Paying Too Much? Price Dispersion in the US Mortgage Market*

Neil Bhutta   Andreas Fuster   Aurel Hizmo†

Preliminary and Incomplete

September 17, 2018

Abstract

We study dispersion in the prices paid for mortgages in the US using new data from a mortgage industry pricing platform. These data allow us to observe the key determinants of mortgage interest rates, including discount points, the time of rate lock, location, and all underwriting information. Conditional on these variables, we find that the gap between the 10th and 90th percentile interest rate is over 50 basis points, which is equivalent to about $7,500 in upfront costs (points) for the average loan. Much of this dispersion occurs within lender, suggesting an important role for price discrimination and negotiation in pricing outcomes. We also find that dispersion differs across borrower types. Most notably, it widens significantly for borrowers with lower credit scores (FICO). We then use data at the lender-market level on lenders' "best offer" for each borrower type, and compare the transacted interest rates to these offer rates. We find that on average borrowers end up with rates that are significantly above what the median lender could offer to identical borrowers in the same market on the same day. This spread is especially wide for lower FICO borrowers, who thus pay more than high FICO borrowers not just because of credit risk, but also because of less effective search and negotiation. However, this spread over the median offer rate compresses when Treasury rates rise, consistent with an increase in borrowing costs encouraging borrowers to search and negotiate more effectively.

*We thank Serafin Grundl as well as seminar participants at the Federal Reserve Board and the Federal Reserve System Microeconomics meeting for helpful comments. The views expressed are those of the authors and do not necessarily reflect those of the Federal Reserve Board, the Federal Reserve System, or the Swiss National Bank.

†Bhutta and Hizmo are at the Federal Reserve Board; Fuster is at the Swiss National Bank. Emails: neil.bhutta@frb.gov; andreas.fuster@snb.ch; aurel.hizmo@frb.gov.

1 Introduction

According to the National Survey of Mortgage Originations, half of the borrowers taking out a mortgage in the US in 2016 only seriously considered one lender, and only three percent of the borrowers considered more than three lenders.[1] Ninety-six percent of the respondents reported that they were satisfied that they received the lowest interest rate for which they could qualify. Taking these facts at face value, one might be led to conclude that either there is not much price dispersion in the mortgage market, or that borrowers are very efficient at searching and finding the most competitive lenders. This might seem a reasonable conclusion especially when considering that the mortgage market appears highly competitive: the majority of mortgages in the US are very standardized and guaranteed by the government (through the GSEs and FHA/VA), and in our data there are over one hundred different lenders offering mortgages in a local market in any given day. However, in contrast to borrowers' perceptions, our data reveal a striking amount of variation in the prices consumers pay for mortgages, especially among certain groups such as borrowers with lower FICO credit scores.

Our data comes from an online platform used by lenders to price mortgages, initiate rate locks, manage pipeline risk, and sell mortgages to investors. The platform provides data on both mortgages offered in each market, and data on mortgages chosen by consumers. First, using data on mortgage interest rate locks, we document dispersion in the rates locked by consumers. We find a large amount of interest rate dispersion: the difference between the 90th and 10th percentile interest rate that identical borrowers pay in the same market, on the same day, and paying the same points, is over 50 basis points. Moreover, a substantial amount of dispersion exists even within lender. Thus, getting a low rate is not simply about "going to the right lender." Instead, it appears that in order to get a low rate, borrowers must be knowledgeable and able to negotiate no matter which lender they end up at.

Next, we draw on the real-time distribution of the *best* interest rates lenders could offer to borrowers with particular characteristics (LTV, DTI, FICO, loan amount, points, etc.) in a given market. We compute the difference between the rate locked by consumers and the median of these best offer rates available in the market for an identical mortgage on a given day. This locked-offered rate gap is positive on average, meaning that borrowers tend to get mortgage rates that are higher than what the median lender could offer for an identical mortgage.

[1] The National Survey of Mortgage Originations is conducted jointly by the Federal Housing Finance Agency (FHFA) and the Consumer Financial Protection Bureau (CFPB).

More importantly, the locked-offered rate gap varies substantially across borrower types. For example, typical borrowers with large ("jumbo") loans pay rates that are 17 basis points below the median of their offer distribution. In contrast, a typical FHA borrower pays a 30 basis points higher interest rate than what the median lender could offer for identical mortgages. Similarly, low-FICO borrowers exhibit a much larger lock-offer gap than high-FICO borrowers. These results are robust even after adding many controls and lender fixed effects in a regression.[2] They imply that low-FICO borrowers pay more for mortgages not only because they present more credit risk, but also because they search and negotiate less effectively.

Price dispersion in *locked* rates is also much higher for the same types of borrowers that appear to overpay relative to what the median lender could offer them, despite the fact that there are little to no differences in the dispersion of *offered* rates between borrower types. Again, this holds even after controlling for lender-location-time fixed effects. Both the higher average locked-offered rate gap and the higher dispersion thus likely reflect differences in borrowers' search process and their ability and willingness to negotiate with lenders.

Lastly, we explore how changes in overall market interest rates affect how borrowers fare relative to what is offered in the market. First, we show that when the level of market interest rates (as measured by Treasury yields) is higher, borrowers are more likely to lock rates from the cheaper end of the offer distribution. Similarly, when Treasury yields increase, the price dispersion of chosen interest rates falls. These results are stronger for low-FICO borrowers, which generally tend to overpay relative to what the median lender offers for identical mortgages. This may partly reflect affordability constraints becoming more binding as rates rise; however, we show that for low-FICO borrowers, even those that appear unconstrained in that dimension exhibit the same relationship. Thus, we conclude that behavioral factors, such as feeling less of a need to shop or negotiate when rates are already low, are likely to be a driver behind these patterns.

Existing work on price dispersion in the US mortgage market is rather sparse relative to the importance of the market. Woodward and Hall (2012) use data on 1,500 FHA loans from 2001 and document that many borrowers seem to overpay (in terms of upfront costs), consistent with suboptimal shopping. Gurun et al. (2016) show substantial dispersion in the reset rates of privately-securitized adjustable-rate mortgages (which were common during the housing boom of the mid-2000s, but much less since) and find that these rates correlate positively with lenders' advertising

[2]One interpretation of including lender fixed effects is that they remove any differences in quality across lenders. It is unlikely therefore, that certain borrower groups are overpaying because they are choosing lenders of higher quality.

expenditures. Perhaps closest to our work is the paper by Alexandrov and Koulayev (2017), who document substantial dispersion in offers based on rate sheets, but do not have data on actual rate locks. They argue that a major factor behind price dispersion in this market is many borrowers' belief that there is no such dispersion, and consider the implications within a structural search model. Our paper advances the literature through its more precise measurement of dispersion in a large transaction dataset containing all relevant variables, and especially through the ability to compare transactions with concurrent market offers. This further enables us to study how well different types of borrowers fare relative to what would be available to them.[3] Finally, to our knowledge this is the first paper documenting how dispersion in contracted rates, and the locked-offered gap, changes with market rate over time.

Other related work comes from different countries or other household financial markets. Allen et al. (2014) study the Canadian market, where there is no dispersion in posted rates, but large dispersion in contracted rates, which they argue arises due to differences bargaining leverage across consumers. Damen and Buyst (2017) provide evidence that mortgage borrowers in Belgium who shop more achieve substantial savings. Turning to other markets, Stango and Zinman (2016) and Argyle et al. (2017) show large dispersion in rates for credit cards and auto loans, respectively, again suggesting limited shopping or negotiation.

The rest of the paper is organized as follows. Section 2 describes our data sources. Section 3 documents price dispersion in the rate lock data. Section 4 explores how locked rates on average compare to the offer distribution, and how this varies across borrowers with different characteristics. Section 5 studies how these patterns evolve over time as market rates change. Section 6 discusses potential explanations of our results and potential policy implications. Section 7 concludes.

2 Data

The data comes from an online platform called Optimal Blue that connects over 600 mortgage lenders with more than 200 whole loan investors. Through the platform, mortgage originators can gather information on mortgage pricing, initiate rate locks, manage pipeline risk, and sell mortgages to investors. Over forty thousand unique users access the system each month to search loan programs and lock in consumer mortgages. There is a variety of lenders using the platform

[3]While we focus on how dispersion and the locked-offered gap vary with borrower financial characteristics such as the FICO score, other work has instead looked at differences in contracted mortgage rates by race or ethnicity (e.g. Bayer et al., 2018; Bhutta and Hizmo, 2018).

such as community banks, mortgage banks, credit unions etc. Many institutions on this platform act as correspondent lenders, meaning that they originate loans intended to be on-sold to other financial institutions such as a large bank like JP Morgan or Wells Fargo (referred to as "investors" in the market). More than $500bn of mortgages were processed through this system in 2017, thus accounting for about 25% of the loan originations nationally.

For this project we use two components of the data generated by the platform: a) data on mortgage products and mortgage prices actually accepted by consumers, and b) data on mortgage products available and mortgage prices offered by lenders in each market.

2.1 Mortgage Rate Lock Data

The first source of data is the universe of "rate lock" agreements for the mortgages processed through the Optimal Blue platform. A mortgage rate lock is a guarantee that the borrower will be issued a mortgage with a specific combination of interest rate and points if the mortgage closes by a specific date. Borrowers typically lock their mortgage rates as a protection against rate increases between the time of the lock and the time when the mortgage closes. A lock can occur at the same time a borrower submits a loan application with a lender, but can also happen at a later time. Not all rate locks ultimately lead to originated mortgages, since the loan application can still be rejected afterwards (e.g. because the appraisal of the home comes in lower than expected) or the borrower could renege.

We have access to all the mortgage locks generated by the platform since late 2013. Since the market coverage increases over the course of 2013-2014, we start using the data from January 2015. The data has a wide geographical coverage of about 280 metropolitan areas as well as rural areas. All of the standard loan characteristics used for underwriting are included: loan-to-value (LTV) ratio, debt-to-income (DTI) ratio, FICO score, loan amount, loan program, purpose (purchase or refinancing), asset documentation, income documentation, employment status, occupancy status, house type, zip code location etc. There are a number of unique features of the data relative to servicing data that is typically used in mortgage research. First, it includes not only the contracted mortgage rate, but also the discount points or credits associated with that rate (meaning additional upfront payments made or received by the borrower). Second, we see the exact time-stamp of when the lock occurred, while in most other datasets, only the day or month of origination is recorded, which can differ from the pricing-relevant lock date by several months. Finally, we have unique identifiers for the lender, the branch and the loan officer that processes each mortgage.

We restrict the sample in various ways to ensure that we study a relatively uniform set of loans that is representative of the type of mortgages originated in recent years. For instance, we only keep 30-year fixed-rate mortgages on single-unit properties, with full documentation of assets and income. We also drop small loans, and those with implausible values for LTV, DTI, or points/credits. Finally, we restrict the sample to purchase mortgages and regular rate/term refinances, meaning that we drop cashout or streamline refinances (which are a relatively small part of the sample but are priced somewhat differently). This leaves us with just over 2.5 million observations.

