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

Neil Bhutta Andreas Fuster Aurel Hizmo†

March 21, 2019

Abstract

We document wide dispersion in the mortgage rates that households get, and assess the role of financial knowledge and shopping in the rates obtained. To study dispersion, we draw on new data from a mortgage industry pricing platform where we observe the "best" rates being offered by lenders, the mortgage rates actually obtained or "locked in" by consumers, and key rate covariates including discount points, rate-lock date, geographic location, and all underwriting information. Looking first at locked rates (conditional on covariates), we estimate a gap between the 10th and 90th percentile of over 50 basis points — equivalent to about $7,500 in upfront costs (points) for the average loan. Much of this dispersion occurs within lender, and low-FICO and low-wealth borrowers experience the most within-lender dispersion, suggesting an important role for price discrimination. Comparing locked rates to the median best offer rate for the same borrower in the same market on the same day, we find that this lock-offer spread is widest for low-FICO and low-wealth borrowers, implying that such borrowers pay more not just because of credit risk, but also because of less effective search and negotiation. However, this spread compresses when Treasury rates rise, suggesting that a rising level of borrowing costs encourages more search and negotiation. Finally, we turn to survey data for direct measures of financial knowledge and shopping. First, using the new National Survey of Mortgage Originations, we provide novel evidence that mortgage rates decline with mortgage knowledge and shopping; that knowledge and shopping increase with FICO score and income; and that shopping activity intensifies in higher interest rate environments. Second, using the Survey of Consumer Finances we document a strong negative relationship between mortgage rates and the Lusardi-Mitchell metric of financial literacy.

*We thank Jason Allen, Serafin Grundl, Katherine Guthrie, Michael Haliassos, John Mondragon, David Zhang, as well as seminar participants at the Federal Reserve Board, the Federal Reserve System Microeconomics meeting, the CEPR European Conference on Household Finance (Ortygia), the FDIC Consumer Research Conference and the Arizona State University 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 (NSMO), 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, in this paper we document a striking amount of variation in the prices consumers pay for mortgages, especially among borrowers who are not likely to be financially sophisticated.

Identifying dispersion in mortgage rates that arises from market inefficiencies is challenging. Even in an efficient market, we would expect to see some variation across consumers in their interest rates due to several factors such as differences in credit risk, day-to-day fluctuations in market rates, and heterogeneity in risk and time preferences that can affect borrowers' choices of various contract terms. To address this challenge, we draw on a unique source of data — 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 the terms being *offered* for specific mortgages in each market and each day, and data on the mortgages *locked*, or obtained, by consumers. The data on locked mortgages include key variables for evaluating mortgage pricing, including several that are unavailable in any other dataset, such as "discount points", exact time of rate lock (as opposed to the closing date), and the lock period (e.g. 30, 45, 60 days, etc.).

Turning first to the data on mortgage interest rate locks, we document a large amount of interest rate dispersion. We find that 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. Given the average point-rate trade-off in our data, 50 basis points is equivalent to paying about 3 points more at closing or $7,500 for a average loan of $250,000. Moreover, a substantial amount of dispersion exists even within lender, especially among borrowers who are likely to be the least financially sophisticated (low FICO, low wealth, low income, or unexperienced home buyers). 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.[2]

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

[2] Notably, including lender fixed effects arguably accounts for differences in lender quality that might have helped explain residual price dispersion.

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 on a given day. For a given consumer, we compute the difference between the rate they locked and the median of these best offer rates available to that consumer on the same day in the same market. 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.

More importantly, the locked-offered rate gap varies substantially across borrower types. For example, while borrowers getting the largest loans ("jumbo" borrowers) typically pay rates that are 25 basis points *below* the median of their offer distribution, FHA borrowers, who tend to have lower income, wealth, and credit scores, typically pay 25 basis points *more* than what the median lender could offer. Overall, our results imply that low-FICO and less-wealthy borrowers pay more for mortgages not simply because they present more credit risk, but also because they search and negotiate less effectively.

We also 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 locked interest rates falls. These results are somewhat stronger for low-FICO borrowers, who tend to overpay the most (relative to the median best offer). 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, likely contribute to these patterns.

Finally, in order to more directly assess the role of borrower search and sophistication on mortgage interest rates, we turn to two survey data sources. The first is the National Survey of Mortgage Originations (NSMO). The NSMO is a new dataset with a unique design that combines detailed administrative records on recent mortgage originations with survey data on the individuals who took out those mortgages. The survey component of the dataset asks several questions about borrowers' shopping behavior and their knowledge of mortgages and interest rates, while the administrative component provides many details on mortgage and underwriting characteristics (interest rate, LTV, FICO, etc.). Using these data, we provide novel evidence that knowledge and shopping are strongly related to lower mortgage rates, controlling for an array of credit risk variables. Moreover, we find that low-FICO and low-income borrowers get higher rates due to less shopping and mortgage knowledge. And, lastly, we show that shopping activity is elevated in higher interest rate environments, consistent with our conjecture that a rise in rates encourages people to shop more.

Second, we draw on the 2016 Survey of Consumer Finances (SCF). We take advantage of new questions aimed at gauging financial literacy that were added to the 2016 SCF. In particular, the SCF added the "big three" questions designed by Annamaria Lusardi and Olivia Mitchell to assess individuals' understanding of basic financial concepts related to saving, borrowing, and investing. We find that recent mortgage borrowers who answered all three questions correctly had mortgage rates that are about 25 basis points lower than borrowers who did not answer any questions correctly, even after including many controls for loan characteristics, credit risk, and demographics.

Overall, our empirical results provide evidence that a large fraction of the borrower population in the US seems to overpay for mortgages likely because of lack of shopping/negotiating and lack of financial sophistication. As most of these borrowers have government guaranteed loans, our results suggest that the GSEs and the FHA could consider policies aimed at limiting price dispersion. Our findings also suggest that search frictions are important for the pass-through of monetary policy to the mortgage market. When treasury rates increase, the amount of overpaying falls as borrowers search and negotiate more effectively, which has a dampening effect on average mortgage rates.

Given the data challenges mentioned earlier, existing work on price dispersion in the US mortgage market is rather sparse, especially relative to the importance of the market. Woodward and Hall (2012) use data on 1,500 FHA loans from 2001 and document wide dispersion in the fees paid to mortgage brokers and argue that this reflects suboptimal shopping and consumer confusion around discount points. Gurun et al. (2016) show substantial dispersion in the reset rates of privately-securitized adjustable-rate mortgages during the housing boom and find that these rates correlate positively with lenders' advertising expenditures. We build on their work by studying dispersion in interest rates, which are more salient than reset rates, on a broader swath of the mortgage market.

Perhaps closest to our work is the paper by Alexandrov and Koulayev (2017), who document substantial dispersion in offers based on lenders' rate sheets. We confirm wide dispersion in offer rates, and add to this an analysis of actual rate locks. Our rate-lock data allow us to document within-lender dispersion in transaction prices that widens for certain groups, and study how well different types of borrowers fare relative to available offers.[3] Also, 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. Finally, we provide novel direct evidence from the new NSMO data on how borrowers' mortgage knowledge and shopping relates to the interest rates they obtained.

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 the Optimal Blue data on rate locks and mortgage offers. 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

[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).

borrowers with different characteristics. Section 5 studies how these patterns evolve over time as market rates change. Section 6 introduces survey data from the NSMO and the SCF and presents direct evidence on the connection between shopping, mortgage knowledge, and interest rate outcomes. Finally, Section 7 discusses potential policy implications.

2 Optimal Blue 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 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 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]

[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.

[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

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. 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

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.

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 implies 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

[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.

median lender could offer for identical mortgages.[9] As shown in Table 4, the average locked-offered rate gap is +11 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 +25bp. This amounts to almost 2% of the mortgage balance in upfront points/fees, which for a typical FHA loan of $200k amounts 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 -6bp for super-conforming mortgages, and even more negative at -25bp 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 740 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 between 640 and 680 on average pay 22bp 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 14bp more than what the median lender could offer them, while repeat homebuyers pay only 7bp 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.

[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.

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:

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

where $rate_{imt}^{lock}$ is the interest rate locked by borrower i in market m on date t, and $\overline{rate}_{X_imt}^{offer}$ 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, or loan amount. 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 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 8bp lower for borrowers with a FICO score of 740 or above than for those with a FICO of between 640 and 680. 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 33bp lower than loans below $200k. Finally, locked-offered gaps tend to be slightly larger for first-time homebuyers.

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 8bp premium that we observe for an otherwise similar borrower with FICO below 680 relative to one with FICO above 740 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 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.

Next we investigate if these patterns also hold within loan program. In column (4) Table 5 we also include indicators for the loan program. Jumbo status is essentially collinear with loan amount, so that including program indicators makes it difficult to independently estimate the effect of loan amount. However, the other coefficients (e.g. on FICO or LTV bins) remain similar if program indicators are added to the regression. The patters we observed in Table 4 is also present here: Jumbo borrowers have the lowest premium in terms of LORG and FHA borrowers have the highest. In column (5) we restrict the sample to FHA loans only and we find similar patterns to the overall sample.

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.

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.[10] 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 Table 6. The first two columns add the 10-year Treasury yields as controls to

[10]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.

the same regression estimated in Table 5. The coefficient in column (2) implies that as the 10-year Treasury yield increases by 1 percentage point, the average locked-offered gap fall by about 12bp. This is sizable, given that we saw earlier that over our sample as a whole, the gap averaged 11bp with a standard deviation of 31bp.

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.[11] To study whether this is likely to be an important factor in the data, we separate borrowers into those with a DTI up to 36 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.[12] While we do not have a good individual measure of being subject to behavioral biases in our data, we use FICO score as 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 36 percent) and repeat the same regression. In column (4), which includes MSA, month, and lender fixed effects, we see that the relationship between treasury yield changes and 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

[11] 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.

