_σ capture the dependence of the interest rate on loan size and price dispersion respectively. We adopt this reduced-form model of the rate menu in the model for simplicity; however, in Appendix D.3, we construct a more detailed microfoundation of the interest rate menu. We assume competitive profit-maximizing lenders make loans, setting prices and LTPs such that they at least break even, given imperfect collateral recovery rates. When price dispersion is higher, lenders must offer a worse rate menu to break even. In a simple calibration of the model, the observed dependence of the interest rate menu on price dispersion in the data can be matched fairly well, under reasonable assumptions for average foreclosure discounts. Since the dependence of the rate menu on price dispersion is quantitatively consistent with this microfoundation in the data, we proceed with the reduced-form model (15), as a simpler linear approximation to the microfounded model.

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 2a): 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 model appraisals in a way that matches these stylized facts. 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:

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

Expression (16) states that, when A_{raw} is above P, 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 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 the appraisal to $A_{raw} + Pb$. This is still useful to the buyer, since any upwards bias in appraisals still allows the buyer to borrow more. We will estimate b based on the distribution of appraisal-to-sale ratios in our data, as we describe in Subsection 5.1 below.[20]

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*

[20]In Appendix D.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.

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 D.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 $\bar{a}(l)$, and will renege otherwise. The cutoff $\bar{a}(l)$ is the value of the appraisal such that the consumer is just indifferent between 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 D.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. Buyers thus optimally choose a target loan size slightly smaller than they would if the house never under-appraised, limiting consumption smoothing in order to decrease interest rates and under-appraisal risk.

5 How Does Price Dispersion Affect Mortgage Outcomes?

While it is intuitive that higher price dispersion leads 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 margins: LTPs, interest rates, and mortgage failures.

We next calibrate the model to show that the model can quantitatively rationalize the observed relationship between price dispersion and mortgage outcomes. The calibrated model helps decompose the channels through which outcomes are affected by price dispersion, setting the stage for the ensuing policy counterfactual analyses in the next section.

5.1 Calibration

We calibrate several parameters externally in relation to the existing literature. 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.[21] The maximum LTV parameter ϕ is set to 0.8, which is the most common regulatory threshold.

We then estimate the remaining parameters by matching model-implied moments to the moments in the data. Table 8 summarizes the estimates.

Appraisal Distribution In our model, raw appraisal values a_{raw} are distorted only when they are below the transaction price, $a_{raw} < 1$, since appraisers have no incentives to further bias appraisals that are above the transaction price. Thus, the distribution of realized

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

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 expression (20), we calculate $\hat{\sigma}_{a,i}$ for each quantile bucket of σ values.

Interest Rate Menu 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 Column 6 in Panel A of Table 5.

Key Parameters Governing Model Tradeoff There are four important parameters in the model that govern homebuyer's tradeoff: under-appraisal penalty (ψ), cost of transaction failure (ζ), appraisal bias (b), and period-2 utlity (u_2'). We estimate these parameters 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.[22] 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 is the appraisal deviation within each σ-bucket:

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

[22]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.

That is the product of the under-appraisal probability and the expectation of the percentage deviation of appraisal prices to sale prices conditional on under-appraisal. Figure 2 visualizes the second set of moments in the data. Overall, $ApprDev_i$ is strongly related to price dispersion in the data. In the model, we calculate $ApprDev_i$ defined in expression (22), conditional on appraisal values that do not result in transaction failure in the model.[23]

The intuition behind the calibration is that the magnitude of under-appraisal pressure depends on the shape of the appraisal distribution, which is controlled largely by the appraisal bias parameter b. Buyers' preferences then determine whether under-appraisals mostly result in transaction failures or under-appraisals with larger down payments. The level of the failure-underappraisal tradeoff and its relationship with σ are affected by the consumption penalty parameter ψ, the cost of transaction failure ζ, and consumers' utility from period-2 consumption u_2'.

5.2 Results and Model Fit

The estimated cost of transaction failure (ζ) 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 reduction ($\psi - 1$) is approximately 61.4%. We view these as roughly reasonable parameter values. We show in Appendix D.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. The high estimated ζ could be due to our assumption that period-2 utility is linear, which increases the amount that consumption must decrease to decrease the utility a given amount.

Figure 6 evaluates the model fit. Panel (a) shows the estimated appraisal standard deviations $\hat{\sigma}_{a,i}$; in the data, the standard deviation of appraisals is monotonically higher for higher σ buckets, and we feed this directly into the model.[24] Panel (b) shows the CDF of appraisals

[23] 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 6, 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 downward pressure that under-appraisals generate for sale prices compared to the simple under-appraisal probability.

[24] We note two features of the appraisal distribution. First, the implied number of houses entering appraisals is

in the model and the data for the fifth σ-percentile bucket. The model matches 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.[25]

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 that the relationship between both outcomes and σ is slightly stronger in the model and in the data.

Panel (e) shows 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 6 but smaller than the IV estimates.

5.3 Decomposition of Channels

Using our model, we evaluate how each of the two channels contributes to driving variation in loan rejections, LTPs, and interest rates. Figure 7 presents our results. In short, we find that the appraisal risk channel has a larger effect on loan-to-price ratios and rejection rates,

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 in houses than we do; this explains 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.

[25] 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 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.

whereas the collateral recovery channel has a large effect on interest rates.

We evaluate the magnitude of the collateral recovery channel by allowing lenders' rate menus to vary according to σ but shutting down the appraisal risk channel by setting appraisal noise constant across σ-buckets – $A_{raw} \sim N(P, \bar{\sigma})$ for all percentile buckets – and re-calculating mortgage market outcomes. Analogously, to evaluate the magnitude of the appraisal risk channel, we shut down the collateral recovery channel, making lenders' rate menus constant across σ-buckets, but assuming that the appraisal distribution varies across σ-buckets.

Panel (a) shows the results on the interest rate-price dispersion relationship. When we shut down the appraisal risk channel, interest rates remain higher in high σ-areas. Surprisingly, when we shut down the collateral recovery channel, the relationship changes signs; consumers in high-dispersion areas in fact receive *lower* interest rates. This is because, when appraisal noise is higher, consumers choose lower target loan sizes to risk the risk of under-appraisal and incidentally receive lower interest rates as a result. Note, however, that the appraisal risk effect is an order of magnitude smaller than the collateral recovery effect: the collateral recovery channel implies that rates increase by 6bps for a 1SD increase in house price dispersion, whereas the appraisal risk channel implies that rates decrease by 0.3bps for each SD increase in price dispersion.[26]

Panel (b) analyzes mortgage failures. Appraisal risk is the main driver of mortgage failures: when appraisals are noisier, under-appraisals and mortgage rejections are more likely. Consumers respond by choosing smaller targeted loan sizes, but not enough to counteract the direct effect of higher appraisal noise. The collateral recovery channel produces a slightly negative effect on rejections: when lenders offer worse rate menus, consumers choose smaller mortgages, limiting the risk of under-appraisals and mortgage failures. However, this effect is quantitatively negligible relative to the appraisal risk effect. Adding both channels, the net effect is that a 1SD increase in house price dispersion produces about a 1pp increase in failure likelihood, which is close to the estimates in Table 3 Panel B column 6.

Panel (c) analyzes loan-to-price ratios. As with mortgage failures, appraisal risk is the main driver of the dispersion-LTP relationship, with the collateral recovery channel playing

[26] The plot shows the effect of every unit of price dispersion. As shown in Table 1, the standard deviation of price dispersion is 0.11.

a smaller role. The collateral recovery channel contributes 0.1pp to the price dispersion effect on LTP. We further divide the effect of appraisals on LTP into two separate effects: an *ex-ante* effect based on borrowers choosing lower-target LTPs and another *ex-post* effect based on realized appraisals. When appraisals are noisier, the gap between l_{final} and l tends to be larger, putting downward pressure on l_{final}. Panel (c) distinguishes between these two effects. We find that ex-post appraisal risk plays a role that is slightly larger than the collateral recovery channel (0.2pp) but much smaller than the ex-ante appraisal risk channel (1.2pp).

6 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.[27] Human appraisals tend to be biased, 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 this may result in under-appraisals being more frequent, especially in areas with high house price dispersion. Automated appraisals thus have the potential to hurt low-income and Black households who tend to live in areas with less predictable house prices.[28]

We evaluate the potential impact of automated appraisals in our calibrated model. We assume automated appraisals would completely eliminate human biases from the appraisal process. Thus, using our calibrated model, we remove appraiser bias, setting $b = 0$, and re-evaluate the model, estimating buyers' optimal decisions, and thus loan-to-price ratios, interest rates, and mortgage rejections.

Figure 8 compares counterfactual mortgage outcomes to the baseline model. We find

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

[28] 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.

that automated appraisals would dramatically increase mortgage failure rates (Panel c) by around 10pp in low-dispersion areas and more than 15pp in high-dispersion areas. Loan size would also decrease by around 2pp of house prices (Panel a). Interest rates would actually decrease by around 5bps (Panel b) due to buyers precautionarily decreasing target loan size to limit under-appraisal risk.

Our results thus illustrate that the biases of human appraisers in fact act to alleviate the effects of price dispersion on mortgage credit availability. Shifting directly to automated appraisals, without compensating for the upwards bias in appraisal prices induced by human appraisers, has the potential to significantly decrease credit provision, particularly in high-dispersion areas.

6.2 Alternative LTV Policy

Our calibration suggests that homebuyers care more about loan size, implied by their inelastic loan size choices to changes in interest rates.[29] Thus, policies targeting under-appraisal risk could be particularly effective because it is the main driver of the extensive margin effect. Motivated by this, we propose an alternative policy which would still qualitatively allow appraisals to affect mortgage credit provision but would give regulators a continuous lever to modulate the effect of appraisals on mortgage credit.

Regulators currently use the smaller of the house sale price and the appraisal value as the denominator, V, for calculating loan-to-value ratios. We consider the following alternative rule:

$$V = \begin{cases} P & A > P \\ tA + (1-t)\,P & A \leq P \end{cases} \tag{23}$$

In words, when the appraisal A is above the transaction price P, V is set equal to the transaction price. When A is below P, V is set to a weighted average of A and P. This implies that appraisals below sale prices affect values — and thus the maximal amount that can be borrowed — less; decreasing the appraisal by \$1 decreases V by only t dollars.[30]

[29] This presumably reflects the fact that most homebuyers are down-payment constrained and like to maximize out the amount of loans they can borrow. This is consistent with findings in Glaeser and Shapiro (2003), and Mabille and Wang (2022).

[30] When $t = 1$, we have the standard rule currently used by regulators and securitizers: the value is equal to the

We quantitatively evaluate how much each alternative appraisal rule would affect mortgage credit provision in two steps. We first present reduced-form statistics, which imposes less model structure. The limitation of the reduced-form approach is that we can only calculate the additional credit received by homebuyers who successfully close the transactions.[31] We then complement the reduced-form statistics with the results derived from our calibrated model through counterfactual analyses.

Reduced-Form Statistics In the Corelogic LLMA data, we observe transaction prices as well as appraisals for each mortgage. Thus, we can calculate the hypothetical V under various alternative LTV policies. From Alternatives 1-3, we increase t, the weight on the appraisal value when the appraisal value is smaller than the transaction, from 20% to 50% to 80%. In each case, we then calculate counterfactual mortgage credit provision by taking each loan's LTV and multiplying it by the counterfactual value V under the alternative rule.

Table 9 presents the changes in credit provision under three alternative appraisal policies. Putting higher weights on sale prices (i.e., a smaller t) relaxes LTV constraints and increases loan size. Quantitatively, when $t = 0.2$, credit provision would increase by \$9,422, or 4%, for an average under-appraised house. High-dispersion areas benefit the most from these policies, since appraisal constraints tend to be most binding in these areas (Figure 2). Areas in the highest price dispersion quartile would receive about twice in percentage terms and 1.4 times in the dollar amount of additional credit than areas in the lowest price dispersion quartile.

These policies also have important distributional consequences. Since houses in low-income and Black-dominant zipcodes tend to have higher price dispersion (Figure 1), relaxed appraisal rules would increase credit provision in these areas proportionally more. Figure 9 shows that, in black-dominant zipcodes, $t = 0.2$ would increase credit provision by 8% for under-appraised properties, which is about twice that in non-black dominant zipcodes.[32] In below-median income zipcodes, setting $t = 0.2$ would increase credit provision by about 4.5% for under-appraised houses, compared to slightly below 3.5% in above-median income

minimum of A and P. Formally, (23) can be thought of as adjusting how concave V is as a function of A. When V is less concave, variance in A affects the mean value of V less.

[31] As shown in both Section 3.1.1 and 5.3, a significant amount of transactions fail because of under-appraisal.

