Collateral Value Uncertainty and Mortgage Credit Provision

Erica Xuewei Jiang and Anthony Lee Zhang*

January 2022

Abstract

Using property transaction and financing data, we document large cross-sectional differences in how effective houses are as collateral for mortgages. Older and less standardized houses tend to have higher price dispersion, and their appraisal values tend to deviate more from transaction prices. Mortgages collateralized by these houses are more likely to be rejected, have lower loan-to-price ratios, and higher risk-adjusted cost menu. This effect is stronger for home buyers with higher default risk, consistent with the collateral channel. We quantify the effect on homeownership gap using a life-cycle model with collateral constraints. We discuss the implications of our findings for FHA mortgage program, the shift from human to automated appraisals, and housing affordability policies.

*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-Shion Lin, and Michael Yip provided outstanding research assistance. We are grateful to Kenneth Ahern, Jason Allen, Eric Budish, Sylvain Catherine, Scarlet Chen, Anthony DeFusco, Jason Donaldson, Marco Giacoletti, Larry Harris, Mark Jansen, Gregor Matvos, Norm Miller, Andrii Parkhomenko, Giorgia Piacentino, Amir Sufi, Selale Tuzel, Joe Vavra, Daniel Greenwald, Jinyuan Zhang, Arthur Korteweg, Rodney Ramcharan, participants at Columbia Workshop in New Empirical Finance, the Guthrie Center Housing and Macroeconomics Conference, USC brownbag, UIUC, and the Bank of Canada for comments. We are grateful for financial support from the Fama-Miller Center for Research in Finance at the University of Chicago Booth School of Business. TransUnion (who provided some of the data for this project) has the right to review the research before dissemination to ensure it accurately describes TransUnion data, does not disclose confidential information, and does not contain material it deems to be misleading or false regarding TransUnion, TransUnion’s partners, affiliates or customer base, or the consumer lending industry. First version: August, 2021
1 Introduction

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

In this paper, we document large cross-sectional differences in how effective houses are as collateral for mortgages. Older and less standardized houses have higher price dispersion. We show that the price dispersion of collateral affects financing: mortgages backed by houses with high price dispersion have lower loan to price ratios, higher interest rates, and are more likely to be rejected. This is true both in the private mortgage market, because value uncertainty makes assets less effective as collateral, as well as in the government-backed segment, due to the way that price dispersion interacts with the housing appraisal process. Through a quantitative life-cycle model, we show that the borrowing constraints induced by house price dispersion has a nontrivial effect on homeownership rates.

Policymakers aiming to encourage homeownership for low-income households have considered interventions in credit markets as well as in housing markets. This paper highlights a link between these two markets: the amount of credit that mortgage lenders provide depends on the value uncertainty of the house used as collateral. We show that low-income households tend to live in areas with older and less standardized houses, which are intrinsically difficult to lend against. Our findings provide a rationale for interventions in the mortgage market, such as the FHA program, which allows low-income households to borrow at higher LTV ratios. These policies alleviates a structural feature of the housing stock which limits low-income households’ credit access even in efficient mortgage markets.

The paper unfolds in two steps. First, we establish that collateral value uncertainty affects home buyers’ mortgage credit access. We begin by using rich residential property

¹See 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.
transaction data from 2000 to 2020 to document substantial cross-sectional variation in the predictability of house prices. We find that older and less standardized houses in terms of the number of bedrooms or square footage have more value uncertainty as measured by the predicted pricing errors of a sophisticated hedonic model. Aggregating price dispersion to zip code level, we show that zip code price dispersion is persistent over time. Thus, differences in price dispersion across regions appear to be mainly driven by characteristics of local housing stocks.

We link property transaction data to mortgage records and show that collateral value uncertainty affects financing at both the extensive margin and the intensive margin. At the regional level, counties with higher price dispersion have more mortgage rejections, lower average loan-to-price ratios conditional on loan approval, and higher interest rates. At property level, comparing two houses which are transacted in the same zip-year at the same transaction price, the house with higher estimated price dispersion tends to receive lower loan-to-price ratios (LTPs). The result holds when we further restrict to comparing houses financed by the same lender. One average, LTPs are 20-46bps lower for houses with one standard deviation higher estimated price dispersion.

Lower LTPs could in principle be caused by differences in credit demand rather than credit supply: borrowers of houses with higher price dispersion could simply have lower demand to borrow, causing them to substitute to smaller loans with lower interest rates. To address this concern, using loan-level data, we estimate the menu of LTP-interest rate pairs that are available in any given zip code-year. We find that, in high-dispersion zip codes, the menu of risk-adjusted interest rate-LTP pairs shifts upwards: for any given LTP, borrowers in high-dispersion zip codes can expect to pay higher risk-adjusted interest rates.

We then show that the relationship between collateral value uncertainty and LTP is stronger for borrowers with higher default risk, consistent with the collateral channel driving our results. A one standard deviation increase in price dispersion induces a 42bp average decrease in LTP for homebuyers with the highest credit scores, compared to a 110bp average effect for homebuyers with the lowest credit scores. The effect of credit scores on the dispersion-relationship holds for both securitized and portfolio loans, but the effect is much stronger for portfolio loans.

We also show that mortgage applications in zip codes with higher house price dispersion are more likely to be rejected. A one standard deviation increase in zip code average price
dispersion is associated with a 1.4-1.8 percentage point increase in the mortgage rejection rate, and an about 50bps increase in rejections explicitly attributable to problems with collateral.

The relationship between price dispersion and LTPs is plausibly driven by fair pricing of debt for portfolio loans that banks keep on their balance sheets. However, it is unclear that lenders have incentives to fairly price the effects of collateral dispersion in loans that they plan to securitize (Hurst et al., 2016; Mian and Sufi, 2009; Keys et al., 2010; Purnanandam, 2011). We show that the dispersion-LTP relationship in the securitized segment of the market works through the residential appraisal process. Securitizers set underwriting policies based on the loan-to-value ratio of a house, which takes the loan amount divided by the minimum value of the transaction price and the appraisal value. Appraisal values are based on recent sale prices of comparable properties. When idiosyncratic price dispersion is higher, appraisal values are noisier, creating downward pressure on the amount of loans that can be borrowed against the underlying properties. Empirically, we show that house transactions receive appraisals farther below their transaction prices in zip codes with higher price dispersion, consistent with the appraisal channel.

We perform three robustness tests of our results. First, we show that homebuyers of houses with higher price dispersion are not more likely to default on their loans, suggesting that our results are not driven by unobserved differences in borrowers’ creditworthiness that are correlated with house price dispersion. Second, our results hold even with lender-zip-year fixed effects, suggesting that the results are not driven by lender market power, or other features of lenders’ behavior which affect all houses within a zipcode uniformly. Third, our findings hold even restricting to a subsample of houses with sale prices below conforming loan limits, suggesting that our findings are not driven by home buyers reducing borrowing amounts to be eligible for GSE or FHA loans.

In the second part of the paper, we build a life-cycle model of housing choice to show how the variation in borrowing constraints induced by collateral value uncertainty influences housing affordability and homeownership rates. In the model, households allocate stochastic labor income between consumption and different qualities of housing over their lifecycles. Households can borrow up to an LTV threshold, which is tied to the quality of the owned housing. We evaluate the effects of collateral value variation by calibrating three versions of the model, to areas with high, moderate, and low house price dispersion. We use our
reduced-form estimates to inform the extent to which maximum mortgage LTVs vary across these three scenarios. We find that moving from a high-dispersion county to a low-dispersion county can increase aggregate homeownership rates by roughly 1.5pp, with larger effects on poorer households (2.6pp), for whom down payment constraints are most binding.

In support of our model results, we show empirically that counties with higher house price dispersion have lower homeownership rates, controlling for household income and other county characteristics. A 1SD increase in price dispersion is associated with approximately a 1.3pp decrease in homeownership rates. Purging the difference in other county characteristics, the effect of 1SD increase in dispersion on homeownership is 1pp greater for households in the lowest income decile.

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

Our results have a number of policy implications. First, we provide a rationale for interventions, such as the FHA loan insurance program, which extend credit to low-income households and first-time homebuyers at loan-to-value ratios much higher than private lenders. We have shown that low-income households face particularly high barriers to homeownership because they tend to live in high-dispersion areas, so lack access to housing with high collateral values. Thus, mortgage credit access is limited precisely for those households who are most down-payment constrained, for whom credit is most valuable. Government interventions to increase credit provision for low-income households, such as the FHA loan program, can alleviate this effect.

Second, our results suggest that the impending shift from human appraisals to automated appraisals, such as the Desktop Appraisal program, may have adverse effects on mortgage credit provision, especially for low-income households. Human appraisals are known to be distorted upwards, so only a small share of houses under-appraise. Automated appraisals are

---

2In 2021, the FHFA announced that banks and mortgage lenders could use automated appraisal software in place of human appraisals. https://www.americanbanker.com/news/fhfa-will-make-desktop-home-appraisals-a-permanent-option
likely to be less distorted, but this may lead to a higher share of under-appraisals, especially in areas where house price dispersion is high. Automating appraisals may thus actually reduce credit access to low-income and minority households, who tend to live in areas with high price dispersion.

Finally, our results suggest that urban policymakers should consider the effects of their policy instruments on how difficult it is to finance house purchases. In many parts of the US, housing construction and renovation is highly regulated: local policymakers impose restrictions on various characteristics of individual houses and also influence the overall rate of renewal of the housing stock. These policies affect the overall value uncertainty of the housing stock and thus how effective houses are as collateral for mortgage. Our results imply that policies that encourage renewal of the housing stock, and promote construction of cheap, homogeneous houses with predictable values could expand credit access, without requiring the government to intervene directly in mortgage markets. These credit increases are disproportionately valuable to low-income households, for whom credit is most important for homeownership.

This paper relates to a number of strands of literature. Broadly, our paper fits into a literature on frictions that affect credit access. (Mian and Sufi, 2011; Agarwal et al., 2017; Beraja et al., 2019; DeFusco et al., 2020; Buchak et al., 2018a; Jiang, 2020) and homeownership (Glaeser and Shapiro, 2003). DeFusco and Mondragon (2020) study two counter-cyclical refinancing frictions – the need to document employment and the need to pay upfront closing costs – and show these frictions prevent borrowers who experience income shocks to refinance. DeFusco (2018) studies how changes in access to housing collateral affect homeowner borrowing behavior and estimate the marginal propensity to borrow out of housing collateral. Greenwald (2016) uses a general equilibrium framework to study how payment-to-income and loan-to-value ratios affect macroeconomic dynamics. We also relate to a classic literature analyzing how collateral values affect the properties of debt contracts collateralized by these assets or firms’ investment decisions (Titman and Wessels, 1988; Shleifer and Vishny, 1992; Bian, 2021). Benmelech and Bergman (2008) analyzes the effect of collateral liquidation values on contract renegotiation, and Benmelech and Bergman (2009) studies how collateral values affect the cost of debt in the context of commercial real estate. Our paper also builds on a literature on idiosyncratic price dispersion in the housing market and its consequences. Case and Shiller (1989) and Giacoletti (2017) analyze idiosyncratic
risk in residential real estate markets. Hartman-Glaser and Mann (2017) documents that lower-income zip codes have more volatile returns to housing than higher-income zip codes. They rationalize the finding with a model where shocks to the representative household’s marginal rate of substitution lead to volatility in the return to housing via the collateral constraint, and lower-incomes have a more volatile marginal rate of substitution, and thus more volatile returns to housing. Sagi (2021) analyzes idiosyncratic risk in commercial real estate. Sklarz and Miller (2016) propose a method to adjust loan-to-value ratios to reflect house value uncertainty.

The contribution of this paper is that we are the first to show that the collateral channel matters in the US residential real estate market: there is substantial cross-sectional heterogeneity in housing collateral values, which affects mortgage credit availability and housing affordability. Our analysis also elucidates the mechanisms through which the collateral channel influences outcomes within the unique structure of the US residential mortgage market: in particular, how house price dispersion interacts with GSE securitization, regulatory constraints on banks, 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 calibrates a model to analyze the effect of collateral values on homeownership rates. We discuss implications of our results in section 5, and conclude in section 6.

2 Data and Measurement Strategy

2.1 Data Sources

House Transaction Data. We use Corelogic Deed & Tax records data on housing transactions in the US. We use data between 2000 and 2020 and restrict the sample to arms-length, non-foreclosure transactions in single family residences. The date set reports each house transaction attached to a specific property and provides information on sale amount, mortgage amount, transaction date, and property location. We exclude transactions with missing sale price, date, property ID, or location information. We merge the transaction records with the Corelogic tax records to get property characteristics like year built and
square footage. We provide detailed description about data cleaning steps in Appendix A.1.

**Mortgage Data.** Besides the mortgage information in Corelogic Deeds, we use two additional sources of mortgage data. First, we use the Corelogic Loan-Level Market Analytics (LLMA) database. The database provides detailed information on mortgage and borrower characteristics at origination – interest rates, down payments, sale prices, credit score, and debt-to-income ratio – and monthly loan performance of the loan, including delinquency status and investor type. Important for our analysis, the LLMA provides both appraised house value and transaction price.

Second, we use the Home Mortgage Disclosure Act (HMDA) for extensive margin analysis on mortgage application rejections. 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. Since we are not allowed to merge the LLMA and the Deeds records, we aggregate the estimated idiosyncratic price dispersion, introduced in detail in the next subsection, to the finest geographic regions in the LLMA (5-digit zip code) and in the HMDA (census tract).

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

Table 1 provides summary statistics.