[12] 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.

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 Shopping, Financial Knowledge, and Mortgage Rates

6.1 Evidence from the NSMO on the Effects of Shopping and Mortgage Knowledge

In this section, we use the National Survey of Mortgage Originations (NSMO) to document how different measures of borrower shopping and financial literacy (in particular knowledge about mortgages) correlate with the mortgage rate a borrower obtains. We also document which borrower types appear to overpay due to lack of shopping and knowledge, and how shopping effort varies with the level of market interest rates. In both cases, our findings align well with our earlier results.

The NSMO is a joint initiative of FHFA and CFPB as part of the "National Mortgage Database" program. It surveys a nationally representative sample of borrowers with newly originated closed-end first-lien residential mortgages in the US, focusing in particular on borrowers' experiences getting a mortgage, their perceptions of the mortgage market, and their future expectations. In November 2018, micro level data for the first 15 survey waves were for the first time made public on the FHFA website, covering originations from January 2013 to December 2016.[13] The NSMO contains a large number of questions, some of which were not asked in all waves, along with administrative information (from matched mortgage servicing and credit records) on borrower characteristics such as FICO credit score at the time of origination, or the spread between a loan's interest rate and the market mortgage interest rate.

The full NSMO dataset contains 24,847 loans. For our analysis, we impose a number of sample

[13] See https://www.fhfa.gov/DataTools/Downloads/Pages/National-Survey-of-Mortgage-Originations-Public-Use-File.aspx. We use the data version as of February 12, 2019.

restrictions. The main ones are that we only consider mortgages on a household's primary residence and drop mobile/manufactured homes as well as 2-4 unit dwellings. In addition, we require the loan term to be either 10, 15, 20, or 30 years, and drop construction loans or those obtained through a builder, mortgages with an associated additional lien, and those with more than two borrowers on the loan. Finally, we drop a few observations where the survey respondent was not a borrower on the loan. This leaves us with 19,906 mortgages for the analysis.

Our analysis in this section will proceed in three parts: first, we estimate the relationship between measures of borrower shopping or knowledge about the mortgage market and the rate they obtain on their loan, controlling for a rich set of borrower and loan characteristics. Second, we study which borrower and loan attributes correlate with lower rate spreads solely due to shopping and knowledge about the mortgage market. Third, we show that shopping effort increases when market interest rates are higher.

6.1.1 The Relationship between Shopping, Knowledge, and Contract Rates

We estimate OLS regressions of the form

$$RateSpread_{ijtw} = \beta X_i + \Gamma Z_{ij} + \alpha_t + \delta_w + \epsilon_{ijtw} \tag{2}$$

where $RateSpread_{ijtw}$ is the spread between the contract rate and the market mortgage rate prior to origination, for borrower i with loan characteristics j, loan origination month t and responding to survey wave w.[14] X_i are different measures of borrower i's shopping effort or knowledge about the mortgage market, as described below. Z_{ij} is a rich set of borrower and mortgage characteristics that could influence the pricing of the loan. The full list of controls is provided in the note to Table 8; it contains for instance flexible controls for FICO and LTV, loan term, program (e.g. GSE or FHA) and purpose (purchase or refinance), as well as borrower income, education, age, and race. We further include origination month fixed effects α_t and survey wave fixed effects δ_w (since there were a few small changes to the wording of questions across waves). In all our NSMO analyses, we use the provided analysis weights, which are based on sampling weights and non-response adjustments.

We consider the following X_i variables:

1. The answer to the question "How many different lenders/mortgage brokers did you seriously consider before choosing where to apply for this mortgage?" 49.5% of respondents answer 1, 35.0% 2, 12.7% 3, 1.8% 4, and 1.0% 5 or more. We combine the last three groups into "3+".

2. The answer to "How many different lenders/mortgage brokers did you end up applying to?" Here, 77.8% answer 1, 18.0% 2, 3.3% 3, 0.6% 4, and 0.3% 5 or more. We combine the last four groups into "2+".

[14] The market mortgage rate is measured through the Freddie Mac Primary Mortgage Market Survey (PMMS), lagged by two weeks relative to the time of loan origination. The gap is truncated at -1.5 and +1.5 percentage points. The interest rate of individual mortgages is not contained in the public dataset.

3. Those who indicated that they applied to two or more lenders are asked which of four non-exclusive reasons were driving the multiple applications. We create an indicator for those who indicate that "Searching for better loan terms" was a reason.

4. A series of questions are asked about nine different possible information sources the borrower could use to get information about mortgages or mortgage lenders. For each of them, a respondent can say they used a source "a lot", "a little", or "not at all". We use the following, which we think of as the best proxies for genuine search effort: "Other lenders or brokers" (32.5% a little, 9.2% a lot); "Websites that provide information on getting a mortgage" (31.5% a little, 20.1% a lot); and "Friends/relatives/co-workers" (30.0% a little, 13.5% a lot).

5. The answer to the question "When you began the process of getting this mortgage, how familiar were you (and any co-signers) with [t]he mortgage interest rates available at that time?" 63.2% respond "Very", 32.1% "Somewhat", and 4.7% "Not at all".

6. An index of "mortgage knowledge" based on 6 responses to the questions "How well could you explain to someone the... Process of taking out a mortgage / Difference between a fixed- and an adjustable-rate mortgage / Difference between a prime and subprime loan / Difference between a mortgage's interest rate and its APR / Amortization of a loan / Consequences of not making required mortgage payments". In each case, the respondent picked from a three point scale from "Not at all" (which we code as 1) to "Very" (3). We take the sum of the 6 responses and standardize it to have mean 0 and standard deviation 1.

7. An indicator for whether a borrower agreed with the statement "Most mortgage lenders would offer me roughly the same rates and fees." This question was only added in Wave 7 and so we only have responses for roughly half of the sample. Of those, 67.8% agree with the statement.

We think of the first four items as capturing shopping effort, while the remaining three capture mortgage market knowledge. We first add these measures to the regression one at a time, and then in a final specification jointly. The results are presented in Table 8. We see that most proxies for intense shopping and better mortgage market knowledge are associated with lower mortgage rates: for instance, considering 3+ lenders rather than just one lender is associated with a 8 basis point (bp) lower rate; while applying to more than one lender in search of better loan terms is associated with a 6bp lower rate. Similarly, more intense use of other lenders/brokers and the web as info sources predicts lower rates, while relying on friends, relatives and co-workers seems to have little effect. A particularly strong predictor is familiarity with available mortgage rates at the beginning of the process of getting the mortgage: those who state they were very familiar on average pay 16bp less than those who say they were not at all familiar. A one-standard-deviation higher value in the mortgage knowledge index is associated with an almost 5bp lower rate, while believing that all lenders offer roughly the same rate is associated with a higher rate.

The final column controls for all X_i jointly. As one might expect, some of the coefficients are attenuated relative to the earlier columns, but many of them remain individually significant, suggesting that there are different dimensions to shopping and knowledge that can contribute to a borrower obtaining a lower rate.[15] For instance, a borrower who is very familiar with market conditions may not need to consider more than one lender, if they can negotiate a good rate purely based on their knowledge. Again, it is important to remember that all of these regressions control finely for other factors that likely influence loan pricing, in order to rule out to the extent possible that these correlations reflect omitted variables that affect loan pricing due to default or prepayment risk.

6.1.2 Who Pays More Because of Lack of Shopping/Knowledge?

The previous subsection strongly suggests that more intense mortgage shopping and more knowledge about the mortgage market is associated with lower contracted rates. We next ask which observable borrower and loan characteristics are associated with stronger reported shopping intensity and mortgage knowledge, and as a result pay lower interest rates. To do this, we first isolate the part of the interest rate spread that can only attributed to shopping and knowledge about the mortgage market. Then, we study how this measure varies with observable characteristics.

We compute the predicted interest rate spread for each borrower specification using a regression almost identical to the one in specification (10) of Table 8. The only changes we make are that we omit the indicator for whether a borrower believed that most mortgage lenders would offer roughly the same rates and fees (since that question is only asked in later waves), and that instead of the "knowledge index" we use each of the six underlying questions individually. All shopping and knowledge variables are thus categorical, and for each of them we use as baseline/omitted value the one that corresponds to the lowest level of shopping or knowledge. We thus compute for each borrower the predicted rate difference relative to a hypothetical borrower that indicates that they did not engage in any shopping-related activities and have a poor understanding of the mortgage market.

We summarize this predicted rate difference in Table 9. Due to shopping/knowledge, the average borrower pays 27bp less than the hypothetical non-shopping, completely clueless borrower. Perhaps more interesting is the magnitude of the difference between the 10th and 90th percentile, which is 21bp. This implies that there are large differences across borrowers in the rate spread that can be explained by shopping behavior and mortgage knowledge.

Because shopping and knowledge about the mortgage market is correlated with various borrower and loan characteristics, the interest savings differ by group. Starting with loan characteristics, borrowers in the jumbo market pay on average about 6bp less in rates than FHA borrowers due to shopping/knowledge. The highest FICO borrowers pay 4bp less than low FICO borrowers, and high LTV borrowers pay more than low LTV borrowers. Also, borrowers with high loan amount pay lower rates than those with low loan amounts.

[15] It is interesting to note that the coefficient on "applied to 2+ lenders" flips sign if we simultaneously control for having applied to 2+ lenders in search of better loan terms. This likely reflects that those who applied to multiple lenders but not in search of better terms got turned down on their previous application (or learned negative news in the process).

Turning to borrower characteristics, borrowers with income of \$175k or higher pay 7bp lower interest rates than borrowers with income of less than \$35k due to their shopping and knowledge about the mortgage market. More educated borrowers on average pay less than their less educated counterparts, and first time homebuyers pay more than repeat homebuyers.