[32] Black-dominant zipcodes are defined as zipcodes with at least 50% population being Black.

zipcodes.

Model Counterfactual We then complement the reduced-form statistics with the results derived from our calibrated model through counterfactual analyses. We re-evaluate the model, changing the distribution of appraisal values to reflect the three alternative rules we consider and calculate the buyers' optimal decisions in each case. We calculate a measure of total credit provision by multiplying the final LTVs by one minus the probability of mortgage rejection. This measure captures all three ways in which price dispersion affects LTPs in the model: buyers' ex-ante precautionary decisions to decrease target loan size, the probability that mortgages are rejected, and the ex-post downwards pressure on mortgage size from under-appraisals. In contrast, the reduced-form analysis only captures the ex-post downwards pressure force. The disadvantage of the model approach is that, since we calibrate only to ten σ-buckets, we cannot finely analyze the distributional consequences of these policies for low-income and high-minority areas in our model.

Figure 10 shows how each alternative policy would increase mortgage credit provision in each σ-bucket. Alternative policies can lead to fairly large changes in credit provision. Total credit expansion is as large as 3.5% in areas with the top quantile price dispersion when we put 20% weight on appraisal values when houses under-appraise. Even when we put 80% weight on appraisal values, which is a much smaller deviation from the status quo policy, credit provision in the highest σ-bin increases by around 1%.

These results illustrate quantitatively that the simple alternative LTV policies could give regulators a continuous lever to modulate under-appraisal risk. It is important to note that this counterfactual does not account for any potential costs of providing too much credit by over-weighing transaction prices; we cannot account for these costs without a richer model taking a stance on the social costs of lending too much. We aim mainly to provide a quantification of how much alternative policies could change the effects of appraisals.

6.3 Housing Affordability and Homeownership Gap

Policymakers aiming to improve homeownership rates for low-income and Black 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. Since low-income and Black households tend to live in areas with higher house price dispersion, they may face difficulty accessing mortgage credit even in an efficient credit market without any form of discrimination.

In Appendix E, we calibrate a standard lifecycle model of housing choices to quantify the effect of the borrowing constraints induced by collateral value uncertainty on homeownership gap. Through counterfactual analyses, we show that 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.[33]

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-income households 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.

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 highlight that the collateral value of the housing stock is an important and

[33]The homeownership gap between above-median income households and below-median income households was 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 was 37% in 2004. https://www.huduser.gov/Publications/pdf/HomeownershipGapsAmongLow-IncomeAndMinority.pdf

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, thus improving the affordability of the housing stock.

An interesting implication of our results is that urban policymakers who regulate the construction and renovation of residential housing may want to 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.

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.

Agarwal, Sumit, Changcheng Song, and Vincent W Yao, 2020, Relational Contracts in the Housing Market, Working paper, Working paper.

Albrecht, James, Pieter A. Gautier, and Susan Vroman, 2016, Directed search in the housing market, *Review of Economic Dynamics* 19, 218 – 231, Special Issue in Honor of Dale Mortensen.

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

Anundsen, André K, Erling Røed Larsen, and Dag Einar Sommervoll, 2020, Strategic price-setting and incentives in the housing market, Technical report, Oslomet Working paper series 2020: 1.

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., Hamilton B. Fout, Mark Palim, and Eric Rosenblatt, 2020a, Contract Price Confirmation Bias: Evidence from Repeat Appraisals, *The Journal of Real Estate Finance and Economics* 60, 77–98.

Eriksen, Michael D., Chun Kuang, and Wenyu Zhu, 2020b, 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* .

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

Genesove, David, and Christopher J Mayer, 1997, Equity and time to sale in the real estate market, *The American Economic Review* 255–269.

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.

Gilbukh, Sonia, and Paul S Goldsmith-Pinkham, 2019, Heterogeneous real estate agents and the housing cycle, *Available at SSRN 3436797* .

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.

Kurlat, Pablo, and Johannes Stroebel, 2015, Testing for information asymmetries in real estate markets, *The Review of Financial Studies* 28, 2429–2461.

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.

Mabille, Pierre, and Olivier Wang, 2022, Intermediary-based loan pricing .

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* .

Pancost, N Aaron, and Garrett Schaller, 2021, Measuring measurement error, *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.

Stroebel, Johannes, 2016, Asymmetric information about collateral values, *The Journal of Finance* 71, 1071–1112.

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.

Wheaton, William C., 1990, Vacancy, Search, and Prices in a Housing Market Matching Model, *Journal of Political Economy* 98, 1270–1292.

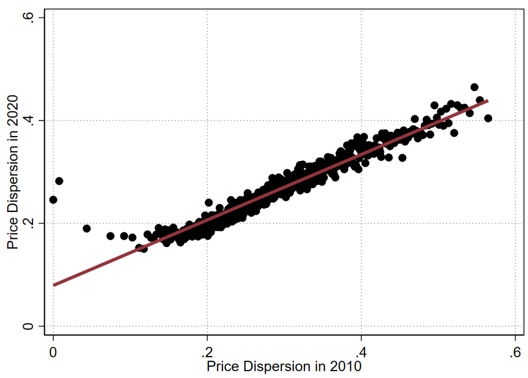
Wong, Arlene, 2019, Refinancing and the transmission of monetary policy to consumption, *Unpublished manuscript* 20.

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

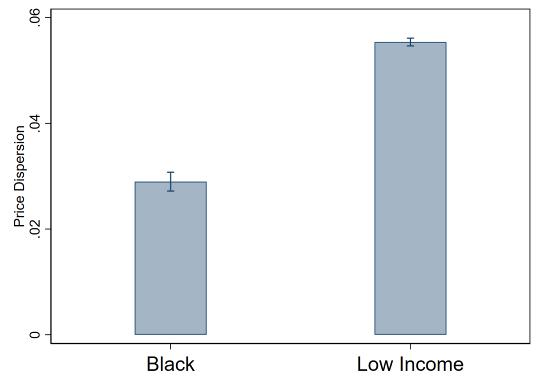
Zhou, Hanqing, Yuan Yuan, Christopher Lako, Michael Sklarz, and Charles McKinney, 2015, Foreclosure discount: definition and dynamic patterns, *Real Estate Economics* 43, 683–718.

Figures

Figure 1. Stylized Facts about Price Dispersion Estimates



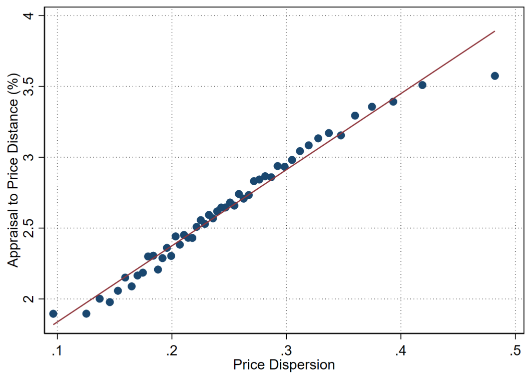
(a) Zipcode Dispersion: 2010 vs 2020



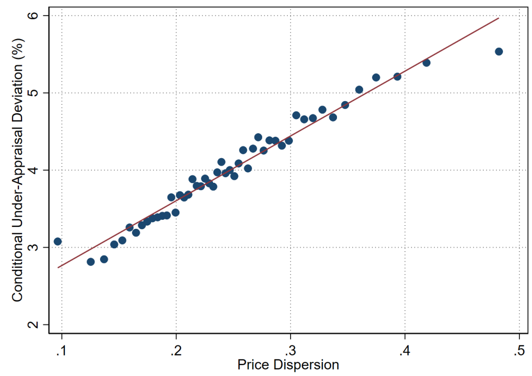
(b) Dispersion by Demographics

Notes: Panel (a) plots zipcode dispersion in 2020 against zipcode dispersion in 2010. Panel (b) shows the price dispersion difference between black-dominant zipcodes (black population share greater than 50%) and non-black dominant zipcodes (black population share less than 50%) conditional on income, as well as the price dispersion difference between high-income zipcodes and low-income zipcodes conditional on race. High and low income zipcodes are defined as above and below yearly median level, respectively. To obtain the values, we regress zipcode price dispersion on dummy variables for a zipcode being black-dominant, and whether the zipcode has below-median income; the figure shows the estimated coefficients and confidence intervals on these dummy variables. The sample includes annual zip level observations from 2000 to 2020. *Source*: Corelogic Deeds and American Community Survey 2008-2012.

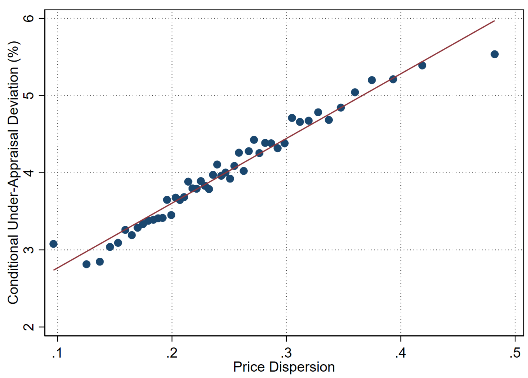
Figure 2. Price Dispersion and Appraisals



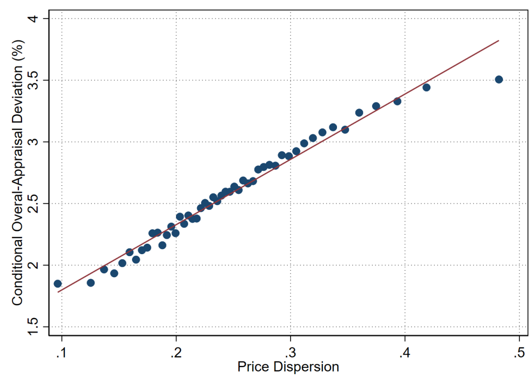
(a) Appraisal to Price Distance



(b) Under-Appraisal Deviation



(c) Under-Appraisal Pressure



(d) Over-Appraisal Deviation

Notes: Panel (a) of this figure shows a binned scatter plot, where the y-variable is Appraisal-to-Price distance, defined as $\frac{|a_i - p_i|}{p_i}$. In panel (b), the y-variable is the average under-appraisal percentage conditional on under-appraisal. In panel (c), the y-variable is under-appraisal pressure, defined as the product of the appraisal gap, and a dummy for under-appraisal, $\frac{|a_i - p_i|}{p_i}\mathbf{1}(a_i < p_i)$. In panel (d), the y-variable is the average over-appraisal percentage conditional on over-appraisal. In all panels, the x-variable is zipcode price dispersion. We divide all loans into 50 buckets based on zipcode house price dispersion. The sample includes loan level observations from 2000 to 2020. *Source*: Corelogic LLMA, Deed and Tax datasets.