Table 1 presents some summary statistics from the lock data sample that we use for the analysis in this paper, separating between the four loan programs in the data, since they differ substantially in terms of borrower and loan characteristics. The four programs are: conforming (with loan amounts below the national conforming loan limit, so they are typically securitized through Fannie Mae or Freddie Mac), super-conforming (with loan amounts above the national conforming limit but below the local limit, so that Fannie Mae or Freddie Mac can still securitize the loan, but at slightly worse prices), jumbo (loan amount above the local conforming limit, meaning the loan cannot be securitized through the government-backed entities), and FHA loans (which require additional mortgage insurance). The table shows that FHA loans are most likely to go to first-time homebuyers with low FICO scores and high LTV and DTI.

2.2 Mortgage Offers Data

Second, we collect data on the menu of mortgage products available and mortgage rates that lenders offer through the platform's pricing engine. Optimal Blue's Pricing Insight allows users to retrieve the real-time distribution of offers for a loan with certain characteristics in a given local market (where an offer consists of a combination of a note rate and upfront fees and points that the borrower pays or receives with this rate). The data is used primarily by mortgage banks to compare their pricing against that of peers. We observe each institution's set of best offers (in terms of lowest required borrower upfront payment for a given interest rate) for a given combination of day, location, and loan characteristics.

A key advantage of these data is that the offers from the Optimal Blue platform are "customer facing," meaning they are the interest rates and fees that would actually be paid by a borrower. The rates and fees data come from lenders who use the Optimal Blue Pricing Engine for their own originations. The rates and fees for these mortgage lenders are determined primarily by: a) the rate sheets of the "investor" who ultimately holds the mortgage, which could be the originator, or any other secondary market investor. These rate sheets are updated at least daily in the platform by investors directly; and b) the markups and fees that the originator charges. The resulting rates and fees are the best offer each lender could make to a consumer who would request a mortgage from them.

We conduct daily searches in one local market (Los Angeles), twice-weekly searches in four markets, and weekly searches for 15 additional markets.[4] We collect offer rate distributions for 100 different loan types, differing across the following dimensions: FICO score, loan-to-value ratio, loan type (conforming, FHA, jumbo), loan purpose (purchase or cash-out refinance), occupancy (owner-occupied or investor), rate type (30-year fixed or 5/1 adjustable), and loan amount. The mortgages require full income, asset and employment documentation, and are used to finance single unit homes.

Two limitations of the offers data are: (i) we are not able to track institutions over time or match them directly to the lenders in the lock data, since there is no fixed lender identifier; (ii) the time series so far is relatively short: we started systematically tracking offers in April 2016.

In the main analysis, we primarily use the offered rates as a benchmark for the rates that borrowers lock. However, in the online appendix we present a separate analysis documenting price dispersion in offers only, which is also substantial and of independent interest.

3 Dispersion in Locked Mortgage Rates

In this section we document the magnitude of price dispersion in mortgage rates that borrowers lock. Dispersion in mortgage rates can arise for multiple reasons, some of which are differences in borrower characteristics, mortgage characteristics and lender characteristics. We are interested in investigating whether identical borrowers who choose the same mortgage product, in the same market, at the same time pay different prices. To investigate this, we regress locked mortgage rates on borrower and loan characteristics, as well as time effects, and then add an increasingly fine set of fixed effects. Our outcome of interest is the remaining dispersion in the residual, which we measure in terms of standard deviations, as well as the gap between 75th-25th or 90th-10th percentiles.

Table 2 shows the results from various specifications, estimated on the same set of 1.94 million

[4]The markets with twice-weekly searches are New York City, Chicago, Denver, and Miami. The markets with weekly searches are Atlanta, Boston, Charlotte, Cleveland, Dallas, Detroit, Las Vegas, Minneapolis, Phoenix, Portland, San Diego, San Francisco, Seattle, Tampa, and Washington DC.

locked loans over the 2015-2018 period.[5] Across all columns, we control for a basic set of variables, which consist of fully interacted bins of values for FICO, LTV, DTI, loan amount, as well as loan program (conforming, super-conforming, jumbo, FHA). The resulting "grid" takes 7,680 unique values. To allow for variation within the grids, we furthermore linearly control of each of the four continuous variables. In addition, we add a fixed effect for whether a locked loan is a refinance, and for the length of the lock period.[6]

Column (1) is our baseline specification, where we only add lock month-by-MSA fixed effects. This is supposed to mimic the regressions one could typically run with a mortgage servicing dataset.[7] We see that the controls explain a sizable share of the raw variation in interest rates—the adjusted R-squared is 0.68—but that substantial dispersion remains: the standard deviation in residuals is 0.27, and the borrower at the 90th percentile of the residual distribution pays 60 basis points (bp) more than the borrower at the 10th percentile.

Columns (2) and (3) add bins for the points paid or received by the borrower to the grid, as well as controlling for the exact day of the lock (rather than just the month). These (usually unobserved) variables indeed explain some of the rate differences across borrowers, but substantial dispersion remains—e.g. the 90th-10th percentile difference is still 54bp. Based on the regression coefficient on discount points (not shown in the table), we can translate interest rates to upfront points. This coefficient implies that 1 discount point changes the interest rate by about 16bp. Therefore, 54bp in rate terms is approximately equivalent to 3 upfront discount points or 3% of the mortgage balance. A borrower with a \$250k mortgage borrowing at the 90^{th} percentile interest rate would thus save the equivalent of \$7500 in upfront points and fees by borrowing at the 10^{th} percentile interest rate.

In column (4), we add lender fixed effects, to allow for the possibility that some of the price differences may reflect differences in lender quality as well as differences in lenders' costs (e.g. convenience of the office location, or service quality). Indeed, the residual dispersion in rates further decreases, but remains substantial. In the remaining columns, we further interact the lender fixed effects with other controls, to allow for the possibility that lenders may differ in how they price certain loan features, or that their (relative) pricing may change over time or across locations.

[5] The estimation drops "singleton" observations that are completely determined by the set of fixed effect. There are more such singletons as we add more fixed effects; to ensure that our results are not driven by changing samples, we use the remaining sample from the most restrictive specification (7) in all specifications.

[6] The lock period typically varies from 15 to 90 days, with 30 and 45 days being the most common choices. A longer lock period leads to a slight increase in the fee (or equivalently the interest rate).

[7] It is already somewhat more precise, since here we observe the month in which a loan is locked, along with the length of the lock period, while in typical dataset loans originated in the same month may have been locked in different months.

The final two columns of the table suggest that indeed such within-lender pricing variation may be important, since the remaining dispersion is roughly 30 percent lower in column (7) than in column (4). Nevertheless, even conditional on these very fine interacted fixed effects, which should come close to looking at nearly-identical borrowers getting a loan from the same lender in the same location at the same time, the 90th-10th percentile difference remains at 32bp, and the interquartile range at 14bp. The remaining dispersion is further illustrated in Figure 1, which compares the distribution of the residualized interest rates from specification (7) of Table 2 with the one from specification (3), which does not feature any lender fixed effects. Adding the lender effects narrows the distribution, but it remains wide. The figure also shows that the distributions are quite symmetric and bell-shaped.

Table 3 shows how the residual dispersion in interest rates from specification (7) of Table 2 varies across different loan programs and characteristics. What stands out is that the dispersion is substantially larger for loan types and borrower characteristics that are associated with being more financially constrained and potentially less sophisticated. For instance, the 90th-10th percentile difference is 44 basis points for borrowers with a FICO below 630, versus only 27 basis points for borrowers with FICO above 750, and the dispersion falls monotonically in between. Similarly, for high-LTV loans, the dispersion is higher than for LTVs below 80. Since most of these high-LTV loans are in the FHA program, it is also not surprising that residual price dispersion there is larger than for other programs. Finally, the last section of the table shows that rates for first-time homebuyers also exhibit larger dispersion than for experienced borrowers.

It is interesting to note that the dispersion in *offered* rates, which is studied in detail in the online appendix, is also substantial, but does not vary much with borrower characteristics. This is illustrated in Figure 2, which plots the interquartile ranges in residualized locked rates (from specification (3), i.e. without lender effects) and in the offer rates for identical mortgages across lenders. This suggests that the differences across FICO scores and LTVs in the locked data arise from differential shopping or negotiating, but not from the supply side per se.

The findings so far have illustrated that there is a large amount of dispersion in the rates that different mortgage borrowers pay. While some of it is explained by different timing, upfront payments, or lender fixed effects, substantial dispersion remains once we control finely for variation in different lenders' pricing over time, across locations, or across loan programs. This implies that two observably identical borrowers may get quite different deals even from the exact same lender at the same time. Furthermore, this appears to be more pronounced for financially less well-off borrowers or those that are inexperienced in the market.

The analysis above has focused on dispersion, or "second moments." We next turn to the question of whether different types of borrowers get good or bad deals *on average* (i.e. the first moment), relative to what is available in the market at the time they lock their mortgage.

4 Comparing the Locked Rates to Lenders' Best Offer

For the analysis in this section, we merge actual transaction interest rates from the mortgage rate lock dataset with the data on lenders' best offer rates (described in Section 2.2). For each observation of a rate lock in our data, we compute the median of the best offer rates in the same market, on the same day for an identical mortgage. We then study the difference between the rate obtained by consumers and the median rate available in the market for an identical mortgage—the *locked-offered rate gap*.[8]

4.1 Summary Statistics

Figure 3 shows the distribution of the locked-offered rate gap for all mortgages in our data. The thick black line denotes the mean of the distribution. The locked-offered rate gap is positive on average, meaning that borrowers end up with mortgage rates that are more expensive than what the median lender could offer for identical mortgages.[9] As shown in Table 4, the average locked-offered rate gap is +15 basis points.

Figure 4 shows the distribution of the locked-offered rate gap for various sub-segments of the mortgage market. The figure shows that the locked-offered rate gap distributions are centered to the right of zero for conventional conforming and FHA loans, meaning that the average borrowers in these segments pays more than the median best offer. The summary statistics for these distributions are given in Table 4. We have fewer observations than in the previous analysis based on lock data only, since here we need to observe the offer side, which is only available for a subset of loan types/characteristics, 20 MSAs, and a shorter time period.

The locked-offered rate gap is largest for FHA loans, with an average of +30bp. This amounts to 2% of the mortgage balance in upfront points/fees, which for a typical FHA loan of $200k amounts

[8]We use the rate at which the median lender offers a loan with zero points and fees from the offers data. To compare to this offer, we adjust the locked rate for points paid or received by the borrower based on the regression coefficient from the previous analysis.

[9]In Figure A-4 in the appendix, we validate that the median rate we use is close to the daily rate that is quoted on Mortgage News Daily, an industry website.

to \$4000. On the other hand, the market for super-conforming mortgages and jumbo mortgages looks very different: the locked-offered rate gap is on average slightly negative at -3bp for super-conforming mortgages, and even more negative at -17bp for jumbo mortgages. Thus, in these two market segments, borrowers pay less than what the median lender in their market could offer them.

