Collateral Value Uncertainty and Mortgage Credit Provision

Erica Xuewei Jiang and Anthony Lee Zhang

Abstract

Using property transaction and financing data, we document large cross-sectional differences in how effective houses are as collateral for mortgages. Older and less standardized houses tend to have higher price dispersion, and their appraisal values tend to deviate more from transaction prices. Mortgages collateralized by these houses 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 urban policies.
1 Introduction

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

In this paper, we document large cross-sectional differences in how effective houses are as collateral for mortgages. Older and less standardized houses have higher price dispersion. Price dispersion affects financing: mortgages backed by high-dispersion houses have lower loan to price ratios, higher interest rates, and are more likely to be rejected. Price dispersion matters for both portfolio loans and securitized loans, where price dispersion affects the latter 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 alleviate a structural feature of the housing stock which limits low-income households’ credit access even in efficient mortgage markets.

We begin by discussing two channels through which house value uncertainty could affect mortgage credit provision. The first is the “fair pricing” channel. The payoff of collateralized debt, upon default, is concave in the resale value of the collateral: lenders do not benefit

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.
if a foreclosed house sells for more than the outstanding mortgage balance, but are on
the hook if the house sells for less. Thus, lenders should be less willing to lend against
collateral which has more variable resale prices. The second, which is unique to the residential
housing market, is the “appraisal” channel. Mortgage loan-to-value limits depend on house
appraisals, which are determined based on the prices of recently sold similar properties.
When house value uncertainty is higher, appraisals will tend to be noisier, generating more
downswards pressure on mortgage amounts. Both these channels predict that, when house
value uncertainty is higher, mortgages should be smaller and mortgage rejections should
be more common. In addition, since collateral values only matter when the homeowner
defaults in the fair pricing channel, the model predicts that the effect of value uncertainty
on mortgage credit should be larger when borrower default rates are higher.

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

We then show that collateral value uncertainty affects financing at both the extensive
margin and the intensive margin. At the regional level, counties with higher price disper-
sion 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 trans-
acted 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. On average, LTPs
are 20-46bps lower for houses with one standard deviation higher estimated price dispersion,
and mortgage applications are 1.4-1.8 percentage points more likely to be rejected and
50 basis-points more likely to be rejected due to collateral reasons. The effect is economi-
cally significantly: given the sample average rejection rate, the effect amounts to about 10%
increase in rejection likelihood.

Our baseline identification strategy exploits within county-year variation by comparing
two properties that are bought in the same county-year, at the same price, and by buyers with the same credit profile and income. To address concerns that house price dispersion is associated with other unobserved characteristics that also affect mortgage credit provision, we construct instruments for price dispersion, based on the heterogeneity of a house relative to its local housing stock. Our results continue to hold using the instrumental variable approach.

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 entire menu shifts upwards: for any given LTP, borrowers in high-dispersion zip codes pay higher risk-adjusted interest rates on average.

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. Benchmark to the baseline LTP-price dispersion sensitivity among Excellent credit history homebuyers, a one standard deviation increase in price dispersion induces a 110bps more reduction in LTP for homebuyers with Poor 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.

While it is plausible that lenders account for collateral price dispersion for their portfolio loans, it is unclear that lenders have incentives to fairly price collateral dispersion for loans that they plan to securitize (Hurst et al., 2016; Mian and Sufi, 2009; Keys et al., 2010; Purnanandam, 2011). We argue that the dispersion-LTP relationship in the securitized segment of the market works through a novel channel, based on the residential appraisal process. Securitized mortgages require houses to undergo appraisals, where the appraisal value is determined based on the prices of other similar houses sold recently. The value of a house, for loan-to-value calculations, is set equal to the smaller of the transaction price and the appraisal price, implying that appraisals are binding constraints on mortgage amounts only when they are lower than the transaction price. Thus, when idiosyncratic price dispersion is higher and appraisals are noisier, under-appraised houses tend to under-appraise by larger amounts, creating more downwards pressure on the size of mortgage
loans. Empirically, we confirm that the deviation of appraisal prices from sale prices for under-appraised houses is larger in zip codes with higher price dispersion.

We perform three main robustness tests of our results. First, we show that the instrumented price dispersion is uncorrelated with default rates, suggesting that our IV results are not driven by unobserved differences in borrowers’ creditworthiness. 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 zip code 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.

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

This paper relates to a number of strands of literature. Broadly, our paper fits into a literature on frictions that affect credit access. (Mian and Sufi, 2011; Agarwal et al., 2017; Beraja et al., 2019; DeFusco et al., 2020; Adelino et al., 2020; Collier et al., 2021; Buchak et al., 2018a; Jiang, 2020) and homeownership (Glaeser and Shapiro, 2003; Gupta et al., 2021). 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. Lang and Nakamura (1993) theoretically argue that the precision of appraisals influences home sales through down payment requirements, leading to sub-optimal lending outcomes. Blackburn and Vermilyea (2007) empirically tests the theories of rational redlining and shows that a low volume of home sales lead to uncertainty in house appraisals, reducing mortgage lending.

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; Pan et al., 2021; 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 (2021) 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 conceptual framework and measurement strategy. Section 3 describes our data and stylized facts on our price dispersion measure. Section 4 studies the effect of price dispersion on mortgage provision. Section 5 calibrates a model to analyze the effect of collateral values on homeownership rates. We discuss implications of our results in section 6, and conclude in section 7.

2 Conceptual Framework and Measurement

2.1 Conceptual Framework

In this section, we discuss why houses have price dispersion, and how price dispersion should affect the provision of mortgage credit. We construct a formal model of dispersion and its effects on mortgage credit provision in Appendix A.

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

House price dispersion can affect credit provision through two channels, both of which imply that houses with higher value uncertainty are more difficult to lend against. 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.\(^2\)

The second channel, which we call the *appraisal channel*, is unique to the residential housing market. Residential properties are appraised almost exclusively through a 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.\(^3\) 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 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. Houses with higher value uncertainty may have noisier appraisal values, which leads to downward pressure on

\(^2\)This 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).

\(^3\)Appraisals 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.
mortgage loan amounts.

Our conceptual framework makes three predictions that we will bring to the data. We derive these predictions formally in Appendix A.

**PREDICTION 1.** For any given interest rate, loan-to-price ratios should be lower for houses with higher value uncertainty.

**PREDICTION 2.** Mortgages are more likely to be rejected when value uncertainty is higher.

The intuition behind prediction 2 is that a mortgage application is rejected if the loan amount that the buyer demands given her cash-on-hand for down payments is higher than the maximum loan amount that the lender is willing to offer given the collateral value. When house value uncertainty is higher, lenders’ loan-to-price limits will be lower, so a larger fraction of mortgage applications will be rejected.

**PREDICTION 3.** The relationship between value uncertainty and mortgage terms should be stronger for borrowers with higher default risks.

Prediction 3 follows because, in the fair pricing channel, the value of collateral matters only if consumers actually default. Hence, price dispersion should affect mortgage credit provision more when default rates are higher.

### 2.2 Measuring Value Uncertainty

We proceed to empirically estimate house price dispersion, at the level of individual house sales, essentially by measuring what kinds of houses tend to have smaller errors when priced with a hedonic regression.\(^4\) Our estimation has two steps. First, we regress transaction prices on house characteristics:

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

\(^4\)A similar methodology is used in Buchak et al. (2020).
We then regress the squared residuals, $\hat{\epsilon}_{it}^2$, from (1), on a flexible function of characteristics and time, to predict which house characteristics make them difficult to price:

$$\hat{\epsilon}_{it}^2 = g_k(x_i, t) + \xi_{it}$$  \hspace{1cm} (2)

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

$$\hat{\sigma}_{it}^2 \equiv \hat{g}_k (x_i, t)$$  \hspace{1cm} (3)

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

\begin{itemize}
\item Note that it is important to use the predicted values of $\hat{\sigma}_{it}^2$ in stage 2 rather than the residuals $\hat{\epsilon}_{it}^2$ in stage 1 directly. This is because the expected value of idiosyncratic dispersion, $\sigma_{it}^2$, 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.
\item According to Piazzesi and Schneider (2016), roughly half of the total volatility in a house price transaction is idiosyncratic.
\end{itemize}
We next discuss why our measurement strategy is robust to two potential concerns. First, we measure value uncertainty as independent shocks at each house sale rather than fluctuations over time that scale with the holding period of a house. This modeling assumption is justified by evidence in Giacoletti (2021) and Sagi (2021), which shows that idiosyncratic component of house price risk has a very flat term structure, scaling very little with the holding period of a house. Also, idiosyncratic price dispersion mainly varies in the cross-section and has relatively small time-series variation (Kotova and Zhang, 2021). Intuitively, search frictions, market thickness, and heterogeneous preference are among the main drivers of the idiosyncratic volatility; these forces tend to generate price shocks that are realized upon sales, rather than a drift term which increases in variance substantially depending on house holding periods.