Figure 3. Home Purchase - Mortgage Origination Diagram

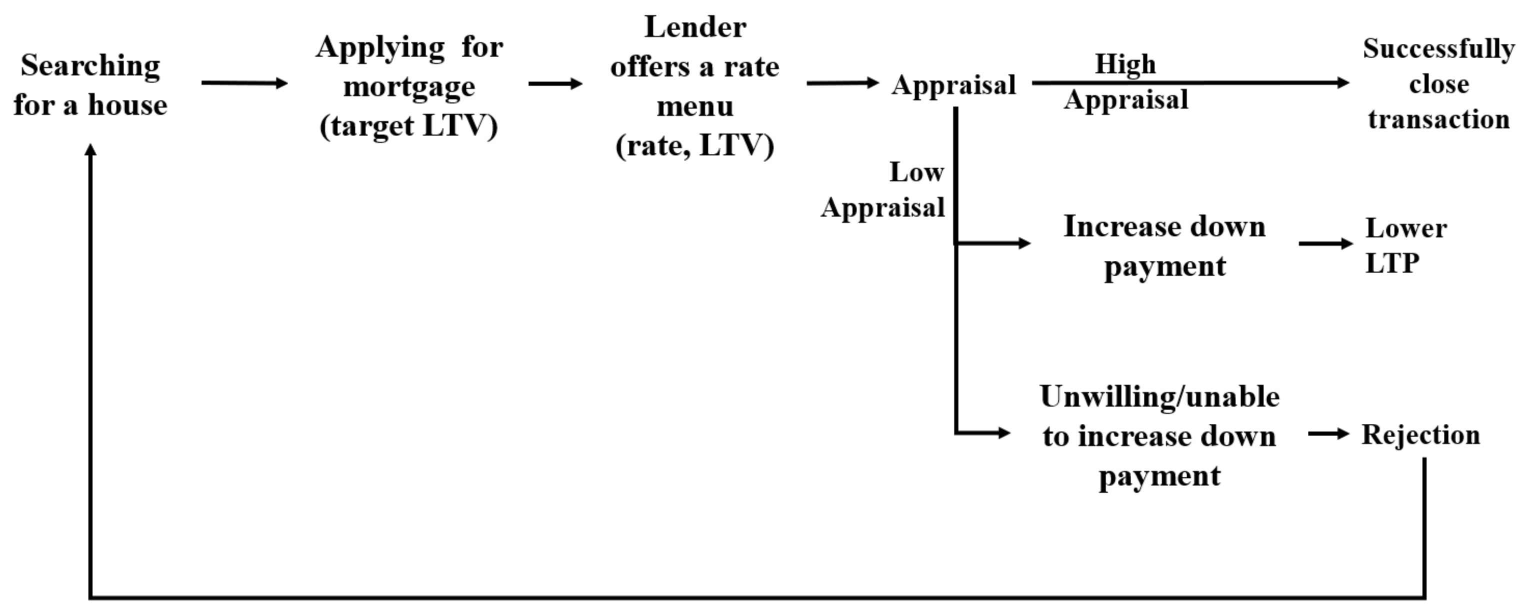
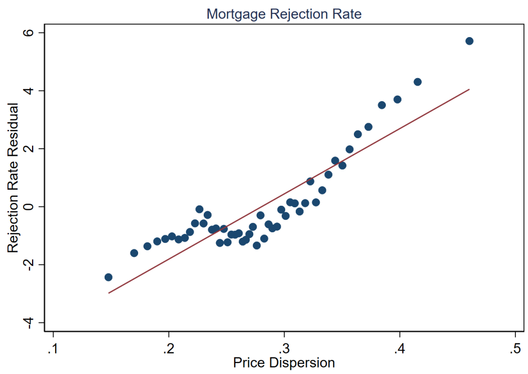
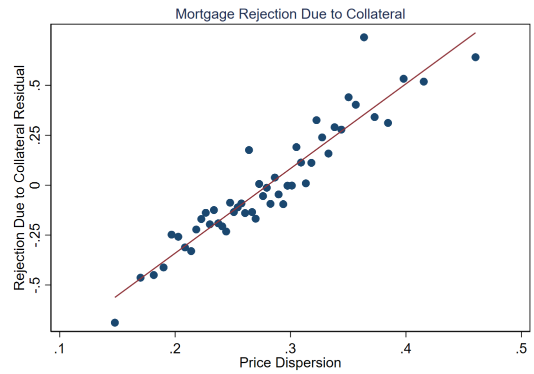


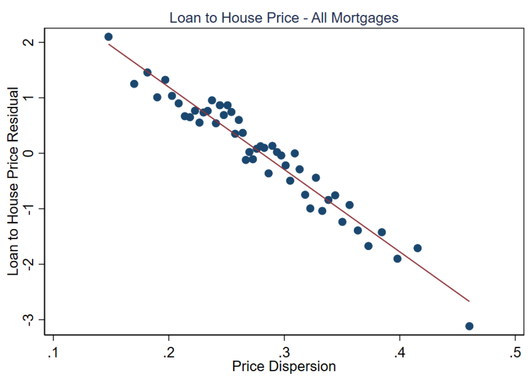
Figure 4. County Level House Price Dispersion and Credit Access



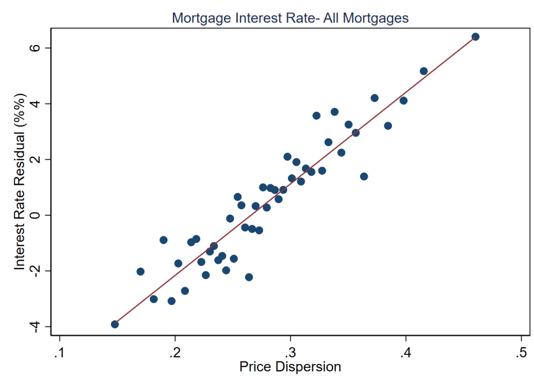
(a) Total Rejection



(b) Rejection due to Collateral



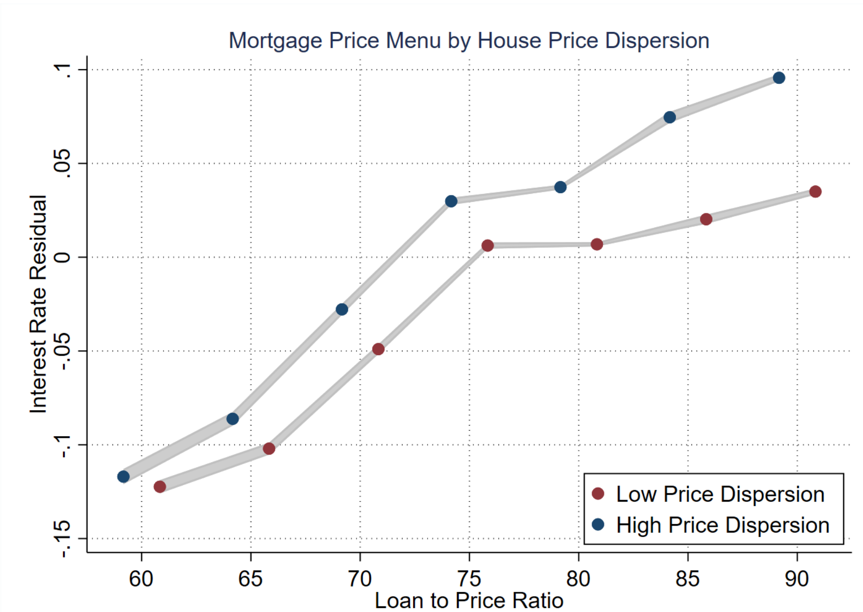
(c) Loan-to-Price



(d) Rate

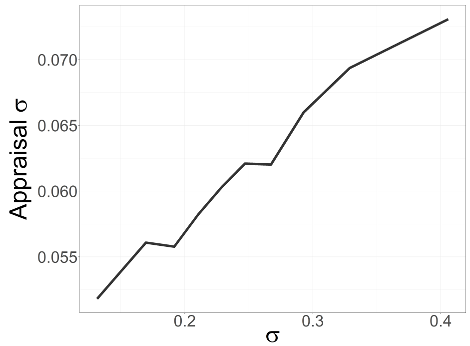
Notes: This figure shows the correlation between county level house price dispersion and various credit access outcomes. Panel (a) plots the total rejection rate in percentage points. Panel (b) plots the rejection rate due to collateral in percentage points. We residualize mortgage rejection rate by taking the residuals of regressions of mortgage rate on county average log house price, credit score, and year fixed effects. In panel (c), the y-variable is the average county-level residual from a regression of county average LTP on county-level house prices. The y-axis values are in percentage points. Panel d 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. The sample includes annual county observations from 2000 to 2017 for panels (a) and (b) and from 2000 to 2020 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 5. Property Level Mortgage Menu by House Price Dispersion

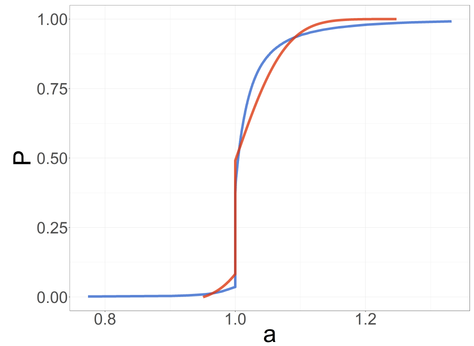


Notes: This figure shows the mortgage price menu (rate-LTP pair) by zip-level house price dispersion. The y values are interest rate residuals from a regression of mortgage rates on borrower fico, fico-squared, DTI, DTI-squared conforming or jumbo indicator, and origination month fixed effects. The dots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

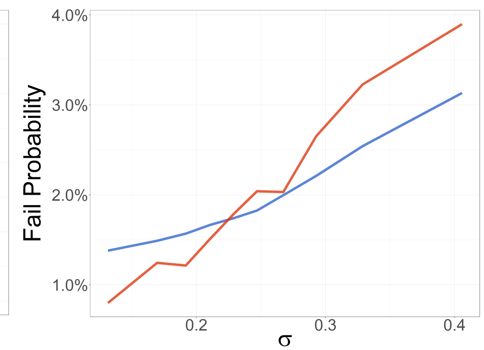
Figure 6. Model Fit



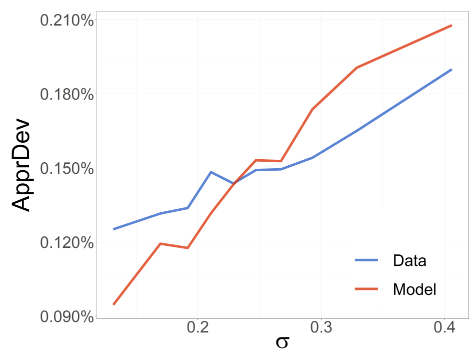
(a) Appraisal Noise



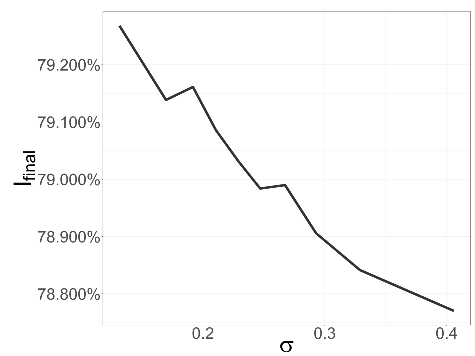
(b) Appraisal-to-Price



(c) Probabilities of Mortgage Failure



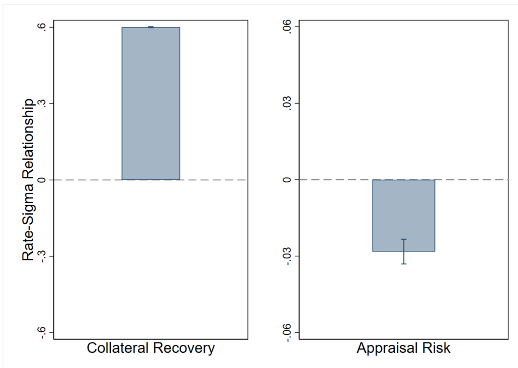
(d) Appraisal Deviation



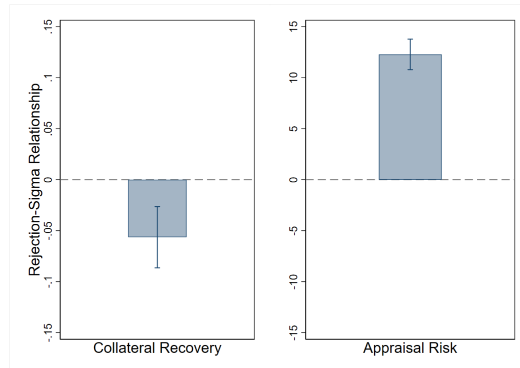
(e) Model Predicted LTP

Notes: 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.

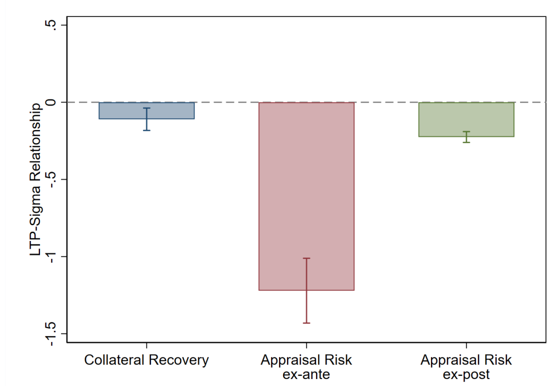
Figure 7. Decomposition of Channels



(a) Interest Rate (pp)



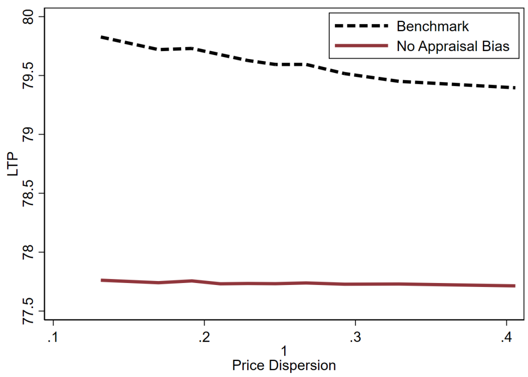
(b) Mortgage Failure (%)



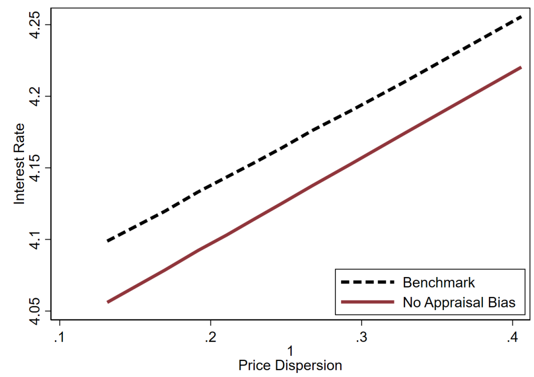
(c) Loan-to-Price (pp)

Notes: Panel (a), (b), and (c) respectively decompose the effect of price dispersion on interest rates, mortgage rejection rates, and loan-to-price ratios, into components attributable to the collateral recovery channel, and the appraisal channel. In all panels, the y-axis is the change in the outcome variable when sigma increases by 1. Note that the units differ from those in our reduced-form results, which standardize sigma: to convert the results in this figure to be comparable to the reduced-form results, all numbers should be multiplied by the standard deviation of sigma, which is 0.11. To calculate the effect of the collateral recovery channel, we shut off the effect of the appraisal channel, by assuming the appraisal distribution does not change with σ. Similarly, to calculate the effect of the appraisal risk channel, we shut off the collateral recovery channel, by assuming the rate menu does not vary with σ. In Panel (c), we further decompose the appraisal risk channel into an ex-ante effect, which measures how target loan size l varies with interest rate buckets, due to buyers' precautionary decisions to decrease loan size; and an ex-post effect, which measures how the gap $l - l_{final}$ changes with σ, which measures how realized under-appraisals limit loan size. Note that, in both panels (a) and (b), the y-axes are not the same in the collateral recovery and appraisal risk panels.