### 2.2 Measuring Value Uncertainty

We measure house value uncertainty as estimated idiosyncratic house price dispersion: which types of houses have smaller pricing errors when priced with a hedonic regression. A similar methodology is used in Buchak et al. (2020). Our methodology has two steps. First, we first regress transaction prices on house characteristics:

\[
p_{it} = \eta_{kt} + f_k(x_i, t) + \epsilon_{it},
\]

where \(i\) indexes properties, \(k\) indexes counties, and \(t\) indexes months. Equation 1 is effectively a hedonic specification for house prices: log prices \(p_{it}\) are determined by a county-month fixed
effect, $\eta_{ct}$, a smooth function $f_c(x_i, t)$ of observable house characteristics $x_i$ and time $t$, and a mean-zero error term $\epsilon_{it}$. $\eta_{ct}$ absorbs parallel shifts in log house prices in a county over time. $f_c(x_i, t)$ allows houses with different observable characteristics $x_i$ to appreciate at different rates. For example, $f_c(x_i, t)$ allows larger houses to appreciate faster than smaller houses, or houses in the east of a certain county to appreciate faster than houses in the west.

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:

$$\hat{\epsilon}_{it}^2 = g_c(x_i, t) + \xi_{it}$$

(2)

where $g_c(x_i, t)$, like $f_c(x_i, t)$, is a generalized additive model in features, which we describe in Appendix A.2. Using the estimates from this equation, we construct the predicted standard deviation of the pricing error for every house:

$$\hat{\sigma}_{it}^2 \equiv \hat{g}_c(x_i, t)$$

(3)

The predicted values of idiosyncratic price dispersion at the house level, $\hat{\sigma}_{it}^2$, are our measure of house value uncertainty. For ease of interpretation, we will generally take the square root of $\hat{\sigma}_{it}^2$, which we call $\hat{\sigma}_{it}$, and use this in our analysis.

Intuitively, specifications (1) and (2) is a heteroskedastic hedonic model of house prices: they allow both the mean and idiosyncratic dispersion in house prices to be a function of house characteristics. Specification (2) estimates, for a given house $i$, whether $\hat{\epsilon}_{it}^2$ terms, the squared errors from (1), tend to be large for houses with similar characteristics to $i$: that is, whether $i$ is priced with low error using a hedonic regression. As an example, suppose large houses are more expensive than small houses, but also have lower price dispersion. $f_c(x_i, t)$ in specification (1) will be higher for large houses, capturing the fact that $p_{it}$ tends to be higher for large houses. $g_c(x_i, t)$ in specification (2) will tend to be smaller for large houses, capturing the fact that the squared deviation of $p_{it}$ from its conditional mean (that is, the hedonic model prediction) tends to be smaller for large houses.\(^3\)

\(^3\)Note that it is important to first run specification (2), using the predicted values of $\hat{\sigma}_{it}^2$, rather than using the squared residuals $\hat{\epsilon}_{it}^2$ in regressions directly. This is because the expected value of idiosyncratic dispersion, $\sigma_{it}^2$, is the analog of $\sigma$ in our model, which is relevant for loan-to-values. Each realization of $\hat{\epsilon}_{it}^2$ is a noisy measure of $\sigma_{it}^2$. 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 $\sigma_{it}$, due to measurement error bias.
We use our estimates of price dispersion at the house level, and also at higher levels of aggregation. For example, we will use $\hat{\sigma}_c$ to denote the empirical estimate of standard deviation of all $\hat{\epsilon}_{it}$ terms in county $c$. $\hat{\sigma}_c$ can be thought of as the log standard deviation of house price residuals, after controlling for features in (1).

2.3 Estimated Value Uncertainty and Housing Market Frictions

We next present some stylized facts about the estimated value uncertainty of the US housing stock over this time period, and discuss what housing market frictions are captured by the estimated value uncertainty.

Zip code level price dispersion is very persistent over time. Figure 1 plots zip-code idiosyncratic price dispersion in 2020 (2010) against zip code dispersion in 2010 (2000). Over both time periods, zip code dispersion in the later year is lined up with the dispersion in the earlier year: regions with high price dispersion in 2000, for example, also have high price dispersion in 2010. This suggests that the differences in price dispersion are driven by persistent characteristics of a local housing stock rather than time-varying local market condition.

To explore this further, Table 2 presents the association between estimated value uncertainty and house characteristics as well as local market conditions. 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. Column 1 shows that older houses have higher price dispersion. 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. Columns 2 and 3 show that, controlling for the age of the building, recently renovated houses – defined as houses with effective year built within 5 years of the transaction date – have lower price dispersion, and houses which were ever renovated have lower price dispersion.

In columns 4-6, we analyze the association between building square footage 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. Similarly, columns 5-6 show that 4-bedroom houses have lower price dispersion than houses with more or less than 4 bedrooms. In terms of local housing market conditions, Panel B of Table 2 shows that
houses in zip codes with larger income inequality, less population density, and more vacancies tend to have higher price dispersion.

Together, Table 2 suggests that house price dispersion is essentially driven by house standardization and market thickness. Controlling for prices, houses which are newer, located in densely populated areas, which are close to the median in terms of characteristics, have lower price dispersion; older and less standardized houses have higher price dispersion.

Figure 2 shows the relationship between price dispersion, average zipcode incomes, and the zipcode minority share. Price dispersion tends to be higher in low-income, high-minority share zipcodes. This is largely driven by the fact that price dispersion is generally decreasing in prices, and house prices are very correlated with average incomes and minority shares across zipcodes.

2.4 Drivers of Idiosyncratic Price Dispersion

Next, 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’

\footnote{This finding is consistent with evidence from other papers: see, for example, Kotova and Zhang (2021) and Andersen et al. (2021).}
equity position (Genesove and Mayer, 1997; Guren, 2018) and homeowners’ nominal losses since purchase (Genesove and Mayer, 2001). 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%, suggesting substantial dispersion in buyers’ values for similar houses. 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. As we argue in section 3.1 below, 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 (2017), 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.

3 House Value Uncertainty and Mortgage Credit

3.1 Theoretical Framework

We present a simple model illustrating how house value uncertainty affects mortgage credit provision in Appendix B, and explain the main intuitions here. In our model, house value uncertainty affects credit provision through two channels. The first channel, which we call the *fair pricing* channel, is that each lender has some maximum amount she is willing to lend against a given house for any given interest rate, which depends on how volatile the foreclosure price of the house is. Formally, if the lender must foreclose and sell the house, her payoff is concave in the house sale price: if the house sells for more than the outstanding debt, the lender cannot keep the surplus, whereas if the house sells for less the lender is on the hook for the difference. Thus, if the house price is more volatile, the lender expects larger losses upon foreclosure, and thus must lower loan-to-value ratios to maintain a given profit margin.⁵

The second channel, which we call the *appraisal channel*, is unique to the residential housing market. Residential properties are appraised almost exclusively through comparable-sales approach: appraisers identify similar properties which were transacted in recent months, called “comps”, adjust the comparable properties’ prices for differences in characteristics between the comp and the property to be appraised, and then take a weighted average of comparables’ prices.⁶ Mortgage securitizers then set underwriting policies based on the loan-to-value ratio of a house, where the value is calculated as the minimum of the transaction price and the appraisal value. This minimum implies that values are a concave function of the

---

⁵This argument is related to a large literature on collateral and debt (Williamson, 1988; Harris and Raviv, 1991; Aghion and Bolton, 1992; Shleifer and Vishny, 1992; Hart and Moore, 1990; Bolton and Scharfstein, 1996; Diamond, 2004).

⁶Appraisals are governed by the Uniform Standards of Professional Appraisal Practice (USPAP). The USPAP identifies three allowable methods for assessment: a “sales comparison” approach, based on comparable sales; a “cost” approach based on the cost of building the property, and an “income” approach based on the rental payment flows from the property. In practice, residential real estate appraisal uses almost exclusively the sales comparison approach.
appraisal price: if the house over-appraises, the transaction price is used to determine loan-to-values, whereas if the house under-appraises the appraisal binds. Thus, when idiosyncratic price dispersion is high and appraisal values are noisier, there is downward pressure on mortgage loan amounts.

Both the fair pricing and appraisal channels imply that houses with higher value uncertainty are more difficult to lend against. Formally, the model makes three predictions. First, for any given interest rate, loan-to-price ratios should be lower for houses with higher value uncertainty. Second, mortgages are more likely to be rejected when value uncertainty is higher. In the model, a mortgage application is rejected if the LTV that the buyer demands, which depends on the buyer’s cash-on-hand for down payments, is higher than the maximum LTV that the lender offers. When house value uncertainty is higher, lenders’ LTV limits will be lower, so a larger fraction of mortgage applications will be rejected. Lastly, the relationship between value uncertainty and LTPs should be stronger for borrowers with higher default risks. This is because, in the fair pricing channel, the value of collateral matters only if consumers actually default: for prime consumers with low default rates, LTPs will tend to be both higher and less sensitive to default rates. We proceed to show that these predictions hold empirically.

### 3.2 County-Level Evidence

We begin with county-level evidence in Figure 3. We show that counties with higher price dispersion have lower average loan-to-price ratios, higher interest rates, and more mortgage rejections.

Panel (a) of Figure 3 plots county average loan to house price (LTP) against the average house value uncertainty after controlling for local house prices. There is a robust negative relationship: counties with higher house value uncertainty have lower average loan-to-price ratios. In Figure A7, we show that the relationship holds for GSE loans, FHA loans, and jumbo loans. Figure A8 confirms the negative relationship using residualized LTP, by taking the residuals of regressions of LTP on mortgage interest rate, debt-to-income ratio (DTI), DTI-square, FICO, FICO-square, log house price, and their interactions with origination years, and origination year fixed effects.

Panel (b) of Figure 3 plots county average residualized mortgage rate against the average
house price dispersion. We residualize rates for individual mortgages on borrower and loan characteristics such as FICO, LTP, DTI, the squared terms, and their interactions with origination year. We find that interest rates are higher in areas with higher price dispersion. In other words, mortgages in high-dispersion areas are worse for borrowers along both the price and LTP dimensions. This is inconsistent with the hypothesis that the relationship between LTP and dispersion is driven by lower demand for credit in areas with lower price dispersion, rather than lenders being less willing to supply credit. If lower LTPs reflect lower credit demand, interest rates should also be lower in areas with higher price dispersion, since home buyers on average are taking out safer, better-collateralized mortgages. Figure A9 shows that the relationship holds for GSE loans, FHA loans, and jumbo loans. We will show further evidence for this in Figure 5 below, which shows that the menu of LTP-interest rate pairs offered in high-dispersion areas appears to be uniformly less favorable for borrowers.

Panels (c) and (d) of Figure 3 plot county-level average mortgage rejection rates against county-level price dispersion. Panel (c) shows all mortgage rejections, and panel (d) considers only mortgage rejections which are tied to collateral quality. We calculate rejection rates by taking the total number of rejected mortgages, and dividing by the total number of mortgages in the HMDA data. We residualize mortgage rejection rates, by taking the residuals from a regression on average log house prices, credit scores, and year fixed effects. Consistent with the extensive margin prediction, mortgage rejection rates are higher in counties with higher price dispersion. In particular, the fraction of mortgages rejected for collateral-related reasons is higher in high-dispersion counties.

3.3 Property-Level Evidence

We next turn to property level evidence, which allows us to exploit within-county variation and better control for borrower characteristics for identification.

3.3.1 Loan-to-Price Ratio

We start with the effect of collateral value uncertainty on property-level LTP. We first visualize the relationship between property-level LTP and collateral value uncertainty in Figure
To compare properties within a county, we regress LTP, and predicted house price dispersion from specification (3), on house prices and county-year fixed effects, and plot the LTP residuals against house price dispersion residuals. Figure 4 shows a clear negative association between LTP and collateral value uncertainty: houses with higher predicted price dispersion receive less credit than other houses in the same county and transacted in the same year.

Formally, we estimate the following property-level specification:

\[ LTP_{ikt} = \alpha + \beta \text{Dispersion}_{ikt} + X_{ikt} \Gamma + \mu_{kt} + \nu_{m} + \epsilon_{ikt} \]  

\( LTP_{ikt} \) and \( \text{Dispersion}_{ikt} \) are the loan-to-sale price ratio and the estimated price dispersion, respectively, of property \( i \) in county \( k \) sold in year \( t \). \( X_{ikt} \) is a set of property and zip-code level controls. Specifically, we include the transaction price of the property, mortgage type, mortgage term, and resale indicator. \( \mu_{kt} \) and \( \nu_{m} \) are county-year and transaction month fixed effects, respectively.

Table 3 reports the results. We first confirm the effect of house price dispersion on mortgage loan-to-price ratio in a less saturated specification by including only loan controls in column 1. For two houses with the same transaction price, the one with higher estimated price dispersion tends to receive a smaller sized loan. The result holds in more saturated specifications with transaction date fixed effects (column 2), plus county-year fixed effects (column 3), plus lender-year fixed effects (column 4), and when we county-year with zip-year fixed effects (column 5). The loan-to-price ratio is about 20-46bps lower for houses with one standard deviation higher estimated price dispersion across these specifications.

### 3.3.2 Mortgage Price Menus

Once again, lower LTPs could be explained by credit demand: borrowers of houses with higher price dispersion could simply have lower demand to borrow, and thus substitute to smaller loans with lower interest rates. Using property-level data, we can evaluate this hypothesis by estimating the entire menu of LTP-interest rate pairs that are available in a market. Figure 5 plots the mortgage price menu, separately for groups of zip codes with high and low price dispersion. Formally, we first residualize interest rates on borrowers’ credit scores, a dummy for conforming vs jumbo loans, and then a time fixed effect; and then plot a menu of average interest rates for different LTPs. Figure 5 shows that, in high-dispersion
zip codes, the entire menu of interest rate-LTP pairs shifts upwards: for any given LTP, borrowers in high-dispersion zip codes can expect to pay higher prices. The difference is about 3bps for loans with LTP below 80 and enlarges to 8bps for loans with LTP above 80.