Table 4 further shows summary statistics of the locked-offered rate gap distribution by splitting the sample by FICO scores, LTV ratios, and whether the borrower is a first-time home-buyer. On average, borrowers with a FICO larger than 750 lock in mortgage rates that are close to the median offer, while borrowers with lower FICO scores lock in rates well above the median offer. For instance, borrowers with FICO scores of less than 660 on average pay 33bp more than what the median lender would offer for identical mortgages. What this means is that low-FICO borrowers on average tend to pay substantially higher rates not just due to additional risk premia embedded in lender offers, but to a large extent due to the fact that they end up with worse rates relative to what is in principle available in the market.

A similar pattern is evident when splitting the sample by LTV: borrowers with LTV less than 90% tend to obtain rates close to the median of the offer distribution, while higher LTV borrowers do worse relative to the median offer. First-time homebuyers also tend to fare worse: on average, first-time buyers pay 20bp more than what the median lender could offer them, while repeat homebuyers pay only 11bp more.

It is worth noting that within each of the groups in Table 4, there is still substantial dispersion in the locked-offered rate gap, as shown in the table's final three column. Thus, even for high-FICO or low-LTV borrowers, which on average have a gap close to zero, there are still a lot of borrowers that lock rates well above what the median lender could offer them. The dispersion tends to be largest for the groups that on average do worst, meaning they have the most positive average gap.

4.2 Regression analysis

Next, we turn to a regression analysis to investigate whether similar patters emerge when controlling for borrower, product, and lender fixed effects simultaneously. We estimate the following specification:

$$rate^{lock}_{imt} - \overline{rate}^{offer}_{X_i mt} = \alpha + f(X_i) + \mu_t + \lambda_l + \xi_m + \varepsilon_{imt} \tag{1}$$

where $rate^{lock}_{imt}$ is the interest rate locked by borrower i in market m on date t, and $\overline{rate}^{offer}_{X_i mt}$ is the median offer in market m at time t for a mortgage with characteristics X_i. X_i denotes observable borrower and mortgage characteristics such as FICO, LTV, DTI, or loan amount. There are some attributes that we use in the calculation of $\overline{rate}^{offer}_{X_i mt}$ that we do not explicitly control for in the regression; in particular, whether a loan is a conforming, super-conforming, jumbo or an FHA mortgage. Especially jumbo status is essentially collinear with loan amount, so that including program indicators makes it difficult to independently estimate the effect of loan amount.[10] Conversely, there are some variables that do not matter for offered rates but that we still use on the right-hand-side as variables of interest, namely whether the borrower is a first-time homebuyer, as well as the age of the borrower. Finally, we control for time fixed effects μ_t, metropolitan area fixed effects ξ_m, and lender fixed effects λ_l, and in some cases the interactions of these fixed effects. We estimate a flexible function f by discretizing each characteristic X and including each group separately. Standard errors are clustered at the MSA, month, and lender level.

The results from the estimation of equation (1) are shown in Table 5. Column (1) controls for MSA and month fixed effects. In line with the summary statistics above, borrowers with higher FICO scores tend to choose lower rates from the offer distribution available to them, even controlling for other observables. The estimated coefficient, also shown graphically in Figure 5, implies that the locked-offered rate gap is about 18bp lower for borrowers with a FICO score of 750 or above than for those with a FICO of below 660. Similarly, the locked-offered rate gap is about 17bp lower for borrowers with a LTV of less then 80% than for those with LTV of 96% or higher. Loan amount is another statistically and economically significant determinant of the gap: the largest loans of $800k or more have a locked-offered gap of 26bp lower than loans below $200k. Finally, locked-offered gaps tend to be larger for borrowers with higher DTI, for first-time homebuyers, and for older borrowers.

The key takeaway from the results in column (1) is that borrower characteristics that are associated with being more financially constrained or less sophisticated strongly correlate with obtaining a mortgage rate that is worse relative to what we know lenders in the market to offer for borrowers with such characteristics. The 18bp premium that we observe for an otherwise similar borrower with FICO below 660 relative to one with FICO above 750 is thus in addition to any premium for higher default risk that is already embedded in lender offers.

One possible explanation is that these coefficients arise from sorting into cheap or expensive lenders. Borrowers might choose expensive lenders because they offer better service or simply

[10] However, the other coefficients (e.g. on FICO or LTV bins) remain almost unchanged if program indicators are added to the regression.

because they spend more on marketing and are more visible. To investigate whether this explains our previous results, we include lender fixed effects in the remaining columns of Table 5. Column (2) shows that just adding constant lender fixed effects increases the R-squared from 23 percent to 40 percent, meaning that lender-specific pricing differences do explain a fair amount of variation in our data. However, most of the coefficients of interest remain unchanged.

The same remains true when we allow for lenders to price differently across MSAs (column 3), or over time within an MSA (column 4). The last specification allows for the possibility that lenders' relative pricing evolves over time (e.g. a new lender may price cheaply to gain market share, but then become more expensive). Adding these interacted fixed effects further increases the explanatory power of the regression, but still leaves the coefficients on borrower characteristics essentially unchanged. Thus, it does not appear that e.g. lower-FICO borrowers end up with higher locked-offered gaps just because they get their loans from (temporarily or constantly) expensive lenders. Instead, the result that they end up with relatively worse deals holds even conditioning on going to the same lender at the same time as an otherwise identical high-FICO borrower.

Robustness. One concern is that most of the lenders making offers in our dataset may be small and hard to find. If that was the case, it would not be surprising that most borrowers pay more than what the median lender is offering. To rule out this potential explanation, we replicate the same findings in our main regression using only offers from high-volume lenders, as designated on the Optimal Blue platform. Our results remain unchanged even for this sub-sample of lenders. [more robustness checks to be added]

5 Time-series Movements in the Locked-Offered Rate Gap and Price Dispersion

The previous two sections explored the cross-sectional patterns in the dispersion of locked rates, and in mean and dispersion of the locked-offered rate gap. In this section, we instead focus on how these measures move over time, with a particular interest in how they respond to changes in market interest rates. Are borrowers more likely to end up with worse rates (relative to what the median lender could offer) when market rates are low, and more likely to get a good deal as rates increase? Does price dispersion change with market interest rates?

Figure 6 plots the average locked-offered rate gap against the 10-year Treasury yield.[11] Between mid-2016 and mid-2018, 10-year Treasury yields increased by almost 1.5 percentage points, but with substantial fluctuations in between. There have also been significant movements in the locked-offered rate gap, which almost mirror movements in the Treasury yields. The locked-offered rate gap is largest when Treasury yields are low and the gap is low when yields are high.

We confirm the statistical significance of the relationship between the locked-offered rate gap and market rates in column (1) of Table 6. We regress the monthly change in the average locked-offered gap (across all mortgages in the sample in a given month) on the monthly change in the 10-year Treasury yield. The coefficient implies that as the 10-year Treasury yield increases by 1 percentage point, the average locked-offered gap fall by about 13bp. This is sizable, given that we saw earlier that over our sample as a whole, the gap averaged 15bp with a standard deviation of 31bp. The R-squared of the regression, at 0.37, also indicates a strong relationship between these variables.

The remaining columns of the table study the strength of this relationship for different subsamples, which may help us shed light on the underlying drivers. One possibility is that the relationship is driven purely by affordability constraints: as market rates increase, the implied monthly mortgage payments increase, and more borrowers may come up against DTI constraints embedded in mortgage underwriting.[12] To study whether this is likely to be an important factor in the data, we separate borrowers into those with a DTI up to 38 percent (which are likely unconstrained by the payment burden) and those with higher DTI (for whom a higher rate may mean that they run up against underwriting constraints). Another possibility is that the relationship is driven more by "behavioral" factors: for instance, when the level of rates is already low, borrowers may feel less compelled to search for a good deal or negotiate hard than when rates are higher, even though in dollar terms the consequences are the same. This might be the case particularly after a recent drop in rates, as borrowers might compare their offer to a higher reference level.[13] While we do not have a good individual measure of being subject to behavioral biases in our data, we use FICO score as

[11] We use the 10-year Treasury yield since it is strongly correlated with the 30-year fixed mortgage rate, but avoids potential endogeneity issues due to the measurement of the latter. However, using the mortgage rate or the current-coupon MBS yield instead leaves our conclusions unchanged.

[12] The relevant debt-to-income ratio in the US is usually the so-called "back-end" ratio, which divides the required monthly payments on all debts (not just the mortgage) by the monthly income. Under the "qualified mortgage" rule that has been in effect in the US since 2014, this back-end DTI ratio is supposed to be below 43 percent (see e.g. DeFusco et al., 2017). However, conforming mortgages guaranteed by Fannie Mae and Freddie Mac are exempt from this requirement; these entities therefore impose their own requirements, which in some cases can be higher.

[13] There are also behavioral factors that might push in the opposite direction: for instance, "relative thinking" would make a 20bp rate saving appear larger when compared against a 3 percent base rate than compared against a 4 percent base rate, and might thus lead borrowers to shop more in the former case.

a proxy for financial sophistication, and check whether the relationship is stronger for borrowers with FICO below 680 than for those with higher FICOs.

We thus form four groups (FICO below/above 680 crossed with DTI below/above 38 percent) and repeat the same regression in monthly changes. Comparing columns (2)-(3) to (4)-(5), we see that the relationship between rate changes and average locked-offered gap tends to be stronger for low-FICO borrowers than for high-FICO borrowers. This might be consistent with a behavioral explanation, though of course it is difficult to rule out other factors. However, at least for low-FICO borrowers, coming up against DTI constraints does not seem to drive the strength of the relationship: the point estimates in columns (2) and (3) are almost identical. In contrast, for higher-FICO borrowers, it is the case that the relationship is stronger for the higher-DTI group, though it also remains significant for the low-DTI (unconstrained) group. Overall, these results suggest a potential role for both behavioral factors and underwriting constraints in driving the strong negative correlation between locked-offered gaps and market rates.

Next, we turn to investigating whether price dispersion also moves with market interest rates. Table 7 regresses the monthly changes in the standard deviation of the residualized locked rate (from specification (4) in Table 2) on changes in market interest rates. We find that dispersion in residualized locked rates falls as interest rates increase. Again, this relationship is stronger for low-FICO borrowers, but within this group, there is little difference based on borrowers' DTI. For these borrowers, as the market rate increases by 1 percent, the standard deviation in residualized rates falls by about 6.5bp (relative to a mean over the sample period of about 23bp). Again, the relationship is strong, as indicated by the R-squared values around 0.4.

Note that when we repeat the same regressions using dispersion in the offer rate distribution (not shown here), we find almost no relationship between price dispersion and market rates. The coefficients are both statistically and economically close to zero and the R-squared is very low. Also, the standard deviation of rates in the offer data changes very little over time.