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

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

3.1 Data Sources

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

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

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

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

Table 1 provides summary statistics.
3.2 Estimated Value Uncertainty and Housing Market Frictions

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

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

To explore this further, Table 2 presents the association between estimated value uncertainty and house characteristics. Panel A analyzes house features. Throughout, we control for linear and squared terms in log house prices, comparing houses with similar prices and different characteristics. Older houses have higher price dispersion (column 1). Controlling for building age, recently renovated houses within 5 years of the transaction date (column 2), or houses that were ever renovated (column 3), have lower price dispersion.\(^7\) Columns 4-6 present the association between property size, measured by square-footage and number of bedrooms, and price dispersion. There is a U-shaped relationship: price dispersion is low for moderately large houses, and higher for houses which are very large or very small. In terms of local housing market conditions, Panel B of Table 2 shows that houses in zip codes with larger income inequality, less population density, and more vacancies tend to have higher price dispersion. Together, Table 2 suggests that house price dispersion is essentially driven by house standardization and market thickness.\(^8\) Lastly, Figure 1 Panel B shows the relationship between price dispersion and average zip-code incomes. Price dispersion tends to be higher in low-income zip codes.

\(^7\)We can partially measure house renovations, as the Corelogic tax data contains an “effective year built” variable, which tracks the last date at which a property was renovated.

\(^8\)This finding is consistent with evidence from other papers: see, for example, Kotova and Zhang (2021) and Andersen et al. (2021).
4 House Value Uncertainty and Mortgage Credit

In this section, we empirically test the relationship between house value uncertainty and mortgage credit provision. We begin by introducing county-level evidence before presenting property-level evidence, followed by the evidence for the appraisal channel. We conclude this section with a number of robustness tests.

4.1 County-Level Evidence

Figure 2 show that counties with higher price dispersion have lower average loan-to-price ratios (LTP), higher interest rates, and more mortgage rejections. Panel A of Figure 2 plots county average LTP against 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. Figure A5 shows that the relationship holds for GSE loans, FHA loans, and jumbo loans. Figure A6 further confirms the negative relationship using residualized LTP. Panel B of Figure 2 plots county average residualized mortgage rate against the average house price dispersion. Interest rates are higher in areas with higher price dispersion. Figure A7 shows that the relationship holds for GSE loans, FHA loans, and jumbo loans.

Panel A and B together suggest that mortgages in high-dispersion areas are less favorable for borrowers along both the price and LTP dimensions. This is consistent with lenders being less willing to supply credit rather than lower demand for credit in areas with lower price dispersion. 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. We will discuss this more formally in Subsection 4.2.3 below, by estimating the menu of LTP-interest rate pairs offered in each area and compare the menus in high-dispersion areas to the menus in low-dispersion areas.

Panels C and D of Figure 2 plot county-level average mortgage rejection rates against county-level price dispersion. Panel (c) shows all mortgage rejections, and panel (d) considers

---

9The residualized LTP takes 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.

10We residualize rates for individual mortgages on borrower and loan characteristics such as FICO, LTP, DTI, the squared terms, and their interactions with origination year.
only rejections where the prospective lender indicates that the reason for rejection is tied to collateral quality.\textsuperscript{11} Consistent with prediction 2, 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.

### 4.2 Property-Level Evidence

We next turn to property level evidence. We explain our identification strategy before presenting evidence supporting each of the predictions in Section 2.1.

#### 4.2.1 Identification

Our first identification strategy exploits within county-year variation by comparing two properties that are bought in the same county-year, at the same price, and by buyers with the same credit profile and income. To implement this strategy, our baseline property-level analyses include county-year fixed effects $\mu_{kt}$ and transaction date fixed effects $\nu_m$ as well as a rich set of borrower and loan characteristics $X_{ikt}$:\textsuperscript{12}

$$
Y_{ikt} = \alpha + \beta \text{Dispersion}_{ikt} + X_{ikt} \Gamma + \mu_{kt} + \nu_m + \epsilon_{ikt},
$$

(4)

$Y_{ikt}$ is one of the outcomes of the mortgage that is collateralized by property $i$ in county $k$ transacted in year $t$. $\text{Dispersion}_{ikt}$ is the estimated price dispersion of the underlying property.\textsuperscript{13}

In the second stage of $\text{Dispersion}_{ikt}$ estimation (Eqn. 2), we project the squared residuals from the first stage hedonic model onto the entire set of house characteristics. In a sense, house characteristics are used as instruments for price dispersion. The identifying assumption of the baseline specification 4 is that conditional on house price and borrower and loan characteristics, characteristics of a house only affect mortgage outcomes, insofar as they affect price dispersion. This assumption can be violated if unobservable borrower characteristics

\textsuperscript{11}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.

\textsuperscript{12}The set of characteristics vary for each outcome variable, which will be detailed later.

\textsuperscript{13}We aggregate property-level dispersion to 5-digit zip code or census tract level for analyses using Corelogic LLMA or HMDA.
are correlated with certain house features. For example, if borrowers who are more likely to purchase smaller, and thus high price dispersion, houses are also less creditworthy after conditioning on hard information like credit score and debt-to-income ratio, then the baseline specifications are subject to omitted variable bias.

We address this concern by instrumenting price dispersion using measures of house heterogeneity relative to the local housing stock. For each of the key house features,

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

we first calculate the average value of all houses transacted in each county \( c \), \( \bar{X}_c^m \). We then construct a set of house-level instruments, for the price dispersion of each individual house \( i \) in county \( c \). For each house, there are 5 instruments, one for each characteristic \( m \). The instrument \( Z_{im} \) is equal to the squared difference between the house’s feature \( m \), and the average value of \( m \) in county \( c \), that is:

\[
Z_{im} = (X_{im} - \bar{X}_c^m)^2, \quad \forall m \in \{age, sqft, bed, bath, geo\}, \tag{5}
\]

Then, we estimate the following 2SLS specification:

\[
\begin{align*}
\text{Stage 1:} & \quad \text{Dispersion}_{it} = \alpha + \beta_1 Z_{iit}^{age} + \beta_2 Z_{iit}^{sqft} + \beta_3 Z_{iit}^{bed} + \beta_4 Z_{iit}^{bath} + \beta_5 Z_{iit}^{geo} \\
& \quad + X_{ikt} \Gamma + \mu_{kt} + \nu_m + \epsilon_{ikt} \tag{6} \\
\text{Stage 2:} & \quad Y_{ikt} = \alpha + \hat{\text{Dispersion}}_{ikt} + X_{ikt} \Gamma + \mu_{kt} + \nu_m + \epsilon_{ikt},
\end{align*}
\]

where \( \hat{\text{Dispersion}}_{ikt} \) is the predicted value from stage 1.

The intuition behind the instruments is that they measure how locally thin the market for a given house \( i \) is, by measuring how nonstandardized house \( i \) is in each characteristics – how different \( i \)’s characteristics are from characteristics of other houses in the county. For example, in a county with mostly large houses, \( Z_{im} \) would be large for small houses. Small houses should have higher price dispersion in areas where most houses are large, since markets for small houses are thinner and there are likely fewer buyers at any given point in time.

Our house-level instruments alleviate the endogeneity concern to the extent that lenders’ willingness to lend against certain house feature — e.g., driven by the clustering of unob-
served borrower creditworthiness and house features — is monotone. We believe this is a reasonable assumption because it is implausible that borrowers purchasing atypical houses in a given county could be systematically less creditworthy than other borrowers. Moreover, we show in Subsection 4.4.1 that, while price dispersion is cross-sectionally correlated with borrower default rates, our instrument is correlated with price dispersion but not with borrower default rates, lending support to the idea that the instrument alleviates endogeneity due to borrower sorting.