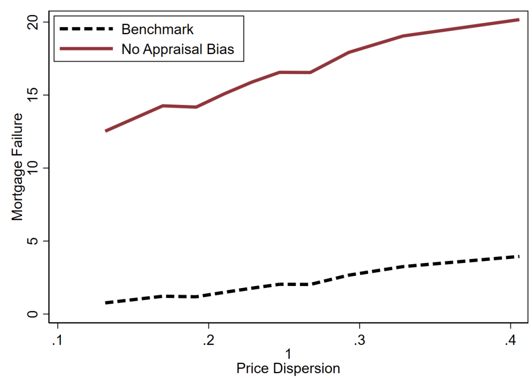
Figure 8. Counterfactual: Desktop Appraisals



(a) Loan-to-Price



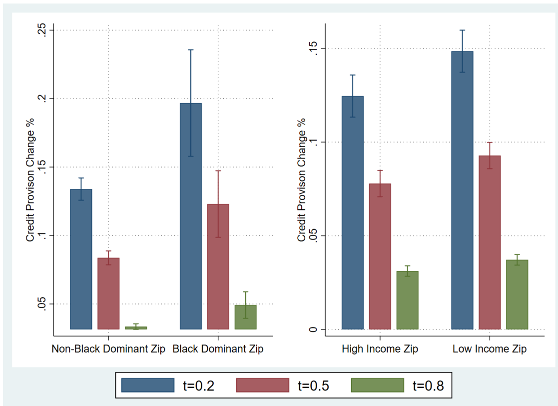
(b) Interest Rate



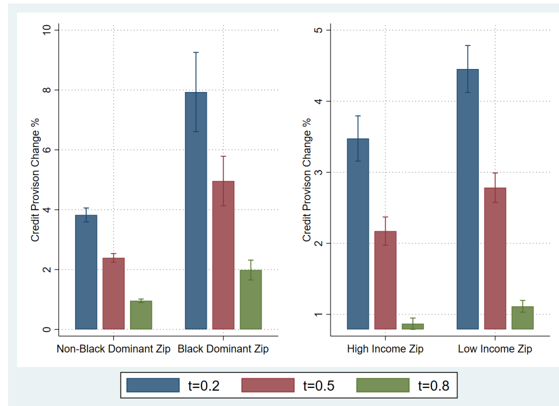
(c) Mortgage Failure Rate

Notes: This figure plots the counterfactual mortgage outcomes in the baseline model, to a counterfactual in which we assume appraisals are totally unbiased, that is, when we set the parameter $b = 0$. Panel (a) compares the benchmark and the counterfactual loan-to-price ratio. Panel (b) compares the benchmark and the counterfactual interest rate. Panel (c) compares the benchmark and the counterfactual mortgage rejections. In all panels, the y-axis is the outcome variable, and the x-axis is house price dispersion.

Figure 9. Alternative LTV Policies



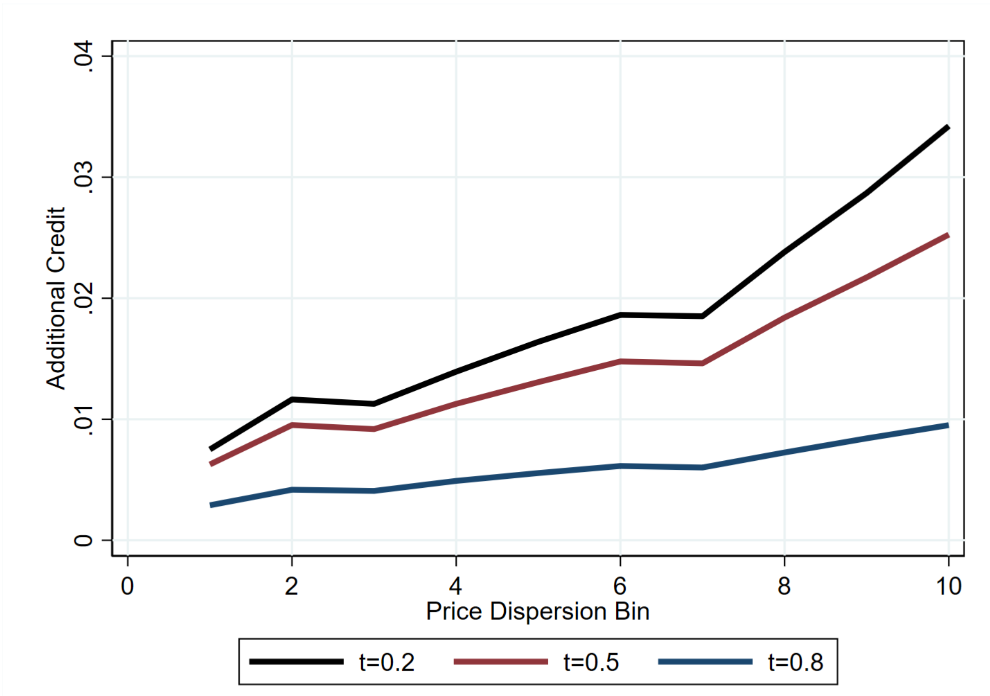
(a) All Borrowers



(b) Conditional on Appraisal < Sale Price

Notes: This figure presents counterfactual credit provision under three alternative appraisal policies. We show the results separately for non-black dominant and black-dominant zipcodes, as well as for above- and below-median income zipcodes. In calculating LTV for various regulation and securitization requirements, regulators currently calculate the denominator V as the smaller of the sale price and the appraisal value. In the alternative appraisal policies, V is instead set to the transaction price if the house over-appraises, and a weighted average of the transaction price and appraisal value if the house under-appraises. The weights on the appraisal value, upon under-appraisal, are 20%, 50%, and 80% in Alternatives 1, 2, and 3, respectively. Panel (a) presents the percentage change of credit for all borrowers. Panel (b) presents the percentage change of credit for borrowers whose appraisal values are less than the sale prices in our data. *Source*: Corelogic LLMA and Deeds and American Community Survey 2008-2012.

Figure 10. Alternative LTV Policies: Model Counterfactual



Notes: This figure presents model counterfactual credit provision under three alternative appraisal policies. Total credit provision is the product of average final mortgage size, and one minus the probability of mortgages failing. In calculating LTV for various regulation and securitization requirements, regulators currently calculate the denominator V as the smaller of the sale price and the appraisal value. In the alternative appraisal policies, V is instead set to the transaction price if the house over-appraises, and a weighted average of the transaction price and appraisal value if the house under-appraises. The weights on the appraisal value, upon under-appraisal, are respectively 20%, 50%, and 80% in Alternatives 1, 2, and 3. The figure shows, for each σ-bucket, the difference between counterfactual credit provision, and credit provision in the baseline model.

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 zipcode market condition (Panel B). All continuous variables are normalized by their standard deviations. We define a house as recently renovated, if it has been renovated within 5 years before 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: Zipcode Market Conditions

	Zipcode 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: Mortgage Rejections and Zip House Price Dispersion

This table presents loan level regression results about 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 zipcode house price dispersion, scaled by its standard deviation. Borrower/Loan controls include zipcode house price, credit score and the squared term, log income, loan type, 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	1.41***	1.42***	0.81***	2.50***	2.56***	2.43***
	(0.093)	(0.103)	(0.116)	(0.179)	(0.189)	(0.632)
Local Controls	✓	✓	✓	✓	✓	✓
County-Year FE	✓	✓	✓	✓	✓	✓
Lender-Year FE	✓	✓	✓	✓	✓	✓
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
Observations	47M	34M	3.6M	47M	34M	3.6M
Underidentification t-stat				83.81	78.17	25.41
Underidentification p-value				0.00	0.00	0.00
Weak identification t-stat				62.28	60.01	16.78
Panel B: Rejection Due to Collateral						
Zip Price Dispersion	0.50***	0.54***	0.38***	0.79***	0.86***	0.79***
	(0.036)	(0.038)	(0.054)	(0.061)	(0.066)	(0.142)
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	47M	34M	3.6M	47M	34M	3.6M
Underidentification t-stat				83.81	78.17	25.41
Underidentification p-value				0.00	0.00	0.00
Weak identification t-stat				62.28	60.01	16.78

Table 4: Mortgage Rejection Reasons

This table presents loan level regression results about mortgage rejection reasons. We restrict the sample to only rejected loans, and estimate Specification 4 using various rejection reason indicators as the outcome variables. The explanatory variable of interest is zipcode house price dispersion, scaled by its standard deviation. Borrower/Loan controls include zipcode house price, credit score and the squared term, log income, loan type, 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***	-0.09***	-0.48***	-0.13***	-0.62***
	(0.139)	(0.020)	(0.068)	(0.017)	(0.053)
Local Controls	✓	✓	✓	✓	✓
County-Year FE	✓	✓	✓	✓	✓
Lender-Year FE	✓	✓	✓	✓	✓
R2	0.16	0.10	0.18	0.05	0.24
Panel B: IV					
Price Dispersion	2.64***	-0.09**	-0.68***	-0.13***	-1.24***
	(0.165)	(0.038)	(0.107)	(0.027)	(0.099)
Local Controls	✓	✓	✓	✓	✓
County-Year FE	✓	✓	✓	✓	✓
Lender-Year FE	✓	✓	✓	✓	✓
Underidentification t-stat	64.68	64.68	64.68	64.68	64.68
Underidentification p-value	0.00	0.00	0.00	0.00	0.00
Weak identification t-stat	70.17	70.17	70.17	70.17	70.17
Sample mean	12.23	4.90	17.08	2.76	21.12
Observations	7.5M	7.5M	7.5M	7.5M	7.5M

Table 5: Price Dispersion and Cost Menu

This table presents loan-level regression results about the "cost menu", that is, interest rates controlling for LTPs. The outcome variable is loan-level interest rate, in 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 zipcode house price dispersion, scaled by its standard deviation. Borrower and loan controls include log house price, FICO score, FICO squared, LTV, LTV 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 6: Property-Level House Price Dispersion and LTP

This table presents property-level regression results on the relationship between price dispersion and mortgage LTPs. 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)
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 7: Ex-Post Performance

This table analyzes the relationship between price dispersion and the 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). The outcome variable is an indicator for default, which is equal to 100 if the loan defaults in two years since origination, and 0 otherwise. The explanatory variable of interest is zipcode house price dispersion, scaled by its standard deviation. Non-reported controls include house price, loan type, and the squared-terms of FICO, DTI, and LTV. 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	0.21***	0.15***	0.19***	-0.04	-0.02	0.04
	(0.054)	(0.036)	(0.063)	(0.096)	(0.098)	(0.118)
Interest Rate	2.18***	2.39***	1.69***	2.18***	2.40***	1.69***
	(0.125)	(0.175)	(0.130)	(0.126)	(0.178)	(0.131)
FICO	-65.98***	-60.94***	-56.81***	-65.98***	-60.94***	-56.82***
	(0.703)	(1.042)	(1.639)	(0.702)	(1.040)	(1.642)
DTI	0.37***	-0.06	0.33***	0.35***	-0.06	0.32***
	(0.056)	(0.062)	(0.055)	(0.054)	(0.060)	(0.054)
LTV	-3.76***	-3.63***	-2.48***	-3.75***	-3.62***	-2.48***
	(0.155)	(0.152)	(0.166)	(0.153)	(0.150)	(0.165)
Origination Month FE	✓	✓	✓	✓	✓	✓
County-Year FE	✓	✓	✓	✓	✓	✓
Property & Loan Controls	✓	✓	✓	✓	✓	✓
R2	0.15	0.13	0.19	0.10	0.07	0.10
Observations	4.3M	2.1M	0.9M	4.3M	2.1M	0.9M
Underidentification test statistic				82.79	86.85	68.21
Underidentification test p-value				0.00	0.00	0.00
Weak identification test statistic				41.87	37.98	34.92

Table 8: Model Estimates

This table presents the estimates of model parameters. Panel A reports values of externally calibrated parameters. Panel B reports values of 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 6
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

Table 9: Counterfactual Appraisal Policy and Credit Provision

This table presents counterfactual credit provision under three alternative appraisal policies, by zipcode idiosyncratic house price dispersion quartile. In calculating LTV for various regulation and securitization requirements, regulators currently calculate the denominator V as the smaller of the sale price and the appraisal value. In the alternative appraisal policies, V is instead set to the transaction price if the house over-appraises, and a weighted average of the transaction price and appraisal value if the house under-appraises. The weights on the appraisal value, upon under-appraisal, are respectively 20%, 50%, and 80% in Alternatives 1, 2, and 3. Fixing LTV, we calculate the additional credit provision under each alternative appraisal policies for each zipcode house price dispersion quartile. *Source*: Corelogic LLMA and Deeds.