Next, we estimate the following loan-level specifications:

\[
Rate_{ikt} = \alpha + \beta LTP_{ikt} + \gamma ZipDispersion_{ikt} + X_{ikt} \Gamma + \mu_{kt} + \nu_m + \epsilon_{ikt}
\]

\[
LTP_{ikt} = \alpha + \beta Rate_{ikt} + \gamma ZipDispersion_{ikt} + X_{ikt} \Gamma + \mu_{kt} + \nu_m + \epsilon_{ikt}
\]

(5)

\(Rate_{ikt}\) is property \(i\)’s mortgage rate. \(LTP_{ikt}\) is the loan-to-sale price ratio of property \(i\) in county \(k\). \(ZipDispersion_{ikt}\) is the zip code average price dispersion. \(X_{ikt}\) is a set of property and zip-code level controls. Specifically, we include the transaction price of the property, borrower credit score, credit score squared, debt-to-income ratio (DTI), DTI-squared, and loan type. \(\mu_{kt}\) and \(\nu_m\) are county-year and transaction month fixed effects, respectively.

Table 4 reports the cost menu results. Panel A shows the interest rate results. Column 1 uses the full sample. We first confirm that higher loan-to-price ratios are associated with higher interest rates. The coefficient on LTP is positive and statistically significant. A one percentage point increase in LTP is associated with about 0.85bps increase in interest rate. Controlling for LTP, houses in zip codes with higher house price dispersion are financed with more expensive mortgages. The mortgage rate increases by 1.1bps in zip codes with one standard deviation higher average house price dispersion. Columns 2 to 4 shows the results for GSE loans, FHA loans, and portfolio loans, respectively. The results hold in all samples. For every one standard deviation higher zip-code average house price dispersion, the mortgage rate of GSE loans increases by 1.24bps, the mortgage rate of FHA loans increases by 1.23bps, and the mortgage rate of the portfolio loans increases by 0.43bps.

Panel B of Table 4 shows that to receive the same mortgage rate, houses with more higher price dispersion require more down payment. Controlling for interest rate, the loan-to-price ratios decrease by about 50bps in zip codes with one standard deviation higher average house price dispersion (column 1). The results hold for GSE loans, FHA loans, and portfolio loans (columns 2-4). For every one standard deviation higher zip-code average house price dispersion, the LTP ratios of GSE loans decrease by 44bps, the LTP ratios of FHA loans decrease by 22bps, and the LTP of portfolio loans decrease by 57bps.
3.3.3 Mortgage Rejection Rates

Next, we study the effect of price dispersion on credit access at the extensive margin. Lenders require larger down payments when appraisal values are more likely to deviate from the sale price. If borrowers are facing budget constraints and thus cannot make bigger down payment, their mortgage applications may be rejected. We test this prediction by estimating the following specification:

\[
Rejection_{ijkt} = \alpha + \beta \text{ZipDispersion}_{ikl} + X_{ikt} \Gamma + \mu_{kt} + \nu_{jt} + \epsilon_{ikt}
\] (6)

\(Rejection_{ikl}\) is an indicator for whether a mortgage application from borrower \(i\) in county \(k\) in year \(t\) is rejected. \(\text{ZipDispersion}_{ikl}\) is the zip code average price dispersion. \(X_{ikt}\) is a set of property and zip-code level controls. Specifically, we include zip code house price, log of borrower income, loan type, county average credit score and its square term, and loan-to-income ratio and its square term. \(\mu_{kt}\) and \(\nu_{jt}\) are county-year and lender-year fixed effects, respectively.

Panel A of Table 5 reports the results. We first confirm the effect of local house price dispersion on mortgage rejection in less saturated specifications by including only origination year fixed effects (column 1) and county-year fixed effects (column 2). Column 3 shows the result with county-year fixed effects and lender-year fixed effects. In all specifications, zip code house price dispersion is positively associated with mortgage rejection in a statistically significant way. The rejection rate increases by about 1.4-1.7 percentage points as zip code house price dispersion increases by one standard deviation.

We can show this collateral appraisal channel more directly by focusing on rejections due to collateral reasons. Panel B of Table 5 reports the results. A mortgage application is about 50bps more likely to be rejected due to collateral reasons in a zip code with one standard deviation higher house price dispersion. Again, the results hold in less saturated specification with only year fixed effects (column 1) and county-year fixed effects (column 2) and also in more saturated specification with county-year and lender-year fixed effects (column 3).

In both panels of Table 5, the results hold for GSE loans (column 4), FHA loans (column 5), and jumbo loans (column 6).
3.3.4 Heterogeneous Effect

Lastly, we show that the relationship between collateral value uncertainty and LTPs is stronger for borrowers with higher default risks. The evidence suggests that the LTP-value uncertainty relationship is driven by credit supply.

We estimate the heterogeneous effect of price dispersion by home buyer credit score:

\[ LTP_{ikt} = \alpha + \beta \text{Rate}_{ikt} + \gamma \text{ZipDispersion}_{ikt} \times \text{CreditScore}_{ikt} + X_{ikt}\Gamma + \nu_m + \epsilon_{ikt} \]  

\( \text{ZipDispersion}_{ikt} \times \text{CreditScore}_{ikt} \) is zip code price dispersion interacted with home buyer’s credit score, which is divided into five groups based on lenders’ common practice: Excellent (800-850), Very Good (740-799), Good (670-739), Fair (580-669), and Poor (300-579). \( X_{ikt} \) includes zip code price dispersion, credit score, and other controls defined in Specification 5. We estimate \( \gamma \), which captures the heterogeneous effect of price dispersion by home buyer credit score, for the full sample, securitized loan sample, and the portfolio loan sample, respectively.

Table 6 presents the result. Among home buyers with Excellent credit score, the loan-to-price ratios do not change with house price dispersion after controlling for loan and home buyer characteristics and interest rates. LTP-price dispersion sensitivity increases as home buyers become less credit worthy. Benchmark to the baseline LTP-price dispersion sensitivity among Excellent credit history home buyers, LTP of houses decrease by 42bps, 72bps, 88bps, and 110 bps more for home buyers with Very Good, Good, Fair, and Poor credit scores, respectively, if house price dispersion increases by one standard deviation (column 1). This is consistent with fair pricing of collateral risk on the credit supply side.

Columns 2 and 3 report \( \gamma \) estimates for securitized loans and portfolio loans, respectively. We define securitized loans as loans sold to GSEs, Ginnie, or other investors indicated in the Corelogic LLMA dataset and define portfolio loans as non-FHA loans that are not sold. Overall, the estimates are consistent with fair pricing of collateral risk in either sample: \( \gamma \) estimate is not statistically significantly different from zero among home buyers with Excellent credit score, and LTP-price dispersion sensitivity increases as home buyers become less credit worthy. Yet, the effect of price dispersion on mortgage credit is much less heterogeneous across home buyer credit worthiness in the securitized loan sample. Figure 6 visualizes the heterogeneity across FICO score buckets by plotting the \( \gamma \) estimates in columns 2 and 3.
The LTP-price dispersion sensitivity difference between Poor and Very Good home buyers is about 30bps for securitized loans and is about 135bps for portfolio loans.

### 3.4 Price Dispersion and Appraisals

While the LTP-price dispersion sensitivity is likely driven by fair-pricing for portfolio loans that banks keep on their balance sheets, lenders may not fairly price such collateral dispersion in loans that they plan to securitize (Hurst et al., 2016; Mian and Sufi, 2009; Keys et al., 2010; Purnanandam, 2011). Yet we have shown that the dispersion-LTP relationship holds in both the securitized and portfolio segments of the market. In this section, we provide evidence that price dispersion affects LTPs through the appraisal channel.

Figure 8 shows binned scatter plots illustrating how under-appraisal is associated with price dispersion across zip codes. For mortgage $i$, let $a_i$ be the appraisal price, and $p_i$ be the transaction price of the house. The dependent variable in panel (a) of Figure 8, which we call the appraisal deviation from sales price, is defined as:

$$
ApprDev_i \equiv \frac{a_i - p_i}{p_i} \mathbf{1}(a_i < p_i)
$$

That is, the percent deviation of appraisal prices from transaction prices, multiplied by an indicator for the house under-appraising (that is, the appraisal price $a_i$ being below the sales price $p_i$). This variable captures the downwards pressure that appraisals produce on mortgage limits, combining the probability of under-appraisal with the average magnitude of under-appraisals. Panel (a) of Figure 8 shows that the appraisal deviation from sales prices is much higher in high-dispersion zipcodes, suggesting that the extent to which under-appraisals put downwards pressure on LTVs is larger in high-dispersion zipcodes.

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

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

---

7Most residential mortgages in the United States are sold to a secondary market after origination. For example, from 2004 to 2006, about only 20 percent of all mortgages stayed on lenders’ balance sheet, while the remaining were securitized (Keys et al., 2013); and post crisis about 80 percent were securitized (Buchak et al., 2018b,a; Jiang, 2020).
Panel (b) shows that the probability that a house under-appraises is similar in high- and low-dispersion zipcodes; in fact, underappraisals are slightly less likely in high-dispersion zipcodes, though this difference is not statistically significant in regression form. However, conditional on under-appraisal, the difference between appraisal and sale prices is much larger in high-dispersion areas. The average magnitude of under-appraisal is around 3% in low-dispersion zipcodes, compared to around 5% in high-dispersion zipcodes.

Table 7 confirms Figure 8 findings in regression settings with origination month fixed effects, county-year fixed effects, and borrower and loan controls. In high-dispersion zipcodes, appraisal deviations tend to be larger: a 1SD increase in dispersion is associated with a 3bp change in the appraisal deviation (columns 1-2). This is mostly because, conditional on under-appraisal, houses under-appraise by larger amounts: a 1SD increase in dispersion is associated with a 50-57bp increase in the conditional appraisal deviation (columns 5-6). The probability of under-appraisal is slightly larger in high-dispersion zipcodes, but this effect is not always statistically significant (columns 3-4).

A remaining question is why, as we showed in Subsection 3.3.4, the dispersion-LTP relationship depends on credit scores in the securitized segment of the market. A potential explanation for this is that interest rate increases that GSEs charge for mortgages with high LTVs are larger for borrowers with lower credit scores. Thus, the effective cost of under-appraisals is higher for low-credit-score borrowers, so they may set LTVs more conservatively as a result.

The appraisal requirement creates several distortions in credit supply, though it partially allows securitized loans to price in collateral risk. We discuss two main distortions in this section and provide additional discussions in Appendix C. First, as shown in Table 6, the LTP-price dispersion sensitivity is more heterogeneous across home buyers’ creditworthiness for portfolio loans than for securitized loans, suggesting that there is a limited extent to which securitized loans can fairly price in collateral risk by deploying the appraisal requirements. If we use portfolio loans as efficient-pricing benchmark – under the assumption that lenders fairly price mortgages on their balance sheets – to assess the distortion in credit supply in the securitized loan segment, the results in Table 6 suggest that less credit-worthy home buyers under-pay for the collateral risk at the cost of more credit-worthy home buyers who overly

---

8See the Fannie Mae and Freddie Mac pricing matrices, which specify interest rate adjustments as a function of LTV and credit score
pay for the collateral risk. Since the main focus of this paper is not the distortion created by securitization on the credit supply side, we stop at showing suggestive evidence for potential credit supply distortion and leave quantification of such distortion to future research.

Moreover, 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 characteristics differences, which will introduce additional variance into appraisals. These estimates of predicted appraisal dispersion have similar magnitude to 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.

### 3.5 Robustness

#### 3.5.1 Unobservable Buyer Creditworthiness

An important identification assumption of our empirical design is that home buyers of houses with high price dispersion are not more likely to default on their mortgage after conditioning on observable borrower and loan characteristics. We acknowledge that unobservable borrower characteristics might vary with observed house choices. We address this possibility by assessing the ex-post performance of mortgage loans. Panel A of Table 8 estimates the specification (5), 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. Column 1 includes the full sample. Column 2 restricts the sample to securitized loans. Column 3 restricts the sample to portfolio loans. All regressions include the full set of borrower and loan characteristics as in our main regression specifications.

---

9 Other papers in the literature have roughly similar estimates of idiosyncratic variance, though the precise numbers depend on the specific methodology used; other methods include repeat-sales terms (Kotova and Zhang, 2021), or adjusting sale prices over time using regional house price indices (Giacolletti, 2017).
The results suggest that home buyers of houses with higher price dispersion are not more likely to default on their loans than home buyers of houses with lower price dispersion. This ex-post loan performance finding helps interpret the LTP-price dispersion relationship that we estimate in the previous section. Although we show that the effect of house price dispersion on mortgage credit is robust to various observable measures of borrower creditworthiness, some unobservable characteristics might vary across home buyers that choose different type of houses. However, unobservable features that could contaminate our identification would cause differences in ex-post performance, which we can observe and test for in Panel A of Table 8. By showing that there are no differences in ex-post delinquency rates across home buyers of houses with higher and lower price dispersion, the findings directly counter the explanation that home buyers of houses with higher price dispersion have higher credit risk after controlling for observable characteristics.

3.5.2 Lender Market Power

The results are not likely driven by lender market power. Firstly, our empirical analysis exploits within county-year variation. Existing literature on local lender market power find local competition at county level. Therefore, it is reasonable to believe that buyers from the same county-year with similar creditworthiness are facing the same credit supply. To address further concerns about the effect of lender market power, we re-estimate specification 5 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 3.3.2 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 B of Table 8 reports the results. The key variable of interest is price dispersion, which is property-level idiosyncratic price dispersion. We first confirm Table 4 results using this sub-sample in column 1. In columns 2-4, we add in more saturated lender fixed effects: lender-year, 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.

We also re-estimate the cost menu in Figure 5 by (1) taking out the effect of lender HHI
on mortgage interest rate and (2) defining High Price Dispersion within a county-year to ensure comparison of houses within the same county-year. The cost menu is not qualitatively different from Figure 5.

### 3.5.3 Bunching Below Conforming Loan Limit

Next, 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.