We also construct instruments for price dispersion at zip-code level by taking geographical averages of \( Z_i^m \), which is a measure of how heterogeneous the housing stock in the zip-code is, along characteristic \( m \). Zip-codes with more heterogeneous housing stocks will tend to have higher price dispersion, since they have thinner local markets for any individual house, so there are likely to be less interested buyers for any given house.\(^{14}\)

### 4.2.2 Loan-to-Price Ratio

We start by analyzing the effects of collateral value uncertainty on property-level LTP. We first visualize the relationship between property-level LTP and collateral value uncertainty in Figure A8. To compare properties within a county, we regress LTP, and the estimated house price dispersion, on house prices and county-year fixed effects, and plot the LTP residuals against house price dispersion residuals. Figure A8 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.

Table 3 presents regression evidence. Column 2 corresponds to Specification 4, while columns 1 and 3 are less saturated specification with only transaction date fixed effects and more saturated specification further add lender-year fixed effects, respectively. For two houses in the same county that are transacted on the same data at the same price, the one with higher estimated price dispersion tends to receive a smaller sized loan. The loan-to-price ratio is about 20-46bps lower for houses with one standard deviation higher estimated price dispersion across these specifications. The result holds in the IV setting in columns 4-6, where column 5 corresponds to Specification 6, and column 4 and 6 are less and

\(^{14}\)The approach of using measures of house nonstandardization as instruments is not new to the literature: similar ideas are used in Andersen et al. (2021), and the approach can be micro-founded in a search and matching framework as done in Guren (2018).
more saturated specifications, respectively. LTP decreases by 47bps for every one standard deviation increase in the estimated price dispersion in the most saturated IV specification.

The estimated effect is economically significant. It is difficult to construct a precise quantitative benchmark for how large the effect of price dispersion on LTVs should be, since this depends on the volatility exposure of the implicit optionality of the mortgage debt contract. However, in Appendix A, we show that, under realistic choices for parameter values, a simple quantitative model of how LTPs depend on volatility can generate effects on LTVs of this magnitude or larger. A further question is whether the induced change in LTVs is large enough to move aggregate outcomes of interest. We address this in Section 5, where we calibrate a lifecycle model of homeownership choice, and show that price dispersion-induced changes in LTV can substantially decrease aggregate homeownership rates.

### 4.2.3 Mortgage Price Menus

We next show that the reduced LTP is not explained by credit demand, that is, borrowers substituting to smaller loans to obtain lower interest rates. We do this by estimating the entire menu of LTP-interest rate pairs that are available in a market using loan-level data, and showing that the entire menu is less favorable to borrowers in areas with higher house price dispersion.

Figure 3 plots the mortgage price menu, separately for groups of zip codes with high and low price dispersion. We residualize interest rates by purging out the effects of borrowers’ credit scores, loan type, and time fixed effects and plot a menu of average interest rates for different LTPs. Figure 3 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.

Table 4 presents the above results in regression settings. Panel A estimates Specification 4 (columns 1-3) and 6 (columns 4-6) with interest rate being the outcome variable while including LTP as one of the explanatory variables.\textsuperscript{15} Column 1 uses the full sample. We

\textsuperscript{15}We use zip-code dispersion instead of property-level dispersion because our price dispersion measure is estimated using Corelogic Deeds, and the data vendor prohibited us from merging loan-level records in LLMA with property-level records in Corelogic Deeds and Tax. We therefore aggregate property-level price dispersion measures to the most granular geographic region in LLMA.
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 a 79bps 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 1.11bps in zip codes with one standard deviation higher average house price dispersion. Columns 2 to 3 show the results for securitized loans and portfolio loans, respectively. The results hold in all samples. For every one standard deviation increase in zip-code average house price dispersion, the mortgage rate of securitized loans increases by 1.38bps, and the mortgage rate of the portfolio loans increases by 1.92bps. We confirm the results in the IV setting in columns 4-6.

Panel B estimates Specification 4 (columns 1-3) and 6 (columns 4-6) using LTP as the outcome variable, while including interest rate as one of the explanatory variables. To receive the same mortgage rate, houses with more higher price dispersion require higher down payments. Controlling for interest rates, the loan-to-price ratios decreases by about 50bps in zip codes with one standard deviation higher average house price dispersion (column 1). The results hold for securitized loans and portfolio loans (columns 2-3). For every one standard deviation higher zip-code average house price dispersion, the LTP ratios of securitized loans decrease by 51bps, and the LTP of portfolio loans decrease by 18bps. Again, we confirm the results in the IV setting in columns 4-6.

4.2.4 Mortgage Rejection Rates

We then study the effect of price dispersion on credit access at the extensive margin by estimating Specification 4 and 6 with mortgage rejection indicator being the outcome variable.\footnote{Again, we use census-tract dispersion instead of property-level dispersion because our price dispersion measure is estimated using Corelogic Deeds, and the data vendor prohibited us from merging loan-level records with property-level records. We therefore aggregate property-level price dispersion measures to the most granular geographic region in HMDA.} Panel A of Table 5 reports the results. Columns 1-3 present OLS results. We first confirm the effect of local house price dispersion on mortgage rejection using full sample (column 1). Census-tract house price dispersion is positively and significantly associated with mortgage rejections. This result holds for both securitized loans (column 2) and portfolio loans (column 3). The rejection rate increases by about 1.4-1.7 percentage points as house price dispersion increases by one standard deviation. The effect is economically significant: given the sample
average rejection rate of about 16%, the estimate amounts to about 10% increase in rejection likelihood.

We provide more direct evidence for the collateral channel by focusing on rejections due to collateral reasons in Panel B of Table 5. A mortgage application is about 50bps more likely to be rejected due to collateral reasons in a zip code with one standard deviation higher house price dispersion, which is about 25% increase in rejection likelihood. Again, the result holds in the full sample (column 1) as well as sub-samples of securitized loans (column 2) and portfolio loans (column 3) and is robust to using our IV specification (columns 4-6).

4.2.5 Heterogeneous Effect

Lastly, we show that the relationship between collateral value uncertainty and LTPs is stronger for borrowers with higher default rates. This supports the hypothesis that the relationship between LTPs and value uncertainty is driven by credit supply.

We estimate the following specification:

$$LTP_{ikt} = \alpha + \beta Rate_{ikt} + \gamma ZipDispersion_{ikt} \times CreditScore_{ikt} + X_{ikt} \Gamma + \mu_{kt} + \nu_m + \epsilon_{ikt} \quad (7)$$

where $ZipDispersion_{ikt} \times CreditScore_{ikt}$ is zip code price dispersion interacted with home buyer’s credit score, which is divided into five groups based on lenders’ common practice: Excellent (800-850), Very Good (740-799), Good (670-739), Fair (580-669), and Poor (300-579). $X_{ikt}$ includes zip code price dispersion, credit score, and other controls defined in Specification ???. 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.

Figure 4 visualizes the heterogeneity in our estimates of $\gamma$ across FICO score buckets. 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.

Table 6 presents the regression evidence. Among home buyers with Excellent credit score, 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. Benchmarking to the baseline LTP-price dispersion
sensitivity among Excellent credit history home buyers, average LTPs decrease by 41bps, 70bps, 86bps, and 105 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 conventional loans sold to GSEs, Ginnie, or other investors indicated in the Corelogic LLMA dataset and define portfolio loans as conventional loans that are not sold. Overall, the estimates are consistent with fair pricing of collateral risk in either sample: the $\gamma$ estimate is not statistically significantly different from zero among home buyers with Excellent credit score, and the LTP-price dispersion sensitivity increases as home buyers become less credit worthy. However, note that the price dispersion-LTP relationship is much less sensitive to buyer creditworthiness in the securitized loan sample. We confirm the results using our house nonstandardness IV in columns 4-6.

### 4.3 Price Dispersion and Appraisals

While the LTP-price dispersion sensitivity is plausibly driven by the fair-pricing channel 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). However, 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: when house prices are more disperse, appraisals are also more disperse, creating downwards pressure on LTPs, since appraisal constraints are only binding when houses under-appraise.