The magnitudes of the differences in the group-specific means may appear relatively small. However, it bears remembering that the right-hand-side variables of the underlying regression are responses to qualitative survey questions, which are likely measured with substantial individual-specific noise, leading to attenuation of the resulting coefficients.[16] Furthermore, we note that the cross-group differences in the 90th percentiles (the "worst" borrowers within each group) are often larger than the differences in means.

Overall, the findings here corroborate the mechanism we postulated in our previous analysis using the rate locks and offers data. All of the evidence points to the hypothesis that borrowers that are more likely to be less financially sophisticated pay more for mortgages for reasons that are unrelated to credit risk.

6.2 Time-series Variation in Shopping Intensity

Earlier, we saw that the locked-offered rate gap in the Optimal Blue data decreases when market interest rates are higher, even for borrowers who do not appear constrained, and speculated that this may partly be driven by increased shopping intensity when interest rates are higher. The NSMO enables us to test this hypothesis directly. We estimate linear probability models of the form:

$$Shopping_{ijtw} = \beta \cdot PMMS_{it} + \Gamma Z_{ij} + \delta_w + \epsilon_{ijtw} \tag{3}$$

where $Shopping_{ijtw}$ is a binary measure of shopping intensity (discussed below) by borrower i with loan characteristics j, loan origination month t and responding to survey wave w. $PMMS_{it}$ is our main variable of interest, the market mortgage rate two weeks prior to loan origination. Z_{ij} are borrower and mortgage characteristics, including the measures of borrowers' mortgage knowledge discussed above. Finally, δ_w are survey wave fixed effects.

As dependent variable, we use binary versions of the four main shopping variables that were associated with lower contract interest rates in Table 8: (i) whether a borrower seriously considered at least two lenders; (ii) whether a borrower applied to at least two lenders in search of better terms; (iii) whether a borrower used other lenders/brokers to get information "a little" or "a lot"; and (iv) whether a borrower used websites that provide information on getting a mortgage "a little" or "a lot". For each of these variables, we report regressions without other covariates (except for survey wave fixed effects) and with the same covariates as in Table 8, except for some variables that seem likely endogenous to the shopping effort itself.[17] Furthermore, we add the knowledge

[16] For instance, respondents likely differ in what they view as using an information source "a lot" vs. "a little", or being "very" vs. "somewhat" familiar with a topic.

[17] These variables are whether a borrower obtained their mortgage through a broker, the term of the loan, and whether it has an adjustable rate.

variables used in Table 8 as well.

Panel A of Table 10 reports the results of these regressions for the full sample. We see that across the different measures, a higher level of market mortgage rates is associated with more shopping effort, in most cases in statistically significant way. For instance, column (1) implies that a 1 percentage point increase in market mortgage rates increases the probability that a borrower considered more than one lender by 4.5 percentage points, relative to a sample average of 51 percent.[18] Column (2) shows that this coefficient is unaffected by the addition of fine borrower- and loan-level control variables, which alleviates concerns that the relationship is driven by variation in the type of borrower that applies at different points in time (and at different levels of market rates).

The effect on the probability of applying to multiple lenders is even substantially larger, especially compared to the sample mean (which is only 19 percent). A higher PMMS rate is also significantly associated with borrowers reporting that they obtained information from other lenders or brokers. The association with using websites to provide information on getting a mortgage is also positive, but not statistically significant.

Panels B to D assess the robustness of these findings in different subsamples. First, panel B shows that the estimated coefficients remain very similar if we restrict the sample to purchase mortgages; this should alleviate concerns that the finding is driven by changing composition between purchase and refinance mortgages as market rates change. Panels C and D then restrict the sample to borrowers that are objectively or subjectively unconstrained by payment-to-income constraints (which, if binding, could "force" borrowers to shop more). In panel C, we only use borrowers whose debt-to-income ratio ends up below 36 percent, suggesting that they had additional room to make larger payments. In panel D, we restrict the sample to those borrowers who responded "not at all" to the question "when you began the process of getting this mortgage, how concerned were you about qualifying for a mortgage?" In both of these subsamples, the estimated coefficients remain positive, and for the first two shopping measures statistically significant. Thus, it does not appear that the positive relationship between market interest rates and shopping is mainly driven by affordability constraints.

In the Appendix, we further complement this analysis by documenting univariate and multivariate correlations between the shopping and knowledge measures, as well as between these measures and various borrower and loan characteristics.

6.3 Evidence from the SCF on the Effects of Financial Literacy

In this section, we draw on data from the longstanding and widely-used Survey of Consumer Finances (SCF). The SCF is a triennial, nationally representative survey of households sponsored by the Federal Reserve Board that broadly covers U.S. families' financial circumstances. It collects detailed information on families' debts, assets, income, expenses, demographics, financial institutions, credit history, and financial decision-making. Notably, for the first time in 2016, the SCF added

[18] Over our sample period, the market mortgage rate as measured by PMMS varied from 3.31% to 4.58%.

three questions designed by Annamaria Lusardi and Olivia Mitchell to gauge individuals' general financial literacy.[19] The three questions assess understanding of basic concepts related to saving, borrowing, and investing:

1. Suppose you had $100 in a savings account and the interest rate was 2% per year. After 5 years, how much do you think you would have in the account if you left the money to grow: more than $102, exactly $102, or less than $102?

2. Imagine that the interest rate on your savings account was 1% per year and inflation was 2% per year. After 1 year, would you be able to buy more than today, exactly the same as today, or less than today with the money in this account?

3. Do you think that the following statement is true or false: buying a single company's stock usually provides a safer return than a stock mutual fund?

For each question, interviewees have the option to respond "do not know," or can refuse to answer. For each respondent, we compute the fraction of questions answered correctly, including "don't know" and "refuse" as not having answered correctly. Across all SCF respondents in 2016, 43% answered all three correctly, 36% answered two correctly, 16% answered one correctly, and 4% answered none correctly.[20]

For our analysis here, we focus on a subsample of SCF households that own their home and recently took out the mortgage on their home (either to refinance or to purchase the property) between 2013 and 2016. In this subsample, 56% answered all three financial literacy questions correctly, 31% answered two correctly, 11% answered one correctly, and 2% answered none correctly.

In Table 11, we provide estimates of the relationship between financial literacy and the interest rate respondents pay on their mortgage (interest rates are self-reported, and we subtract out the average prime rate for the month when the loan was taken out). Column 1 indicates that moving from none correct to getting all three questions correct is associated with a lower interest rate of 25 basis points. This magnitude is largely robust to adding controls. It drops a little in column 2 after controlling for credit history[21], loan characteristics, race, income, age, and education, but then rises back to about 25 basis points in column 3 after controlling for state fixed effects.

In addition to this measure of financial literacy, the SCF also asks respondents about how much they shop when trying to get a loan: "When making major decisions about borrowing money or obtaining credit, some people search for the very best terms while others don't. On a scale from zero to ten, where zero is no searching and ten is a great deal of searching, what number would you (and your husband/wife/partner) be on the scale?"

[19] A growing literature has explored the relationship between various financial outcomes and this and other metrics of financial literacy. For a review, see Lusardi and Mitchell (2014).

[20] Note that these statistics and all other results reported in this section use the SCF sampling weights to adjust for the sampling design of the SCF, which oversamples high wealth households.

[21] Unlike the NSMO, we do not observe credit scores in the SCF. However, we control for any late payment in the past year, bankruptcy in the last 4 years, and foreclosure in the last 5 years.

Table 11 shows how shopping relates to mortgage rates in the SCF, where we have divided the numerical responses by 10 so that the shopping variable ranges from zero to one. The results indicate that those who report shopping the most intensely have mortgage rates that are about 25 basis points lower than those who do no shopping. And, again, this result is robust to including a number of controls that help explain a considerable amount of the variation in reported rates.

In sum, data from the 2016 SCF are consistent with the message from the NSMO data: borrowers with higher financial knowledge and those who shop more tend to obtain better mortgage rates.

7 Discussion of Potential Policy Implications

Our empirical results provide evidence that many borrowers from the most vulnerable part of the borrower population in the US seem 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 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.[22] 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).

[22] 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).

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.

Lusardi, Annamaria and Olivia S. Mitchell, "The Economic Importance of Financial Literacy: Theory and Evidence," *Journal of Economic Literature*, March 2014, *52* (1), 5–44.

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	66,719	0.11	0.31	-0.07	0.25
Program					
Conforming	37,754	0.08	0.24	-0.06	0.19
Super-Conforming	7,337	-0.06	0.25	-0.21	0.07
Jumbo	2,061	-0.25	0.42	-0.39	-0.07
FHA	19,567	0.25	0.35	0.03	0.44
FICO					
640-679	12,680	0.22	0.36	0.00	0.42
680-719	16,256	0.14	0.33	-0.05	0.31
720-739	7,931	0.10	0.28	-0.07	0.24
740+	29,852	0.04	0.25	-0.10	0.16
LTV					
60-80	21,204	0.01	0.25	-0.12	0.14
81-90	10,480	0.03	0.30	-0.11	0.18
91-95	14,368	0.09	0.27	-0.07	0.22
96-97	20,667	0.25	0.34	0.03	0.43
First-Time Homebuyer					
No	33,635	0.07	0.28	-0.09	0.20
Yes	33,083	0.14	0.32	-0.05	0.30