One possible explanation of our finding is that home buyers purposefully lower their borrowing amounts to receive subsidized credit. 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 C of Table 8 reports the results. The results show that our main finding is not driven by home buyers’ incentive to keep their loan amount below the conforming loan limit. Among houses with prices below the conforming loan limit, houses with higher price dispersion are financed with smaller loans given the same interest rates than houses with lower price dispersion. The result holds in more saturated specifications in columns 2-4 with lender-year, lender-county-year, and lender-zip-year fixed effects, respectively, which control for lender market power.

### 4 Implications for Homeownership: A Quantitative Model and Reduced Form Evidence

In this section, we quantify the effects of the borrowing constraints induced by collateral value uncertainty on homeownership rates in a life-cycle model of housing choices.
4.1 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 $\tau$ is the income tax rate, and:

$$\log (y_t) = \chi_t + \zeta_t$$ (10)

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

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

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

**Housing.** There is a discrete grid of house qualities $h_i \in S = \{s_1, s_2 \ldots s_H\}$, ordered in increasing order. There is a cutoff $s_R$, for $R < H$. All house qualities below $s_1 \ldots s_R$ are available for rent only, and all house qualities $s_{R+1} \ldots s_H$ are available to purchase only. Thus, the household can only rent low-quality houses, and must purchase a house to receive housing services above $s_R$. Rental housing has a flow cost of $p^r h_i$, that is, $p^r$ per unit of housing services rented. The price of an owned house of quality $h_i$ is $p^h h_i$. Homeowners pay a depreciation cost of $\delta^h$ 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 $\phi(h_i)$ 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 4.2 below, and Appendix D.1. Let $a_t$ represent cash-on-hand; homeowners’ borrowing constraint is thus:

$$a_t \geq -\phi(h_t)p^h h_t$$

(11)

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:

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

discounting at rate $\beta$. Per-period utility is:

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

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 $\zeta_t$. The household’s value function is:

$$V_t(h_t, a_t, \zeta_t) = \max \left\{ V_{t}^{\text{renter}}(h_t, a_t, \zeta_t), V_{t}^{\text{purchase}}(h_t, a_t, \zeta_t) \right\}$$
If the household decides to rent in period $t$, it solves:

$$V_{\text{renter}}(h_t, a_t, \zeta_t) = \max_{c_t, a_{t+1}, h_{t+1}} u(c_t, h_{t+1}) + \beta E \left[ V_{t+1}(h_{t+1}, a_{t+1}, \zeta_{t+1}) \mid \zeta_t \right]$$  \hspace{1cm} (12)

subject to

$$c_t + \frac{a_{t+1}}{1 + r_t} = a_t + y_t + p^h h_t 1(h_t > s_R) - p_r h_{t+1}$$  \hspace{1cm} (13)

$\zeta_t$ is set to 0.25. For retired households, $\chi_t$ is set to $15,000$ annually, which is approximately the average social security payout in the US.\textsuperscript{10}

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:

$$V_{\text{purchase}} = \max_{c_t, a_{t+1}, h_{t+1}} u(c_t, h_{t+1}) + \beta E \left[ V(h_{t+1}, a_{t+1}, \zeta_{t+1}) \mid \zeta_t \right]$$  \hspace{1cm} (14)

subject to

$$c_t + \frac{a_{t+1}}{1 + r_t} = a_t + y_t + p^h h_t 1(h_t > s_R) - (1 + \delta^h + F_{\text{pur}} 1(h_{t+1} \neq h_t)) p^h h_{t+1}$$  \hspace{1cm} (15)

$\zeta_t$ is set to 0.25. For retired households, $\chi_t$ is set to $15,000$ annually, which is approximately the average social security payout in the US.\textsuperscript{10} We use standard values of $\beta, \sigma, \alpha$ in the literature. Housing transaction costs $F_{\text{pur}}$ are set to 0.05, which is the typical fee charged by real estate brokers in the US. This value is also used in

\textsuperscript{10}See Table A in the Social Security Program Fact Sheet.

4.2 Calibration

The model period is annual. Average log earnings over the lifecycle, $\chi_t$, are from the 2016 SCF. The income tax rate $\tau$ is set to 0.25. For retired households, $\chi_t$ is set to $15,000$ annually, which is approximately the average social security payout in the US.\textsuperscript{10} We use standard values of $\beta, \sigma, \alpha$ in the literature. Housing transaction costs $F_{\text{pur}}$ 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 $\beta$ and depreciation rates $\delta^h$, multiplied by an adjustment parameter $K^H$ which influences how attractive homeownership is compared to rental. We set the initial distribution of $\zeta_t$, the idiosyncratic income shock, for 25-year-olds such that probabilities are log-linear in the level of $\zeta_t$, that is:

$$P_{25}(\zeta) \propto \exp(K\zeta)$$

where $k_\zeta$ controls whether probability weights are higher for high or low values of $\zeta$.

We calibrate the persistence of idiosyncratic income shocks $\rho_\zeta$ to 0.91, and the standard deviation of shocks $\sigma_\epsilon$ to 0.21, following Floden and Lindé (2001).

The core input into our model is $\phi(h)$, 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 Figure 10, and describe details of how we construct these functions in Appendix D.1. We essentially estimate the relationship between prices and average price dispersion $\sigma$ in each group of counties, and then calculate LTVs by multiplying the differences in $\sigma$ by the coefficient from specification 1 in Table 6, which is the reduced-form relationship between price dispersion and LTVs, controlling for other observable features that may affect LTV. The average difference in $\sigma$ between high- and low-dispersion counties is roughly 2.7SD. From Table 6 column 1, a 1SD change in $\sigma$ 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%.

We choose the set of house qualities, the bequest parameter $K_B$, the housing attractiveness parameter $K^H$, and the initial income shock distribution slope parameter $K^B$ to match the level and path of homeownership and debt-to-assets from the 2016 SCF, as well as the ratio of of median net worth at age 75 to net worth at age 50 of 1.51, as in Kaplan 11.

Without adjusting the initial distribution of $\zeta$, we found that homeownership rates rose too quickly in the model relative to the data.
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. Additional details on how we numerically solve the model are in Appendix D.2. Table 9 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. Figure 11 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.

### 4.3 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.\(^\text{12}\)

Table 10 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.\(^\text{13}\)

\(^{12}\)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.

\(^{13}\)Since incomes are fairly persistent in lifecycle models, initial incomes have persistent effects on wealth and income at later ages.
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. Therefore, our results suggest that, in a standard calibrated lifecycle model of housing choice, LTV differences induced by price dispersion can have sizable effects on aggregate homeownership rates, and the homeownership gap between high- and low-income households.

4.4 Empirical Evidence

To provide suggestive empirical support for our model results, we empirically analyze the relationship between collateral value uncertainty and homeownership. Using individual 5-year American Community Survey data from 2010, 2015, and 2019, we calculate the homeownership rates of low incomes (bottom income decile) and high incomes (upper nine income deciles) within each county-year, and plot them against county-level house price dispersion in Figure 9. Overall homeownership rates are lower in areas with higher price dispersion. Moreover, the difference between high- and low-income homeownership rates also increases as local house price dispersion increases.

Table 11 presents this fact in a formal regression setting:

\[
HomeOwner_{ikt} = \alpha + \beta PriceDispersion_{kt} + X_{ikt}\Gamma + \mu_t + \epsilon_{ikt}
\]  

The outcome variable is an indicator for whether the survey participant is a homeowner. We restrict the sample to households at least 25-year old. \( PriceDispersion_{kt} \) is the average

\footnote{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}
house price dispersion in county \( k \) in year \( t \). \( X_{ikt} \) is a set of household and county controls, including county price dispersion, household income, income-squared, age, age-squared, local house price, and the square of local house prices. \( \mu_t \) represents year fixed effects. We estimate this specification for all households (column 1), low income households in the bottom income decile within each county-year (column 2), and high income households in the upper nine income deciles within each county-year (column 3), respectively. We then include an interaction term of price dispersion and a low income indicator in column 4 using the full sample.

Overall, counties with higher house price dispersion tend to have lower homeownership rates, and the effect is stronger for low income households. A one standard deviation increase in county house price dispersion is associated with a 1.3pp decrease in homeownership rates. For a household in the bottom income decile, a 1SD increase in dispersion is associated with a 1.46pp decrease in homeownership. Purging the difference in other county characteristics by including county-year fixed effects, the effect of 1SD increase in dispersion on homeownership is 1pp greater for low incomes than for high incomes.

These effects are quantitatively similar to, and in fact slightly larger than, the predicted effects of dispersion on homeownership in our model. This provides empirical evidence that housing collateral values have an effect on aggregate homeownership rates, especially among low-income households.

We view a 1-2pp change in homeownership rates as a fairly large effect, since homeownership rates are quite hard to move. For example, over the period 1960 to the present, homeownership rates hovered mostly between 63% to 66%, rising to 69% during the 2008 housing boom and decreasing back to around 63% thereafter.\(^{15}\)

\(^{15}\)See FRED.
5 Discussion and Policy Implications

5.1 Homeownership Gap and Government Mortgage Programs

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

As we show in Figure 2, low-income households tend to live in areas with higher house price dispersion on average, and thus likely receive lower mortgage LTVs as a result. This is not a market inefficiency, or a form of credit market discrimination: it is a rational response of lenders to the fact that more volatile assets are worse as collateral for debt.

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

5.2 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. It is known in the literature that human appraisals tend to be distorted, so that they are generally equal to or higher than transac-

---

16 See, for example, the discussion in Herbert et al. (2005).
17 https://www.americanbanker.com/news/fhfa-will-make-desktop-home-appraisals-a-permanent-option

31
tion prices (Calem et al., 2015; Eriksen et al., 2019; Bogen and Shui, 2020; Conklin et al., 2020; Calem et al., 2021; Kruger and Maturana, 2021). Automated appraisals are likely to be less distorted, but as a result, under-appraisals will be more frequent, especially in areas with high house price dispersion. Automated appraisal thus have the potential to hurt low incomes and minorities who tend to live in areas with less predictable house prices.\footnote{Blattner and Nelson (2021) and Fuster et al. (2020) have made similar arguments that low incomes and minorities tend to have noisier hard information, and the development of FinTech is going to increase statistical discrimination in mortgage lending.}

5.3 Housing Affordability

Our results suggest that urban policymakers, who regulate the construction and renovation of residential housing, should consider the effects of policies on the collateral value of the housing stock. Policies which affect the ease of housing rebuilding and renovation, and which control floor area ratios, lot sizes, construction materials and standards, all have the potential to affect aggregate housing value uncertainty. In particular, policies which tend to make the housing stock newer and more homogeneous will tend to increase mortgage credit provision in competitive lending markets. This could contribute 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.

6 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 risk, and partly through the effect of idiosyncratic price dispersion on appraisal noise. We have shown theoretically and empirically that the effects of price dispersion on mortgage credit availability matter for aggregate homeownership rates. Our results have implications for policy interventions in mortgage and housing markets aimed at improving credit access and homeownership,
especially among low-income individuals and first-time homebuyers.
References


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


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


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


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


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


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


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


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

Figures

**Figure 1.** Price Dispersion Estimates

(a) Zip-Code Dispersion: 2000 vs 2010

(b) Zip-Code Dispersion: 2010 vs 2020

*Note:* This figure shows how stable estimated price dispersion measures are over time. Panel (a) plots zip-code dispersion measures in 2010 against zip-code dispersion measures in 2000. Panel (b) plots zip-code dispersion measures in 2020 against zip-code dispersion measures in 2010. *Source:* House price dispersion is estimated using Corelogic Deeds records.
Figure 2. House Price Dispersion, Incomes and Minority Shares

Note: Panel (a) of this figure shows the association between house price dispersion and zip-code household income prices. Panel (b) of this figure shows the association between house price dispersion and zip-code non-white population share. We divide all zip codes into five buckets based on local median household income or non-white share and plot the average values in each bucket. The sample includes annual zip level observations from 2000 to 2020. Source: Corelogic Deeds and American Community Survey 2008-2012.
Figure 3. County Level House Price Dispersion and Credit Access

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

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

Note: This figure shows the correlation between county level house price dispersion and various credit access outcomes. Panel a plots county average LTP. Panel b plots county average residualized mortgage interest rate. Individual mortgage interest rates are residualized using borrower and loan characteristics, such as FICO, LTP, DTI, the squared terms, and their interactions with origination year. We then take the county-average of residualized mortgage rates. Panels c and d plot mortgage rejection rate. Panel c plots the total rejection rate. Panel d plots the rejection rate due to collateral. We residualize mortgage rejection rate by taking the residuals of regressions of mortgage rate on county average log house price, credit score, and year fixed effects. The sample includes annual county observations from 2000 to 2020 for panels (a) and (b) and from 2000 to 2017 for panels (c) and (d). Source: County house price dispersion is estimated using Corelogic Deeds records. Mortgage data are from Corelogic LLMA and HMDA.
Figure 4. Property Level House Price Dispersion and Mortgage LTP

Note: This figure plots the binned scatterplot of mortgage LTP and house price dispersion *within* a county-year. To exploit within county-year variation, we regress mortgage LTP on house prices and county-year fixed effects and regress house price dispersion on house prices and county-year fixed effects; and we then plot the LTP residuals against house price dispersion residuals. The sample includes property transaction level observations from 2000 to 2020. Source: Corelogic Deeds records.
Figure 5. Property Level Mortgage Menu by House Price Dispersion

Note: This figure shows mortgage price menu (rate-LTP pair) by zip-level house price dispersion. The y values are interest rate residuals from a regression of mortgage rates on borrower fico, fico-squared, DTI, DTI-squared conforming or jumbo indicator, and origination month fixed effects. The dots represent the average mortgage rate in each LTP bucket. The shaded area indicates 95% confidence interval. The sample includes loan level observations of conventional loans from 2000 to 2020. Source: Corelogic LLMA and Deeds.
Figure 6. Heterogeneous Effect of Price Dispersion by FICO