---

17 Most 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 sheets, while the remaining were securitized (Keys et al., 2013); and post crisis about 80 percent were securitized (Buchak et al., 2018b,a; Jiang, 2020).

18 One concern is that multiple comparable sales are used to appraise any given house, and the idiosyncratic error terms may average out somewhat across comps. Quantitatively, this is unlikely to eliminate all idiosyncratic appraisal dispersion. 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. The variance of appraisal prices induced by idiosyncratic errors in comps should thus range from $26\%\sqrt{7}$ to $26\%\sqrt{3}$, or 9.83% to 15.01%, depending on the number of comps used. This back-of-envelope estimate of predicted appraisal dispersion has a similar magnitude to estimates in the literature; for example, Agarwal et al. (2020) find that appraisal prices have a standard deviation of 13.4% relative to AVM prices.
Figure 5 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 5, which we call the appraisal deviation from sales price, is defined as:

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

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 5 shows that the appraisal deviation from sales prices is much higher in high-dispersion zip codes, suggesting that the extent to which under-appraisals put downwards pressure on LTVs is larger in high-dispersion zip codes.

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

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

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

Table 7 confirms Figure 5 findings in regression settings with origination month fixed effects, county-year fixed effects, and borrower and loan controls. In high-dispersion zip codes, appraisal deviations tend to be larger: a 1SD increase in dispersion is associated with a 2bp change in the appraisal deviation (column 1). This is mostly because, conditional on under-appraisal, houses under-appraise by larger amounts: a 1SD increase in dispersion is associated with a 53bp increase in the conditional appraisal deviation (column 3). The probability of under-appraisal is statistically significant (column 2). The results are robust
in the IV setting as in columns 4-6.

A remaining question is why, as we showed in Subsection 4.2.5, 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.\textsuperscript{19} Thus, the effective cost of underappraisals is higher for low-credit-score borrowers, so they may set LTVs more conservatively as a result.

4.4 Robustness

4.4.1 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. To address this, we assessing the ex-post performance of mortgage loans, to test whether ex-post default rates are associated with house price dispersion. Table 8 Panel A estimates the specifications 4 (columns 1-3) and 6 (columns 4-6) but sets the outcome variable equal to 100 for loans that become 60 or more day-delinquent within 2 years after origination and zero otherwise. Columns 1 and 4 include the full sample. Columns 2 and 5 restricts the sample to securitized loans. Columns 3 and 6 restricts the sample to portfolio loans. All regressions include the full set of borrower and loan characteristics as in our main regression specifications.

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

Note that, in our OLS specifications, the coefficient estimate on price dispersion is positive and statistically significant. This could be because certain house characteristics, which are associated with higher house price dispersion, also tend to attract homeowners who have

\textsuperscript{19}See the Fannie Mae and Freddie Mac pricing matrices, which specify interest rate adjustments as a function of LTV and credit score
higher default rates. If this were the case, it would upwards bias our OLS estimates of
the effect of price dispersion on LTPs. This further validates the importance of using our
instrument, which is associated with price dispersion, but is not associated with homeowners’
default rates.

4.4.2 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 4 and
6 with lender-zip-year fixed effects using a sub-sample of house transactions in Corelogic
Deeds records that we also observe the mortgage interest rates. Note that we cannot do this
robustness check using Corelogic LLMA data as we did in Section 4.2.3 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 2-3, we add in more saturated lender fixed effects:
lender-county-year and lender-zip-year fixed effects, respectively. The results hold in all
specifications, confirming that the effect of house price dispersion on mortgage credit is not
driven by lender market power.

4.4.3 Bunching Below Conforming Loan Limit

Lastly, we test whether the effect of price dispersion on mortgage LTP and cost menu is driven
by home buyers lowering the loan-to-price ratio to be eligible for securitization with the par-
ticipation of government-sponsored enterprises (GSEs). Specifically, conforming mortgages
must be below the conforming loan limits, which vary across regions and time. Conforming
loans are much easier to sell than non-conforming loans, also known as jumbo loans, because
of the participation of GSEs. GSEs insure default risks of loans they purchase and securitize,
providing subsidized credit to GSE mortgage borrowers.
We test if our main findings are robust to the sub-sample of house transactions with sale prices below local conforming loan limits. These house transactions are not subject to the concern about bunching below conforming loan limit as the transaction prices are already below the conforming loan limit.

Panel 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 both OLS and IV settings.

5 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.

5.1 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 exogenous 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 \( F^\text{pur} \) of the value of the house, or \( F^\text{pur} p^h h_t \); this can be thought as representing realtor fees and other costs of buying a house.

Households can borrow up to a fraction \( \phi (h_t) \) of the house’s value, that is, at mortgage rate \( r^h > r^b \). \( \phi (h_t) \) can depend on \( h_t \), so lower quality houses can have different LTV requirements, in a way disciplined by data; we describe in detail how we calibrate \( \phi (h_t) \) in Subsection 5.2 below, and Appendix D.2. Let \( a_t \) represent cash-on-hand; homeowners’ borrowing constraint is thus:

\[ a_t \geq -\phi (h_t) p^h h_t \tag{11} \]

The household faces a mortgage rate \( r^m > r^h \). 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:

$$V_t(h_t, a_t, \zeta_t) = \max \big\{ V^\text{renter}_t(h_t, a_t, \zeta_t), V^\text{purchase}_t(h_t, a_t, \zeta_t) \big\}$$

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

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

where $\mathbb{E}$ denotes the expectation. The constraints for renting are:

$$s.t. \quad c_t + \frac{a_{t+1}}{1 + r_t} = a_t + y_t + \left\{ \begin{array}{ll} p^h h_{t+1} (h_t > s_R) - p_r h_{t+1} \\ Selling old house \end{array} \right.$$  

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

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

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

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

subject to

\[
c_t + \frac{a_{t+1}}{1 + r_t} = a_t + y_t + p^h h_t 1(h_t > s_R) - \left( 1 + \delta^h + F^{\text{pur}} 1(h_{t+1} \neq h_t) \right) p^h h_{t+1}
\] (15)

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

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

### 5.2 Calibration

The model period is annual. Most of our choices for parameter calibrations are standard, and we discuss them in Appendix D.1. The core way in which we deviates from the standard lifecycle model calibration is in the \( \phi(h) \) function, which determines the relationship between house qualities and average LTV. We calibrate three different versions of \( \phi(h) \), to represent the loan-to-price ratios available to households in counties with high (top decile), medium (median decile), and low (bottom decile) average price dispersion. We plot these functions in Appendix Figure A9, and describe details of how we construct these functions in Appendix D.2. 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%.

Additional details on how we numerically solve the model are in Appendix D.3. 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. Appendix Figure A10 evaluates the fit of the model, comparing homeownership rates
and debt-to-assets in the model to data from the 2016 SCF. We are able to match the path of homeownership rates very well, and the path of debt-to-assets over the lifecycle fairly well.

5.3 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.\textsuperscript{20}

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.\textsuperscript{21}

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

\textsuperscript{20}Note that we showed in Subsection 3.2 that price dispersion is lower for houses that are newer. It is important also that house prices are held fixed: in practice, rebuilding houses would likely change the level of average house prices, and this would also affect homeownership rates. We disregard this effect in the calibration, though it may be important in practice.

\textsuperscript{21}Since incomes are fairly persistent in lifecycle models, initial incomes have persistent effects on wealth and income at later ages.
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.

5.4 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 6. 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:

$$HomeOwner_{ikt} = \alpha + \beta PriceDispersion_{kt} + X_{ikt}\Gamma + \mu_t + \epsilon_{ikt}$$ \hspace{1cm} (16)

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 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. 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
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.23

6 Discussion and Policy Implications

6.1 The 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.24 Policymakers aiming to improve homeownership rates for low-income households have considered interventions in credit markets as well as in housing markets. Our analysis highlights a link between these two markets: the amount of credit that mortgage lenders provide depends on the value uncertainty of the house used as collateral, even in a fully competitive mortgage market.

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

23See FRED.
24See, for example, the discussion in Herbert et al. (2005).
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.