To summarize, when interest rates increase, borrowers obtain mortgage rates that are lower relative to the mean of the offer distribution, i.e. the locked-offered rate gap decreases. The price dispersion in the lock data also drops. These effects are stronger for borrowers with low FICO scores; DTI does not matter for the strength of the relationship within the low-FICO group, though it does matter somewhat for higher-FICO borrowers (where the more constrained ones respond more to market rates). It appears likely that behavioral factors play at least some role behind these patterns, although affordability constraints may also matter at least for some borrowers.

6 Discussion of Potential Explanations and Policy Implications

6.1 Potential Explanations

There are different potential explanations for the results in our paper. In this section we discuss and assess different mechanisms that could generate the patterns we have documented so far.

Differences in lender quality. It is possible that borrowers choose mortgages with positive locked-offered gap because the lenders that offer these provide a better product overall. Some lenders may provide better customer service, are less prone to commit errors, and are more likely to close quickly. While this is all true, it is implausible that differences in lender quality can explain the patterns observed in our data. First, the amount of price dispersion in both offered and locked rates appears too large, especially when considering identical and very standardized mortgages that are guaranteed by the government. Second, and more importantly, the regression results from Table 5 directly control for lender fixed effects. The last specification in that table even allows for the possibility that lender quality varies across locations and over time. The results remain unchanged: what seem to be more sophisticated borrowers obtain mortgages from the lower end of the rate distribution. This suggests that our results are not driven by differences in lender quality.[14]

Another interpretation for why borrowers who are likely financially better off (higher FICO scores, larger loan amounts, etc.) obtain better terms is that banks want to engage in cross-selling of other financial products to them. However, most of the lenders in our dataset are mortgage companies that do not offer other financial services (such as credit cards or savings accounts), therefore cross-selling is highly unlikely to explain our results.

Shopping and negotiation. One hypothesis explored in the literature is that while savvy borrowers fare well in the mortgage shopping process, most borrowers search for mortgages suboptimally, perhaps because of lack of sophistication and understanding of the structure of the mortgage market (e.g. Woodward and Hall, 2012; Alexandrov and Koulayev, 2017).

In our data, we find evidence that is consistent with this explanation: borrowers that are likely to be more sophisticated (based on their high FICO score, or on being repeat homebuyers) obtain lower rates (relative to market offers) than their less savvy counterparts.[15]

[14] It might also be possible that borrowers choose expensive lenders because these lenders specialize in offering the product the borrower is looking for. In order to rule this out, we have replicated our main results using a sub-sample of lenders that offer a proportional mix of mortgage products available in the market (not shown).

[15] We also find that borrowers with higher loan amounts (mostly jumbo borrowers) obtain better rates. While this

However, we again emphasize that this is for the most part *not* driven by less sophisticated borrowers simply "going to the wrong lender": the findings above hold after conditioning on lender (or lender-location-time) fixed effects. Furthermore, looking only at the locked rates, there is substantial residual rate dispersion for identical-looking borrowers going to the same lender at the same time (and this dispersion is smaller for higher-FICO, lower-LTV borrowers).

This suggests that an important part of getting a good rate lies in the negotiation process. For instance, more sophisticated borrowers may know what rate/point combinations various lenders offer (either from obtaining direct quotes, or from shopping online), and ask their "preferred" lender to match these prices. Another borrower may go to the same lender but end up with a less favorable rate, because the lender has no need to compete.[16]

Our findings are similar to those of Allen et al. (2014) for the Canadian mortgage market. There, lenders post a weekly indicative rate (with essentially no price dispersion across the main six lenders that dominate the market) and borrowers then negotiate discounts, which results in substantial dispersion in the rates that borrowers end up paying. The authors argue that this dispersion reflects differences in consumer bargaining leverage, although they note that some of the price differences could reflect unobserved lending cost heterogeneity. In our setting, we can control for potential lending cost heterogeneity (e.g. differential credit risk premia) since we observe all the relevant underwriting variables. Thus, the remaining dispersion within lender is likely to indeed reflect differential bargaining leverage across otherwise identical borrowers (e.g. some having shopped for outside offers). Relative to the Canadian market, where the "discount negotiation" process is presumably quite transparent, it may be more surprising that negotiation plays such a big role in the US market, where there is already much dispersion and loan-characteristics-specific pricing in offered rates.

6.2 Policy Implications

Our empirical results provide evidence that many borrowers from the most vulnerable part of the borrower population in the US seems to overpay for mortgages: those that are most likely to be relatively low income, low net worth, and more likely to be first-time homebuyers. These are the exact borrowers that various government programs effectively subsidize. If they were to obtain

may partly reflect differential sophistication, this relationship would also be predicted by a fully rational search model where borrowers face fixed costs (independent of loan amount) of searching for other offers.

[16]One channel through which this might work is that the lender in our data—e.g. a mortgage company—could choose which investor to work with depending on how competitive the rate needs to be. It is conceivable that some investors offer more convenience to lenders, but higher rates.

mortgages from the lower end of the offer distribution, this would make their mortgage payments more affordable and leave them with more disposable income. Alternatively, the FHA and the GSEs could afford to raise their guarantee fees substantially without affecting final cost to borrowers.

Thus, it might be worth at least considering policies that would help borrowers search and negotiate more effectively. This could take the form of required information disclosure to borrowers of the rates available to them across different lenders in the same market (for instance at the time they lock their rate). We recognize that this is not a straightforward endeavor given the multi-dimensional nature of mortgage pricing in the US, but advances in technology may make this more feasible than in the past. Alternatively, the guaranteeing agencies could impose requirements on the maximum locked-offered gap they allow for loans to be securitized. Of course, one would also want to consider general equilibrium effects on the offers that lenders make (see also Alexandrov and Koulayev, 2017).

The negative relationships between average locked-offered rate gap and rate dispersion with the level of market rates that we document in Section 5 also matter for monetary policy transmission. Our findings imply that as rates fall (e.g. in response to central bank actions), borrowers tend to do worse relative to the distribution of offered rates, perhaps due to less shopping or negotiation. It follows that the contract rates they end up with do not fall as much as they could, based on lenders offers, adding another friction to the pass-through of monetary policy to the mortgage market.[17] Furthermore, this relationship is stronger for low-FICO borrowers, whose spending and default hazard might respond most strongly to a larger drop in their mortgage rate (e.g. Abel and Fuster, 2018).

7 Conclusion

Using new data that includes both offered and transacted mortgages, we document new facts on price dispersion in the US mortgage market. First, we document a large amount of price dispersion in locked mortgage rates, even after very finely controlling for borrower characteristics. Most of this dispersion persists within lender, and even within lender in a given location at a given time. Furthermore, comparing locked rates to the distribution of lenders' best offers, we show that the

[17] Existing work has shown that offers (as measured from investor rate sheets) respond less to increases in MBS prices than to decreases, and less so when borrower demand is already high, which happens after falls in rates (Fuster et al., 2017). Limited competition may also limit pass-through (Agarwal et al., 2017; Scharfstein and Sunderam, 2016). Finally, many borrowers fail to refinance when it is in their financial interest to do so (e.g., Campbell, 2006; Andersen et al., 2015; Keys et al., 2016).

typical borrower ends up with a rate higher than what the median lender could offer them.

Not all borrowers overpay relative to the median lender, however. Borrowers that are more likely to be more financially constrained and less sophisticated do relatively worse. This again persists after conditioning on lender fixed effects, meaning that it is not just due to these borrowers going to higher-priced lenders. Rather, we argue that it likely reflects differences in negotiating/bargaining, perhaps due to having shopped around for other offers.

Finally, both price dispersion and the average gap between locked rates and offers vary over time: when interest rates increase, borrowers end up toward the lower end of the offer distribution, and price dispersion also falls, especially for low-FICO borrowers. This may reflect that when market rates are very low, borrowers feel less of a need to shop around or negotiate.

If these findings are indeed generated mostly by differences in how effectively borrowers search for mortgages and negotiate with their lender, there may be large scope for public policy to help borrowers that overpay relative to the median lender. In future versions of this paper, we intend to consider the benefits from such policies through counterfactual simulations.

References

Abel, Joshua and Andreas Fuster, "How Do Mortgage Refinances Affect Debt, Default, and Spending? Evidence from HARP," Staff Report 841, Federal Reserve Bank of New York 2018.

Agarwal, Sumit, Gene Amromin, Souphala Chomsisengphet, Tim Landvoigt, Tomasz Piskorski, Amit Seru, and Vincent Yao, "Mortgage Refinancing, Consumer Spending, and Competition: Evidence From the Home Affordable Refinance Program," Working Paper 21512, National Bureau of Economic Research 2017.

Alexandrov, Alexei and Sergei Koulayev, "No Shopping in the U.S. Mortgage Market: Direct and Strategic Effects of Providing Information," Working Paper, CFPB 2017.

Allen, Jason, Robert Clark, and Jean-Francois Houde, "Price Dispersion in Mortgage Markets," *The Journal of Industrial Economics*, 2014, *62* (3), 377–416.

Andersen, Steffen, John Y. Campbell, Kasper Meisner Nielsen, and Tarun Ramadorai, "Inattention and Inertia in Household Finance: Evidence from the Danish Mortgage Market," Working Paper 21386, National Bureau of Economic Research July 2015.

Argyle, Bronson, Taylor Nadauld, and Christopher Palmer, "Real Effects of Search Frictions in Consumer Credit Markets," Working Paper 2017.

Bayer, Patrick, Fernando Ferreira, and Stephen L. Ross, "What Drives Racial and Ethnic Differences in High-Cost Mortgages? The Role of High-Risk Lenders," *Review of Financial Studies*, 2018, *31* (1), 175–205.

Bhutta, Neil and Aurel Hizmo, "Do Minorities Pay More for Mortgages?," Working Paper 2018.

Campbell, John Y, "Household Finance," *Journal of Finance*, 2006, *61* (4), 1553 – 1604.

Damen, Sven and Erik Buyst, "Mortgage Shoppers: How Much Do They Save?," *Real Estate Economics*, 2017, *45* (4), 898–929.

DeFusco, Anthony A., Stephanie Johnson, and John Mondragon, "Regulating Household Leverage," Working Paper, Northwestern University 2017.

Fuster, Andreas, Stephanie H. Lo, and Paul S. Willen, "The Time-Varying Price of Financial Intermediation in the Mortgage Market," Staff Report 805, Federal Reserve Bank of New York 2017.

Gurun, Umit G., Gregor Matvos, and Amit Seru, "Advertising Expensive Mortgages," *The Journal of Finance*, 2016, *71* (5), 2371–2416.

Keys, Benjamin, Devin Pope, and Jaren Pope, "Failure to Refinance," *Journal of Financial Economics*, 2016, *122* (3), 482 – 499.

Scharfstein, David and Adi Sunderam, "Market Power in Mortgage Lending and the Transmission of Monetary Policy," Working Paper, Harvard University 2016.

Stango, Victor and Jonathan Zinman, "Borrowing High versus Borrowing Higher: Price Dispersion and Shopping Behavior in the U.S. Credit Card Market," *The Review of Financial Studies*, 2016, *29* (4), 979–1006.

