DISCUSSION PAPER SERIES

DP14924
(v. 3)

Paying Too Much? Borrower Sophistication and Overpayment in the US Mortgage Market

Neil Bhutta, Andreas Fuster and Aurel Hizmo

FINANCIAL ECONOMICS

CEPR

ISSN 0265-8003

Paying Too Much? Borrower Sophistication and Overpayment in the US Mortgage Market

Neil Bhutta, Andreas Fuster and Aurel Hizmo

Discussion Paper DP14924
First Published 24 June 2020
This Revision 29 January 2021

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

Paying Too Much? Borrower Sophistication and Overpayment in the US Mortgage Market

Abstract

Using administrative data on mortgage rates that borrowers obtain and on rates that lenders could provide for the same loan, we find that many homebuyers significantly overpay for their mortgage, particularly those least likely to be financially savvy. For example, one quarter of government-insured borrowers, who tend to be lower-income first-time homebuyers, pay at least 45 basis points more than their median available rate—equivalent to an upfront payment (in points) of $5,400 for a typical loan. We further document considerable dispersion in the interest rates that identical borrowers get, even from the same lender and loan officer, suggesting important roles for shopping and negotiation. Across time, we find that overpayment tends to decrease as market rates rise and, using new survey data, provide direct evidence that borrowers shop more when market rates are higher, suggesting that behavioral forces affect search effort. More generally, we use the survey data to demonstrate that a significant amount of the mortgage rate variation across consumers can indeed be traced to heterogeneity in financial knowledge and shopping behavior.

JEL Classification: G21, G51, G53, E43

Keywords: mortgage market, household finance, price dispersion, financial literacy

Neil Bhutta - neil.bhutta@frb.gov
Federal Reserve Board

Andreas Fuster - andreas.fuster@gmail.com
Swiss National Bank and CEPR

Aurel Hizmo - aurel.hizmo@frb.gov
Federal Reserve Board

Acknowledgements
The paper previously circulated as“Paying Too Much? Price Dispersion in the US Mortgage Market.” We thank Jason Allen, Robert Avery, John Campbell, Nick Embrey, Serafin Grundl, Katherine Guthrie, Haj Hadeishe, Michael Haliassos, Chris Hansman, Gregor Matvos, Raven Molloy, John Mondragon, Christopher Palmer, Saty Patrabansh, Chad Redmer, James Rowe, David Zhang, as well as seminar and conference participants at the American University (Kodog), Arizona State University, Baruch College (Zicklin), EPFL, Federal Reserve Bank of Atlanta, Federal Reserve Board, Freddie Mac, Norges Bank, NYU Stern, Oxford Saïd, AFA Annual Meeting, AREUEA National Conference, CEPR European Conference on Household Finance (Ortygia), Cherry Blossom Financial Education Institute, FCA- Imperial Household Finance Conference, FDIC Consumer Research Conference, University of Copenhagen and the NBER Summer Institute (Real Estate) for helpful comments. Special thanks to Jay Shultz and Saty Patrabansh for helping us access and use data at FHFA. 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.

1 Introduction

Recent survey data indicate that half of the borrowers taking out a mortgage in the US in 2016 only seriously considered one lender, and just 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 there is little variation in mortgage pricing, or that borrowers are very efficient at finding the best rates. This might seem a reasonable conclusion especially when considering that the mortgage market appears highly competitive: the majority of mortgages in the US are highly standardized and guaranteed by the government, and there are hundreds of lenders offering mortgages on any given day. However, in contrast to borrowers' perceptions, we find that many borrowers substantially overpay for their mortgage, and that overpayment varies systematically across borrower types and over time.

To assess overpayment, we draw on a unique source of data—an industry platform used by lenders to price mortgages and conduct transactions with borrowers. The platform provides data on both *available* rates—the rates that lenders could offer for specific mortgages/borrowers in each market and each day—and data on the mortgages *locked*, or obtained, by consumers. The available rates are inclusive of the fees and markups that a borrower would pay if they chose a particular lender. Although they are often referred to in the industry as "offer" rates, it is important to recognize that these rates are lender's private information rather than publicly posted, and may not be automatically offered to prospective borrowers.[2] 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 or 60 days).