6.2 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 transaction prices (Calem et al., 2015; Eriksen et al., 2019; Bogin and Shui, 2020; Conklin et al., 2020; Calem et al., 2021; Kruger and Maturana, 2021). Automated appraisals are likely to be less distorted, but as a result, under-appraisals will be more frequent, especially in areas with high house price dispersion. Automated appraisal thus have the potential to hurt low-income households who tend to live in areas with less predictable house prices.

6.3 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


\[26\text{Blattner and Nelson (2021) and Fuster et al. (2020) have made similar arguments that low-income households tend to have nosier hard information, and the development of FinTech is going to increase statistical discrimination in mortgage lending.}\]
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.

7 Conclusion

In this paper, we have shown that house value uncertainty affects mortgage credit provision in the US residential real estate market. Houses differ substantially in their degree of idiosyncratic price dispersion, which affects their value as collateral and thus the availability of mortgage credit. This effect is partially due to fair pricing of collateral 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*.


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


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.** Stylized Facts about Price Dispersion Estimates

(a) Zip-Code Dispersion: 2010 vs 2020  

(b) Dispersion and Income

*Note:* Panel (a) plots zip-code dispersion measures in 2020 against zip-code dispersion measures in 2010. Panel (b) shows the association between house price dispersion and zip-code household income prices. We divide all zip codes into five buckets based on local median household income and plot the average values in each bucket. The sample includes annual zip level observations from 2000 to 2020. *Source:* Corelogic Deeds and American Community Survey 2008-2012.
Figure 2. County Level House Price Dispersion and Credit Access

(a) Loan-to-Price

(b) Rate

(c) Total Rejection

(d) Rejection due to Collateral

Note: This figure shows the correlation between county level house price dispersion and various credit access outcomes. Panel a plots county average LTP. 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 3. Property Level Mortgage Menu by House Price Dispersion

Note: This figure shows mortgage price menu (rate-LTP pair) by zip-level house price dispersion. The y values are interest rate residuals from a regression of mortgage rates on borrower fico, fico-squared, DTI, DTI-squared conforming or jumbo indicator, and origination month fixed effects. The dots represent the average mortgage rate in each LTP bucket. The shaded area indicates 95% confidence interval. The sample includes loan level observations of conventional loans from 2000 to 2020. Source: Corelogic LLMA and Deeds.
Figure 4. 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 indicate 95% confidence intervals. The sample includes loan level observations of conventional loans from 2000 to 2020. Source: Corelogic LLMA and Deeds.
Figure 5. Price Dispersion and Appraisals

Note: Panel (a) of this figure shows a binned scatter plot, where the y-variable is $\text{ApprDev}_i$, 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 6.** House Price Dispersion, Income, and Homeownership

*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.
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.

### Property Level Sample

<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>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>
</tbody>
</table>

### Loan Level Sample

<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>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>
</tbody>
</table>

### Mortgage Application Sample

<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>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>Feature</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Building Age</td>
<td>0.04***</td>
<td>0.04***</td>
<td>0.05***</td>
<td>0.04***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.002)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recent Renovation</td>
<td>-0.01***</td>
<td>-0.01***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ever Renovate</td>
<td>-0.01***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Benchmark: Square-Footage**

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

### Panel B: Zip Code Market Condition

<table>
<thead>
<tr>
<th>Variable</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zip Code Price Dispersion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gini Index</td>
<td>0.01***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population Density</td>
<td>-0.01***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vacancy Share</td>
<td>0.03***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Year FE**

- ✓
- ✓
- ✓
- ✓

**R2**

- 0.33
- 0.34
- 0.35
- 0.27

**Observations**

- 39M
- 18M
- 18M
- 39M
- 29M
- 15M

---

46
This table presents property-level regression results. Columns 1-3 present OLS results. Columns 4-6 present IV results. In all columns, the outcome variable is the loan level loan-to-sale price ratio. The explanatory variable of interest in columns 1-3 is property-level house price dispersion, scaled by its standard deviation, and is the predicted price dispersion in columns 4-6. Controls include the transaction price of the property, mortgage type, mortgage term, and resale indicator. The sample includes property transaction level observations from 2000 to 2020. Standard errors are clustered at county level. ***, **, * represent 1%, 5%, and 10% significance, respectively.

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>2SLS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Price Dispersion</td>
<td>-0.45***</td>
<td>-0.22***</td>
</tr>
<tr>
<td></td>
<td>(0.044)</td>
<td>(0.037)</td>
</tr>
<tr>
<td>Log House Price</td>
<td>-4.30***</td>
<td>-4.59***</td>
</tr>
<tr>
<td></td>
<td>(0.144)</td>
<td>(0.184)</td>
</tr>
<tr>
<td>Loan Controls</td>
<td>✓ ✓ ✓ ✓</td>
<td>✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td>Transaction Date FE</td>
<td>✓ ✓ ✓ ✓</td>
<td>✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td>County-Year FE</td>
<td>✓ ✓ ✓ ✓</td>
<td>✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td>Lender-Year FE</td>
<td>✓ ✓ ✓</td>
<td>✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td>R2</td>
<td>0.34</td>
<td>0.36</td>
</tr>
<tr>
<td>Observations</td>
<td>28M</td>
<td>28M</td>
</tr>
<tr>
<td>Underidentification test statistic</td>
<td>166.62</td>
<td>165.51</td>
</tr>
<tr>
<td>Weak identification test statistic</td>
<td>77.99</td>
<td>93.27</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. In Panel A, the outcome variables are loan-level interest rate. In Panel B, the outcome variables are loan-to-sale price ratio. In both panels, columns 1-3 present OLS results, and columns 4-6 present 2SLS results. Column 1 (4) uses the full sample. Columns 2-3 (5-6) use securitized conventional loans (i.e., non-FHA loans that are securitized) and portfolio conventional loans (i.e., non-FHA loans that are held on lenders’ balance sheets), respectively. The explanatory variable of interest is zip-code house price dispersion, scaled by its standard deviation. Borrower and loan controls in the top tables of both panels include log house price, FICO score, FICO squared, LTP, LTP squared, DTI, DTI-squared, and loan type. Borrower and loan controls in the bottom tables of both panels 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>OLS</th>
<th></th>
<th>2SLS</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td></td>
<td>Full</td>
<td>Securitized</td>
<td>Portfolio</td>
<td>Full</td>
</tr>
<tr>
<td>Zip Price Dispersion</td>
<td>1.11***</td>
<td>1.38***</td>
<td>1.92***</td>
<td>2.18***</td>
</tr>
<tr>
<td></td>
<td>(0.120)</td>
<td>(0.098)</td>
<td>(0.397)</td>
<td>(0.357)</td>
</tr>
<tr>
<td>LTP</td>
<td>0.79***</td>
<td>0.56***</td>
<td>2.30***</td>
<td>0.80***</td>
</tr>
<tr>
<td></td>
<td>(0.096)</td>
<td>(0.034)</td>
<td>(0.208)</td>
<td>(0.113)</td>
</tr>
<tr>
<td>R²</td>
<td>0.85</td>
<td>0.87</td>
<td>0.82</td>
<td>0.08</td>
</tr>
<tr>
<td>Underidentification test statistic</td>
<td>92.54</td>
<td>91.88</td>
<td>76.70</td>
<td>0.00</td>
</tr>
<tr>
<td>Underidentification test p-value</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weak identification test statistic</td>
<td>13851.77</td>
<td>17250.99</td>
<td>7337.60</td>
<td></td>
</tr>
</tbody>
</table>

**Panel B: Loan-to-Price**

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th></th>
<th>2SLS</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td></td>
<td>Full</td>
<td>Securitized</td>
<td>Portfolio</td>
<td>Full</td>
</tr>
<tr>
<td>Zip Price Dispersion</td>
<td>-0.50***</td>
<td>-0.51***</td>
<td>-0.18***</td>
<td>-0.38***</td>
</tr>
<tr>
<td></td>
<td>(0.051)</td>
<td>(0.055)</td>
<td>(0.067)</td>
<td>(0.091)</td>
</tr>
<tr>
<td>Interest Rate</td>
<td>0.01***</td>
<td>0.03***</td>
<td>-0.01***</td>
<td>0.01***</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.001)</td>
<td>(0.001)</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>0.25</td>
<td>0.33</td>
</tr>
<tr>
<td>Observations</td>
<td>6M</td>
<td>2.8M</td>
<td>1.3M</td>
<td>5M</td>
</tr>
<tr>
<td>Underidentification test statistic</td>
<td>92.52</td>
<td>91.87</td>
<td>76.73</td>
<td></td>
</tr>
<tr>
<td>Underidentification test p-value</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weak identification test statistic</td>
<td>15461.51</td>
<td>18154.96</td>
<td>8167.13</td>
<td></td>
</tr>
</tbody>
</table>