Woodward, Susan E. and Robert E. Hall, "Diagnosing Consumer Confusion and Sub-optimal Shopping Effort: Theory and Mortgage-Market Evidence," *American Economic Review*, 2012, *102* (7), 3249–76.

Table 1: Summary Statistics of the Rate Lock Data

	Conforming		Super-Conforming		Jumbo		FHA	
	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Loan Amount ($000)	256	92	528	65	715	242	222	93
Interest Rate	4.28	0.45	4.27	0.41	4.15	0.44	4.21	0.56
Discout Points Paid	-0.08	1.02	0.08	0.99	-0.04	0.86	-0.19	1.18
FICO	738	43	748	36	762	28	674	44
LTV	83	13	80	12	77	9	96	5
DTI	35	9	36	9	31	9	42	9
First-time Homebuyer %	23		20		8		51	
Refinance Share %	18		25		25		5	
N. Observations	1371329		82814		50048		776587	

Table 2: Dispersion in Locked Interest Rates After Controlling for Borrower and Loan Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Residual Dispersion							
St. Deviation	0.27	0.26	0.25	0.23	0.22	0.19	0.16
75-25th percentile	0.29	0.27	0.25	0.22	0.20	0.17	0.14
90-10th percentile	0.60	0.57	0.54	0.48	0.44	0.38	0.32
FICO x LTV x DTI x Loan Amount x Program grid	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lock month x MSA F.E.	Yes	Yes					
Discount points/credits added to grid		Yes	Yes	Yes	Yes	Yes	Yes
Lock date x MSA F.E.			Yes	Yes	Yes	Yes	Yes
Lender F.E.				Yes			
Lender x FICO x LTV x Program F.E.					Yes	Yes	
Lender x Points F.E.					Yes	Yes	
Lender x MSA x Week F.E.						Yes	Yes
Lender x FICO x LTV x Program x Week F.E.							Yes
Lender x Points x Week F.E.							Yes
Observations	1939237	1939237	1939237	1939237	1939237	1939237	1939237
Adusted R-squared	0.68	0.71	0.72	0.76	0.79	0.82	0.84

Notes: The dependent variable is the mortgage interest rate locked. The data covers mortgage rates locked for 277 metropolitan areas during the period between 2015-2018. We focus on 30 year, fixed rate, fully documented mortgages. All specifications include Lock period f.e., refinance f.e., as well linear controls for all the variables in the grid.

Table 3: Summary Statistics of the Residualized Locked Rate

	Observations	St. Deviation	Percentile Differences	
			$75^{th} - 25^{th}$	$90^{th} - 10^{th}$
All Mortgages	1,939,237	0.16	0.14	0.32
Program				
Conforming	1,182,667	0.14	0.13	0.29
Super-Conforming	62,404	0.12	0.13	0.29
Jumbo	35,437	0.14	0.14	0.29
FHA	658,729	0.20	0.17	0.38
FICO				
<630	102,797	0.20	0.21	0.44
630-660	222,588	0.19	0.18	0.39
660-690	289,280	0.19	0.16	0.36
690-720	326,162	0.17	0.15	0.34
720-750	340,604	0.15	0.13	0.30
>750	657,806	0.13	0.12	0.27
LTV (%)				
<80	317,010	0.12	0.12	0.26
80-90	416,425	0.13	0.13	0.27
90-94	220,893	0.15	0.14	0.32
94-96	337,412	0.16	0.15	0.33
>96	647,497	0.20	0.17	0.39
First-Time Homebuyer				
No	1,301,344	0.15	0.13	0.30
Yes	637,770	0.19	0.17	0.37

Note - This table summarizes the residualized locked mortgage rate from specification (7) of table (2).

Table 4: Summary Statistics of the Rate Locked Minus the Median Offered Rate for Identical Mortgages

	Observations	Mean	St. Deviation	Percentiles	
				25^{th}	75^{th}
All Mortgages	65,165	0.15	0.31	-0.04	0.29
Program					
Conforming	36,126	0.10	0.24	-0.04	0.22
Super-Conforming	6,151	-0.03	0.23	-0.18	0.08
Jumbo	1,846	-0.17	0.34	-0.30	-0.03
FHA	21,042	0.30	0.36	0.07	0.50
FICO					
FICO$<$660	9,341	0.33	0.38	0.08	0.55
660$\leq$FICO$<$690	9,828	0.23	0.33	0.02	0.41
690$\leq$FICO$<$720	11,572	0.18	0.30	0.00	0.33
720$\leq$FICO$<$750	11,163	0.11	0.27	-0.06	0.24
FICO$\geq$750	23,261	0.04	0.24	-0.10	0.16
LTV					
70$\leq$LTV$<$80	7,436	0.04	0.22	-0.09	0.16
80$\leq$LTV$<$90	16,813	0.04	0.25	-0.10	0.18
90$\leq$LTV$<$96	19,558	0.12	0.28	-0.05	0.25
LTV$\geq$96	21,358	0.29	0.35	0.06	0.48
First-Time Homebuyer					
No	35,547	0.11	0.28	-0.06	0.24
Yes	29,614	0.20	0.33	-0.02	0.37

Note - For each mortgage rate locked by borrowers in our data, we compute the median rate offered by lenders in the same market on the same day for an identical mortgage. This table summarizes the difference between each locked rate and the median offered rate.

Table 5: Explaining the gap between the rates consumers lock and the rates offered by the median lender for identical mortgages

	(1)	(2)	(3)	(4)
FICO groups				
$I_{660 \leq FICO < 690}$	-0.081***	-0.079***	-0.076***	-0.075***
	(0.013)	(0.010)	(0.010)	(0.010)
$I_{690 \leq FICO < 720}$	-0.097***	-0.096***	-0.094***	-0.093***
	(0.018)	(0.015)	(0.014)	(0.015)
$I_{720 \leq FICO < 750}$	-0.147***	-0.142***	-0.141***	-0.136***
	(0.020)	(0.017)	(0.016)	(0.016)
$I_{FICO \geq 750}$	-0.182***	-0.172***	-0.169***	-0.165***
	(0.017)	(0.014)	(0.013)	(0.013)
LTV groups				
$I_{80 \leq LTV < 90}$	0.003	0.002	0.004	0.007
	(0.004)	(0.004)	(0.005)	(0.004)
$I_{90 \leq LTV < 95}$	0.057***	0.051***	0.055***	0.056***
	(0.006)	(0.008)	(0.008)	(0.009)
$I_{LTV \geq 96}$	0.167***	0.156***	0.158***	0.158***
	(0.011)	(0.015)	(0.016)	(0.017)
DTI groups				
$I_{20 \leq DTI < 40}$	0.024***	0.009	0.004	0.003
	(0.006)	(0.009)	(0.008)	(0.005)
$I_{DTI \geq 40}$	0.050***	0.025***	0.016*	0.019***
	(0.007)	(0.008)	(0.008)	(0.005)
Loan Amount				
$I_{\$200k \leq Loan < \$400k}$	-0.064***	-0.056***	-0.059***	-0.051***
	(0.007)	(0.006)	(0.005)	(0.006)
$I_{\$400k \leq Loan < \$600k}$	-0.088***	-0.077***	-0.085***	-0.076***
	(0.005)	(0.008)	(0.009)	(0.008)
$I_{\$600k \leq Loan < \$800k}$	-0.154***	-0.140***	-0.154***	-0.142***
	(0.009)	(0.013)	(0.014)	(0.012)
$I_{Loan \geq \$800k}$	-0.262***	-0.237***	-0.254***	-0.252***
	(0.026)	(0.033)	(0.034)	(0.030)
First Time Homebuyer	0.019***	0.015***	0.008*	0.010**
	(0.005)	(0.004)	(0.004)	(0.004)
Age	0.001**	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
MSA F.E.	Yes	Yes	Yes	Yes
Month F.E.	Yes	Yes	Yes	Yes
Lender F.E.		Yes		
Lender x MSA F.E.			Yes	
Lender x MSA x Month F.E.				Yes
Observations	58609	58570	58238	54911
Adj. R-squared	0.231	0.396	0.418	0.458

Note - The dependent variable is the mortgage interest rate locked minus the median offer rate in the same market and day for an identical mortgage. The data covers mortgage rates for 20 metropolitan areas during the period between 2016-2018. We focus on 30 year, fixed rate, fully documented mortgages. The standard errors are clustered at the MSA, month, and lender level.

Table 6: Relationship between Changes in the Locked-Offered Rate Gap and Changes in Treasury Yields

Dep. Var.: ΔLocked-Offered Rate$_t$	All Data	$FICO \leq 680$ $DTI \leq .38$	$FICO \leq 680$ $DTI > .38$	$FICO > 680$ $DTI \leq .38$	$FICO > 680$ $DTI > .38$
	(1)	(2)	(3)	(4)	(5)
Δ10 Year Treasury Yield$_t$	-0.127***	-0.162**	-0.175**	-0.081**	-0.130***
	(0.03)	(0.06)	(0.06)	(0.03)	(0.03)
Observations	28	28	28	28	28
R-Squared	0.37	0.21	0.23	0.20	0.35

Note - The dependent variable is the change in monthly average of rate locked minus median offered rate for identical mortgages after controlling for borrower characteristics. Huber/White robust standard errors shown in parentheses.

Table 7: Relationship between Changes in the Dispersion of Residualized Locked Rates and Changes in Treasury Yields

Dep. Var.: ΔStd. Residual of Locked Rate$_t$	All Data	$FICO \leq 680$ $DTI \leq .38$	$FICO \leq 680$ $DTI > .38$	$FICO > 680$ $DTI \leq .38$	$FICO > 680$ $DTI > .38$
	(1)	(2)	(3)	(4)	(5)
Δ10 Year Treasury Yield$_t$	-0.036***	-0.064***	-0.067***	-0.016**	-0.039***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Observations	43	43	43	43	43
R-Squared	0.28	0.40	0.35	0.09	0.33

Note - The dependent variable is the month to month change in standard deviation of the residualized locked rates. Huber/White robust standard errors shown in parentheses.

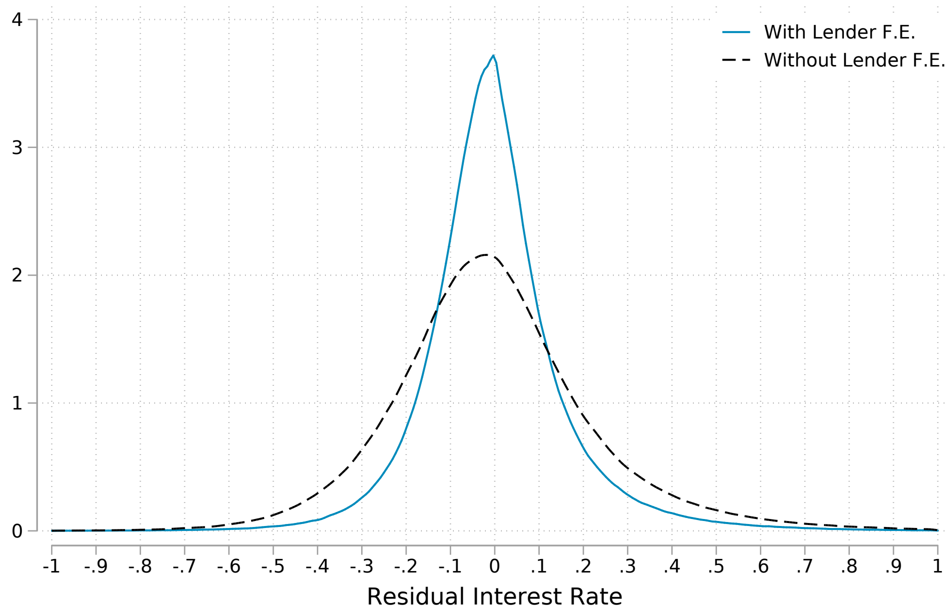

Figure 1: Residualized Locked Mortgage Rates Controlling for Borrower and Loan Characteristics, and Allowing for Differential Pricing of Loans by Lender-Location-Loan Characteristics-Time

Note: This figure plots fitted distributions of the residuals from the regression in columns (3) and (7) of Table 2.