		Idiosyncratic Price Dispersion Quartile			
	Total	Q1	Q2	Q3	Q4
Total Δ Credit Provision (Million)					
Counterfactual 1	389	379	386	355	437
Counterfactual 2	243	237	241	222	273
Counterfactual 3	97	95	97	89	109
Average Percentage Change in Credit Provision (%)					
Counterfactual 1	0.14	0.11	0.13	0.13	0.17
Counterfactual 2	0.09	0.07	0.08	0.08	0.11
Counterfactual 3	0.03	0.03	0.03	0.03	0.04
Δ Credit Provision per Borrower | (Appraisal<Sale Price)					
Counterfactual 1	9422	8169	9075	9160	11282
Counterfactual 2	5889	5106	5672	5725	7052
Counterfactual 3	2355	2042	2269	2290	2821
Percentage Change in Credit Provision (%) | (Appraisal<Sale Price)					
Counterfactual 1	4.00	2.90	3.79	3.92	5.40
Counterfactual 2	2.50	1.81	2.37	2.45	3.37
Counterfactual 3	1.00	0.72	0.95	0.98	1.35

Internet 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:

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

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 Drivers of Idiosyncratic Price Dispersion

We discuss a number of factors and theoretical forces that may drive dispersion, and explain why these theories have similar implications for mortgage credit provision.

Information asymmetry. Lenders of secured loans must be concerned about adverse selection. This is especially the case in the consumer credit market, where houses and used cars, for example, have diverse characteristics, some of which are difficult to measure, and homeowners have better information about these characteristics (Kurlat and Stroebel, 2015; Stroebel, 2016). Houses with more hard-to-measure characteristics tend to have higher value uncertainty. Thus, lenders who lend against houses with higher value uncertainty may worry more about adverse selection because the owners have more information advantage about the house than the lenders.

Search frictions. The housing search literature has argued that house transaction prices are not determined in a fully competitive and frictionless market. Prices appear to depend not only on house characteristics: the transaction price of a house appears to be causally influenced by characteristics of the buyer and seller. Sellers who are more patient achieve higher sale prices, by setting higher list prices and keeping houses on the market for longer; this has been shown using instruments for seller patience, such as homeowners' equity position (Genesove and Mayer, 1997; Guren, 2018) and homeowners' nominal losses since purchase (Genesove and Mayer, 2001). Dispersion in different buyers' values for the same house may also drive house price dispersion: using data from Norwegian housing auctions, Anundsen et al. (2020) shows that the standard deviation of the ratio between buyers' bid prices and appraisal values is approximately 7.9%. Other factors, such as the experience of the realtor selling the house, also appear to affect house sale probabilities and prices (Gilbukh and Goldsmith-Pinkham, 2019).

Other factors. We also note that there are other possible housing market frictions which generate price dispersion. The literature has studied many different models, such as random search (Wheaton, 1990), directed search (Albrecht et al., 2016), and price posting (Guren, 2018).

We do not take a stance on this paper on the particular theoretical microfoundation of price dispersion, since it is not crucial for studying the effects of dispersion on credit provision. Price dispersion decreases credit provision by increasing lenders' expected losses upon foreclosure, and by making appraisals noisier and thus appraisal constraints more binding. Both effects occur regardless of the particular theoretical microfoundation of prices dispersion.

If we observed all characteristics of houses that market participants observed, and our functional forms for house prices were fully flexible, our measurement strategy would fully filter out the effects of house characteristics, capturing only price dispersion generated by housing market frictions. In practice, in addition to frictional price dispersion, 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. Second, our functional forms in (1) may not be flexible enough to capture the true conditional expectation function; model misspecification will thus contribute to our estimates of price dispersion. Both of these effects serve as confounds we would like to filter out from our analysis, since if lenders use the correct price model with the full set of observables, frictional price dispersion should affect mortgage lending decisions, but not errors attributable to unobservables or model misspecification.

We believe these confounds are unlikely to drive our main results, for the following reasons. We observe a rich set of characteristics, which are essentially all the features that mortgage lenders observe for houses. A limitation of our data is that we only have time-invariant do not observe renovations and time variation in house characteristics. However, Giacoletti (2021), using data on remodeling expenditures for houses in California appears to have quantitatively small effects on estimated price: accounting for renovations decreases the estimated standard deviation of returns by only around 2% of house prices.

C Robustness Checks

C.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:

$$p_{it} = \eta_{kt} + \mu_i + \epsilon_{it} \tag{A1}$$

where i indexes properties, k indexes counties, and t indexes months. Equation (A1) 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 (A1) 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, (A1) 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.[34] 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:

$$\hat{\epsilon}_{it}^2 = h_k\left(sales_i, tbs_i\right) + \zeta_{it} \tag{A2}$$

Where, $h_k\left(sales_i, tbs_i\right)$ interacts a vector of $sales_i$ dummies with a fifth-order polynomial in 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 $\hat{\epsilon}_{it}^2$ within county k:

$$\hat{\epsilon}_{TBSadj,it}^2 = \hat{\zeta}_{it} + E_k\left[\hat{\epsilon}_{it}^2\right] \tag{A3}$$

$\hat{\epsilon}_{TBSadj,it}^2$ can be interpreted as the baseline estimates, $\hat{\epsilon}_{it}^2$, nonparametrically purged of all

[34] 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.

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 (A1) 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 (A1) are that it is unable to capture any features of houses which have time-varying effects on house prices.

Figure A1 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.

C.2 Implications of Controlling for House Prices

We briefly discuss the implications of controlling for house prices in our main empirical specifications. Conceptually, when we control for prices, β is identified by comparing, for example, zipcode A to zipcode B, which have similar average prices, but B has higher price dispersion. However, in most models, price dispersion affects the level of house prices: if two zipcodes have similar house quality, but one has higher price dispersion, average prices should be lower in the high-dispersion zipcode. Thus, in our example, if zipcode A and B have identical average prices, zipcode B should have *higher* average house quality than zipcode A.

Regressing LTP on price dispersion controlling for prices, and not controlling for house quality, makes sense in a model in which the distribution of house prices captures all features of houses that are relevant for lenders' decision problem. We construct a simple model which links lending decisions to the mean and variance of house prices in Appendix D.3. Since lenders do not directly interact with the house they lend against, in principle they should only care about house characteristics to the extent that they change the level or dispersion of house prices. Controlling for house prices alleviates the possibility that lenders may have a preference to have systematically higher or lower LTPs for high-priced houses. It is not necessary to control for house quality in addition to house prices, since any two houses with the same mean and variance of house prices are equivalent to lenders, regardless of the particular characteristics of the two houses.

A simpler reason why controlling for prices should not substantially matter for our results is that, when house prices are higher, *loan size* should increase, but it is not obvious whether *loan-to-price* ratios should increase or decrease. In most lending models, such as our model in Appendix D.3, the overall level of prices has no effect on loan-to-price ratios. To test that our results are not driven by this choice, in Table A6, we estimate our main specifications without controls for prices; all results are qualitatively and quantitatively similar to the main text.

C.3 Lender Market Power

Our results are not likely to be driven by lender market power. Firstly, this is because 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 the main empirical specifications 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 A1 reports the results. The key variable of interest is price dispersion, which is property-level idiosyncratic price dispersion. We first confirm Table 5 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.

C.4 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 A1 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.

D Proofs and Supplementary Material for Section 4

D.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:

$$V(a,l) \equiv \max\left[V(a,l,continue), V(a,l,renege)\right] \tag{A4}$$

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

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

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

$$V(a,l,continue) = \frac{c_1^{1-\eta}-1}{1-\eta} + \beta^T u_2' c_2 \tag{A5}$$

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

$$c_1 = \underbrace{W_1 - P(1-l)}_{Targeted\ consumption} - \underbrace{\psi P \max(0, l - \phi a)}_{Appraisal\ shortfall} \tag{A6}$$

$$c_2 = -\left(1 + r(l)\right)^T P\left(l - \max\left[0, l - \phi a\right]\right) \tag{A7}$$

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, (A6) 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 (A6) and (A7) into (A5), we have:

$$\begin{aligned}V\left(a,l,continue\right)=&u_{1}\left(W_{1}-P\left(1-l\right)-\psi P\max\left(0,l-\phi a\right)\right)\\&-u_{2}'\beta^{T}\left(1+r\left(l\right)\right)^{T}Pl\\&+u_{2}'\beta^{T}\left(1+r\left(l\right)\right)^{T}P\max\left[0,l-\phi a\right]\end{aligned}\tag{A8}$$

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:

$$V\left(a,l,continue\right)=-u_{2}'\beta^{T}\left(1+r\left(l\right)\right)^{T}Pl+\omega\left(a,l\right)\tag{A9}$$

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

If the buyer reneges, she receives:

$$V\left(a,l,renege\right)=-\beta^{T}u_{2}'\zeta P+E_{a}\left(V\left(a,l\right)\right)\tag{A10}$$

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:

$$E_{a}\left(V\left(a,l\right)\right)=\int_{0}^{\infty}\max\left(V\left(a,l,continue\right),V\left(a,l,renege\right)\right)dF_{a}\left(a\right)\tag{A11}$$

Now, note that $V\left(a,l,renege\right)$, is independent of a, whereas from (A8), $V\left(a,l,continue\right)$ 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:

$$V\left(\bar{a},l,renege\right)=V\left(\bar{a},l,continue\right)\tag{A12}$$

Substituting for $V(\bar{a}, l, continue)$ using (A9), we have:

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

Substituting into (A11), we have:

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

$$E_a\left(V(a, l)\right) = -\beta^T \left(1 + r(l)\right)^T u_2' Pl + \int_0^\infty \max\left(\omega(a, l), \omega(\bar{a}, l)\right) dF_a(a) \tag{A13}$$

Substituting into (A10), we have:

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

D.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 (A12), we have:

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

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

$$\omega(\bar{a}, l) = -\beta^T u_2' \zeta P + \int_0^\infty \max\left(\omega(a, l), \omega(\bar{a}, l)\right) dF_a(a) \tag{A15}$$

This is (17) of Theorem 1. Equation (A15) characterizes $\bar{a}(l)$. In words, the LHS of (A15) 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 (A15) to:

$$\int_{a>\bar{a}} \big(\omega(a,l) - \omega(\bar{a},l)\big)\, dF_a(a) = \beta^T u_2' \zeta P \tag{A16}$$

Since ω is increasing in a, the LHS of (A16) is strictly decreasing in $\bar{a}$, hence for any parameters, there is at most one value of $\bar{a}$ which solves (A16). Note also that (A16) 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 (A16) is 0 for all $\bar{a} > \frac{l}{\phi}$; the RHS is positive, so it can never be optimal to set $\bar{a} > \frac{l}{\phi}$.

D.1.4 Optimal Loan Choice

Repeating (A13), 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:

$$E\Big(V\big(\bar{a}(l),l\big)\Big) = -\beta^T \big(1+r(l)\big)^T u_2' P l + \int_0^\infty \max\Big(\omega(a,l), \omega\big(\bar{a}(l),l\big)\Big)\, dF_a(a) \tag{A17}$$

The buyer picks l to maximize (A17); this is (19).