48
Table 5: Mortgage Rejections and Zip House Price Dispersion

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

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>2SLS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td></td>
<td>Full</td>
<td>Securitized</td>
</tr>
<tr>
<td>Zip Price Dispersion</td>
<td>1.38***  (0.093)</td>
<td>1.37***  (0.102)</td>
</tr>
<tr>
<td>R2</td>
<td>15.9%</td>
<td>16.5%</td>
</tr>
<tr>
<td>Underidentification test statistic</td>
<td>93.16  &lt;0.005</td>
<td>87.19  &lt;0.005</td>
</tr>
<tr>
<td>Weak identification test statistic</td>
<td>1732.51  58.02</td>
<td>2973.48  58.02</td>
</tr>
<tr>
<td>Panel B: Rejection Due to Collateral</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zip Price Dispersion</td>
<td>0.50***  (0.035)</td>
<td>0.53***  (0.037)</td>
</tr>
<tr>
<td>Local Controls</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>County-Year FE</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Lender-Year FE</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Rejection due to Collateral Mean</td>
<td>1.9% 2.0% 2.2%</td>
<td>1.9% 2.0% 2.2%</td>
</tr>
<tr>
<td>R2</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>Observations</td>
<td>49M</td>
<td>35M</td>
</tr>
<tr>
<td>Underidentification test statistic</td>
<td>93.16  26.04</td>
<td>87.19  26.04</td>
</tr>
<tr>
<td>Weak identification test statistic</td>
<td>1732.51  58.02</td>
<td>2973.48  58.02</td>
</tr>
</tbody>
</table>

49
Table 6: Heterogeneous Effect by FICO

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

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>2SLS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1) Full</td>
<td>(2) Securitized</td>
</tr>
<tr>
<td>Zip Price Dispersion</td>
<td>0.06</td>
<td>0.02</td>
</tr>
<tr>
<td>(0.081)</td>
<td>(0.083)</td>
<td>(0.107)</td>
</tr>
<tr>
<td>Baseline: Excellent FICO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zip Price Dispersion × Very Good</td>
<td>-0.41***</td>
<td>-0.42***</td>
</tr>
<tr>
<td>(0.040)</td>
<td>(0.045)</td>
<td>(0.063)</td>
</tr>
<tr>
<td>Zip Price Dispersion × Good</td>
<td>-0.70***</td>
<td>-0.63***</td>
</tr>
<tr>
<td>(0.054)</td>
<td>(0.063)</td>
<td>(0.080)</td>
</tr>
<tr>
<td>Zip Price Dispersion × Fair</td>
<td>-0.86***</td>
<td>-0.51***</td>
</tr>
<tr>
<td>(0.069)</td>
<td>(0.082)</td>
<td>(0.107)</td>
</tr>
<tr>
<td>Zip Price Dispersion × Poor</td>
<td>-1.05***</td>
<td>-0.63***</td>
</tr>
<tr>
<td>(0.108)</td>
<td>(0.171)</td>
<td>(0.171)</td>
</tr>
<tr>
<td>Origination Month FE</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>County-Year FE</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Borrower/Loan Controls</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>R2</td>
<td>0.40</td>
<td>0.27</td>
</tr>
<tr>
<td>Observations</td>
<td>6M</td>
<td>28M</td>
</tr>
<tr>
<td>Underidentification test statistic</td>
<td>140.04</td>
<td>145.01</td>
</tr>
<tr>
<td>Underidentification test p-value</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Weak identification test statistic</td>
<td>91.78</td>
<td>59.46</td>
</tr>
</tbody>
</table>
## Table 7: Price Dispersion and Appraisals

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

<table>
<thead>
<tr>
<th></th>
<th>OLS (1)</th>
<th>OLS (2)</th>
<th>OLS (3)</th>
<th>OLS (4)</th>
<th>OLS (5)</th>
<th>OLS (6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zip Price Dispersion</td>
<td>0.02***</td>
<td>0.00</td>
<td>0.53***</td>
<td>0.01**</td>
<td>-0.00***</td>
<td>0.53***</td>
</tr>
<tr>
<td>(0.002)</td>
<td>(0.000)</td>
<td>(0.033)</td>
<td>(0.005)</td>
<td>(0.001)</td>
<td>(0.068)</td>
<td></td>
</tr>
<tr>
<td>Origination Month 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>Property &amp; Loan Controls</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>R²</td>
<td>0.02</td>
<td>0.03</td>
<td>0.16</td>
<td>0.00</td>
<td>0.00</td>
<td>0.02</td>
</tr>
<tr>
<td>Observations</td>
<td>5M</td>
<td>5M</td>
<td>0.2M</td>
<td>5M</td>
<td>5M</td>
<td>0.2M</td>
</tr>
<tr>
<td>Underidentification test statistic</td>
<td>92.24</td>
<td>92.26</td>
<td>60.65</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Underidentification test p-value</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weak identification test statistic</td>
<td>15971.93</td>
<td>15974.44</td>
<td>1350.66</td>
<td></td>
<td></td>
<td></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>Panel A: Default</th>
<th>OLS</th>
<th>2SLS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Full</td>
<td>GSE&amp;FHA</td>
</tr>
<tr>
<td>Zip Price Dispersion</td>
<td>0.16***</td>
<td>0.10***</td>
</tr>
<tr>
<td></td>
<td>(0.056)</td>
<td>(0.038)</td>
</tr>
<tr>
<td>Interest Rate</td>
<td>2.25***</td>
<td>2.61***</td>
</tr>
<tr>
<td></td>
<td>(0.123)</td>
<td>(0.185)</td>
</tr>
<tr>
<td></td>
<td>(0.238)</td>
<td>(0.346)</td>
</tr>
<tr>
<td>DTI</td>
<td>0.06***</td>
<td>0.03***</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.006)</td>
</tr>
</tbody>
</table>

- Origination Month FE ✓ ✓ ✓ ✓ ✓ ✓
- County-Year FE ✓ ✓ ✓ ✓ ✓ ✓
- Property & Loan Controls ✓ ✓ ✓ ✓ ✓ ✓
- R2 0.15 0.13 0.19 0.09 0.06 0.09
- Observations 4.3M 2.1M 0.9M 4.3M 2.1M 0.9M
- Underidentification test statistic 91.51 90.77 74.96
- Underidentification test p-value 0.00 0.00 0.00
- Weak identification test statistic 15681.45 16059.38 8185.28

52
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></th>
<th>OLS</th>
<th>2SLS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Price Dispersion</td>
<td>-0.58*** (0.119)</td>
<td>-0.50*** (0.087)</td>
</tr>
<tr>
<td>Interest Rate</td>
<td>1.24*** (0.071)</td>
<td>1.06*** (0.060)</td>
</tr>
<tr>
<td>Loan Controls</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Origination Month FE</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>County-Year FE</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Lender-County-Year FE</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Lender-Zip-Year FE</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>R2</td>
<td>0.48 0.2M</td>
<td>0.59 0.2M</td>
</tr>
<tr>
<td>Observations</td>
<td>0.065 0.2M</td>
<td>0.048 0.2M</td>
</tr>
<tr>
<td>Underidentification test statistic</td>
<td>12.27</td>
<td>10.75</td>
</tr>
<tr>
<td>Weak identification test statistic</td>
<td>61.80</td>
<td>69.64</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>2SLS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Price Dispersion</td>
<td>-0.36*** (0.065)</td>
<td>-0.28*** (0.048)</td>
</tr>
<tr>
<td>Interest Rate</td>
<td>1.20*** (0.066)</td>
<td>0.97*** (0.040)</td>
</tr>
<tr>
<td>Loan Controls</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Origination Month FE</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>County-Year FE</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Lender-Year FE</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>R2</td>
<td>0.42 0.2M</td>
<td>0.52 0.2M</td>
</tr>
<tr>
<td>Observations</td>
<td>0.065 0.2M</td>
<td>0.048 0.2M</td>
</tr>
<tr>
<td>Underidentification test statistic</td>
<td>13.96</td>
<td>13.31</td>
</tr>
<tr>
<td>Weak identification test statistic</td>
<td>32.70</td>
<td>33.50</td>
</tr>
</tbody>
</table>
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>$\sigma_\varepsilon$</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>$F_{out}$</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>2.1</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>