Note: This figure shows heterogeneous effect of price dispersion by FICO score. 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 7. Robustness: Realized Default Probability

Note: This figure shows realized mortgage default probability in two years since origination for high versus low dispersion zip codes. The y values are probability of 60+days delinquent in 2 years since origination. 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. Source: Corelogic LLMA, Deed and Tax datasets.
Figure 8. Appraisal Channel

(a) Appraisal to Price Distance

(b) Appraisal below Sale

(c) Conditional Appraisal to Price Distance

Note: Panel (a) of this figure shows a binned scatter plot, where the y-variable is $ApprDev_i$, the product of the percentage deviation of appraisal prices to sale prices with a dummy for a house under-appraising, defined in (8). In panel (b), the y-variable is the probability that appraisals are below transaction prices. In panel (c), the y-variable is the average under-appraisal percentage conditional on under-appraisal, defined in (9). In all panels, the x-variable is zip code price dispersion. We divide all loans into 50 buckets based on zip code house price dispersion. The sample includes loan level observations from 2000 to 2020. Source: Corelogic LLMA, Deed and Tax datasets.
Figure 9. House Price Dispersion, Income, and Home Ownership

Note: This figure plots the homeownership rates across counties by income. Low Income is defined as the bottom income decile within each county-year, and High Income is defined as all other deciles. Homeownership rates are on the Y axis, and county average house price dispersion is on the X axis. The sample includes annual county level observations from 2000 to 2020. Source: Corelogic Deed and Tax data and American Community Survey 2008-2012.
Figure 10. $\sigma(p)$ and $\phi(p)$ functions

Note: The left panel shows $\sigma(p)$, the average of price dispersion $\sigma$ 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 6. 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 11. Model Fit**

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

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.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>Stdev</th>
<th>P25</th>
<th>Median</th>
<th>P75</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Property Level Sample</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loan to Price</td>
<td>39M</td>
<td>85.43</td>
<td>15.75</td>
<td>80.00</td>
<td>89.82</td>
<td>98.19</td>
</tr>
<tr>
<td>Price Dispersion</td>
<td>39M</td>
<td>0.24</td>
<td>0.11</td>
<td>0.18</td>
<td>0.24</td>
<td>0.31</td>
</tr>
<tr>
<td>Sale Price</td>
<td>39M</td>
<td>270281</td>
<td>216652</td>
<td>140000</td>
<td>215000</td>
<td>331250</td>
</tr>
<tr>
<td>Mortgage Amount</td>
<td>39M</td>
<td>220474</td>
<td>158371</td>
<td>121539</td>
<td>182100</td>
<td>275400</td>
</tr>
<tr>
<td>Building Age</td>
<td>39M</td>
<td>29.65</td>
<td>28.19</td>
<td>6.00</td>
<td>22.00</td>
<td>47.00</td>
</tr>
<tr>
<td>Square Footage</td>
<td>39M</td>
<td>1975</td>
<td>2677</td>
<td>1370</td>
<td>1782</td>
<td>2374.00</td>
</tr>
<tr>
<td><strong>Loan Level Sample</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loan to Price</td>
<td>6M</td>
<td>85.35</td>
<td>15.03</td>
<td>80.00</td>
<td>90.00</td>
<td>98.19</td>
</tr>
<tr>
<td>Zip Price Dispersion</td>
<td>6M</td>
<td>0.25</td>
<td>0.08</td>
<td>0.19</td>
<td>0.24</td>
<td>0.29</td>
</tr>
<tr>
<td>Sale Price</td>
<td>6M</td>
<td>280797</td>
<td>240336</td>
<td>144500</td>
<td>220000</td>
<td>341915</td>
</tr>
<tr>
<td>Appraised to Price Ratio</td>
<td>6M</td>
<td>1.03</td>
<td>0.20</td>
<td>1.00</td>
<td>1.00</td>
<td>1.02</td>
</tr>
<tr>
<td>Mortgage Amount</td>
<td>6M</td>
<td>227546</td>
<td>168739</td>
<td>124053</td>
<td>186456</td>
<td>284255</td>
</tr>
<tr>
<td>FICO</td>
<td>6M</td>
<td>725.40</td>
<td>61.42</td>
<td>681.00</td>
<td>735.00</td>
<td>778.00</td>
</tr>
<tr>
<td>Debt-to-Income</td>
<td>6M</td>
<td>37.09</td>
<td>11.35</td>
<td>29.63</td>
<td>37.86</td>
<td>44.60</td>
</tr>
<tr>
<td><strong>Mortgage Application Sample</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rejection Rate</td>
<td>62M</td>
<td>16.05</td>
<td>36.71</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Rejection due to Collateral Reasons</td>
<td>62M</td>
<td>1.95</td>
<td>13.84</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Zip Price Dispersion</td>
<td>62M</td>
<td>0.26</td>
<td>0.08</td>
<td>0.20</td>
<td>0.25</td>
<td>0.31</td>
</tr>
<tr>
<td>Applicant Income (Thousand)</td>
<td>62M</td>
<td>102.01</td>
<td>191.47</td>
<td>47.00</td>
<td>72.00</td>
<td>114.00</td>
</tr>
<tr>
<td>Loan-to-Income</td>
<td>62M</td>
<td>241.78</td>
<td>6263.60</td>
<td>135.54</td>
<td>227.35</td>
<td>316.24</td>
</tr>
<tr>
<td>County Credit Score</td>
<td>62M</td>
<td>667.38</td>
<td>22.48</td>
<td>650.95</td>
<td>666.18</td>
<td>684.07</td>
</tr>
</tbody>
</table>
**Table 2:** Determinants of House Price Dispersion

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

### Panel A: House Features

<table>
<thead>
<tr>
<th></th>
<th>Estimated Price Dispersion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>Building Age</td>
<td>0.04***</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
</tr>
<tr>
<td>Recent Renovation</td>
<td>-0.01***</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
</tr>
<tr>
<td>Ever Renovate</td>
<td>-0.01***</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
</tr>
</tbody>
</table>

**Benchmark: Square-Footage < 1286**

- [1286,1604]
  - -0.03*** -0.03*** -0.02***
  - (0.002) (0.002) (0.002)
- [1605,1970]
  - -0.03*** -0.03*** -0.02***
  - (0.003) (0.003) (0.004)
- [1971,2540]
  - -0.02*** -0.02*** -0.02***
  - (0.003) (0.004) (0.004)
> 2540
  - 0.01*** 0.01** 0.03***
  - (0.004) (0.004) (0.005)

**Benchmark: Bedrooms < 4**

- =4
  - -0.01*** -0.01***
  - (0.001) (0.001)
- >4
  - 0.01*** 0.01***
  - (0.002) (0.002)

Log House Price

- -0.50*** -0.45*** -0.46*** -0.51*** -0.51*** -0.33***
  - (0.023) (0.032) (0.031) (0.020) (0.025) (0.024)

Log House Price Squared

- 0.52*** 0.47*** 0.47*** 0.51*** 0.51*** 0.34***
  - (0.023) (0.033) (0.032) (0.020) (0.025) (0.025)

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>County-Year FE</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>R2</td>
<td>0.33</td>
<td>0.34</td>
<td>0.35</td>
<td>0.27</td>
<td>0.27</td>
<td>0.38</td>
</tr>
<tr>
<td>Observations</td>
<td>39M</td>
<td>18M</td>
<td>18M</td>
<td>39M</td>
<td>29M</td>
<td>15M</td>
</tr>
</tbody>
</table>

52
### Table 3: Property-Level House Price Dispersion and LTP

This table presents property-level regression results. The outcome variable is the loan level loan-to-sale price ratio. The explanatory variable of interest is property-level house price dispersion, scaled by its standard deviation. 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.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Price Dispersion</strong></td>
<td>-0.44***</td>
<td>-0.46***</td>
<td>-0.21***</td>
<td>-0.23***</td>
<td>-0.20***</td>
</tr>
<tr>
<td></td>
<td>(0.042)</td>
<td>(0.041)</td>
<td>(0.032)</td>
<td>(0.030)</td>
<td>(0.024)</td>
</tr>
<tr>
<td><strong>Log House Price</strong></td>
<td>-4.04***</td>
<td>-4.24***</td>
<td>-4.54***</td>
<td>-4.27***</td>
<td>-4.14***</td>
</tr>
<tr>
<td></td>
<td>(0.127)</td>
<td>(0.121)</td>
<td>(0.152)</td>
<td>(0.137)</td>
<td>(0.118)</td>
</tr>
<tr>
<td><strong>Loan Controls</strong></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Transaction Date FE</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>County-Year FE</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Lender-Year FE</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Zip-Year FE</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>R2</td>
<td>0.34</td>
<td>0.34</td>
<td>0.37</td>
<td>0.40</td>
<td>0.42</td>
</tr>
<tr>
<td>Observations</td>
<td>37M</td>
<td>37M</td>
<td>37M</td>
<td>37M</td>
<td>36M</td>
</tr>
</tbody>
</table>
Table 4: Cost Menu: Not about Interest Rate-LTP Substitution

This table presents loan level regression results of the cost menu. The outcome variable is the loan-level interest rate in Panel A and loan-to-sale price ratio in Panel B. Column 1 uses the full sample. Columns 2-4 use GSE loans (i.e., non-FHA loans sold to GSE), FHA loans, and jumbo conventional loans (i.e., non-FHA and non-GSE eligible loans), respectively. The explanatory variable of interest is zip-code house price dispersion, scaled by its standard deviation. Borrower and loan controls in Panel A include log house price, FICO score, FICO squared, LTP, LTP squared, DTI, DTI-squared, and loan type. Borrower and loan controls in Panel A include log house price, FICO score, FICO-squared, interest rate, 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.

<table>
<thead>
<tr>
<th></th>
<th>(1) Full</th>
<th>(2) GSE</th>
<th>(3) FHA</th>
<th>(4) Jumbo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: Interest Rate (bps)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zip Price Dispersion</td>
<td>1.10***</td>
<td>1.24***</td>
<td>1.23***</td>
<td>0.43**</td>
</tr>
<tr>
<td></td>
<td>(0.117)</td>
<td>(0.091)</td>
<td>(0.106)</td>
<td>(0.199)</td>
</tr>
<tr>
<td>LTP</td>
<td>0.85***</td>
<td>-0.00</td>
<td>0.73***</td>
<td>0.21**</td>
</tr>
<tr>
<td></td>
<td>(0.097)</td>
<td>(0.025)</td>
<td>(0.122)</td>
<td>(0.103)</td>
</tr>
<tr>
<td>Borrower and Loan Controls</td>
<td>✓ ✓ ✓ ✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Origination Month FE</td>
<td>✓ ✓ ✓ ✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>County-Year FE</td>
<td>✓ ✓ ✓ ✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R2</td>
<td>0.85</td>
<td>0.88</td>
<td>0.89</td>
<td>0.88</td>
</tr>
<tr>
<td>Observations</td>
<td>6M</td>
<td>2.8M</td>
<td>1.7M</td>
<td>0.3M</td>
</tr>
<tr>
<td>Panel B: Loan-to-Price</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zip Price Dispersion</td>
<td>-0.50***</td>
<td>-0.44***</td>
<td>-0.22***</td>
<td>-0.57***</td>
</tr>
<tr>
<td></td>
<td>(0.050)</td>
<td>(0.053)</td>
<td>(0.022)</td>
<td>(0.093)</td>
</tr>
<tr>
<td>Interest Rate</td>
<td>0.01***</td>
<td>0.04***</td>
<td>-0.00***</td>
<td>0.01***</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.001)</td>
<td>(0.000)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>Borrower and Loan Controls</td>
<td>✓ ✓ ✓ ✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Origination Month FE</td>
<td>✓ ✓ ✓ ✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>County-Year FE</td>
<td>✓ ✓ ✓ ✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R2</td>
<td>0.40</td>
<td>0.22</td>
<td>0.07</td>
<td>0.29</td>
</tr>
<tr>
<td>Observations</td>
<td>6M</td>
<td>2.8M</td>
<td>1.7M</td>
<td>0.3M</td>
</tr>
</tbody>
</table>
Table 5: Mortgage Rejections and Zipcode House Price Dispersion

This table presents loan level regression results of mortgage rejections. The outcome variable is an indicator that equals 100 if a loan is rejected and 0 otherwise. The explanatory variable of interest is zip code house price dispersion, scaled by its standard deviation. Borrower/Loan controls include zip code house price, log income, loan type, county average credit score and its square term, and loan to income ratio and its square term. The sample includes loan level observations from 2001 to 2016. Standard errors are clustered at county level. ***, **, * represent 1%, 5%, and 10% significance, respectively.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Full</td>
<td>Full</td>
<td>Full</td>
<td>GSE</td>
<td>FHA</td>
<td>Jumbo</td>
</tr>
<tr>
<td>Panel A: Rejection</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zip Price Dispersion</td>
<td>1.42***</td>
<td>1.77***</td>
<td>1.36***</td>
<td>1.36***</td>
<td>1.10***</td>
<td>0.81***</td>
</tr>
<tr>
<td></td>
<td>(0.131)</td>
<td>(0.117)</td>
<td>(0.086)</td>
<td>(0.093)</td>
<td>(0.084)</td>
<td>(0.102)</td>
</tr>
<tr>
<td>Rejection Mean</td>
<td>16%</td>
<td>16%</td>
<td>16%</td>
<td>16.7%</td>
<td>14.5%</td>
<td>16.3%</td>
</tr>
<tr>
<td>R2</td>
<td>0.02</td>
<td>0.04</td>
<td>0.16</td>
<td>0.19</td>
<td>0.12</td>
<td>0.17</td>
</tr>
<tr>
<td>Panel B: Rejection Due to Collateral</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zip Price Dispersion</td>
<td>0.45***</td>
<td>0.51***</td>
<td>0.48***</td>
<td>0.52***</td>
<td>0.42***</td>
<td>0.37***</td>
</tr>
<tr>
<td></td>
<td>(0.033)</td>
<td>(0.029)</td>
<td>(0.030)</td>
<td>(0.032)</td>
<td>(0.027)</td>
<td>(0.043)</td>
</tr>
<tr>
<td>Rejection due to Collateral Mean</td>
<td>2%</td>
<td>2%</td>
<td>2%</td>
<td>2%</td>
<td>1.7%</td>
<td>2.3%</td>
</tr>
<tr>
<td>Dispersion Mean</td>
<td>0.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dispersion Stdev</td>
<td>0.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local Controls</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Origination Year FE</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>County-Year FE</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Lender-Year FE</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>R2</td>
<td>&lt;0.01</td>
<td>0.01</td>
<td>0.04</td>
<td>0.05</td>
<td>0.04</td>
<td>0.08</td>
</tr>
<tr>
<td>Observations</td>
<td>59M</td>
<td>59M</td>
<td>59M</td>
<td>42M</td>
<td>9.2M</td>
<td>4.4M</td>
</tr>
</tbody>
</table>
Table 6: Heterogeneous Effect by FICO

This table presents heterogeneous effects of price dispersion on LTPs by FICO scores. The outcome variable is the loan to price ratio. The explanatory variable of interest is the interaction between zip-code house price dispersion, scaled by its standard deviation, and FICO score buckets. The omitted benchmark credit score bucket is Excellent, including FICO score of 800 or above. Borrower/Loan controls include zip price dispersion, FICO score, FICO-squared, mortgage interest rate, and loan type. Column 1 uses the full sample. Column 2 uses securitized loans. Column 3 uses non-FHA portfolio 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.