D.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:

$$\Gamma(l) = \max_{\bar{a}} \left[\int_{\bar{a}}^{\infty} \left[-\beta^T \left(1 + r(l)\right)^T u_2' Pl + \omega(a, l)\right] dF_a(a) + F_a(\bar{a}) \left[\Gamma(l) - P\beta^T u_2' \zeta\right]\right] \tag{A18}$$

In words, the buyer receives $-\beta^T \left(1 + r(l)\right)^T u_2' Pl + \omega(a, l)$ 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:

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

Hence, the envelope theorem applies; we have:

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

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

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

Now, note that:

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

Hence, we can write (A19) as:

$$\begin{aligned}\Gamma(l) = \max_{\bar{a}} \int_{\bar{a}}^{\infty} & \left[-\beta^T \left(1+r(l)\right)^T u_2' P \min\left(l, \phi a\right)\right. \\ & \left. + u_1 \left(W_1 - P(1-l) - \psi P \max\left(0, l-\phi a\right)\right)\right] dF_a(a) \\ & + F_a(\bar{a}) \left[\Gamma(l) - P\beta^T u_2' \zeta\right]\end{aligned} \tag{A20}$$

Differentiating with respect to l, we have:

$$\begin{aligned}\frac{d\Gamma}{dl} = & \frac{\partial}{\partial l}\left[\int_{\bar{a}}^{\infty} -\beta^T \left(1+r(l)\right)^T u_2' P \min\left(l, \phi a\right)\right. \\ & \left. + u_1 \left(W_1 - P(1-l) - \psi P \max\left(0, l-\phi a\right)\right) dF_a(a)\right] + F_a(\bar{a}^*) \frac{d\Gamma}{dl}\end{aligned} \tag{A21}$$

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

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:

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

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

$$\left(1-F_a\left(\frac{l}{\phi}\right)\right)\left(\underbrace{-T\beta^T\left(1+r\left(l\right)\right)^{T-1} r'\left(l\right) u_2' Pl}_{Rate\ Change} + \underbrace{Pu_1'\left(W_1-P\left(1-l\right)\right) - P\beta^T\left(1+r\left(l\right)\right)^T u_2'}_{Consumption\ Smoothing}\right)$$
$$-\underbrace{\left[-\beta^T\left(1+r\left(l\right)\right)^T u_2' Pl + u_1\left(W_1-P\left(1-l\right)\right)\right] f\left(\frac{l}{\phi}\right)}_{Nuisance\ Term} \quad \text{(A23)}$$

The "rate change" term in (A23) 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:

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

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

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

$$+\underbrace{\left[-\beta^T\left(1+r\left(l\right)\right)^T u_2' P\phi\left(\frac{l}{\phi}\right)+u_1\left(W_1-P\left(1-l\right)-\psi P\left(l-\phi\left(\frac{l}{\phi}\right)\right)\right)\right]f\left(\frac{l}{\phi}\right)}_{Nuisance\ \ Term} \tag{A24}$$

The "rate increase" term is analogous to (A23). 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 (A23) once we add the two components.

Combining (A23) and (A24), we have:

$$\frac{\partial}{\partial l}\int_{\bar{a}}^{\infty} u_1\left(W_1-P\left(1-l\right)-\psi P\max\left(0,l-\phi a\right)\right)dF_a\left(a\right)=$$
$$\left(1-F_a\left(\bar{a}\right)\right)\left(-T\beta^T\left(1+r\left(l\right)\right)^{T-1} r'\left(l\right)u_2' Pl\right)+$$
$$\left(1-F_a\left(\frac{l}{\phi}\right)\right)P\left(u_1'\left(W_1-P\left(1-l\right)\right)-\beta^T\left(1+r\left(l\right)\right)^T u_2'\right)-$$
$$\int_{\bar{a}}^{\frac{l}{\phi}}\left(1-\psi\right)Pu_1'\left(W_1-P\left(1-l+\psi\left(l-\phi a\right)\right)\right)dF_a\left(a\right) \tag{A25}$$

Finally, combining (A25) with (A22), we have:

$$
\begin{aligned}
\frac{d\Gamma}{dl} = &\left(-T\beta^{T}\left(1+r\left(l\right)\right)^{T-1}r'\left(l\right)u_2'Pl\right)+\\
&\frac{1}{\left(1-F_a\left(\bar{a}^{*}\right)\right)}\left[\left(1-F_a\left(\frac{l}{\phi}\right)\right)P\left(u_1'\left(W_1-P\left(1-l\right)\right)-\beta^{T}\left(1+r\left(l\right)\right)^{T}u_2'\right)-\right.\\
&\left.\int_{\bar{a}}^{\frac{l}{\phi}}\left(1-\psi\right)Pu_1'\left(W_1-P\left(1-l+\psi\left(l-\phi a\right)\right)\right)dF_a\left(a\right)\right]
\end{aligned}
$$

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

$$
\begin{aligned}
&\underbrace{\left(1-F_a\left(\frac{l}{\phi}\right)\right)P\left(u_1'\left(W_1-P\left(1-l\right)\right)-\beta^{T}\left(1+r\left(l\right)\right)^{T}u_2'\right)}_{Consumption\ Smoothing}=\\
&\underbrace{\left(1-F_a\left(\bar{a}^{*}\right)\right)\left(T\beta^{T}\left(1+r\left(l\right)\right)^{T-1}r'\left(l\right)u_2'Pl\right)}_{Rate\ Change}+\\
&\underbrace{\int_{\bar{a}}^{\frac{l}{\phi}}\left(1-\psi\right)Pu_1'\left(W_1-P\left(1-l+\psi\left(l-\phi a\right)\right)\right)dF_a\left(a\right)}_{Under-Appraisal\ Penalty}
\end{aligned}
\tag{A26}
$$

The LHS of (A26) 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.

D.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:

$$F = P\left(1 - c + \epsilon_F\right) \tag{A27}$$

Thus, for a non-recourse mortgage, lender's expected loss conditional on default is:[35]

$$Loss = E\left[P\left(l - \max\left[l, 1 - c + \epsilon_F\right]\right)\right] = PE\left[\max\left[0, l - \left(1 - c + \epsilon_F\right)\right]\right] \tag{A28}$$

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.[36] 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 $Pl\left(r - \rho\right)$. 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:

$$Pl\left(1 - \delta\right)\left(r - \rho\right) = \delta PE\left[\max\left[0, l - \left(1 - c + \epsilon_F\right)\right]\right] \tag{A29}$$

The LHS of (A29) 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'

[35] 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.

[36] 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: $Loss = E\left[P\left(l(1+\rho) - \max\left[l(1+r), 1 - c + \epsilon_F\right]\right)\right]$

expected losses conditional on default, multiplied by the default probability δ.

Expression (A29) 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 (A29) is always increasing in σ_F.[37] Thus, holding l fixed, increasing σ_F must cause r to increase. This rationalizes our observations in Figure 5 and Table 5. Expression (15) in the main text can be thought of as a linear approximation to this menu.

D.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 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:

$$\tilde{r}_{fld} = r_{fld} - \bar{r}_f \tag{A30}$$

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

[37] Note that the RHS of (A29) 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.

points in the interest rate menu of Figure 5, 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 C.1, separately for high-dispersion and low-dispersion counties.[38] 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 $r^{model}_{fld} - \rho$ using (A29):

$$r^{model}_{fld} - \rho = \frac{\delta E\left[\max\left[0, l - (1 - c + \epsilon_F)\right]\right]}{l(1-\delta)} \tag{A31}$$

where the expectation on the RHS of (A31) can be analytically calculated, since we assumed ϵ_F is normally distributed. We can then calculate the model counterpart of the interest rate residuals (A30), by subtracting the mean interest rate in each FICO bucket f:

$$\begin{aligned}\tilde{r}^{model}_{fld}(c) = r(l,\delta,c,\sigma_F) - \frac{\sum_l \sum_f r(l,\delta,c,\sigma_F)}{\sum_l \sum_f 1} = \\ \left(r(l,\delta,c,\sigma_F) - \rho\right) - \frac{\sum_l \sum_f r(l,\delta,c,\sigma_F) - \rho}{\sum_l \sum_f 1}\end{aligned} \tag{A32}$$

Note that (A32) implies that $\tilde{r}^{model}_{fld}$ does not depend on the choice of ρ, so we set an arbitrary value of ρ in calculating $\tilde{r}^{model}_{fld}(c)$. We then choose a value of the foreclosure discount c through generalized method of moments, to minimizes the squared distance between the

[38] 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.

data residuals $\tilde{r}_{fld}$, and the model residuals $\tilde{r}^{model}_{fld}$:

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

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 $\tilde{r}^{model}_{fld}$, 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:

$$r_{lf,d=H} - r_{lf,d=L} \tag{A33}$$

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, (A33) 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 A5. 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.

D.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:

$$c(A, A_{raw}) = \gamma (A - A_{raw})^2 \tag{A34}$$

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., 2020b) 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:

$$\max_A U_{appr}(A) = \chi L_{max}(A) - \gamma (A - A_{raw})^2 \tag{A35}$$

The optimization problem in (A35) 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:

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

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:

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

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:

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

which is exactly (16) in the main text.

D.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.

E 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.

E.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:

$$\log(y_t) = \chi_t + \zeta_t \tag{A36}$$

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 $h_i \in S = \{s_1, s_2 \dots s_H\}$, ordered in increasing order. There is a cutoff s_R, for $R < H$. All house qualities below $s_1 \dots s_R$ are available for rent only, and all house qualities $s_{R+1} \dots s_H$ 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 E.2 below, and Appendix E.4.2. Let a_t represent cash-on-hand; homeowners' borrowing constraint is thus:

$$a_t \geq -\phi(h_t) p^h h_t \tag{A37}$$

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:

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

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

$$V^{renter} (h_t, a_t, \zeta_t) = \max_{c_t, a_{t+1}, h_{t+1}} u (c_t, h_{t+1}) + \beta E \left[V_{t+1} (h_{t+1}, a_{t+1}, \zeta_{t+1}) \mid \zeta_t \right] \tag{A38}$$

$$s.t.\ c_t + \frac{a_{t+1}}{1 + r_t} = a_t + y_t + \underbrace{p^h h_t 1 (h_t > s_R)}_{Selling\ old\ house} - p_r h_{t+1} \tag{A39}$$

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

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

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:

$$V^{purchase} = \max_{c_t, a_{t+1}, h_{t+1}} u (c_t, h_{t+1}) + \beta E \left[V (h_{t+1}, a_{t+1}, \zeta_{t+1}) \mid \zeta_t \right] \tag{A40}$$

$$s.t.\ c_t + \frac{a_{t+1}}{1 + r_t} = a_t + y_t + \underbrace{p^h h_t 1 (h_t > s_R)}_{Selling\ old\ house} - \underbrace{\left(1 + \delta^h + F^{pur} 1 (h_{t+1} \neq h_t)\right) p^h h_{t+1}}_{Buying\ new\ house} \tag{A41}$$

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

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

E.2 Calibration

The model period is annual. Most of our choices for parameter calibrations are standard, and we discuss them in Appendix E.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 E.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 A4, 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 A4 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 E.4.3. Table A7 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.

E.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.[39]

Table A8 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.[40]

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.[41] Therefore, our results suggest that, in a standard

[39] Note that we showed in Subsection 2.3 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.

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

[41] 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

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.

E.4 Additional Calibration Details

E.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.[42] 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.[43] 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

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

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

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 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.

E.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:

$$\sigma_{med}\left(p\right) \equiv E\left[\sigma_{ict} \mid p_{ict} = p, c \in \mathcal{C}_{mid}\right] \tag{A42}$$

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 $\sigma_{high}\left(p\right)$ and $\sigma_{low}\left(p\right)$ analogously to (A42), for the high- and low-dispersion set of counties. The three curves $\sigma\left(p\right)$ 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 A4. 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:

$$\phi_{med}(p) = 0.8 + \beta_{LTV\_\sigma}\left(\sigma_{med}(p) - \sigma_{med}(p_{min\_\sigma})\right) \tag{A43}$$

Where β_{LTV_σ} is the coefficient from regressing LTV on price dispersion, from column 1 of Table A4. In words, (A43) 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:

$$\phi_{high}(p) = 0.8 + \beta_{LTV\_\sigma}\left(\sigma_{high}(p) - \sigma_{med}(p_{min\_\sigma})\right) \tag{A44}$$

That is, analogous to (A43), $\phi_{high}(p)$ 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:

$$\phi_{low}(p) = 0.8 + \beta_{LTV\_\sigma}\left(\sigma_{low}(p) - \sigma_{med}(p_{min\_\sigma})\right) \tag{A45}$$

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.