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

					FHA Only
	(1)	(2)	(3)	(4)	(5)
FICO groups					
$I_{680 \leq FICO<720}$	-0.041***	-0.040***	-0.039***	-0.029***	-0.037***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)
$I_{720 \leq FICO<740}$	-0.057***	-0.050***	-0.049***	-0.031***	-0.056***
	(0.008)	(0.008)	(0.008)	(0.007)	(0.009)
$I_{FICO \geq 740}$	-0.082***	-0.070***	-0.067***	-0.042***	-0.065***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.012)
LTV groups					
$I_{80<LTV \leq 90}$	0.006	-0.001	-0.001	-0.007	-0.007
	(0.006)	(0.005)	(0.006)	(0.006)	(0.021)
$I_{90<LTV \leq 95}$	0.056***	0.048***	0.049***	0.036***	-0.003
	(0.007)	(0.006)	(0.006)	(0.006)	(0.017)
$I_{LTV>95}$	0.169***	0.153***	0.151***	0.109***	0.061***
	(0.012)	(0.011)	(0.012)	(0.008)	(0.018)
Loan Amount					
$I_{\$200k \leq Loan<\$400k}$	-0.063***	-0.059***	-0.054***	-0.057***	-0.066***
	(0.006)	(0.005)	(0.005)	(0.005)	(0.009)
$I_{\$400k \leq Loan<\$600k}$	-0.090***	-0.083***	-0.082***	-0.033***	-0.010
	(0.013)	(0.011)	(0.012)	(0.011)	(0.029)
$I_{\$600k \leq Loan<\$800k}$	-0.165***	-0.154***	-0.156***	-0.012	0.008
	(0.021)	(0.019)	(0.020)	(0.017)	(0.043)
$I_{Loan \geq \$800k}$	-0.334***	-0.305***	-0.330***	-0.049	
	(0.032)	(0.030)	(0.033)	(0.036)	
First Time Homebuyer	0.014**	0.010***	0.003	0.002	0.009
	(0.006)	(0.004)	(0.004)	(0.004)	(0.006)
Program					
Super-Conforming				-0.146***	
				(0.016)	
Jumbo				-0.303***	
				(0.033)	
FHA				0.037**	
				(0.017)	
MSA F.E.	Yes	Yes			
Month F.E.	Yes	Yes			
Lender F.E.		Yes			
Lender x MSA x Month F.E.			Yes	Yes	Yes
Observations	66718	66666	61520	61520	15841
R-squared	0.185	0.360	0.498	0.519	0.595

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: The Relationship Between the Lock-Offered Gap and Treasury Yields

	(1)	(2)	(3)	(4)
Treasury Yield	-0.086***	-0.117***		
	(0.010)	(0.017)		
Treasury Yield $\times$				
$FICO \leq 680$, $DTI \leq 36\%$			-0.103***	-0.134***
			(0.016)	(0.021)
$FICO \leq 680$, $DTI > 36\%$			-0.122***	-0.154***
			(0.017)	(0.020)
$FICO > 680$, $DTI \leq 36\%$			-0.067***	-0.099***
			(0.011)	(0.020)
$FICO > 680$, $DTI > 36\%$			-0.084***	-0.116***
			(0.011)	(0.017)
Borrower Controls	Yes	Yes	Yes	Yes
MSA F.E.	Yes	Yes	Yes	Yes
Lender F.E.	Yes	Yes	Yes	Yes
Month F.E.		Yes		Yes
Observations	66271	66271	66271	66271
R-squared	0.381	0.382	0.381	0.383

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. All specifications include controls for FICO, LTV, loan amount and loan program. The data covers mortgage rates for 20 metropolitan areas during the period between 2016-2019. We focus on 30 year, fixed rate, fully documented purchase mortgages. The standard errors are clustered at the month and lender level.

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

Dep. Var.:		$FICO \leq 680$	$FICO \leq 680$	$FICO > 680$	$FICO > 680$
ΔStd. Residual of	All Data	$DTI \leq .36$	$DTI > .36$	$DTI \leq .36$	$DTI > .36$
Locked Rate$_t$	(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.

Table 8: Relationship between mortgage rates and measures of shopping and knowledge. Dependent variable: spread between a borrower's mortgage interest rate and the market mortgage rate prior to origination, in percentage points (censored at -1.5 and +1.5)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Seriously considered 2 lenders	-0.039***									-0.020**
	(0.009)									(0.010)
Seriously considered 3+ lenders	-0.078***									-0.043***
	(0.012)									(0.014)
Applied to 2+ lenders		-0.039***								0.051**
		(0.010)								(0.022)
Applied to 2+ lenders in search of better loan terms			-0.057***							-0.079***
			(0.010)							(0.023)
Used other lenders/brokers to get info? A little				-0.024***						0.000
				(0.009)						(0.010)
Used other lenders/brokers to get info? A lot				-0.061***						-0.021
				(0.014)						(0.015)
Used web to get info? A little					-0.042***					-0.033***
					(0.010)					(0.010)
Used web to get info? A lot					-0.065***					-0.036***
					(0.011)					(0.012)
Used friends/relatives to get info? A little						-0.001				0.001
						(0.010)				(0.010)
Used friends/relatives to get info? A lot						0.010				0.013
						(0.013)				(0.013)
Familiar with mortgage rates? Somewhat							-0.075***			-0.060**
							(0.024)			(0.024)
Familiar with mortgage rates? Very							-0.159***			-0.122***
							(0.024)			(0.024)
Index of mortgage knowledge (Std)								-0.046***		-0.031***
								(0.005)		(0.005)
Most lenders offer same rate? Yes									0.033***	0.026**
									(0.012)	(0.012)
Adj. R2	0.18	0.18	0.18	0.18	0.18	0.17	0.18	0.18	0.18	0.19
Obs.	19906	19906	19906	19906	19906	19906	19906	19906	19906	19906

Sample restricted to first-lien loans (without a junior lien) for single-family principal residence properties, with no more than two borrowers, and a loan term of 10, 15, 20 or 30 years. Observations weighted by NSMO sample weights. All regressions control for origination month fixed effects, survey wave fixed effects, FICO score (linear term plus dummies for 11 FICO bins), LTV (linear term plus dummies for each percentage point from 79-98), indicators for loan purpose (purchase, refinance, or cash-out refinance), 9 loan amount categories, loan program (Freddie, Fannie, FHA, VA, FSA/RHS, other), loan term, first-time homebuyer status, single borrowers, using a mortgage broker, whether the loan has an adjustable rate, jumbo status, 6 borrower income categories, 6 borrower education categories, whether the household owns 4 different types of financial assets, race and ethnicity, metropolitan CRA low-to-moderate income tract status, borrower age and gender, and self-assessed creditworthiness, likelihood of moving, selling, or refinancing, and risk aversion. Robust standard errors in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 9: Summary statistics of the interest rate spread that can be attributed to shopping and mortgage knowledge

	Observations	Mean	Percentiles	
			10^{th}	90^{th}
All Mortgages	19,906	-0.27	-0.37	-0.16
Program				
Conforming	11,103	-0.28	-0.37	-0.17
Jumbo	679	-0.31	-0.38	-0.22
FHA	2,734	-0.25	-0.35	-0.13
FICO				
$\leq$ 600	411	-0.24	-0.35	-0.14
601-640	1,089	-0.25	-0.35	-0.14
641-680	2,195	-0.26	-0.36	-0.15
681-740	4,784	-0.26	-0.36	-0.16
$>$ 740	11,427	-0.28	-0.37	-0.18
LTV				
$\leq$ 75	8,216	-0.28	-0.37	-0.18
76-80	3,551	-0.28	-0.37	-0.18
81-95	4,551	-0.27	-0.37	-0.16
96-97	1,805	-0.24	-0.35	-0.12
Loan Amount				
<100k	3,011	-0.25	-0.35	-0.13
[100k, 200k)	7,736	-0.27	-0.36	-0.16
[200k, 300k)	4,656	-0.28	-0.37	-0.17
[300k, 400k)	2,405	-0.29	-0.38	-0.19
$\geq$ 400k	2,098	-0.30	-0.38	-0.21
First-Time Homebuyer				
No	16,717	-0.28	-0.37	-0.18
Yes	3,189	-0.24	-0.36	-0.12
Income				
<35k	1,189	-0.23	-0.34	-0.11
[35k, 75k)	6,014	-0.25	-0.35	-0.14
[75k, 175k)	9,752	-0.28	-0.37	-0.18
$\geq$ 175k	2,951	-0.30	-0.38	-0.21
Education				
Less than college	3,322	-0.24	-0.34	-0.13
Some college	3,975	-0.26	-0.36	-0.16
College grad	7,017	-0.28	-0.37	-0.17
Postgrad	5,592	-0.29	-0.38	-0.19

The variable we are summarizing here is the interest rate spread that is only due to shopping and knowledge about the mortgage market. We compute the predicted value of the interest rate spread using only the displayed variables on shopping behavior and knowledge about mortgages, in a way similar to specification (10) of Table 8. (see text for details)

Table 10: Relationship between various binary measures of mortgage shopping and mortage market interest rates (PMMS).