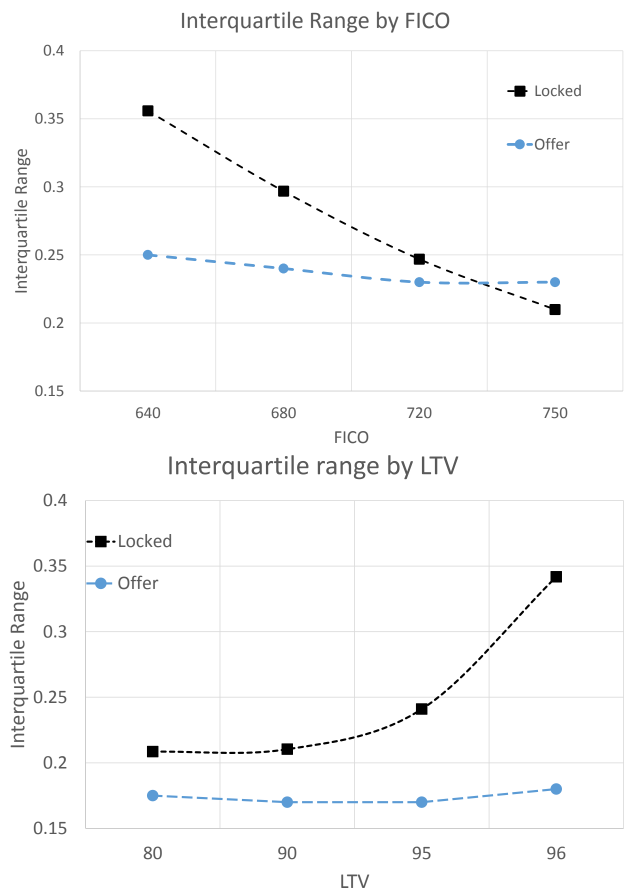

Figure 2: Comparing the Mortgage Rate Dispersion in Offer and Rate Lock Data

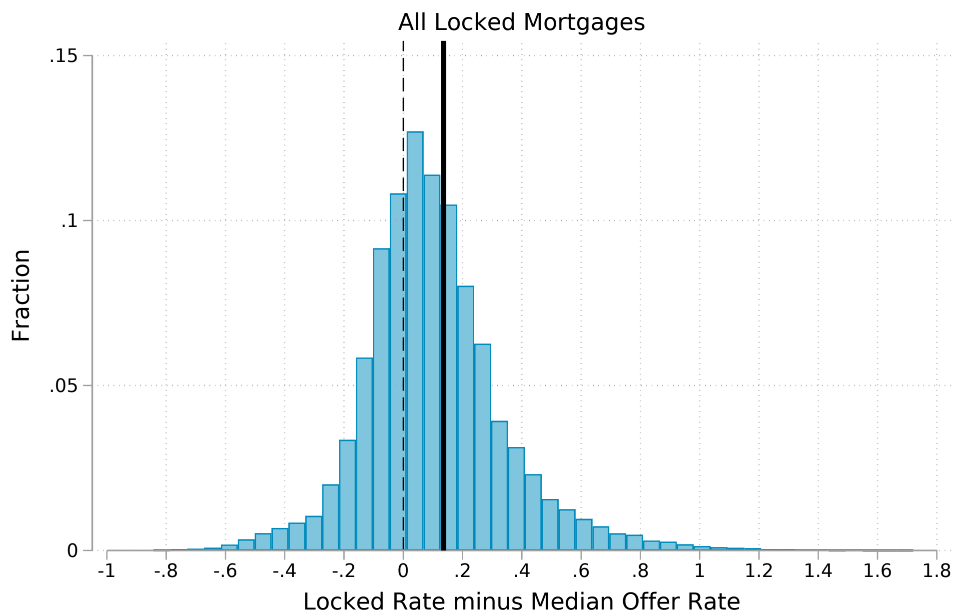

Figure 3: Distribution of Rate Locked Minus the Median Best Offered Rate for Identical Mortgages

Note: For each mortgage rate locked by borrowers in our data, we compute the median best offer by lenders in the same market on the same day for an identical mortgage. This figure shows the distribution of the difference between each locked rate and the median offered rate. The solid black line denotes the mean of the distribution.

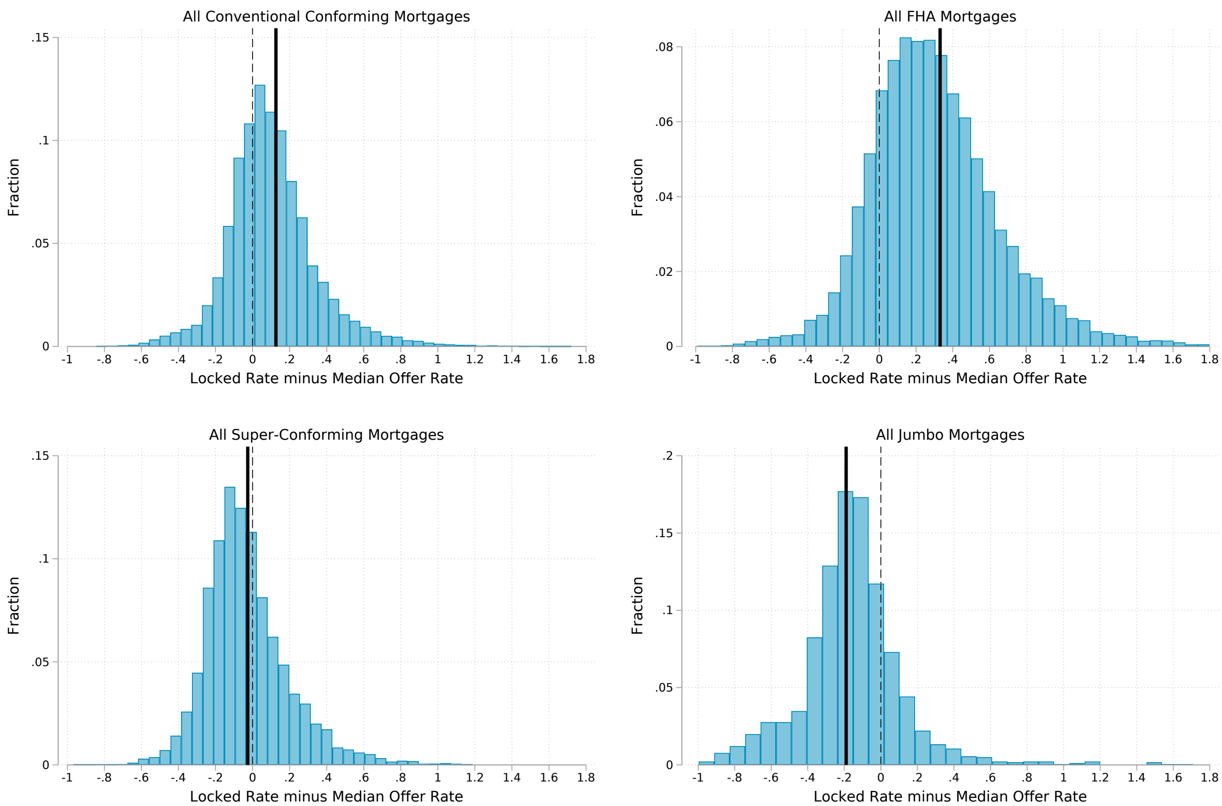

Figure 4: Distribution of Rate Locked Minus the Average Offered Rate for Identical Mortgages

Note: For each mortgage rate locked by borrowers in our data, we compute the average rate offered by lenders in the same market on the same day for an identical mortgage. This figure shows the distribution of the difference between each locked rate and the average offered rate. The solid black line denotes the mean of the distribution.

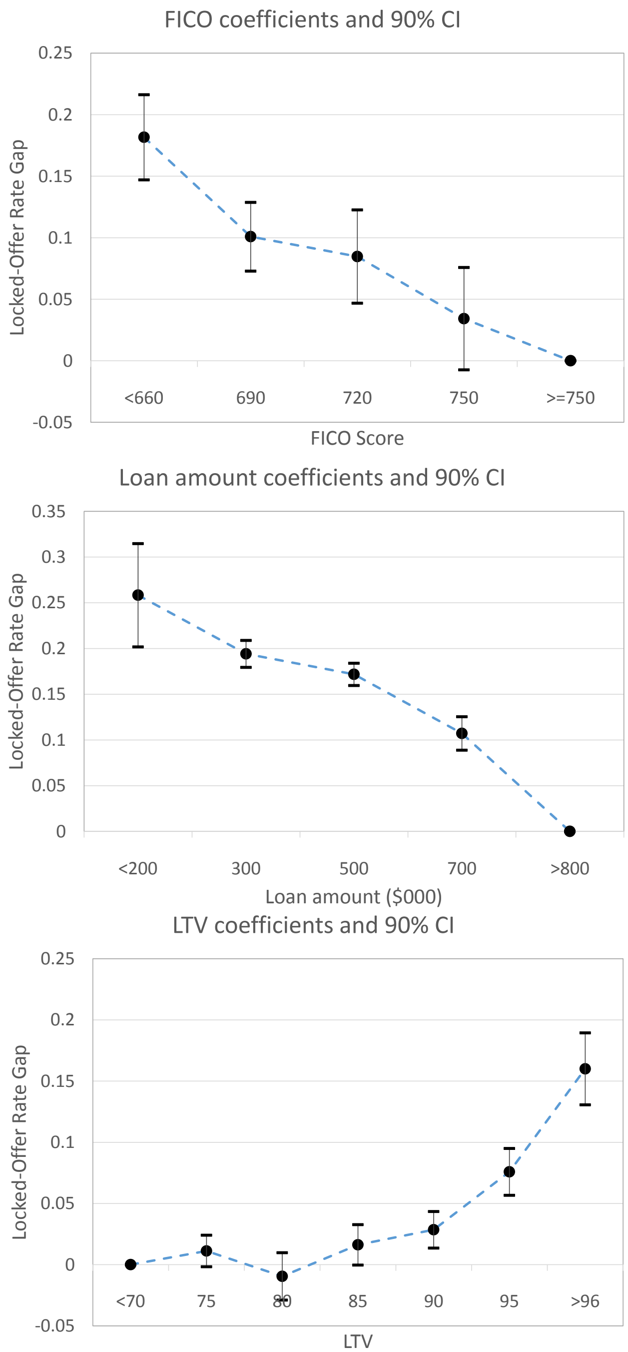

Figure 5: The Effects of Observables on the Locked-Offered Rate Gap

Note: These are plots of the coefficients from specification (1) in Table 5.