E.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 (A37), 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, (A39) and (A41), and rewriting expressions in terms of wealth, we can write the budget constraint equation as:

$$\begin{aligned} w_{t+1} = \\ (1+r_t)\left(w_t + y_t - c_t - \underbrace{1\left(h_{t+1} > s_R\right)\left(1 + \delta^h + F^{pur} 1\left(h_{t+1} \neq h_t\right)\right) p^h h_t}_{Buying\ new\ house} - \underbrace{1\left(h_t < s_R\right) p^r}_{Rent} \right) + \\ 1\left(h_{t+1} > s_R\right) p^h h_t \end{aligned} \tag{A46}$$

Using (A46), we eliminate consumption c_t from the household's optimization problem, (A38) and (A38). 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.

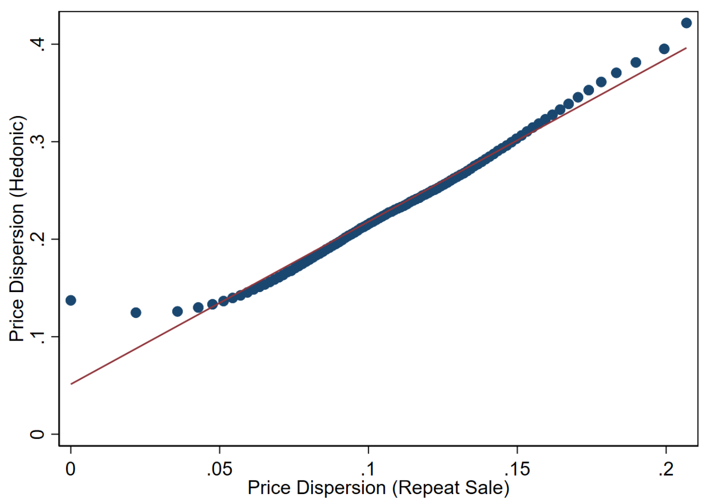
To solve the problem, we discretize ζ_t into 8 states using the Tauchen (1986) method. We use a 150-point approximately exponential grid for w_t, and a 7-point grid for house qualities.

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.

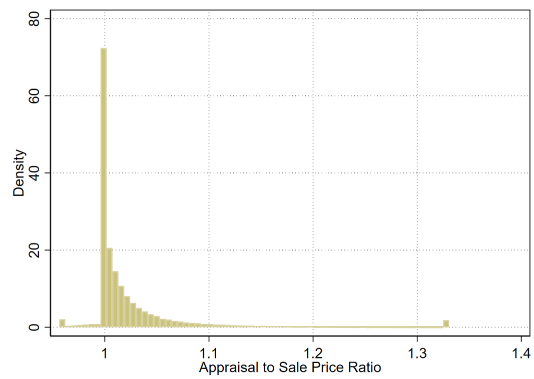
F Appendix Figures and Tables

Figure A1. Repeat-Sales Estimates and Hedonic Estimates



Notes: 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.

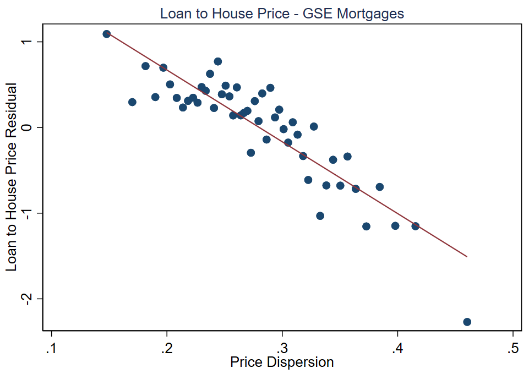
Figure A2. Appraisal Distribution



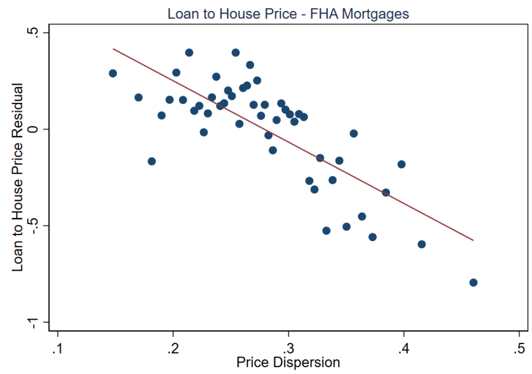
(a) Appraisal to Price Ratio

Notes: This figure shows the histogram of appraisal-to-transaction price ratios, winsorized at 1% to remove outliers. The sample includes loan level observations from 2000 to 2020. *Source*: Corelogic LLMA.

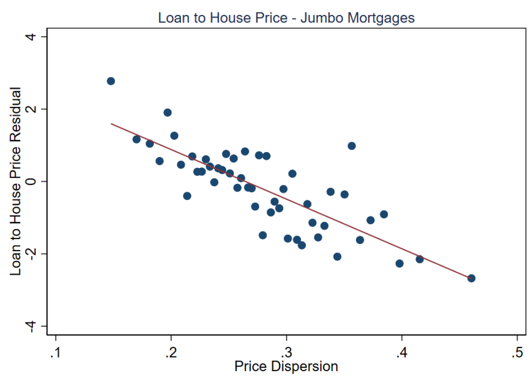
Figure A3. County Level House Price Dispersion and LTP



(a) GSE



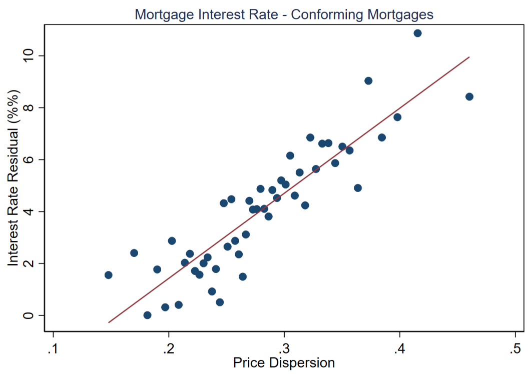
(b) FHA



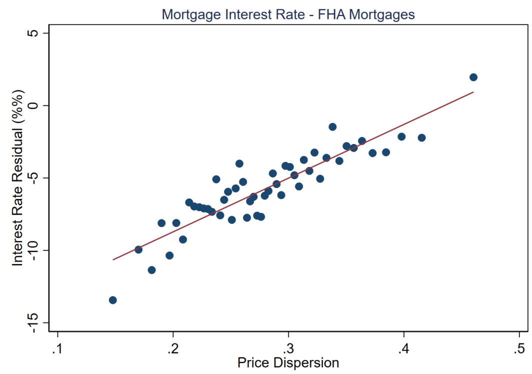
(c) Jumbo

Notes: 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.

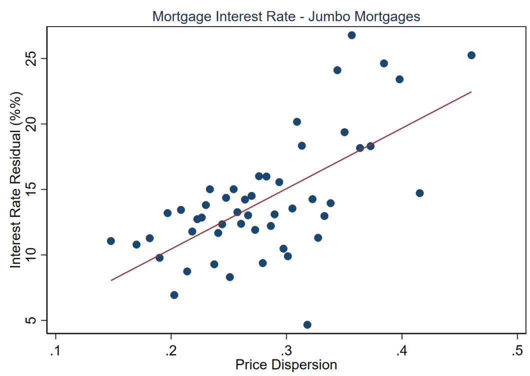
Figure A4. County Level House Price Dispersion and Mortgage Rate



(a) GSE



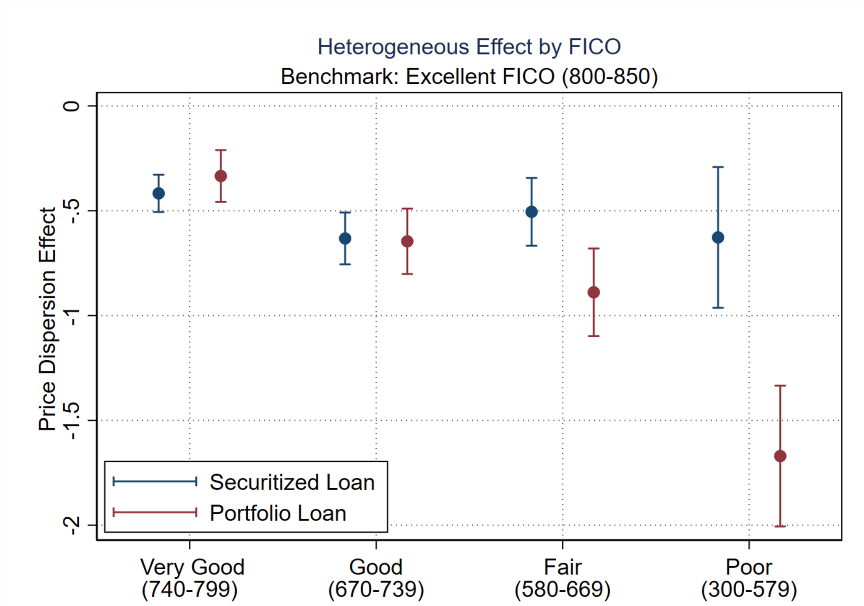
(b) FHA



(c) Jumbo

Notes: 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 A5. Heterogeneous Effect of Price Dispersion by FICO

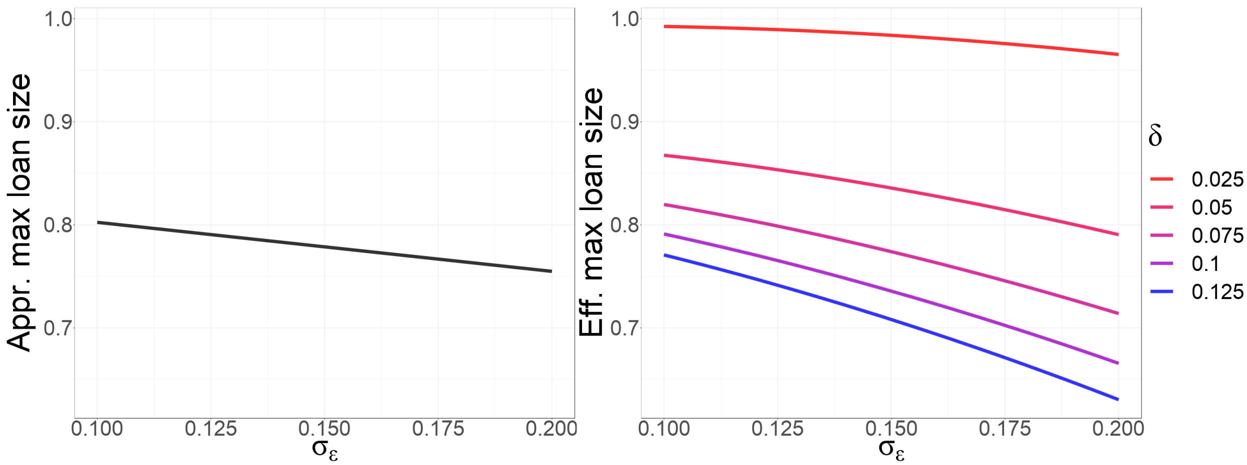


Notes: This figure shows heterogeneous effect of price dispersion by FICO score. We estimate the following specification:

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

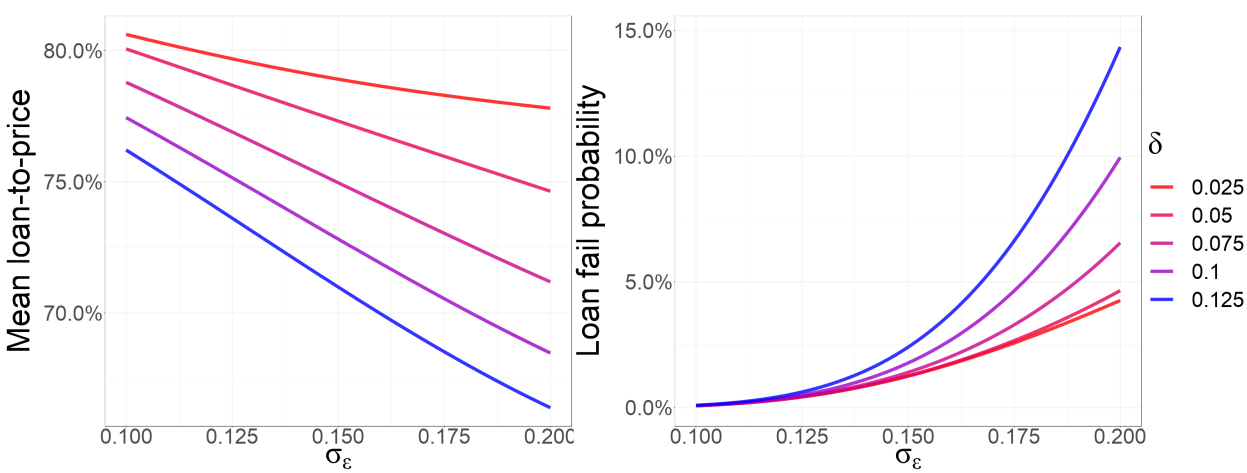
where $ZipDispersion_{ikt} \times CreditScore_{ikt}$ is zipcode price dispersion interacted with home buyer's credit score, which is divided into five groups based on lenders' common practice: Excellent (800-850), Very Good (740-799), Good (670-739), Fair (580-669), and Poor (300-579). X_{ikt} includes zipcode price dispersion, credit score, and other controls in Table 5. 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 A6. Behavior of $\bar{L}_{appr}$ and $\bar{L}_{fair}$