	Considered 2+ lenders		Applied to 2+ lenders for better terms		Used other lenders to get info		Used web to get info	
A. Full sample	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PMMS rate	0.045**	0.045**	0.069***	0.062***	0.048***	0.050***	0.019	0.026
	(0.018)	(0.018)	(0.014)	(0.014)	(0.018)	(0.018)	(0.018)	(0.018)
Controls?	No	Yes	No	Yes	No	Yes	No	Yes
Mean of Dependent Variable	0.510	0.510	0.190	0.190	0.418	0.418	0.533	0.533
Obs.	19906	19906	19906	19906	19906	19906	19906	19906
B. Purchase loans								
PMMS rate	0.060**	0.054*	0.077***	0.072***	0.049*	0.041	0.014	0.009
	(0.029)	(0.029)	(0.024)	(0.024)	(0.029)	(0.028)	(0.029)	(0.027)
Controls?	No	Yes	No	Yes	No	Yes	No	Yes
Mean of Dependent Variable	0.534	0.534	0.223	0.223	0.430	0.430	0.550	0.550
Obs.	9254	9254	9254	9254	9254	9254	9254	9254
C. DTI ≤ 36								
PMMS rate	0.039	0.045*	0.081***	0.074***	0.029	0.036	0.003	0.016
	(0.025)	(0.025)	(0.019)	(0.019)	(0.025)	(0.025)	(0.025)	(0.024)
Controls?	No	Yes	No	Yes	No	Yes	No	Yes
Mean of Dependent Variable	0.503	0.503	0.176	0.176	0.411	0.411	0.541	0.541
Obs.	10590	10590	10590	10590	10590	10590	10590	10590
D. Not concerned about qualif.								
PMMS rate	0.041*	0.045*	0.060***	0.052***	0.023	0.031	-0.005	0.014
	(0.024)	(0.024)	(0.018)	(0.017)	(0.024)	(0.023)	(0.024)	(0.023)
Controls?	No	Yes	No	Yes	No	Yes	No	Yes
Mean of Dependent Variable	0.488	0.488	0.165	0.165	0.387	0.387	0.499	0.499
Obs.	11203	11203	11203	11203	11203	11203	11203	11203

Sample restricted to first-lien loans (without a junior lien) for single-family principal residence properties, with no more than two borrowers, and a loan term of 10, 15, 20 or 30 years. All four dependent variables are binary. All regressions control for survey wave fixed effects and use NSMO analysis weights. The multivariate regressions (even columns) further control for FICO score (linear term plus dummies for 11 FICO bins), LTV (linear term plus dummies for each percentage point from 79-98), indicators for loan purpose (purchase, refinance, or cash-out refinance), 9 loan amount categories, loan program (Freddie, Fannie, FHA, VA, FSA/RHS, other), first-time homebuyer status, single borrowers, jumbo status, 6 borrower income categories, 6 borrower education categories, whether the household owns 4 different types of financial assets, race and ethnicity, metropolitan CRA low-to-moderate income tract status, borrower age and gender, and self-assessed creditworthiness, likelihood of moving, selling, or refinancing, and risk aversion. Robust standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table 11: Relationship between Interest Rate Spreads and Measures of Financial Literacy and Shopping in the Survey of Counsumer Finances

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Financial Literacy (Fraction Correct)	-0.247 *	-0.198 *	-0.245 *				-0.243 *
	(0.110)	(0.095)	(0.096)				(0.099)
Shops Around for Credit				-0.262 **	-0.262 **	-0.236 **	-0.228 *
				(0.090)	(0.086)	(0.086)	(0.089)
Loan Characteristics		Yes	Yes		Yes	Yes	Yes
Borrower Characteristics		Yes	Yes		Yes	Yes	Yes
State Fixed Effects			Yes			Yes	Yes
Observations	820	816	816	821	817	817	816
R-squared	0.011	0.15	0.225	0.009	0.151	0.222	0.229

Data source: 2016 Survey of Consumer Finances (SCF)

Notes: Sample comprised of households that took out a 15 year or 30 year fixed-rate home purchase or refinance mortgage in 2013-2016 for their principal residence. Outcome variable is the interest rate (self-reported) on the first lien mortgage relative to the average Freddie Mac PMMS prime rate for a loan of the same term in the month the mortgage was taken out. The Financial Literacy variable refers to the fraction correct on three questions designed by Lusardi and Mitchell and asked in the 2016 SCF. The Shopping Around variable is a self-reported value between 0 and 10 gauging the degree to which respondents shop for credit; we divide responses by 10 so that the range is 0 to 1. The loan characteristics we control for in specifications (2), (3) and (5)-(7) include loan program, loan term, and loan purpose (purchase, refinance or cash out). Borrower controls include indicators of whether they were late on any payment in the past year, had a bankruptcy in the last 4 years, had a foreclosure in the last 5 years, as well as controls for income, education, age and race/ethnicity.

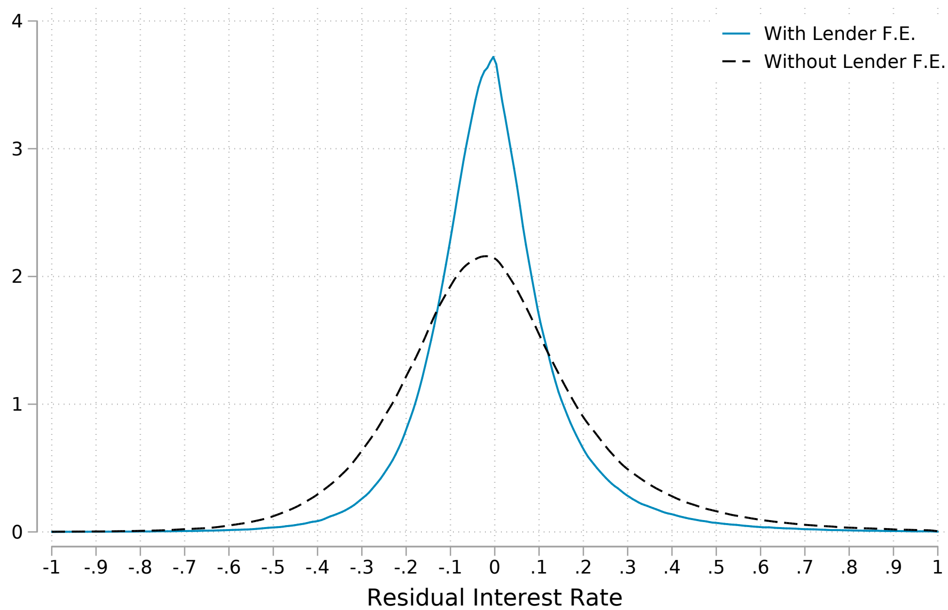

Figure 1: Residualized Locked Mortgage Rates Controlling for Borrower and Loan Characteristics, and Allowing for Differential Pricing of Loans by Lender-Location-Loan Characteristics-Time

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

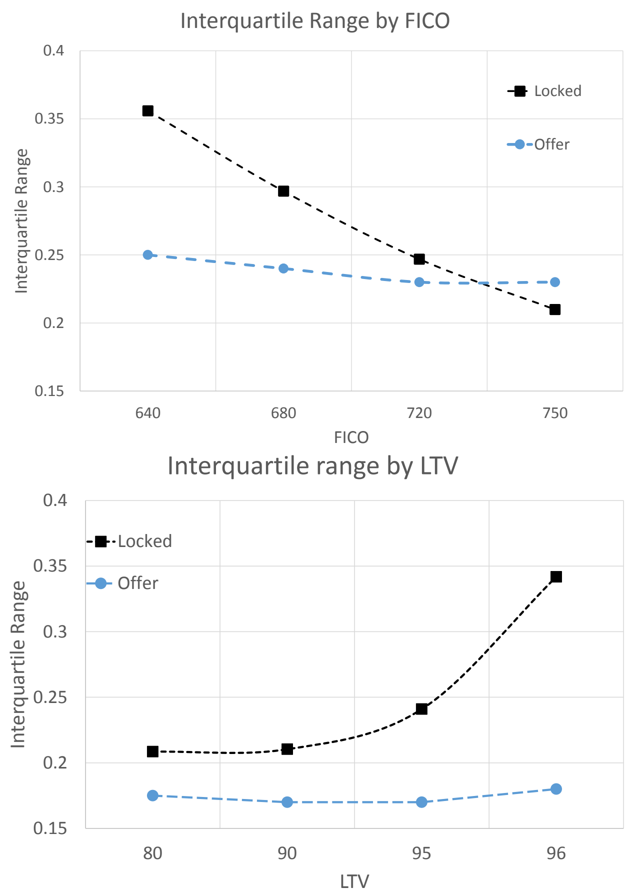

Figure 2: Comparing the Mortgage Rate Dispersion in Offer and Rate Lock Data

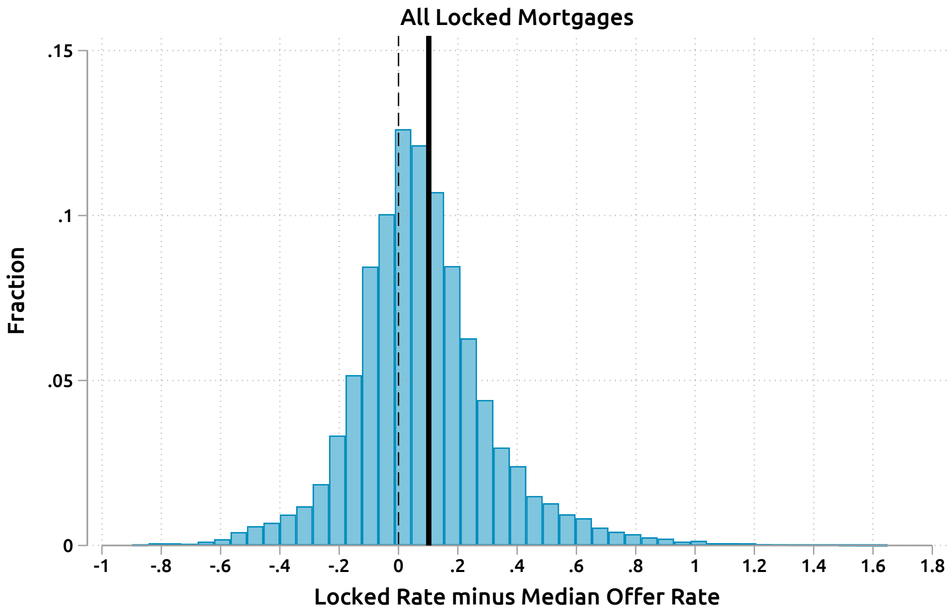

Figure 3: Distribution of Rate Locked Minus the Median Best Offered Rate for Identical Mortgages

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.