54
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>Home Ownership</th>
<th>(1) Full Sample</th>
<th>(2) Low Income</th>
<th>(3) High Income</th>
<th>(4) Full Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>County Price Dispersion</td>
<td>-1.29** (0.551)</td>
<td>-1.46** (0.662)</td>
<td>-1.07** (0.531)</td>
<td></td>
</tr>
<tr>
<td>County Price Dispersion × Lowest Income Decile</td>
<td>-1.04*** (0.304)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Household Controls</td>
<td>✓ ✓ ✓ ✓</td>
<td>✓ ✓ ✓ ✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year 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>0.21 0.12 0.19 0.16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>17M 17M 17M 17M</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix

A Model

In this appendix, we construct a simple 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

\[ P = 1 + \epsilon_P \]  

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.\(^{27}\) 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:

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

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 \( \bar{L} \),

\(^{27}\)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.
a fraction:
\[ F_\theta \left( 1 - \frac{L}{P} \right) \]  
(19)
of transactions will fail, due to borrowers having insufficient funds for down payments.

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.\(^{28}\) 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:

\[ Loss = E \left[ L - \max[L, 1 - c + \epsilon_F] \right] = E \left[ \max[0, L - (1 - c + \epsilon_F)] \right] \]  
(21)

\(^{28}\)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.
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:

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

(22)

The LHS of (22) is lenders’ expected profit: the mortgage spread $r - \rho$ multiplied by the probability the borrower repays, $(1 - \delta)$. 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 $\bar{L}_{\text{fair}}$. 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. 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.\(^{29}\)

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.

---

\(^{29}\)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.
and the appraisal price, so we must have:

\[ L \leq \bar{L}_{appr} \equiv \phi \min [P, A] \quad (24) \]

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 \( \frac{A}{F} \) 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 5 shows empirically that a larger share of houses under-appraise in counties with higher value uncertainty. Moreover, Figure 2 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:

\[
\bar{L}_{fair} = \left\{ L : L (1 - \delta) (r - \rho) = \delta E \left[ \max \left[ 0, L - (1 + c + \epsilon_F) \right] \right] \right\} \quad (25)
\]
There is an appraisal constraint, which satisfies:

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

(26)

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

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

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

\[ P = 1 + \epsilon_P \]

The application succeeds if:

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

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. While we do not formally calibrate this model to the data, we choose values of parameters to approximately match the data. We vary price dispersion \( \sigma_{\epsilon} \) from 10% to 20% of house prices, and default rates \( \delta \) from 2.5% to 12.5%. We set the interest rate spread on mortgages to \( r - \rho = 0.005 \).\(^{30}\) We set \( c = 0.2 \), reflecting foreclosure discounts of approximately 20%, which is similar to values from the empirical literature on foreclosure discounts (Campbell et al., 2011). We set \( \phi \), the loan-to-value ratio, to 85%.

Figure A1 shows how \( \bar{L}_{appr} \) and \( \bar{L}_{fair} \) 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 4.** When \( \sigma_{\epsilon} \) increases, holding fixed loan interest rates, the average loan-
price ratio, \( \frac{L}{P} \), decreases.

Prediction 4 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 \( L_{\text{appr}} \) 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.

Note also that the magnitudes of the effects in Figure A1 under the fair-pricing channel are fairly large: even with \( \delta = 2.5\% \), a 10% increase in \( \sigma_{\text{epsilon}} \) can shift maximum loan size by 2%-3% of house prices. We emphasize this is not a full calibration: it is a numerical illustration of the fact that the effects of price dispersion on maximum loan size can be quantitatively large.

**PREDICTION 5. When \( \sigma_\epsilon \) is higher, a larger fraction of mortgage applications fail.**

Prediction 5 is shown in the right panel of figure A2. It follows immediately from prediction 4: 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 6. The effect of \( \sigma_\epsilon \) on LTVs is larger when the default rate \( \delta \) is higher.**

Prediction 6 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.\(^{31}\) Thus, it is possible for the appraisal channel to also affect borrowers of different credit scores heterogeneously.

\(^{31}\)See the Fannie Mae and Freddie Mac pricing matrices, which specify interest rate adjustments as a function of LTV and credit score.
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:

$$r_{fair} = \left\{ r : L(1 - \delta)(r - \rho) = \delta E\left[\max[0, L - (1 + c + \epsilon_F)]\right]\right\} \quad (27)$$

In (27), the RHS is increasing in $\sigma_\epsilon$ and the LHS is increasing in $r$, so $r_{fair}$ 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 4.2.3, where we show that the entire menu of LTV-interest rate pairs shifts monotonically with $\sigma_\epsilon$.

B Data Cleaning and Measure Estimation

B.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.

### B.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:

\[
f_c(x_i, t) = h_{c,latlong}^f(t, lat_i, long_i) + h_{c,sqft}^f(t, sqft_i) + h_{c,yrbuilt}^f(t, yrbuilt_i) + h_{c,bedrooms}^f(t, bedrooms_i) + h_{c,bathrooms}^f(t, bathrooms_i)
\]

The functions \(h_{c,latlong}^f\), \(h_{c,sqft}^f\), and \(h_{c,yrbuilt}^f\) are tensor products of 5-dimensional cubic splines in their constituent components: hence, for example, the \(h_{c,latlong}^f(t, lat_i, long_i)\) is a three-dimensional spline tensor product, with a total of \(5^3 = 125\) degrees of freedom. To combat overfitting, the spline terms also includes a shrinkage penalty term on the second derivative of the spline functions, with the smoothing penalty determined through generalized cross-validation. The functions \(h_{c,bedrooms}^f\) and \(h_{c,bathrooms}^f\) 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 \(f_c(x_i, t)\):

\[
g_c(x_i, t) = h_{c,latlong}^g(t, lat_i, long_i) + h_{c,sqft}^g(t, sqft_i) + h_{c,yrbuilt}^g(t, yrbuilt_i) + h_{c,bedrooms}^g(t, bedrooms_i) + h_{c,bathrooms}^g(t, bathrooms_i)
\]
C Repeat-Sale Estimation and Results

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

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

(28)

where $i$ indexes properties, $k$ indexes counties, and $t$ indexes months. Equation (28) is a repeat-sales model for house prices: log prices $p_{it}$ are determined by county-month fixed effects $\eta_{kt}$, time-invariant house fixed effects $\mu_i$, and a mean-zero error term $\epsilon_{it}$. Specification (28) thus models log house prices as following parallel trends, plus error terms: if house A sells for twice the price of house B in June of 2011, house A should sell for twice as much as house B in June of 2017, and any deviation from this is attributed to the error term $\epsilon_{it}$.