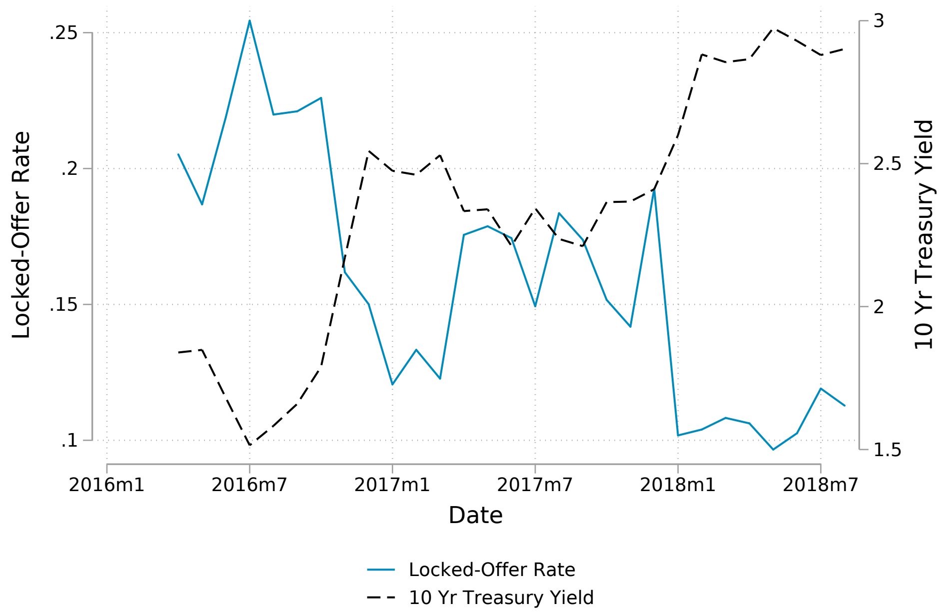

Figure 6: The Evolution of Rate Locked Minus the Average Offered Rate and Treasury Yields

Note: For each mortgage rate locked by borrowers in our data, we compute the average rate offered by lenders in the same market on the same day for an identical mortgage. The solid blue line is the average difference between each locked rate and the average offered rate. The dashed line is the 10 year treasury yield.

Online Appendix for
“Paying Too Much? Price Dispersion in the US Mortgage Market”

A.1 Price Dispersion in Mortgage Offers

In this appendix, we study price dispersion in offered mortgage rates across different lenders offering the same mortgage product in the same location at the same time, as observed on the Optimal Blue Insights platform. We show that, similar to the large dispersion in locked rates documented in the main text, there is also large dispersion in offered rates.

There are two things to consider when thinking about the what the price of a mortgage means in this context. First of all, lenders do not offer a single mortgage rate to borrowers but rather a menu with different combinations of mortgage rates and discount points to choose from. Borrowers can pay discount points, each equal to one percent of their mortgage balance, in order to lower their mortgage interest rate by roughly 15bp per discount point paid. Borrowers can also choose negative points, known as lender credits, in return for a higher mortgage rate of roughly 15bp per point. In this case, borrowers receive cash from the lender which can be used toward closing costs.

Secondly, lenders also charge origination fees. While fees are not typically considered as part of the price of the mortgage, they are part for the total cost of securing the mortgage. In a way we can think of lender fees and discount points as interchangeable. From the borrowers perspective, a lender that charges an origination fee of one percent to originate a mortgage at 4% interest is equivalent to a lender that charges no fees but requires the borrower to pay one discount point for a mortgage rate of 4%.

In light of the above considerations, there are two ways in which we quantify price dispersion. First, we look at the dispersion in mortgage rates for identical mortgages offered with no points and fees. While most lenders charge fees, the platform reports the rate at which lender credits (or negative discount points) would be equal to lender fees. A borrower would not have to pay the lender anything to lock this mortgage rate. Computing the dispersion in offer rates with no points and fees is not possible with any other data set we are aware of since one would need to know both the rate/point trade offs and the lender fees.

The second way we quantify price dispersion is by looking at the total points and fees a borrower would have to pay at different lenders in order to borrow at a given median interest rate. Since points and fees are paid by the borrower upfront, this makes it possible to quantify the price dispersion in terms of dollars and cents without having to engage in any present value calculations.

A.1.1 Dispersion in Offered Rates

We start by documenting the dispersion in mortgage rates available from different lenders for identical mortgages in Los Angeles. We only compare identical real-time mortgage offers with no points or fees, with exactly the same FICO, loan-to-value ratio, debt-to-income ratio, loan amount and location. We also focus on fixed rate mortgages with a 30 years term, with fully documented income, assets and employment, and mortgages secured by a single property for this analysis. The first panel of Figure A-1 shows the distribution of rates offered by different lenders for conforming mortgages with an amount of \$300k, FICO=750, LTV=80 and DTI=36 in Los Angeles. There are about 120 different lenders offering this mortgage in Los Angeles in any given day. The histogram shows the daily offered rates after subtracting the median (for the same day) over the period of April 2016 to April 2018. Figure A-1 uncovers a wide distribution of mortgage rates offered by different lenders even for the same mortgage product in the same location, at the same time. There is almost a full percentage point difference between the cheapest and the most expensive lender. Moreover, even though much of the mass is in the middle of the distribution, the tails of the distribution are rather fat. These patterns can also be seen in the other two panels of Figure A-1, which plot the dispersion in a typical FHA mortgage and a Jumbo mortgage. The exact shape of the distribution does look different across these different mortgages, however the amount of dispersion is similar.

Figure A-2 shows the dispersion in mortgage rates available from different lenders in all of the 20 metropolitan areas. To make the distribution comparable across time and locations, we demean the offered rates for each mortgage type in each market and day. Even in our pooled data, the amount of price dispersion seems similar to that in Los Angeles.

Table A-1 shows more detailed summary statistics of the rate dispersion in our offer data, broken down by mortgage types. There are typically about 120 unique lenders in any given day making offers for each mortgage type in each location. The median mortgage rate is higher for jumbo loans than for conforming loans reflecting in part the fact that conforming loans are guaranteed by Fannie or Freddie in exchange for a low guarantee fee, which is rolled into the mortgage rate. FHA mortgages have lower interest rates than other products since borrowers also have to pay upfront (175bp) and ongoing mortgage insurance premia (85bp) which are not part of the quoted mortgage rate. The price dispersion appears similar across different programs, FICO scores and loan-to-value ratios. Generally, the price dispersion is a bit higher for mortgages with low FICO scores, high LTVs and FHA mortgages. Overall, there is about a 75 basis point difference in mortgage rates between the 1^{st} percentile lender and the 99^{th} percentile lender.

Table A-2 compares the rate dispersion for a "plain vanilla" conforming mortgage with LTV of 80 and FICO of 750 across MSAs. We see that, while there are some differences in the exact amount of dispersion across MSAs, the qualitative points from above generalize across all of the cities, and Los Angeles is not an outlier.

A.1.2 Dispersion in Offered Points and Fees

In this subsection we focus on the points and fees charged by lenders to originate a mortgage with a median interest rate. The median interest rate for each mortgage type is defined exactly as in the previous subsection: it is the median interest rate across lenders for an identical mortgage offered with no points and fees. Figure A-3 shows the distribution of points and fees charged by different lenders to originate this median interest rate mortgage, with discount points and fees measured as a percent of the mortgage balance. Table A-3 summarizes this dispersion for different mortgage types. The differences in the upfront costs of a mortgage across lenders seems staggering. The difference between the 99^{th} percentile and 1^{st} percentile lender is close to 4% of the mortgage balance. For a typical conforming loan of $250K that amounts to a $9000 difference in upfront costs between these lenders. Even going from the 75^{th} percentile to the 25^{th} percentile lender would save about $3000 for a typical borrower with a $250k loan.

Table A-1: The real-time interest rate dispersion for offered mortgage products with no points and fees

	Median No. Offers	Median Rate	Standard Deviation	Percentile Differences $75^{th}-25^{th}$	$90^{th}-10^{th}$	$99^{th}-1^{st}$
Program						
Conforming	115	4.24	0.17	0.23	0.44	0.73
Super-Conforming	143	4.47	0.18	0.25	0.47	0.75
Jumbo	97	4.59	0.17	0.22	0.45	0.77
FHA	112	3.79	0.19	0.29	0.53	0.77
FICO						
640	113	4.62	0.18	0.25	0.48	0.76
680	110	4.34	0.17	0.24	0.46	0.75
720	118	4.20	0.17	0.23	0.45	0.75
750	118	4.13	0.17	0.23	0.45	0.75
LTV (%)						
70	118	4.24	0.17	0.23	0.45	0.75
80	116	4.31	0.18	0.24	0.47	0.76
90	105	4.51	0.17	0.23	0.45	0.75
95	125	4.34	0.17	0.23	0.45	0.73
96	112	4.01	0.18	0.26	0.49	0.76

Note - This table compares real-time interest rates for identical offered mortgages (same FICO, LTV, DTI, loan amount, location, time etc.) with no points and fees. Column 1 shows the median number of lenders offering each mortgage product in a location on a specific day. Columns 3-5 show the difference between various percentiles of the offer distribution.

Table A-2: The real-time interest rate dispersion for offered conforming mortgages with no points and fees

	Median	Median	Percentile Differences		
	No. Offers	Rate	$75^{th} - 25^{th}$	$90^{th} - 10^{th}$	$99^{th} - 1^{st}$
Atlanta, GA	104	4.20	0.20	0.39	0.64
Boston-Worcester-Lawrence, MA-NH-ME-CT	68	4.15	0.22	0.47	0.77
Charlotte-Gastonia-Rock Hill, NC-SC	81	4.22	0.19	0.39	0.70
Chicago-Gary-Kenosha, IL-IN-WI	109	4.14	0.23	0.42	0.69
Cleveland-Akron, OH	47	4.24	0.24	0.46	0.70
Dallas-Fort Worth, TX	129	4.19	0.22	0.40	0.68
Denver-Boulder-Greeley, CO	122	4.20	0.19	0.37	0.62
Detroit-Ann Arbor-Flint, MI	70	4.14	0.22	0.42	0.75
Las Vegas, NV	77	4.33	0.21	0.42	0.72
Los Angeles-Riverside-Orange County, CA	154	4.20	0.23	0.44	0.72
Miami-Fort Lauderdale, FL	102	4.21	0.27	0.44	0.73
Minneapolis-St. Paul, MN	72	4.18	0.19	0.37	0.71
New York-Northern New Jersey-Long Island	97	4.17	0.25	0.47	0.75
Phoenix-Mesa, AZ	107	4.23	0.22	0.40	0.69
Portland-Salem, OR	87	4.22	0.21	0.39	0.71
San Diego, CA	113	4.20	0.21	0.40	0.66
San Francisco-Oakland-San Jose, CA	120	4.20	0.21	0.40	0.69
Seattle-Tacoma-Bremerton, WA	103	4.20	0.21	0.35	0.66
Tampa-St. Petersburg-Clearwater, FL	116	4.22	0.22	0.41	0.68
Washington-Baltimore, DC-MD-VA	114	4.18	0.21	0.40	0.68

Note - This table compares real-time interest rates for 30 year fixed rate conforming mortgages with a LTV=80, FICO=750, DTI=36, and with no points and fees. Column 1 shows the median number of lenders offering mortgages in a location on a specific day. Columns 3-5 show the difference between various percentiles of the offer distribution.