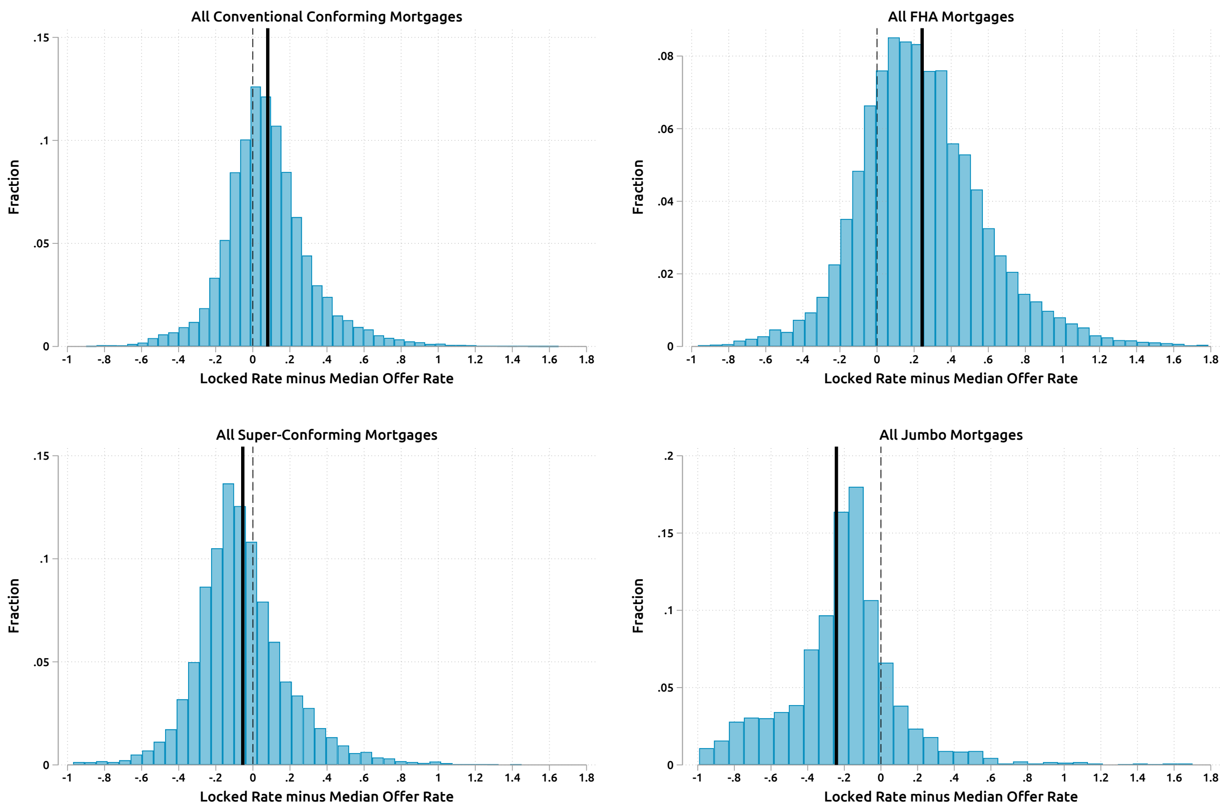

Figure 4: Distribution of Rate Locked Minus the Average Offered Rate for Identical Mortgages

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.

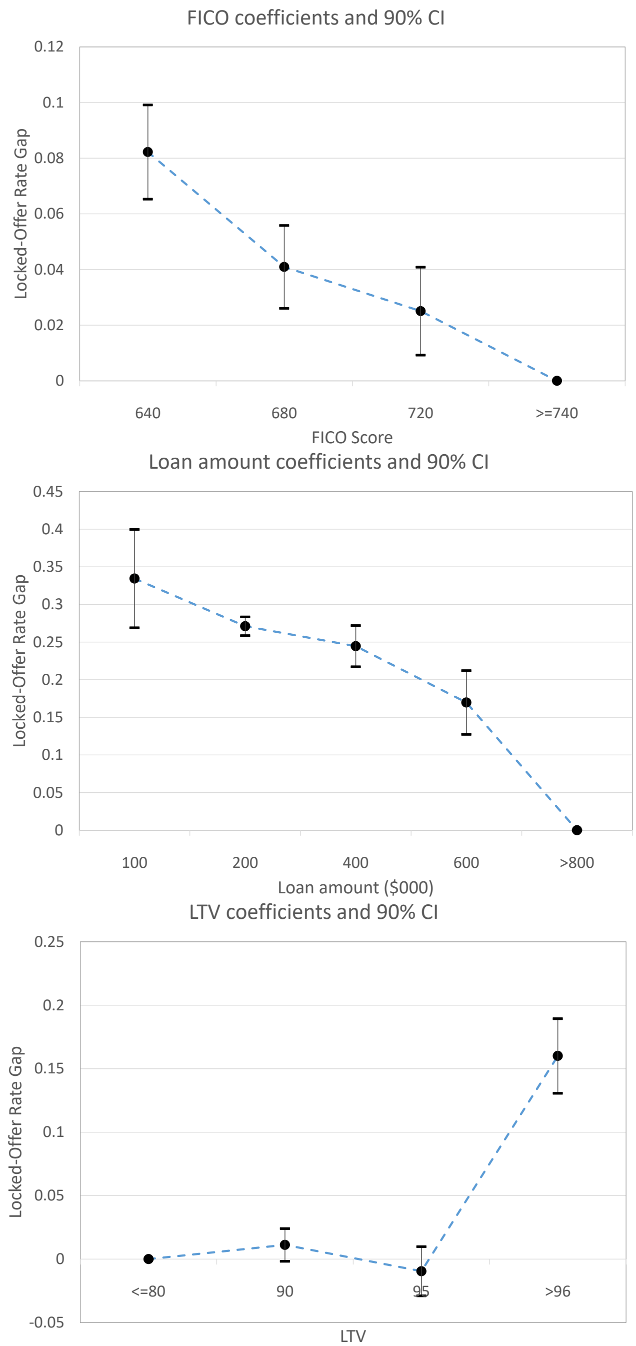

Figure 5: The Effects of Observables on the Locked-Offered Rate Gap

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

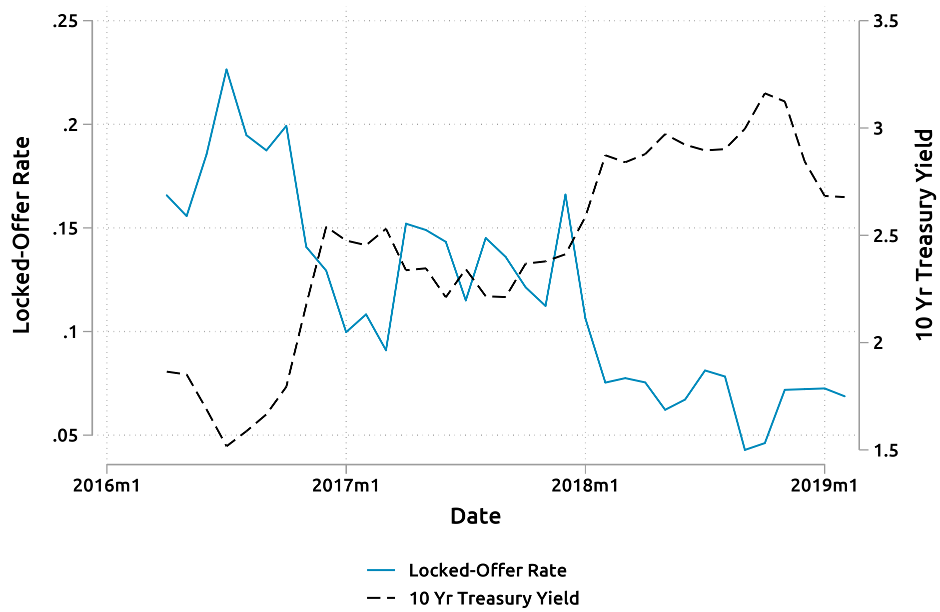

Figure 6: The Evolution of Rate Locked Minus the Average Offered Rate and Treasury Yields

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 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 year 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.

A.2 Correlates of Shopping Intensity and Knowledge

Section 6 strongly suggests that more intense mortgage shopping and better knowledge of the mortgage market are associated with lower contracted rates. In this appendix, we document how different shopping and knowledge measures are correlated with one another, and also study which observable borrower and loan characteristics are associated with stronger reported shopping intensity and higher knowledge.

In Table A-4, we report results from regressions of the four binary shopping measures already used in Section 6.2 on the three mortgage knowledge measures introduced in Section 6.1.1, as well as various other loan and borrower characteristics, most of which we turn into binary variables for ease of interpretation. We run regressions with one covariate of interest at a time (with survey wave fixed effects as the only additional control), or controlling for all of them jointly and further controlling for other factors that may also affect shopping intensity (for instance, a stronger expectation of selling the property soon). The former type of regression is called "univar." in Table A-4 while the latter type is called "multivar."

In Table A-5, we report similar regressions but with the knowledge measures as dependent variables (and only the borrower and loan characteristics as independent variables). Note that for the first two of the three outcomes in that table, higher values correspond to more knowledge, while for the last one, the opposite is true. We discuss the results from both tables jointly, since in some cases they contrast in interesting ways.

The first three rows of Table A-4 indicate that borrowers that are more knowledgable also shop more. Of course, in this case it is difficult to rule out reverse causality, namely that the additional shopping made them more knowledgable (for instance, about price differences across lenders). The fourth coefficient shows that people who say that they were "not at all concerned about qualifying for a mortgage when they began the process of getting this mortgage" also report shopping less.[1] This suggests that less confidence in one's ability to qualify for a loan can have the beneficial side effect of inducing additional shopping.

Next, we reproduce the positive relationship between PMMS and shopping measures documented in Table 10.[2] We further see that mortgage knowledge tends to be slightly lower when PMMS is higher, although the relationship is no longer significant once other variables are controlled for.