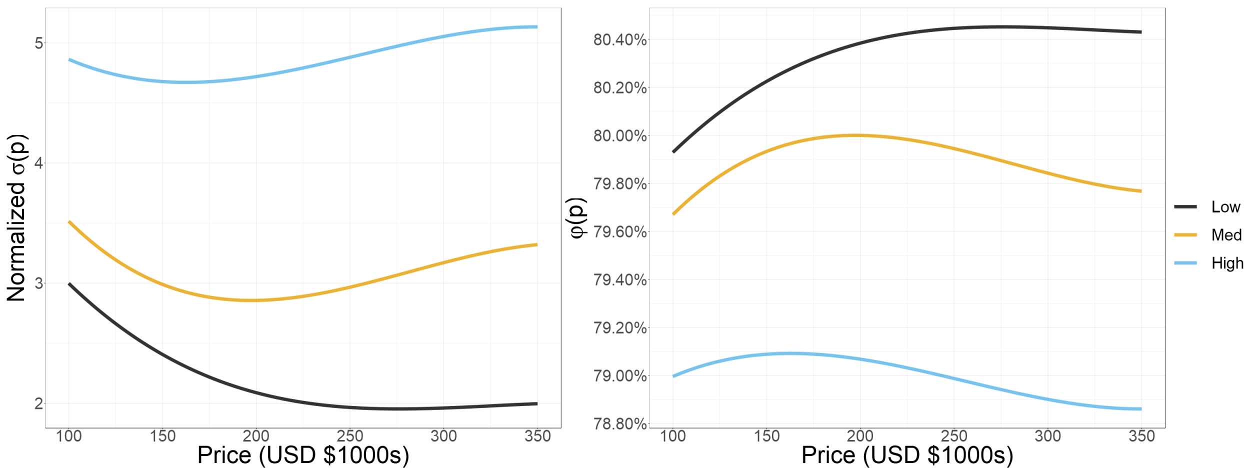
Notes: In the above figure, the left panel shows the behavior of the average value of $\bar{L}_{appr}$ for successful loans (which does not depend on δ), and the right panel shows the average value of $\bar{L}_{fair}$, as σ_ϵ varies, for different values of δ. Throughout, we set $\phi = 0.85$, $c = 0.2$,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

Figure A7. LTP and fail probabilities



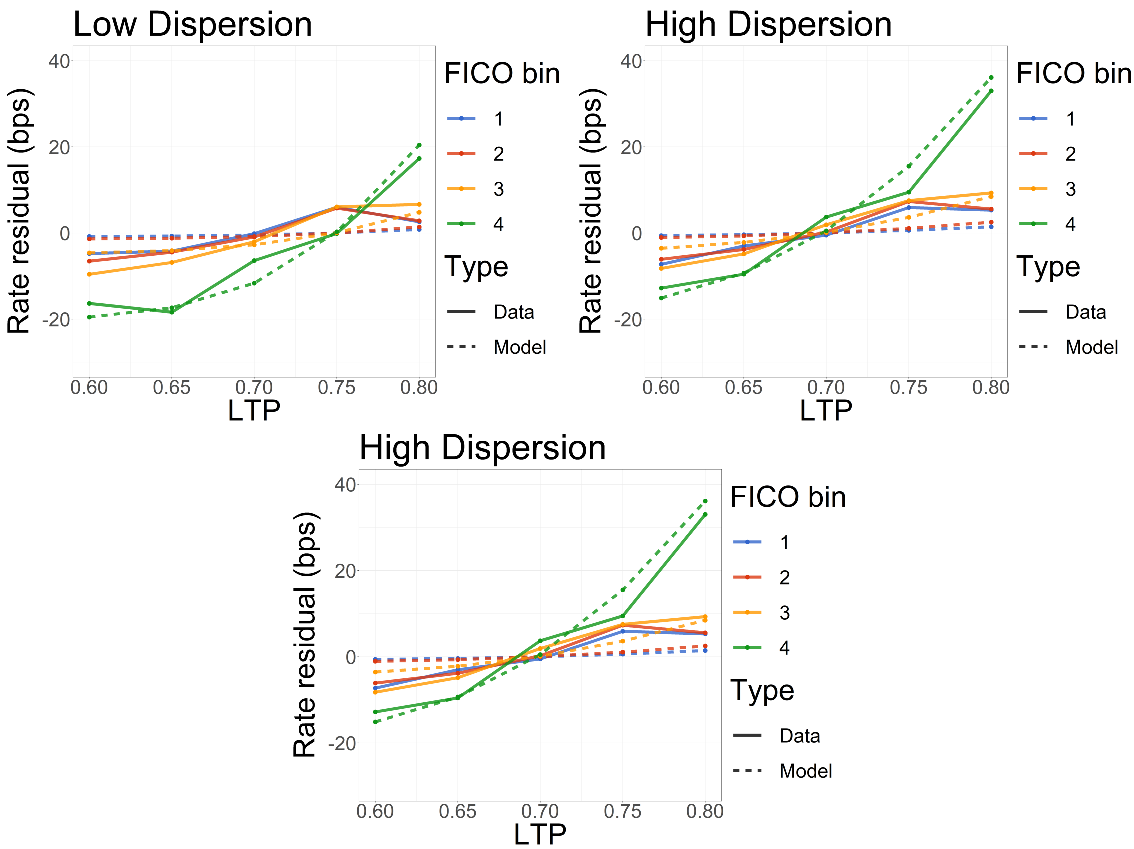
Notes: In the above figure, the left panel shows the mean loan-to-price ratio. The right panel shows the probability of loans failing. Colored lines represent different values of δ. Throughout, we set $\phi = 0.85$, $c = 0.2$, $r - \rho = 0.005$.

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



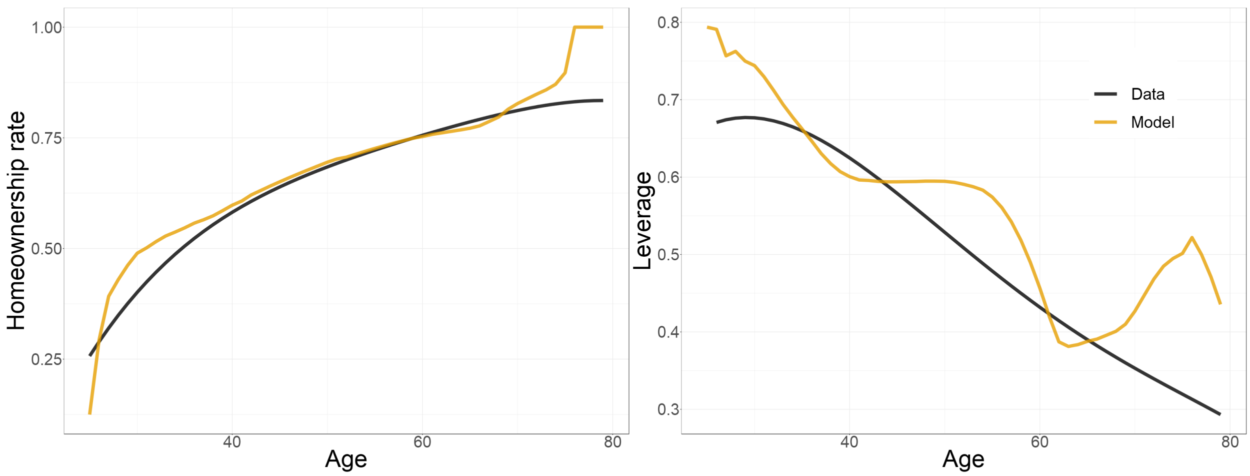
Notes: The left panel shows $\sigma(p)$, the average of price dispersion σ conditional on house prices, for the low, medium, and high dispersion versions of our calibration. We normalize $\sigma(p)$ by its standard deviation across houses, the same units used in Table A4. The right panel shows the resultant $\phi(p)$ 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



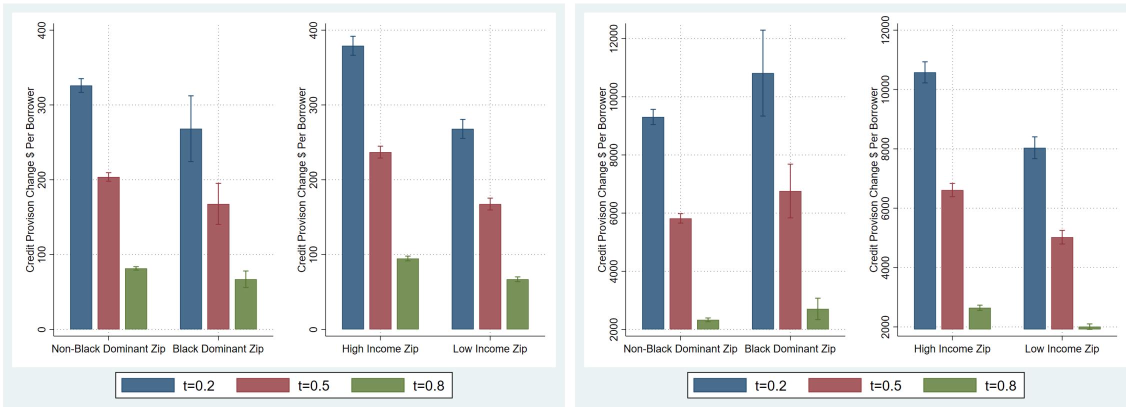
Notes: This figure shows our model fit of the rate menu. The top two panel shows empirical interest rate residuals $\tilde{r}_{fld}$ (solid lines), from (A30), and model-predicted rate residuals $\tilde{r}^{model}_{fld}(c)$ (dashed lines), from (A32), in the fitted model. LTP ratios are shown on the 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



Notes: 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.

Figure A11. Alternative LTV Policies — Dollar Amount



(a) All Borrowers **(b)** Conditional on Appraisal < Sale Price

Notes: This figure presents counterfactual credit provision under three alternative appraisal policies by zipcode minority population. In calculating LTV for various regulation and securitization requirements, the baseline policy takes house sale price as the value of house when the sale price is smaller than the appraisal value and takes appraisal as the value of house when the appraisal value is smaller than the sale price. In the three alternative appraisal policies, we apply the same rule when the sale price is smaller than the appraisal value and change the formula used to calculate LTV when the appraisal value is smaller than the sale price. From Alternatives 1-3, we gradually increase the weight on the appraisal value: in Alternative 1, the house value is 80% of sale price plus 20% of appraisal value; in Alternative 2, the house value is 50% of sale price plus 50% of appraisal; and in Alternative 3, the house value is 20% of sale price plus 80% of appraisal. Panel (a) presents additional dollar amount for all borrowers. Panel (b) presents additional dollar amount for borrowers whose appraisal values are less than the sale prices. *Source*: Corelogic LLMA and Deeds and American Community Survey 2008-2012.

Table A1: 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 A2: 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 A3: 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 A2. The outcome variable in column 2 corresponds to column 1 in Table A2, in column 3 corresponds to column 2 in Table A2, and in column 4 corresponds to column 3 in Table A2. 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 zipcode 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 A4: 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 zipcode 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 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: Property-Level House Price Dispersion and LTP — Without House Price as a Control

This table presents property-level regression results without house price as a control. Columns 1-2 present OLS results. Columns 3-4 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 1-2 is property-level house price dispersion, scaled by its standard deviation, and is the predicted price dispersion in columns 3-4. Controls include 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)
Price Dispersion	-0.30***	-0.33***	-1.34***	-1.36***
	(0.055)	(0.046)	(0.159)	(0.147)
Controls	✓	✓	✓	✓
Transaction Date FE	✓	✓	✓	✓
County-Year FE	✓	✓	✓	✓
Lender-Year FE		✓		✓
R2	0.32	0.37	0.25	0.22
Observations	28M	28M	28M	28M
Underidentification test statistic			159.98	163.34
Underidentification test p-value			0.00	0.00
Weak identification test statistic			232.70	230.86

Table A7: 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 A8: 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