<table>
<thead>
<tr>
<th></th>
<th>LTV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td></td>
<td>Full</td>
</tr>
<tr>
<td>Zip Price Dispersion</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>(0.080)</td>
</tr>
<tr>
<td>Baseline: Excellent FICO</td>
<td></td>
</tr>
<tr>
<td>Zip Price Dispersion × Very Good</td>
<td>-0.42***</td>
</tr>
<tr>
<td></td>
<td>(0.040)</td>
</tr>
<tr>
<td>Zip Price Dispersion × Good</td>
<td>-0.72***</td>
</tr>
<tr>
<td></td>
<td>(0.054)</td>
</tr>
<tr>
<td>Zip Price Dispersion × Fair</td>
<td>-0.88***</td>
</tr>
<tr>
<td></td>
<td>(0.068)</td>
</tr>
<tr>
<td>Zip Price Dispersion × Poor</td>
<td>-1.10***</td>
</tr>
<tr>
<td></td>
<td>(0.108)</td>
</tr>
<tr>
<td>Origination Month FE</td>
<td>✓</td>
</tr>
<tr>
<td>County-Year FE</td>
<td>✓</td>
</tr>
<tr>
<td>Borrower/Loan Controls</td>
<td>✓</td>
</tr>
<tr>
<td>R2</td>
<td>0.40</td>
</tr>
<tr>
<td>Observations</td>
<td>6M</td>
</tr>
</tbody>
</table>
Table 7: Appraisal Channel

This table presents evidence showing that price dispersion is associated with under-appraisals. The outcome variable in columns 1-2 is the appraisal deviation $ApprDev_i$, which is the product of the percentage deviation of appraisal prices to sale prices with an under-appraisal dummy, defined in (8). The outcome variable in columns 3-4 is a dummy for appraisals being below transaction prices. The outcome variable in columns 5-6 is the percentage difference between appraisal prices and sale prices, conditional on under-appraisal. The explanatory variable is zip code price dispersion scaled by its sample standard deviation. Borrowers and loan controls include mortgage rate, log house price, FICO, FICO-squared, DTI, DTI-squared, LTV, LTV-squared, GSE indicator, and loan type. The sample includes all loans originated from 2000 to 2020. Standard errors are clustered at county level. ***, **, * represent 1%, 5%, and 10% significance, respectively.

<table>
<thead>
<tr>
<th></th>
<th>Appraisal Deviation</th>
<th>I(Appraisal&lt;Price)</th>
<th>Conditional Appraisal Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Zip Price Dispersion</td>
<td>0.03***</td>
<td>0.02***</td>
<td>0.09*</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.002)</td>
<td>(0.052)</td>
</tr>
<tr>
<td>Origination Month FE</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>County-Year FE</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Borrower/Loan Controls</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>R2</td>
<td>0.01</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>Observations</td>
<td>6M</td>
<td>6M</td>
<td>6M</td>
</tr>
</tbody>
</table>
Table 8: Robustness: Ex-Post Performance, Lender Market Power, Bunching below Conforming Loan Limit, Appraisal

This table presents robustness tests. Panel A analyzes ex-post performance of mortgage loans. Outcome variable is 100 if the loan defaults in two years since origination and 0 otherwise. The explanatory variable of interest is zip-code house price dispersion, scaled by its standard deviation. Other controls include house price and loan type. The sample includes all loans originated from 2000 to 2018. Since we need at least two-year performance to define default, we remove loans originated after 2018 from the full sample for this analysis. Standard errors are clustered at county level. ***, **, * represent 1%, 5%, and 10% significance, respectively.

<table>
<thead>
<tr>
<th></th>
<th>Default</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Full GSE&amp;FHA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zip Price Dispersion</td>
<td>-0.02</td>
<td>-0.05</td>
<td>-0.31***</td>
</tr>
<tr>
<td></td>
<td>(0.068)</td>
<td>(0.061)</td>
<td>(0.086)</td>
</tr>
<tr>
<td>Interest Rate</td>
<td>2.76***</td>
<td>3.98***</td>
<td>2.33***</td>
</tr>
<tr>
<td></td>
<td>(0.114)</td>
<td>(0.129)</td>
<td>(0.263)</td>
</tr>
<tr>
<td>Loan to Price</td>
<td>0.08***</td>
<td>0.07***</td>
<td>-0.03***</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.003)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>Debt to Income</td>
<td>0.08***</td>
<td>0.07***</td>
<td>0.04***</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>Baseline: Excellent FICO</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very Good</td>
<td>-0.27***</td>
<td>-0.26***</td>
<td>0.26*</td>
</tr>
<tr>
<td></td>
<td>(0.054)</td>
<td>(0.064)</td>
<td>(0.147)</td>
</tr>
<tr>
<td>Good</td>
<td>1.70***</td>
<td>1.29***</td>
<td>1.90***</td>
</tr>
<tr>
<td></td>
<td>(0.087)</td>
<td>(0.085)</td>
<td>(0.215)</td>
</tr>
<tr>
<td>Fair</td>
<td>11.47***</td>
<td>10.38***</td>
<td>7.36***</td>
</tr>
<tr>
<td></td>
<td>(0.138)</td>
<td>(0.169)</td>
<td>(0.544)</td>
</tr>
<tr>
<td>Poor</td>
<td>32.21***</td>
<td>30.88***</td>
<td>27.08***</td>
</tr>
<tr>
<td></td>
<td>(0.340)</td>
<td>(0.356)</td>
<td>(3.296)</td>
</tr>
<tr>
<td>Controls</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Origination Month FE</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>County-Year FE</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>R2</td>
<td>0.14</td>
<td>0.15</td>
<td>0.15</td>
</tr>
<tr>
<td>Observations</td>
<td>5M</td>
<td>4M</td>
<td>0.2M</td>
</tr>
</tbody>
</table>
Panel B presents robustness test 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 C 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 B: Not about Lender Market Power

<table>
<thead>
<tr>
<th>LTP</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price Dispersion</td>
<td>-0.35***</td>
<td>-0.30***</td>
<td>-0.30***</td>
<td>-0.22***</td>
</tr>
<tr>
<td></td>
<td>(0.027)</td>
<td>(0.025)</td>
<td>(0.024)</td>
<td>(0.028)</td>
</tr>
<tr>
<td>Interest Rate</td>
<td>1.08***</td>
<td>0.88***</td>
<td>0.95***</td>
<td>1.00***</td>
</tr>
<tr>
<td></td>
<td>(0.057)</td>
<td>(0.044)</td>
<td>(0.040)</td>
<td>(0.054)</td>
</tr>
</tbody>
</table>

Loan Controls | ✓ ✔ ✔ ✔
Origination Month FE | ✓ ✔ ✔ ✔
County-Year FE | ✓ ✔
Lender-Year FE | ✓
Lender-County-Year FE | ✓
Lender-Zip-Year FE | ✓
R2 | 0.47 | 0.53 | 0.59 | 0.67
Observations | 7M | 7M | 6.6M | 5M

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

<table>
<thead>
<tr>
<th>LTP</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price Dispersion</td>
<td>-0.23***</td>
<td>-0.18***</td>
<td>-0.15***</td>
<td>-0.13***</td>
</tr>
<tr>
<td></td>
<td>(0.019)</td>
<td>(0.017)</td>
<td>(0.017)</td>
<td>(0.019)</td>
</tr>
<tr>
<td>Interest Rate</td>
<td>1.07***</td>
<td>0.89***</td>
<td>0.95***</td>
<td>0.99***</td>
</tr>
<tr>
<td></td>
<td>(0.068)</td>
<td>(0.051)</td>
<td>(0.047)</td>
<td>(0.063)</td>
</tr>
</tbody>
</table>

Loan Controls | ✓ ✔ ✔ ✔
Origination Month FE | ✓ ✔ ✔ ✔
County-Year FE | ✓ ✔
Lender-Year FE | ✓
Lender-County-Year FE | ✓
Lender-Zip-Year FE | ✓
R2 | 0.43 | 0.50 | 0.56 | 0.64
Observations | 6M | 6M | 5M | 4M

59
Table 9: 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.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Symbol</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discount factor</td>
<td>$\beta$</td>
<td>0.96</td>
</tr>
<tr>
<td>Intertemporal elasticity of substitution parameter</td>
<td>$\sigma$</td>
<td>2</td>
</tr>
<tr>
<td>Housing budget share</td>
<td>$\alpha$</td>
<td>0.4</td>
</tr>
<tr>
<td>Bequest parameter</td>
<td>$K_B$</td>
<td>300</td>
</tr>
<tr>
<td>Earning persistence</td>
<td>$\rho_\zeta$</td>
<td>0.91</td>
</tr>
<tr>
<td>Standard deviation of earnings shocks</td>
<td>$\sigma_\zeta$</td>
<td>0.21</td>
</tr>
<tr>
<td>Income tax rate</td>
<td>$\tau$</td>
<td>0.25</td>
</tr>
<tr>
<td>Saving rate</td>
<td>$r^B$</td>
<td>0.02</td>
</tr>
<tr>
<td>Mortgage rate</td>
<td>$r^M$</td>
<td>0.04</td>
</tr>
<tr>
<td>House transaction cost</td>
<td>$F^{pur}$</td>
<td>0.05</td>
</tr>
<tr>
<td>House depreciation rate</td>
<td>$\delta^h$</td>
<td>0.01</td>
</tr>
<tr>
<td>Rent price</td>
<td>$p^r$</td>
<td>12</td>
</tr>
<tr>
<td>House price</td>
<td>$p^h$</td>
<td>192</td>
</tr>
</tbody>
</table>

Table 10: 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.

<table>
<thead>
<tr>
<th>Age</th>
<th>Total</th>
<th>Low Income</th>
<th>High Income</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;30</td>
<td>3.5</td>
<td>2.4</td>
<td>4.6</td>
</tr>
<tr>
<td>30-40</td>
<td>0.8</td>
<td>1.6</td>
<td>0</td>
</tr>
<tr>
<td>40-50</td>
<td>1.1</td>
<td>2.3</td>
<td>0</td>
</tr>
<tr>
<td>50-60</td>
<td>1.4</td>
<td>2.8</td>
<td>0</td>
</tr>
<tr>
<td>60-70</td>
<td>1.6</td>
<td>3.2</td>
<td>0</td>
</tr>
<tr>
<td>Overall</td>
<td>1.5</td>
<td>2.6</td>
<td>0.5</td>
</tr>
</tbody>
</table>
Table 11: Home Ownership and Local Price Dispersion

This table presents individual home ownership results using 5-year American Community Survey in 2010, 2015, and 2019. The dependent variable is 100 if the survey participant is a home owner and 0 otherwise. We restrict the sample to households 25 year old or above. County Price Dispersion is scaled by sample standard deviation. In all specifications except for column 4, we control for log household income, log income squared, age, age squared, log local house price, and log local house price squared. Lowest Income Decile is defined as 1 if the household’s income is among the 1st decile among all households in a county. Columns 1 and 4 use the full sample. Columns 2 and 3 use the low income sample (lowest income decile) and the high income sample (2nd-10th income deciles), respectively. Regressions are weighted by the population weight in the ACS dataset, which indicates how representative each survey participant is in the population. Standard errors are clustered at county level. ***, **, * represent 1%, 5%, and 10% significance, respectively.