Turning to borrower and loan characteristics, we see that borrowers with higher FICO scores are more likely to have seriously considered more than one lender, although for the other shopping measures the evidence is more mixed. However, high-FICO borrowers tend to be substantially more knowledgable, especially when considering the univariate correlations with mortgage-rate familiarity and the knowledge index. There is no significant relation between FICO and the propensity to think that all lenders offer similar terms.

Borrowers with higher LTVs tend to shop more, but are less knowledgable. Similarly, FHA borrowers do not appear to shop less, but tend to be significantly less knowledgable than other borrowers (except that they do have a slightly higher propensity to believe in price dispersion). Given that our earlier Optimal Blue analysis found that these groups see substantially higher locked-offered rate gaps, these patterns suggest that knowledge may be the key differential driver of those patterns. Similarly, we also see that borrowers with purchase loans, and especially first-time homebuyers, report higher shopping intensity, but are substantially less knowledgable than refinancers (which makes sense, since the latter likely have more experience with the process). Borrowers with larger loan amounts, and especially jumbo borrowers, both shop more and are more knowledgable – in line with their lower rate spreads.

Finally, in terms of borrower demographics, more educated respondents are much more likely to shop, and have better mortgage knowledge. Income appears to have little effect on shopping once other factors are controlled for, but still correlates significantly with knowledge. Finally, we see that minorities appear to shop more that Non-Hispanic White borrowers (the omitted category), but were less familiar with mortgage rates and have a lower knowledge index. However, they are more likely to believe in price dispersion.

[1] This self-assessed creditworthiness was also used as a control variable in Table 8.

[2] The coefficients differ slightly because in this section, we use less fine control variables.

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 No. Offers	Median Rate	Percentile Differences $75^{th} - 25^{th}$	$90^{th} - 10^{th}$	$99^{th} - 1^{st}$
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.

Table A-4: Relationship between various binary measures of mortgage shopping and characteristics of borrower and loan.

	Considered 2+ lenders		Applied to 2+ lenders for better terms		Used other lenders to get info		Used web to get info	
	Univar. (1)	Multivar. (2)	Univar. (3)	Multivar. (4)	Univar. (5)	Multivar. (6)	Univar. (7)	Multivar. (8)
Very familiar with mortgage rates	0.057***	0.045***	-0.007	0.009	0.022***	0.010	0.003	0.008
	(0.008)	(0.009)	(0.007)	(0.007)	(0.008)	(0.009)	(0.008)	(0.009)
Index of mortgage knowledge (Std)	0.047***	0.033***	0.005*	0.006	0.030***	0.022***	0.038***	0.041***
	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Most lenders offer same rate? Yes	-0.085***	-0.076***	-0.052***	-0.051***	-0.071***	-0.063***	-0.017	-0.014
	(0.012)	(0.012)	(0.010)	(0.010)	(0.012)	(0.012)	(0.012)	(0.011)
Not concerned about qualifying for mtg.	-0.047***	-0.076***	-0.052***	-0.044***	-0.068***	-0.095***	-0.073***	-0.092***
	(0.008)	(0.009)	(0.006)	(0.007)	(0.008)	(0.009)	(0.008)	(0.009)
Market mortgage rate (PMMS)	0.045**	0.046**	0.069***	0.063***	0.048***	0.050***	0.019	0.027
	(0.018)	(0.018)	(0.014)	(0.014)	(0.018)	(0.018)	(0.018)	(0.018)
FICO/100	0.015**	0.017**	-0.015***	-0.002	0.008	0.013*	-0.005	0.017**
	(0.006)	(0.007)	(0.005)	(0.006)	(0.006)	(0.007)	(0.006)	(0.007)
LTV/100	0.051**	0.007	0.130***	0.052***	0.049**	0.045*	0.187***	0.088***
	(0.020)	(0.025)	(0.015)	(0.019)	(0.020)	(0.025)	(0.020)	(0.025)
Loan amount > 200k	0.081***	0.034***	0.029***	0.018**	0.083***	0.049***	0.061***	0.009
	(0.008)	(0.009)	(0.006)	(0.008)	(0.008)	(0.009)	(0.008)	(0.009)
Jumbo	0.116***	0.042**	0.017	0.000	0.116***	0.047**	-0.018	-0.073***
	(0.020)	(0.020)	(0.016)	(0.017)	(0.020)	(0.021)	(0.020)	(0.020)
FHA	-0.004	-0.000	0.031***	-0.007	-0.010	-0.005	0.031***	0.005
	(0.011)	(0.013)	(0.010)	(0.011)	(0.011)	(0.013)	(0.011)	(0.013)
VA/FSA	-0.005	0.002	0.005	-0.013	0.009	0.014	0.003	0.019
	(0.012)	(0.014)	(0.010)	(0.011)	(0.012)	(0.014)	(0.012)	(0.014)
Purpose = home purchase	0.045***	0.037***	0.058***	0.041***	0.023***	0.013	0.030***	-0.064***
	(0.008)	(0.010)	(0.006)	(0.008)	(0.008)	(0.010)	(0.008)	(0.010)
First-time homebuyer	0.048***	0.023*	0.067***	0.016	0.019*	0.003	0.148***	0.110***
	(0.011)	(0.013)	(0.009)	(0.012)	(0.011)	(0.013)	(0.010)	(0.013)
At least college degree	0.087***	0.053***	0.028***	0.018**	0.076***	0.053***	0.133***	0.090***
	(0.008)	(0.009)	(0.006)	(0.007)	(0.008)	(0.009)	(0.008)	(0.009)
Household income > 100k	0.060***	0.004	0.004	-0.010	0.050***	-0.000	0.057***	0.015
	(0.008)	(0.010)	(0.006)	(0.008)	(0.008)	(0.010)	(0.008)	(0.010)
White Hispanic	0.033**	0.032**	0.057***	0.042***	0.014	0.010	0.052***	0.043***
	(0.016)	(0.016)	(0.013)	(0.014)	(0.016)	(0.016)	(0.016)	(0.016)
Black	0.061***	0.067***	0.060***	0.052***	-0.000	-0.010	0.055***	0.052***
	(0.017)	(0.017)	(0.014)	(0.015)	(0.016)	(0.017)	(0.017)	(0.016)
Asian	0.115***	0.061***	0.030**	0.003	0.119***	0.071***	0.149***	0.088***
	(0.017)	(0.017)	(0.014)	(0.015)	(0.017)	(0.017)	(0.016)	(0.016)
Other race	0.063***	0.055**	0.046**	0.034*	0.038	0.028	0.057**	0.041*
	(0.024)	(0.024)	(0.020)	(0.020)	(0.024)	(0.024)	(0.024)	(0.022)
Mean of Dependent Variable		0.510		0.190		0.418		0.533
Adj. R2		0.04		0.03		0.03		0.07
Obs.		19906		19906		19906		19906

Sample restricted to first-lien loans (without a junior lien) for single-family principal residence properties, with no more than two borrowers, and a loan term of 10, 15, 20 or 30 years. All four dependent variables are binary. Observations weighted by NSMO sample weights. The univariate regressions (odd columns) only feature one of the covariates in the table, along with survey wave fixed effects. The multivariate regressions (even columns) simultaneously control for all the variables listed in the table, survey wave fixed effects, and the following additonal variables: indicators for single borrowers, cash-out refinances, whether the household owns 4 different types of financial assets, metropolitan CRA low-to-moderate income tract status, borrower age and gender, and self-assessed likelihood of moving, selling, or refinancing, as well as risk aversion. Robust standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table A-5: Relationship between various measures of mortgage knowledge and characteristics of borrower and loan.

	Very familiar with mortgage rates		Knowledge Index (std)		Thinks all lenders offer same terms	
	Univar.	Multivar.	Univar.	Multivar.	Univar.	Multivar.
	(1)	(2)	(3)	(4)	(5)	(6)
Market mortgage rate (PMMS)	-0.061***	-0.022	-0.074**	-0.003	-0.017	-0.024
	(0.018)	(0.017)	(0.037)	(0.034)	(0.040)	(0.039)
FICO/100	0.113***	0.046***	0.179***	0.018	0.002	0.001
	(0.006)	(0.007)	(0.013)	(0.014)	(0.008)	(0.009)
LTV/100	-0.398***	-0.049**	-0.658***	-0.096**	0.117***	0.081**
	(0.019)	(0.023)	(0.041)	(0.049)	(0.027)	(0.035)
Loan amount > 200k	0.117***	0.023***	0.331***	0.079***	-0.020**	-0.015
	(0.008)	(0.009)	(0.016)	(0.018)	(0.010)	(0.012)
Jumbo	0.173***	0.023	0.501***	0.103***	-0.124***	-0.121***
	(0.017)	(0.017)	(0.037)	(0.037)	(0.027)	(0.028)
FHA	-0.189***	-0.031**	-0.344***	-0.063**	-0.022	-0.040**
	(0.011)	(0.013)	(0.023)	(0.025)	(0.015)	(0.017)
VA/FSA	-0.055***	0.001	-0.116***	-0.047*	0.021	0.009
	(0.012)	(0.013)	(0.025)	(0.026)	(0.015)	(0.017)
Purpose = home purchase	-0.168***	-0.051***	-0.181***	0.009	0.044***	0.043***
	(0.008)	(0.009)	(0.016)	(0.019)	(0.010)	(0.014)
First-time homebuyer	-0.322***	-0.206***	-0.413***	-0.156***	0.012	-0.043**
	(0.010)	(0.013)	(0.021)	(0.025)	(0.014)	(0.017)
At least college degree	0.067***	0.014*	0.285***	0.147***	0.006	-0.000
	(0.008)	(0.008)	(0.016)	(0.017)	(0.011)	(0.012)
Household income > 100k	0.180***	0.067***	0.457***	0.174***	-0.010	0.001
	(0.008)	(0.009)	(0.015)	(0.018)	(0.010)	(0.013)
White Hispanic	-0.104***	-0.021	-0.224***	-0.061**	-0.075***	-0.066***
	(0.016)	(0.015)	(0.032)	(0.030)	(0.021)	(0.021)
Black	-0.102***	-0.027	-0.074**	0.059*	-0.131***	-0.116***
	(0.017)	(0.017)	(0.032)	(0.032)	(0.022)	(0.023)
Asian	-0.042**	-0.070***	-0.086**	-0.230***	-0.102***	-0.079***
	(0.017)	(0.016)	(0.035)	(0.034)	(0.022)	(0.023)
Other race	-0.076***	-0.029	-0.070	-0.004	-0.115***	-0.110***
	(0.024)	(0.023)	(0.051)	(0.048)	(0.033)	(0.032)
Mean of Dependent Variable		0.617		-0.025		0.682
Adj. R2		0.14		0.16		0.02
Obs.		19906		19906		10275