There are two additional concerns with measuring idiosyncratic dispersion using a repeat-sales specification. First, the number of data points used to estimate each house fixed effect is very low; thus, the estimated residuals $\hat{\epsilon}_{it}^2$ will tend to be larger for houses which are sold more times, because the house fixed effect $\gamma_i$ is estimated more precisely. Second, (28) implicitly assumes that idiosyncratic price dispersion does not depend on the house holding period; a concern is that there is a idiosyncratic price dispersion behaves partially like a random walk, so the error terms may be systematically larger for houses that are sold less frequently. To alleviate the concern that our estimates of $\hat{\epsilon}_{it}^2$ are mechanically driven by sale frequency and time-between-sales, we purge $\hat{\epsilon}_{it}^2$ of any variation which can be explained by $tbs_i$ and $sales_i$. First, we filter to houses sold at most four times over the whole sample period, with estimated values of $\hat{\epsilon}_{it}^2$ below 0.25. We then run the following regression, separately for each county:

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

(29)

Where, $h_k(sales_i, tbs_i)$ interacts a vector of $sales_i$ dummies with a fifth-order polynomial in $tbs_i$. The residual $\hat{\epsilon}_{it}$ from this regression can be interpreted as the component of the house's price variance which is not explainable by $sales_i$ and $tbs_i$. We then add back the mean of $\hat{\epsilon}_{it}^2$

\[\text{Note that Giacoletti (2021) and Sagi (2021) show that a large component of idiosyncratic dispersion does not scale with holding period, for both residential and commercial real estate transactions.}\]
within county $k$:

$$
\hat{\epsilon}_{TBSadj,it}^2 = \hat{\zeta}_{it} + E_k [\epsilon_{it}^2] \tag{30}
$$

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

In comparison to the hedonic model, the repeat-sales model in (28) is able to capture observable and unobservable features of houses that have time-invariant effects on house prices. Moreover, house fixed effects allow us to capture time-invariant house quality components in a fully nonparametric way, alleviating concerns that the specific functional form we use in (1) is driving our results. A weakness of specification (28) are that it is unable to capture any features of houses which have time-varying effects on house prices.

Figure A4 shows a binscatter of $\hat{\sigma}_{RS,it}^2$ against our baseline estimates $\hat{\sigma}_{it}^2$. There is a very strong positive relationship. The repeat-sales and hedonic methodologies for measuring house value uncertainty are econometrically quite different; the fact that they produce very correlated results at the house level suggests that both measurement strategies are picking up fundamental value uncertainty among properties, rather than simply reflecting misspecification in the model we use for house prices.

Next, we repeat our regression specifications utilizing $\hat{\sigma}_{RS,it}^2$ as our measure of house price dispersion. Table A1 shows the results; all of our baseline results continue to hold, using $\hat{\sigma}_{RS,it}^2$ as our measure of house price dispersion.

D Additional Details for Section 5

D.1 Parameter choices for calibration

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 $15,000 annually, which is approximately the average social security payout in the US.$^{33}$ We use standard values of $\beta, \sigma, \alpha$ in the

$^{33}$See Table A in the Social Security Program Fact Sheet.
literature. Housing transaction costs $F^{pur}$ are set to 0.05, which is the typical fee charged by real estate brokers in the US. This value is also used in Berger et al. (2018) and Wong (2019), among other papers. We set the depreciation rate to 0.01, approximately matching the depreciation rate in BEA data. We set house prices $p^h$ to:

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

that is, $p^h$ is rent adjusted for discount rates $\beta$ 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 $\zeta$, the idiosyncratic income shock, for 25-year-olds such that probabilities are log-linear in the level of $\zeta$, that is:

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

where $k_\zeta$ controls whether probability weights are higher for high or low values of $\zeta$.\textsuperscript{34} We calibrate the persistence of idiosyncratic income shocks $\rho_\zeta$ to 0.91, and the standard deviation of shocks $\sigma_\varepsilon$ to 0.21, following Floden and Lindé (2001).

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

\textsuperscript{34}Without adjusting the initial distribution of $\zeta$, we found that homeownership rates rose too quickly in the model relative to the data
D.2 Calibrating the $\phi(h)$ Functions

We calibrate $\phi(h)$ based on the average price dispersion for each level of house prices and the relationship between price dispersion and LTV that we empirically identified. Our goal for calibrating $\phi(h)$ is to match the relationship between house prices and $\sigma$ 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:

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

where we used $c$ to index counties, and $c \in C_{\text{mid}}$ means that county $c$ is in the middle quantile of counties by price dispersion. We define $\sigma_{\text{high}}(p)$ and $\sigma_{\text{low}}(p)$ analogously to (31), for the high- and low-dispersion set of counties. The three curves $\sigma(p)$ curves are shown in the left panel of Figure A9. 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:

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

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

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

Where $\beta_{\text{LTV},\sigma}$ is the coefficient from regressing LTV on price dispersion, from column 1 of Table 6. In words, (32) 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 $p_{\text{min}}$, plus an adjustment which is the difference between price dispersion at $p$, and price dispersion at $p_{\text{min}}$, multiplied by $\beta_{\text{LTV}}$, the effect of price dispersion on LTVs identified in our reduced-form results. Note that we adjust using $\beta_{\text{LTV}}$, 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 $\beta_{\text{LTV}}$.

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

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

That is, analogous to (32), $\phi_{\text{high}}(p)$ 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 $\beta_{\text{LTV}}$.

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

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

Figure A9 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 $\beta_{\text{LTV}}$ 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.3 Numerically Solving the Model

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

$$w_t = a_t + 1(h_t > s_R) p^h h_t$$
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:

\[
 w_{t+1} = \\
(1 + r_t) \left( w_t + y_t - c_t - 1 (h_{t+1} > s_R) \left( 1 + \delta^h + F^{par} 1 (h_{t+1} \neq h_t) \right) p^h h_t - 1 (h_t < s_R) p^r \right) + \\
1 (h_{t+1} > s_R) p^h h_t \tag{35}
\]

Using (35), 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 Appendix Tables and Figures

Figure A1. Behavior of $\bar{L}_{appr}$ and $\bar{L}_{fair}$

Note: In the above figure, the left panel shows the behavior of the average value of $\bar{L}_{appr}$ for successful loans (which does not depend on $\delta$), 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

Note: In the above figure, the left panel shows the mean loan-to-price ratio. The right panel shows the probability of loans failing. Colored lines represent different values of $\delta$. Throughout, we set $\phi = 0.85$, $c = 0.2$, $r - \rho = 0.005$. 71
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 $\sigma = 0.15$, $\phi = 0.85$, $c = 0.2$, $r - \rho = 0.005$. 
Figure A4. Repeat-Sales Estimates and Hedonic Estimates

Note: This figure compares the repeat-sale estimates and the hedonic estimates by making the binned scatterplot. The x-axis is the repeat-sale estimates, and the y-axis is the hedonic estimates used in the main analysis. The sample includes property-level observations from 2000 to 2020. Source: Corelogic Deeds.
Figure A5. County Level House Price Dispersion and LTP

(a) GSE

(b) FHA

(c) Jumbo

Note: This figure shows the correlation between county level house price dispersion and residualized county average LTP. Panels a-c plot GSE loans, FHA loans, and jumbo loans, respectively. The sample includes annual county observations from 2000 to 2020. Source: County house price dispersion is estimated using Corelogic Deeds records. Mortgage data are from Corelogic LLMA.
**Figure A6.** 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 A7. 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.
Figure A8. 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.
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 A10. Model Fit**

*Note:* The left plot shows homeownership rates in the model and in the data. The right plot shows debt-to-assets in the model and in the data. The data is from the 2016 SCF. For both SCF data series, we smooth the input series by projecting values on a fourth-degree polynomial in age and taking the predicted values.
Table A1: Property-Level House Price Dispersion and LTP - Repeat Sale

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

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>2SLS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Price Dispersion</td>
<td>-0.77***</td>
<td>-0.36***</td>
</tr>
<tr>
<td></td>
<td>(0.090)</td>
<td>(0.038)</td>
</tr>
<tr>
<td>Interest Rate</td>
<td>0.83***</td>
<td>0.91***</td>
</tr>
<tr>
<td></td>
<td>(0.073)</td>
<td>(0.058)</td>
</tr>
<tr>
<td>Log House Price</td>
<td>-3.29***</td>
<td>-3.41***</td>
</tr>
<tr>
<td></td>
<td>(0.112)</td>
<td>(0.119)</td>
</tr>
<tr>
<td>Loan Controls</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Transaction Date FE</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>County-Year FE</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Lender-Year FE</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>R^2</td>
<td>0.44</td>
<td>0.47</td>
</tr>
<tr>
<td>Observations</td>
<td>3M</td>
<td>3M</td>
</tr>
<tr>
<td>Underidentification test statistic</td>
<td>77.81</td>
<td>84.67</td>
</tr>
<tr>
<td>Underidentification test p-value</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Weak identification test statistic</td>
<td>47.31</td>
<td>73.13</td>
</tr>
</tbody>
</table>

80