<table>
<thead>
<tr>
<th></th>
<th>Home Ownership</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td></td>
<td>Full Sample</td>
</tr>
<tr>
<td>County Price Dispersion</td>
<td>-1.29**</td>
</tr>
<tr>
<td></td>
<td>(0.551)</td>
</tr>
<tr>
<td>County Price Dispersion ×</td>
<td></td>
</tr>
<tr>
<td>Lowest Income Decile</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Household Controls</td>
<td>✓</td>
</tr>
<tr>
<td>Year FE</td>
<td>✓</td>
</tr>
<tr>
<td>County-Year FE</td>
<td></td>
</tr>
<tr>
<td>R2</td>
<td>0.21</td>
</tr>
<tr>
<td>Observations</td>
<td>17M</td>
</tr>
</tbody>
</table>
Appendix

A Supplementary Material for Section 2

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 Details on $f_c$ 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:
\[
f_c(x_i, t) = h_c^{f,\text{latlong}}(t, \text{lat}_i, \text{long}_i) + h_c^{f,\text{sqft}}(t, \text{sqft}_i) + h_c^{f,\text{yrbuilt}}(t, \text{yrbuilt}_i) + h_c^{f,\text{bedrooms}}(t, \text{bedrooms}_i) + h_c^{f,\text{bathrooms}}(t, \text{bathrooms}_i)
\]

The functions \(h_c^{f,\text{latlong}}, h_c^{f,\text{sqft}}, \) and \(h_c^{f,\text{yrbuilt}}\) are tensor products of 5-dimensional cubic splines in their constituent components: hence, for example, \(h_c^{f,\text{latlong}}(t, \text{lat}_i, \text{long}_i)\) 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 \(h_c^{f,\text{bedrooms}}\) and \(h_c^{f,\text{bathrooms}}\) 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 \(f_c(x_i, t)\):

\[
g_c(x_i, t) = h_c^{g,\text{latlong}}(t, \text{lat}_i, \text{long}_i) + h_c^{g,\text{sqft}}(t, \text{sqft}_i) + h_c^{g,\text{yrbuilt}}(t, \text{yrbuilt}_i) + h_c^{g,\text{bedrooms}}(t, \text{bedrooms}_i) + h_c^{g,\text{bathrooms}}(t, \text{bathrooms}_i)
\]
B Model

We construct a model to show how house value uncertainty affects loan-to-values and the probability of mortgages failing, through the “fair pricing” and appraisal channels.

**Transaction prices.** Consider a buyer and seller, who negotiate to purchase a house. We normalize the mean value of the house to 1. We assume the buyer and seller have agreed on a price

\[ P = 1 + \epsilon_P \]  

(17)

for the house, where \( \epsilon_P \) is a random variable, with variance \( \sigma_\epsilon \). \( \epsilon_P \) reflects house value uncertainty: precisely, unpredictable variation in house prices. It may result from differences in buyers’ and sellers’ preferences, relative to their outside options.\(^{19}\) When \( \sigma_\epsilon \) is larger, house prices are more disperse. Other possible drivers of \( \sigma_\epsilon \) include asymmetric information between the lender and the borrower about the quality of the house, or frictions in bargaining between buyers and sellers. For the purposes of our model, any force which causes sale prices of houses to have a larger unpredictable component has equivalent effects on mortgage credit provision.

We assume buyers purchase houses using mortgages. Buyers are cash-constrained: each buyer has enough liquid wealth to pay a fraction \( \theta \) of the price \( P \), where \( \theta \sim F_\theta (\cdot) \), and \( \theta \) is independent of \( P \). Hence, the buyer can only buy if:

\[ \theta \geq 1 - \frac{L}{P} \]  

(18)

that is, if one minus the loan-to-price ratio the buyer is offered is less than the buyer’s maximum down payment \( \theta \). If the buyer and seller agree to transact at price \( P \), but the lender is not willing to lend the borrower enough to purchase the house, the transaction will fail. From (18), Hence, for transactions at price \( P \), if the bank is willing to lend at most \( L \), a fraction:

\[ F_\theta \left( 1 - \frac{L}{P} \right) \]  

(19)

of transactions will fail, due to borrowers having insufficient funds for down payments.

\(^{19}\)Using a simple bargaining model of price formation, one can show that \( \epsilon_P \) will be larger when buyers’ and sellers’ values are more disperse.
Implicitly, we are assuming buyers and sellers cannot renegotiate transactions if the loan approval fails. In practice, while re-negotiation sometimes occurs, if the house appraisal is far below the attempted transaction price, sellers may prefer to go with a different buyer who is willing to pay a higher down payment. Moreover, anecdotally and in our data, a nontrivial fraction of transactions fail because $L$ is too low.

Mortgages are subject to two constraints. The first is a fair-pricing constraint: conditional on the lender’s belief of the house’s value, the lender’s willingness to lend depends on loan size and the volatility $\sigma_\epsilon$. The second is an appraisal constraint: there is an exogeneous regulatory constraint on the mortgage’s loan-to-value, where the value is set as the minimum of the price and a noisy appraisal-based estimate of the house’s value.

**Fair pricing.** We assume lenders set LTPs based on default rates, to maintain some minimum expected profit margin. Once the buyer has purchased the house, the buyer will default at some rate $\delta$. If the borrower defaults, the lender sells the house. We assume the lender faces some exogeneous cost $c$, reflecting foreclosure discounts and other hassle costs of foreclosing, and also sells at some random price $\epsilon_F$, which has standard deviation $\sigma_\epsilon$, identical to the standard deviation of sale prices $P$. Thus, the foreclosure price is:

$$F = 1 - c + \epsilon_F$$

For simplicity, we assume mortgages are non-recourse.\footnote{Mortgages are recourse in some states, but wage garnishment and other methods for collecting debt from borrowers after the house has been sold are expensive, and borrowers cannot be collected from if they file Chapter 7 bankruptcy.} If the lender has lent $L$ to the buyer, her payoff upon default is thus:

$$\max[L, 1 - c + \epsilon_F]$$

(20)

Hence, if a lender lends $L$ on the property, her expected loss conditional on default is:

$$Loss = E[L - \max[L, 1 - c + \epsilon_F]] = E\left[\max[0, L - (1 - c + \epsilon_F)]\right]$$

(21)

Expression (21) implies that the lender’s expected loss is increasing in $\sigma_\epsilon$. This is because the lender can recover at most $L$, and bears the cost when the foreclosure price is less than $L$. Thus, when the variance of the foreclosure price is larger, the lender’s expected losses on loans is higher. Let $\rho$ represent the lender’s cost of funds and $r$ represent the interest rate,
and fix the lender’s mortgage spread $r - \rho$. We assume lenders are willing to make any loan for which their expected profits are nonnegative:

$$\left(1 - \delta\right) \left( r - \rho \right) \geq \delta E \left[ \max \left[ 0, L - 1 + c - \epsilon_F \right] \right]$$

(22)

The LHS of (22) is lenders’ expected profit: the mortgage spread $r - \rho$ multiplied by the probability the borrower repays, $\left(1 - \delta\right)$. The RHS is lenders’ expected losses (21), multiplied by the default probability $\delta$. Expression (22) defines a maximum loan size the lender is willing to make, conditional on the parameters $\delta, r, \rho, c, \sigma, \epsilon$. We will call this cap $\bar{L}_{fair}$. When the variance $\sigma_\epsilon$ increases, holding fixed other parameters, (22) implies that lenders’ maximum loan size will decrease, pushing loan-to-price ratios downwards. This will also increase down payment requirements for buyers, causing more transactions to fail.

**Appraisals.** We assume that institutions are subject to an appraisal constraint on loans. We highlight the main features of the appraisal process that lead to the economic forces in our model and provide a detailed description in Appendix C.1. Since the majority of residential sales appraisals are based on the comparable sales method, we will model the appraisal price as having the same mean as the transaction price, but an independent idiosyncratic error term. That is, we have:

$$A = 1 + \epsilon_A$$

(23)

where we assume $\epsilon_A$ and $\epsilon_P$ are independently and identically distributed; thus, $A$ and $P$ are also i.i.d., and the variation of $\epsilon_A$ is $\sigma_\epsilon$. Intuitively, it is reasonable to think that $A$ and $P$ have similar distributions, since houses chosen as comparables for appraisal purposes are chosen to be comparable, in terms of geographic proximity and house features, to the house being purchased.\(^{21}\)

Loans are subject to an LTV constraint, imposed on the minimum of the appraisal and the purchase price. The loan must be less than $\phi$ times the minimum of the purchase price and the appraisal price, so we must have:

$$L \leq \bar{L}_{appr} \equiv \phi \min \left[ P, A \right]$$

(24)

\(^{21}\)Since the appraisal generally uses a number of comparable sales, in a more realistic model, the idiosyncratic variance in $A$ would be somewhat lower than the variance of $P$. However, appraisals generally use around 3-7 sales in practice, so appraisals should be nontrivially noisy estimates of fundamental values, and the noise in appraisals should be directly related to the noise in sale prices.
The result of (24) is that appraisals create a random ceiling on loan sizes. The ceiling is binding if \( \bar{L}_A \) is below the fair-market loan limit. When \( \sigma_\epsilon \) increases, the appraisal-to-sale price ratio \( \frac{A}{P} \) becomes more volatile; it becomes more likely that the property will appraise at a price much lower than the attempted sales price, imposing a binding cap on the loan size. As a result, the loan-to-price ratio \( \frac{L}{P} \) will tend to be lower when \( \sigma_\epsilon \) is higher.

Note that, in practice, it is known that appraisals appear to be distorted such that they are rarely lower than house sales 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). Appraisers appear to use a variety of methods, such as misreporting house attributes (Eriksen et al., 2020b) and strategically changing the weights on different comparable properties’ transaction prices (Eriksen et al., 2019).

We do not explicitly model these forces. In practice, if appraisers were able to distort appraisal prices by arbitrary amounts costlessly, the appraisal channel we describe here would not work. However, a nontrivial share of appraisals do fail in practice, suggesting that appraisers may face some costs of distorting appraisals by large amounts. If this is the case, the appraisal channel should still work: when comparable sales prices are noisier, it becomes more likely that the fair appraisal price of the house is sufficiently far below the transaction price that the appraiser finds it too costly to distort the appraisal upwards, and the house thus under-appraises. Supporting this interpretation, Figure 8 shows empirically that a larger share of houses under-appraise in counties with higher value uncertainty. Moreover, Figure 3 shows that a larger share of mortgages are rejected, and more mortgages are rejected explicitly for collateral-related reasons, in areas with higher value uncertainty.

The following proposition collects the equations determining model outcomes.

**Proposition 1.** Suppose a buyer attempts a transaction with parameters \( \epsilon_P, \epsilon_A, \delta, \theta, c, \sigma_\epsilon \). There are two constraints on loan size. First, for any interest rate \( r \), there is a fair-pricing maximum loan size \( \bar{L}_{fair} \), which is the value of \( L \) which satisfies:

\[
\bar{L}_{fair} = \left\{ L : L (1 - \delta) (r - \rho) = \delta E \left[ \max \left[ 0, L - (1 + c + \epsilon_F) \right] \right] \right\}
\]

(25)

There is an appraisal constraint, which satisfies:

\[
L \leq \bar{L}_{appr} = \phi \min \left[ 1 + \epsilon_P, 1 + \epsilon_A \right]
\]

(26)
where \( \phi \) is an exogeneous parameter. Thus, the borrower faces a loan size cap:

\[
\bar{L} = \min (\bar{L}_{\text{fair}}, \bar{L}_{\text{appr}})
\]

If the mortgage application succeeds, the transaction occurs at price:

\[
P = 1 + \epsilon_P
\]

The application succeeds if:

\[
\theta \geq 1 - \frac{\bar{L}}{P}
\]

where \( \theta \sim F_\theta (\cdot) \) is the maximum down payment fraction the buyer can afford. If the max loan size is \( \bar{L} \), a fraction

\[
F_\theta \left( 1 - \frac{\bar{L}}{P} \right)
\]

of loans will fail.

We proceed to numerically solve the model to illustrate results. Figure A1 shows how \( \bar{L}_{\text{appr}} \) and \( \bar{L}_{\text{fair}} \) vary with \( \sigma_\epsilon \) and \( \delta \). Figure A2 shows how LTP and fail probabilities vary. The first prediction of the model concerns how \( \sigma_\epsilon \) affects loan-to-price ratios.

**PREDICTION 1.** When \( \sigma_\epsilon \) increases, holding fixed loan interest rates, the average loan-price ratio, \( \frac{L}{P} \), decreases.

Prediction 1 follows from Figure A1. The intuition is that, under both channels, loan limits are lower when \( \sigma_\epsilon \) is higher. However, the mechanisms are slightly different in the two cases. For the appraisal channel, increasing \( \sigma_\epsilon \) depresses \( \bar{L}_{\text{appr}} \) because appraisals are more likely to be far below house prices, making appraisal-based loan size caps more binding. For the fair pricing channel, lenders’ expected losses conditional on default are increasing in \( \sigma_\epsilon \). When \( \sigma_\epsilon \) is higher, holding fixed loan interest rates, banks must lower loan sizes to break even. As a result, from the left panel of Figure A2, loan-to-price ratios are decreasing as \( \sigma_\epsilon \) increases.

**PREDICTION 2.** When \( \sigma_\epsilon \) is higher, a larger fraction of mortgage applications fail.

Prediction 2 is shown in the right panel of figure A2. It follows immediately from prediction 1: since loan sizes are smaller when \( \sigma_\epsilon \) is larger, buyers’ down payment constraints
are more likely to bind, so a larger fraction of transactions will fail.

**PREDICTION 3.** The effect of $\sigma_\epsilon$ on LTVs is larger when the default rate $\delta$ is higher.

Prediction 3 follows from the left panel of Figure A1: the slope of the relationship between LTPs and $\sigma_\epsilon$ is steeper for consumers with higher default rates. This is because, in the fair pricing channel, the recovery value of collateral only matters if consumers actually default. Thus, lenders’ maximum LTPs are higher and less sensitive to collateral price dispersion for borrowers with lower default rates.

Note that the constraint imposed by the appraisal channel, in our model, is fully independent of $\delta$. In practice, however, appraisal constraints for GSE loans are set not as hard limits, but as interest rate adjustments, and these adjustments tend to be larger for consumers with lower credit scores.\(^{22}\) Thus, it is possible for the appraisal channel to also affect borrowers of different credit scores heterogeneously.

Our main empirical specifications test the effects of value uncertainty on interest rates. Thus, for simplicity, we have assumed that that interest rates are exogeneous, and lenders set LTV limits so that they break even in expectation. We could alternatively assume that LTVs are exogeneous and interest rates vary. From arguments analogous to (25), fixing a loan size $L$, the interest rate that makes lenders break even is:

$$r_{\text{fair}} = \left\{ r : L (1 - \delta) (r - \rho) = \delta E \left[ \max \left[ 0, L - (1 + c + \epsilon F) \right] \right] \right\}$$

(27)

In (27), the RHS is increasing in $\sigma_\epsilon$ and the LHS is increasing in $r$, so $r_{\text{fair}}$ must be increasing in $\sigma_\epsilon$. In words, fixing a loan-to-value ratio, interest rates will be higher when $\sigma_\epsilon$ is higher. This is consistent with our evidence in subsection 3.3.2, where we show that the entire menu of LTV-interest rate pairs shifts monotonically with $\sigma_\epsilon$.