Sample restricted to first-lien loans (without a junior lien) for single-family principal residence properties, with no more than two borrowers, and a loan term of 10, 15, 20 or 30 years. The dependent variables are binary except in columns (3)-(4), where the knowledge index is standardized to have mean 0 and standard deviation 1 (in unweighted sample). Observations weighted by NSMO sample weights. The univariate regressions (odd columns) only feature one of the covariates in the table, along with survey wave fixed effects. The multivariate regressions (even columns) simultaneously control for all the variables listed in the table, survey wave fixed effects, and the following additonal variables: indicators for single borrowers, cash-out refinances, whether the household owns 4 different types of financial assets, metropolitan CRA low-to-moderate income tract status, borrower age and gender, and self-assessed likelihood of moving, selling, or refinancing, as well as risk aversion. Robust standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

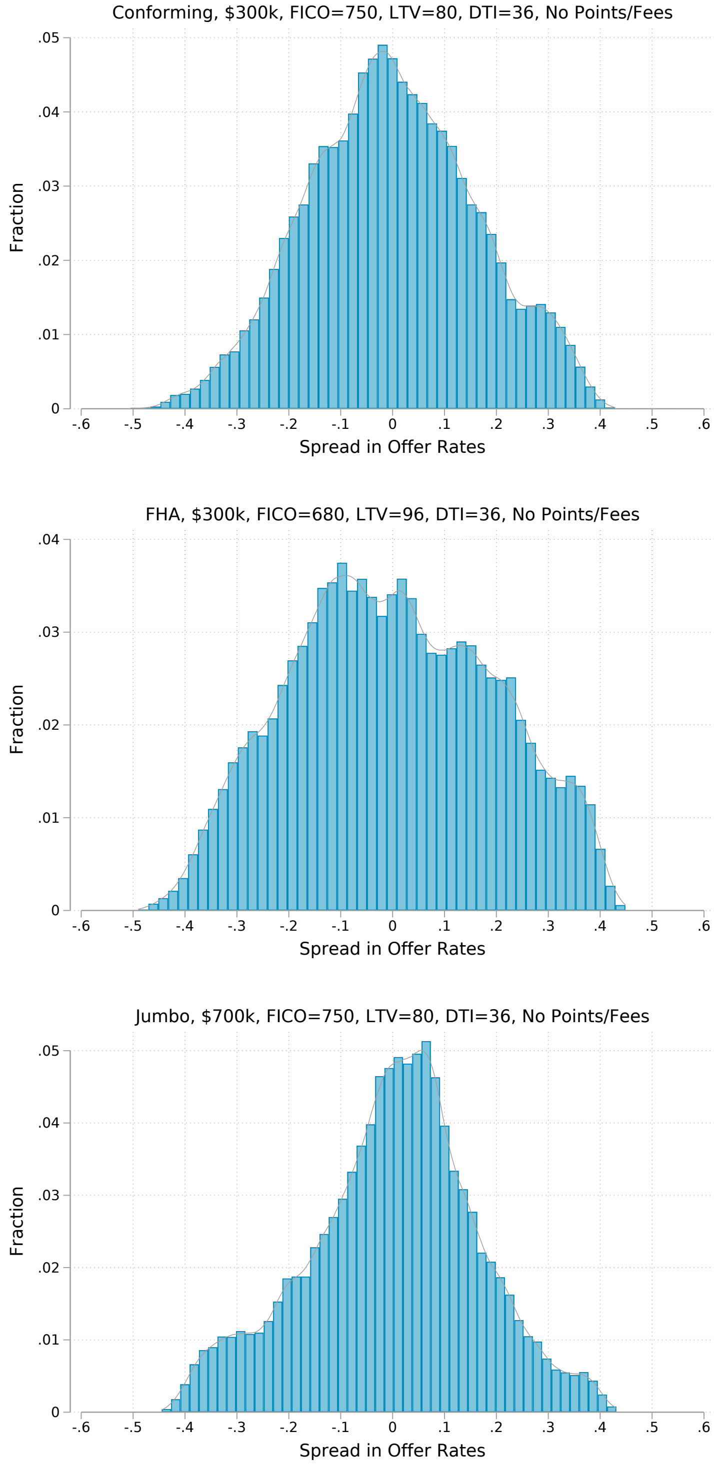

Figure A-1: Interest Rate Offer Dispersion for Identical Mortgages in Los Angeles

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.

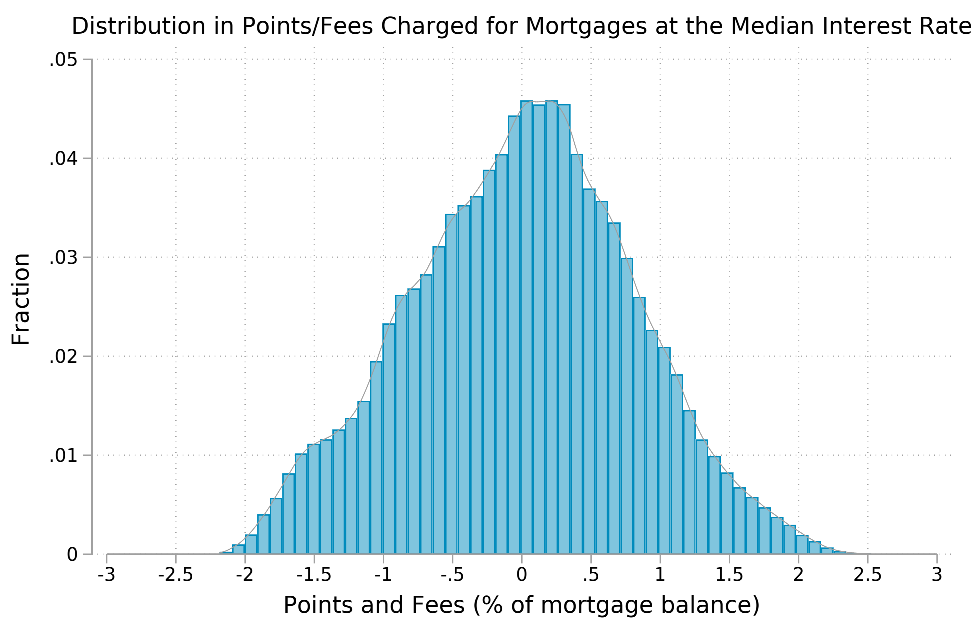

Figure A-3: Dispersion in Points and Fees Lenders Charge for Identical Mortgages at the Median Interest Rate

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.

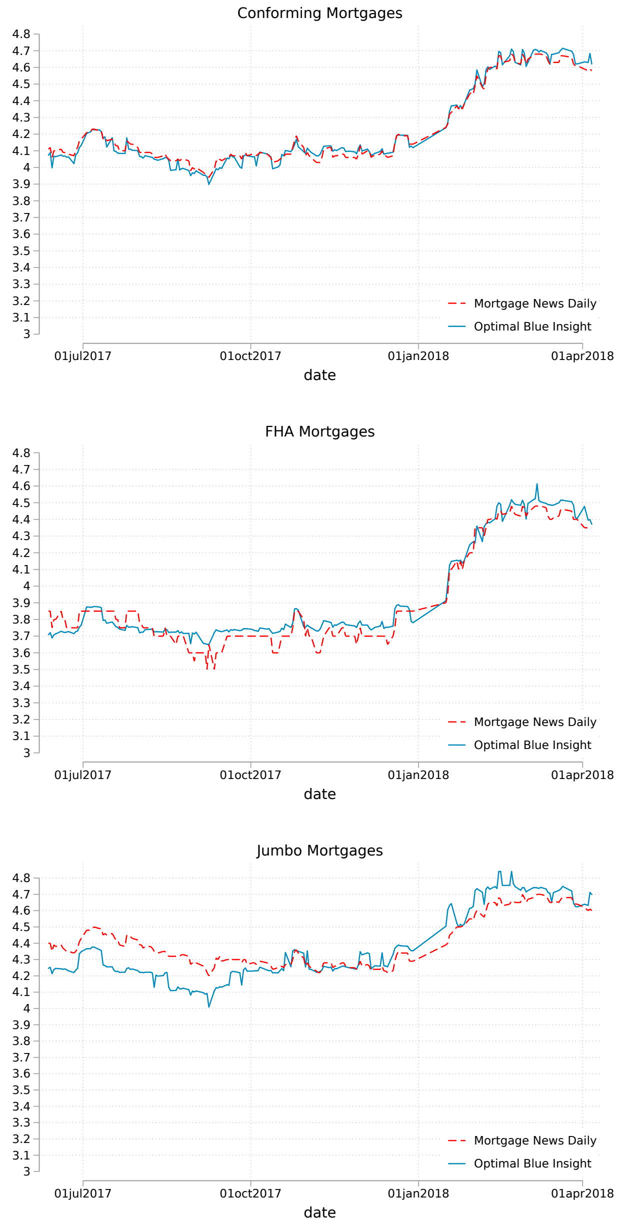

Figure A-4: Comparison of average offer rate from Optimal Blue with Mortgage News Daily data

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.