Table A-3: Dispersion in points and fees that lenders charge to originate at the median interest rate

	Percentile Differences		
	$75^{th} - 25^{th}$	$90^{th} - 10^{th}$	$99^{th} - 1^{st}$
Program			
Conforming	1.15	2.19	3.63
Super-Conforming	1.24	2.37	3.76
Jumbo	1.08	2.24	3.86
FHA	1.43	2.63	3.85
FICO			
640	1.23	2.39	3.82
680	1.19	2.32	3.76
720	1.16	2.26	3.76
750	1.17	2.27	3.75
LTV			
70	1.14	2.24	3.75
80	1.21	2.34	3.80
90	1.15	2.24	3.76
95	1.17	2.25	3.67
96	1.32	2.46	3.80

Note - This table compares real-time points and fees charged by different lenders to originate identical mortgages at the median interest rate. Points and fees are given as percent of the mortgage balance. The median interest rate is the same as in Table 1, and the average lender charges no points and fees at this interest rate.

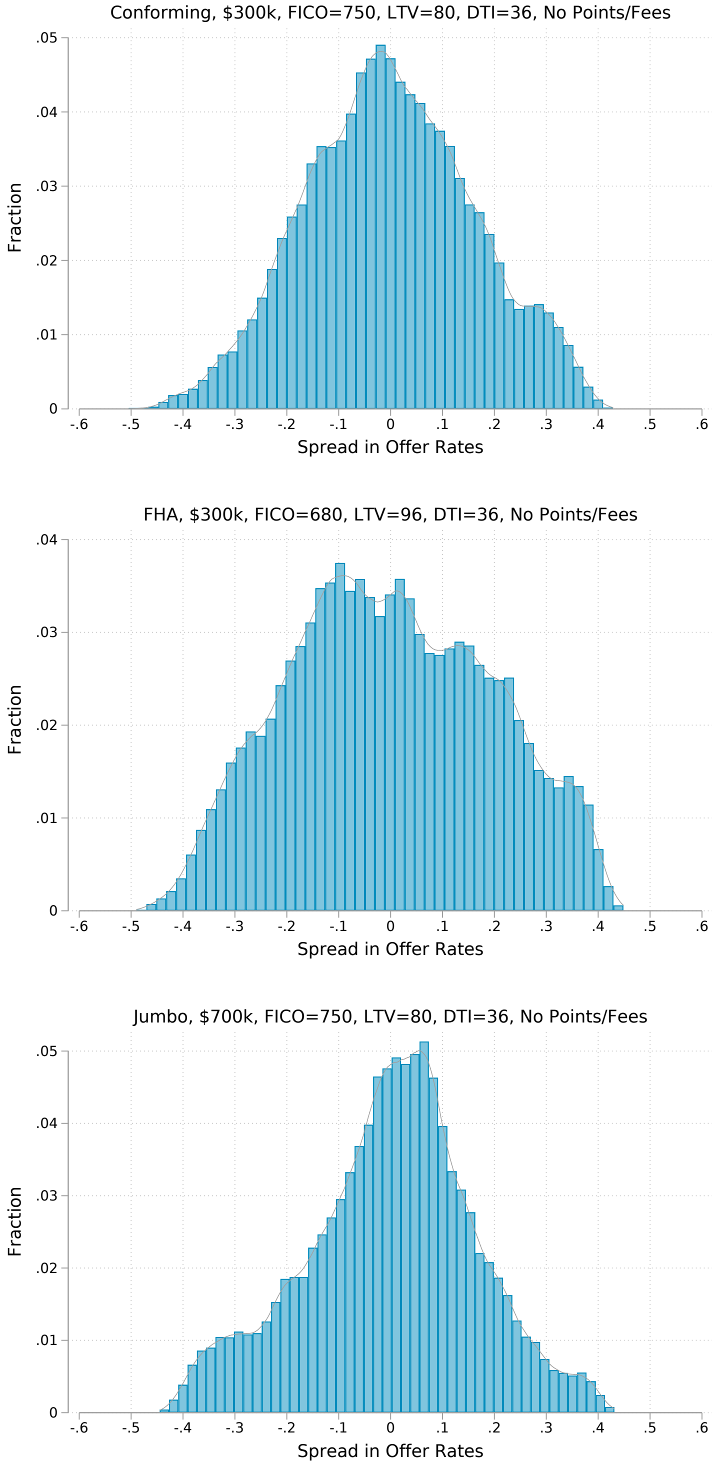

Figure A-1: Interest Rate Offer Dispersion for Identical Mortgages in Los Angeles

Note: The spread is defined as the difference between real-time mortgage rate offers and the median offered rate for identical mortgage products. The histogram includes daily data between April 2016 and August 2018.

Figure A-2: Interest Rate Offer Dispersion for Identical Mortgages in 20 Metropolitan Areas

Note: The spread is defined as the difference between real-time mortgage rate offers and the median offered rate for identical mortgage products. The histogram includes data between April 2016 and August 2018.

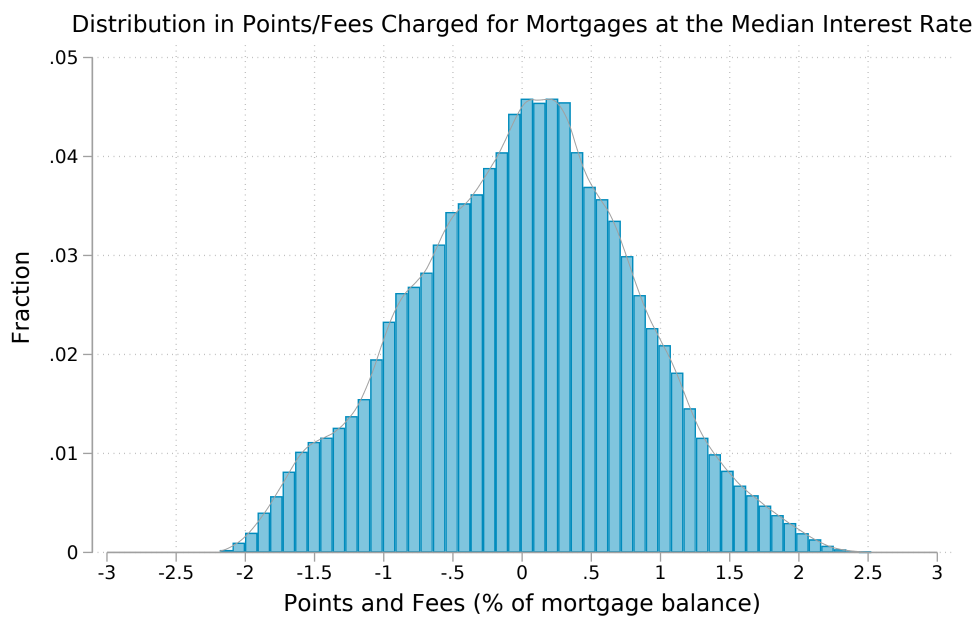

Figure A-3: Dispersion in Points and Fees Lenders Charge for Identical Mortgages at the Median Interest Rate

Note: Points and fees are given as percent of the mortgage balance. The median interest rate is calculated as the rate at which the average lender charges no points and fees.

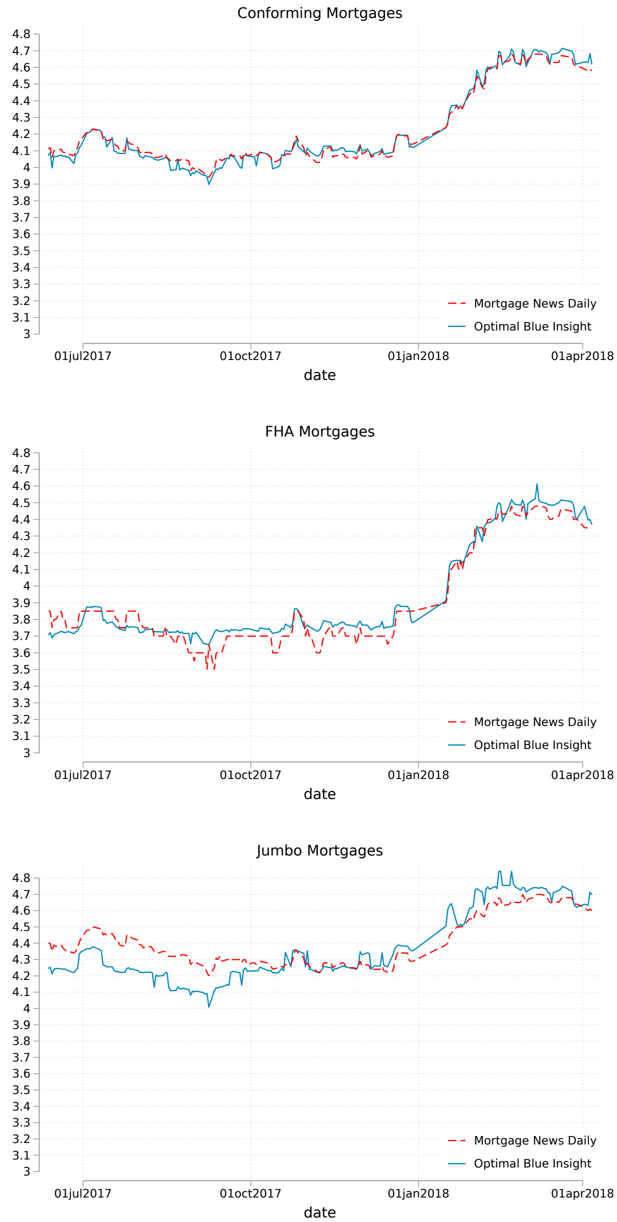

Figure A-4: Comparison of average offer rate from Optimal Blue with Mortgage News Daily data

Note: The Optimal Blue Data is for borrowers with LTV=80, FICO=750, DTI=36, with no points/fees. The Mortgage News Daily data is a survey that does not control for loan characteristics and points and fees. To make the data comparable we assume that MND data is quoted for 0.5% points and fees.