\(^{22}\)See the Fannie Mae and Freddie Mac pricing matrices, which specify interest rate adjustments as a function of LTV and credit score.
**Figure A1.** Behavior of $\bar{L}_{appr}$ and $\bar{L}_{fair}$

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 $\delta$), and the right panel shows the average value of $\bar{L}_{fair}$, as $\sigma_\epsilon$ varies, for different values of $\delta$. Throughout, we set $\phi = 0.85$, $c = 0.2$, $r - \rho = 0.005$.

**Figure A2.** LTP and fail probabilities

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 $\delta$. Throughout, we set $\phi = 0.85$, $c = 0.2$, $r - \rho = 0.005$. 
Figure A3. Binding constraints

Note: The above figure shows the fraction of successful mortgages for which the appraisal constraint is binding, $\bar{L}_{appr} < \bar{L}_{fair}$, as we vary $\delta$. We set $\sigma_e = 0.15$, $\phi = 0.85$, $c = 0.2$, $r - \rho = 0.005$. 
C Appraisal Distortion

C.1 Distorted Appraiser Incentives

There is evidence in the literature that appraisals are not simply averages of comparable sales prices: a large literature has argued that appraisals are biased upwards, and bunched above sale 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). There are a few mechanisms through which this occurs. First, there is some selection bias, since when appraisals are low, transactions may fail (Ding and Nakamura, 2016) or prices may be negotiated downwards (Fout et al., 2021). Appraisers also appear to shift prices upwards when they are below contract prices, for example, by misreporting comp or house characteristics (Eriksen et al., 2020b) or changing the weightings on comparable sales (Eriksen et al., 2019).

A simple question is, if appraisers can simply shift appraisal prices, whether appraisal and house price variance indeed constrain credit provision as our model suggests. We found in subsection 3.3.3 that price dispersion is associated with higher rates of collateral-related mortgage rejections, suggesting that dispersion does affect the likelihood that appraisals fail. Moreover, as figure A4 shows, while the appraisal values on most properties are bigger or equal to their transaction prices, there are about 3.5% houses receiving appraisals below their transaction prices. There is also regional heterogeneity in the share of houses that are under-appraised or over-appraised (Figure A5). These results could be explained by, for example, a model in which appraisers have some ability to bias appraisals upwards, but are unable (or unwilling) to raise appraisals to the level of contract prices when the gap is too large.
Figure A4. Distribution of Appraisal to Sale Price

Note: This figure shows the distribution of appraisal to sale price ratios. Panel (a) plots the histogram of property level appraisal to price ratios. Panels (b) and (c) plots the zip code share of houses with appraisal value below and above sale price, respectively. We keep zip codes with at least 20 transactions from 2001 to 2016. Source: Corelogic LLMA.
Figure A5. Regional Heterogeneity of Appraisal to Sale Price

Note: This figure shows the regional heterogeneity of appraisal to sale price ratios. Panel (a) plots the zip code share of houses with appraisal value below sale price. Panel (b) plots the zip code share of houses with appraisal value above sale price. We keep zip codes with at least 20 transactions from 2001 to 2016. Source: Corelogic LLMA.
C.2 GSE Appraisal Policies

The appraisal channel applies to only securitized loans. We exploit this feature to identify the distortion created by appraisal channel by comparing GSE loans and portfolio loans.

We have already shown the difference in the LTP to price dispersion sensitivity between GSE loans and portfolio loans in Table 4. Specifically, we regress LTP on zip code price dispersion and other controls for GSE loans and portfolio loans separately. The estimated coefficient on zip code price dispersion is more negative in the GSE sample. For every standard deviation increase in price dispersion, the LTP of a GSE loan decreases by 38bps, while the LTP of a portfolio loan decreases by 28bps. The results suggest that the LTV rule of securitization further restricts credit provision in areas with high price dispersion.

Borrowers of conforming GSE loans may have different characteristics than borrowers of jumbo portfolio loans. To conduct a tighter comparison, we exploit a size discontinuity in GSE eligibility at the conforming loan limit, similar to Hurst et al. (2016). Jumbo loans exceed the limits set by the Federal Housing Finance Agency and cannot be securitized by GSEs. We explore the differential LTP to price dispersion sensitivity using a regression discontinuity approach:

\[
LTP_{ikt} = \alpha + \beta_0 \text{Dispersion}_{it} + \beta_1 \text{Bin}_{it} + \delta \text{Bin}_{it} \times \text{Dispersion}_{it} + (X_{it}\Gamma)D_{it}^{\text{jumbo}} + \mu_{kt} + \nu_{m} + \epsilon_{ikt}
\] (28)

We pool the GSE sample and the jumbo sample for the years 2001 to 2016. \(D_{it}^{\text{jumbo}}\) is an indicator for whether the loan is a jumbo loan. The specification allows the responsiveness of LTP to \(X_{it}\) (house price and borrower credit score) to differ across the types of loans. The key additions to this specification are the interaction of \(\text{Dispersion}_{it}\) and \(\text{Bin}_{it}\). For each loan, we compute the ratio of loan amount to the county conforming loan limit. \(\text{Bin}_{it}\) is an indicator for the distance between the loan size and the conforming loan limit. Specifically, \(\text{Bin}_{it}\) is defined in 0.05 unit intervals of the ratio of the loan size to the conforming loan limit, ranging from 0.4 to 1.6. We remove loans with ratios in range 0.95 to 1.05 to alleviate the concerns that (1) borrowers bunch at the conforming loan limit and bunching of consumers below the conforming loan limit and (2) loans at the conforming loan limit may be misclassified due to rounding of loan amount.
Figure A6 shows the estimations of $\delta$. Benchmarking to the highest ratio (1.6), the LTP’s of loans below the conforming loan limits are more responsive to the house price dispersion. The results show that credit provision with respect to house price dispersion changes discretely between the GSE sample and the jumbo sample. Table A1 shows the regression-discontinuity estimates of the differences in LTP-price dispersion sensitivity. Columns 1-4 gradually narrow down the sample from loans with size 0.4-1.6 times of the conforming limit to loans with size 0.8-1.2 times of the conforming limit. Within the narrowest size-to-conforming limit range, LTP decreases by 16bps more for conforming loans as the zip-code idiosyncratic price dispersion increases by one standard deviation. Our RD estimate is similar to the estimated coefficient difference in Table 4, confirming the baseline OLS result.

**Figure A6. Discontinuity at Conforming Loan Limit**

*Note:* This figure shows the LTP-price dispersion sensitivity for conforming and jumbo loans. We group all loans based on their loan to conforming limit ratios. Each bucket $r$ contains a 0.05 ratio range. Dots are the estimated $\beta$ in expression 28. The omitted baseline bucket is 1.65. Thus, each dot represents the LTP-price dispersion sensitivity relative to the sensitivity of loans falling into the 1.6 ratio bucket. We remove loans with ratios in range 0.95 to 1.05 to alleviate the concerns that (1) borrowers bunch at the conforming loan limit and bunching of consumers below the conforming loan limit and (2) loans at the conforming loan limit may be misclassified due to rounding of loan amount. The sample includes property level observations from 2001 to 2016. *Source:* Corelogic Deeds.
Table A1: Appraisal Channel: Discontinuity at Conforming Loan Limit

This table presents the regression discontinuity results. The outcome variable is loan to price ratio. The explanatory variable of interest is property level price dispersion interacted with the conforming indicator. The conforming indicator is 1 if the loan amount to conforming loan limit is below 1 and 0 otherwise. Column 1 uses the sample of loans whose ratios are from 0.4 to 1.6. Column 2 uses the sample of loans whose ratios are from 0.6 to 1.4. Column 3 uses the sample of loans whose ratios are from 0.8 to 1.2. We remove loans at the conforming loan limit. Borrower/Loan controls include county average fico score, fico-squared, and log of house price. The sample includes property transaction level observations from 2001 to 2016. Standard errors are clustered at county level. ***, **, * represent 1%, 5%, and 10% significance, respectively.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price Dispersion × Conforming</td>
<td>-0.77***</td>
<td>-0.43***</td>
<td>-0.16***</td>
</tr>
<tr>
<td></td>
<td>(0.056)</td>
<td>(0.045)</td>
<td>(0.023)</td>
</tr>
<tr>
<td>Dispersion Mean</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>Dispersion Stdev</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>Origination Month FE</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>County-Year FE</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Borrower/Loan Controls</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>R2</td>
<td>0.39</td>
<td>0.57</td>
<td>0.83</td>
</tr>
<tr>
<td>Observations</td>
<td>6.2M</td>
<td>3.1M</td>
<td>1.2M</td>
</tr>
</tbody>
</table>
D Additional Details for Section 4

D.1 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 \( \sigma \) 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 \( \phi \) as a function of house prices rather than qualities.

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

\[
\sigma_{med}(p) \equiv E[\sigma_{ict} | p_{ict} = p, c \in C_{mid}]
\]

where we used \( c \) to index counties, and \( c \in C_{mid} \) means that county \( c \) is in the middle quantile of counties by price dispersion. We define \( \sigma_{high}(p) \) and \( \sigma_{low}(p) \) analogously to (29), for the high- and low-dispersion set of counties. The three curves \( \sigma(p) \) curves are shown in the left panel of Figure 10. We normalized \( \sigma \) by its standard deviation across houses, so the units are identical to those of Table 6. High-dispersion counties have roughly a standard deviation higher values of \( \sigma \) 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 \( \sigma \), 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:

\[
\phi_{med}(p) = 0.8 + \beta_{LTV,\sigma} (\sigma_{med}(p) - \sigma_{med}(p_{min,\sigma}))
\]

(30)
Where $\beta_{LTV,\sigma}$ is the coefficient from regressing LTV on price dispersion, from column 1 of Table 6. In words, (30) 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 $p_{\text{min,}\sigma}$, plus an adjustment which is the difference between price dispersion at $p$, and price dispersion at $p_{\text{min,}\sigma}$, multiplied by $\beta_{LTV,\sigma}$, the effect of price dispersion on LTVs identified in our reduced-form results. Note that we adjust using $\beta_{LTV,\sigma}$, 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 $\beta_{LTV,\sigma}$.

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

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

That is, analogous to (30), $\phi_{\text{high}}(p)$ is set so that, for any price $p$, the difference $\phi_{\text{high}}(p) - \phi_{\text{med}}p_{\text{min,}\sigma}$ is equal to the dispersion difference, $\sigma_{\text{high}}(p) - \sigma_{\text{med}}(p_{\text{min,}\sigma})$, multiplied by $\beta_{LTV,\sigma}$.

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

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

Figure 10 shows the resultant $\phi_{\text{low}}(p), \phi_{\text{med}}, \phi_{\text{high}}$ functions. The left panel shows that high and low-dispersion groups differ by around 1SD of $\sigma$; multiplying by the $\beta_{LTV,\sigma}$ coefficient, we get an average difference in LTVs of approximately 1.1% between $\phi_{\text{low}}(p)$ and $\phi_{\text{high}}(p)$ 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.
D.2 Numerically Solving the Model

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

\[ w_t = a_t + 1 (h_t > s_R) p^h h_t \]

From (11), the leverage constraint then becomes:

\[ w_t \geq (1 - \phi (h_t)) p^h h_t \]

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

Combining the owner and renter budget constraints, (13) and (15), and rewriting expressions in terms of wealth, we can write the budget constraint equation as:

\[
\begin{align*}
    w_{t+1} = & \\
    (1 + r_t) \left( w_t + y_t - c_t - 1 (h_{t+1} > s_R) \left( 1 + \delta^h + F^{par} 1 (h_{t+1} \neq h_t) \right) p^h h_t - 1 (h_t < s_R) p^r \right) + & \\
    & 1 (h_{t+1} > s_R) p^h h_t \quad (33)
\end{align*}
\]

Using (33), we eliminate consumption \(c_t\) from the household’s optimization problem, (12) and (12). The household thus chooses end-of-period wealth \(w_{t+1}\) and house quality \(h_{t+1}\) each period, where the state variables are \(w_t, h_t, \zeta_t\).

To solve the problem, we discretize \(\zeta_t\) 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 $\zeta_t$ at its stationary distribution. We then simulate 1,000,000 households over their lifespan, and take average quantities over all households.
E  Additional Empirical Analysis

Figure A7. County Level House Price Dispersion and LTP

Note: This figure shows the correlation between county level house price dispersion and residualized county average LTP. Panels a-c plot GSE loans, FHA loans, and jumbo loans, respectively. The sample includes annual county observations from 2000 to 2020. Source: County house price dispersion is estimated using Corelogic Deeds records. Mortgage data are from Corelogic LLMA.
Figure A8. County Level House Price Dispersion and LTP Residuals

Note: This figure shows the correlation between county level house price dispersion and residualized county average LTP. Panel a plots the full sample. Panels b-d plot GSE loans, FHA loans, and jumbo loans, respectively. We residualize LTP values by taking the residuals of regressions of LTP on mortgage interest rate, debt-to-income ratio (DTI), DTI-square, FICO, FICO-square, log house price, and their interactions with origination years, and origination year fixed effects. We then take the county-average of residualized LTP. The sample includes annual county observations from 2000 to 2020. Source: County house price dispersion is estimated using Corelogic Deeds records. Mortgage data are from Corelogic LLMA.
Figure A9. County Level House Price Dispersion and Mortgage Rate

Note: This figure shows the correlation between county level house price dispersion and residualized county average mortgage interest rate. Panels a-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.