For a given borrower, we compute the difference between the rate they locked and the median rate available for the same type of loan and borrower (same loan-to-value ratio [LTV], credit score [FICO], points, etc.) on the same day in the same market.[3] We find that this "locked-offer rate gap" varies substantially across borrower types. For example, "jumbo" borrowers, who tend to have relatively high incomes, on average obtain a rate that is 21 basis points (bp) *below* the median available rate for their loan type, suggesting that such borrowers are able to find relatively good deals. In contrast, FHA borrowers, who tend to have lower income, wealth, and credit scores, on average pay 25bp *more* than what the median lender could offer for their exact loan. Remarkably, one quarter of FHA borrowers pay in excess of 45bp more than the median available rate for their exact same loan. Using the average point-rate trade-off in our data, a 45bp rate difference is equivalent to an upfront payment of 2.1 points, or about $5,400 for a typical loan of $250,000.

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

[2] See Duncan (2019) for a discussion of what makes comparison shopping in the mortgage market more complicated and time-consuming than shopping for ordinary goods.

[3] Our results are qualitatively unchanged when we alternatively consider the expected gains from one extra search as a measure of overpayment.

We also document that low-FICO and high-LTV borrowers have relatively high locked-offer rate gaps even within the same lender branch and controlling for loan amount.[4] One possible reason such borrowers might pay more would be if they tend to require more attention and service from loan officers. However, using data on loan officer compensation, which we observe for a subset of lenders, we do not find support for such a story. Overall, our analysis yields the novel insight that the higher rates paid by low-FICO and high-LTV borrowers are not simply due to risk-based pricing, but also reflect less effective shopping and negotiation by these borrowers.

In addition to the cross-section, we also study the time-series variation in locked-offer rate gaps. If driven simply by borrowers' time-invariant search costs, this gap should not vary over time. However, we find that average overpayment declines when the level of market interest rates rises. This may partly reflect affordability constraints becoming more binding as rates rise; however, we show that even borrowers that appear unconstrained (based on their debt-to-income ratio) 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 influence search effort.[5]

We provide supporting evidence of ineffective shopping and negotiation by analyzing price dispersion in the locks data alone. While previous work has also documented price dispersion in this market, the additional details available in our data (e.g. exact lock date, points, as well as lender, branch, and loan officer identifiers) allow us to exhaustively control for potential drivers of rate differences and thus quantify their relative importance in explaining dispersion. It also worth noting that most loans today are government-backed, and thus lenders and investors have minimal exposure to credit risk and limited incentive to price unobserved borrower risk characteristics. We find that the difference between the 90th and 10th percentile interest rate that identical borrowers lock in for the same (30-year fixed-rate fully-documented) loan in the same market, on the same day, and paying the same points, is 54bp; for a typical loan of $250,000, this corresponds to an upfront payment of $6,500. The largest residual dispersion occurs for borrowers who are likely to be the least financially sophisticated (e.g. low credit score and inexperienced homebuyers).

Additionally, we show that time-invariant lender fixed effects explain little of this dispersion, suggesting limited explanatory power of factors such as lender reputation or quality. However, allowing for lender-specific time-varying pricing and branch-by-month fixed effects cuts the residual dispersion by almost one-half. In other words, different lenders, as well as branches within lenders, set different prices, and the cheapest lenders and branches change over time. Still, significant dispersion remains within branch and even within loan officer, consistent with a role for negotiation. Notably, the lending platform providing our data is mostly used by monoline nonbank mortgage originators, and our results are unchanged when we limit the sample exclusively to such lenders.

[4] Looking across locations, locked-offer rate gaps are highest in ZIP codes with low median household incomes, fewer college-educated households, and high minority shares. We find little association of locked-offer gaps with local mortgage market concentration.

[5] In line with this, we also find evidence suggesting that FHA and jumbo borrowers may be "anchoring" to the average prime conforming rate, which is the rate most often advertised and reported in the media: they obtain relatively better rates (a lower locked-offer rate gap) when the difference between their offer rate and the prime conforming offer rate increases.

Thus we can rule out explanations related to cross-selling or bundling of other services (Hortaçsu and Syverson, 2004) as factors that could explain observed price dispersion in this market.

Finally, we provide direct support for the importance of borrower sophistication using new data from the National Survey of Mortgage Originations (NSMO). The NSMO combines detailed administrative records on recent mortgage originations with survey data on the individuals who took out those mortgages. The survey component focuses on borrowers' shopping behavior and their knowledge of mortgages and interest rates. Using these data, we show that shopping and financial knowledge are predictive of borrowers getting lower mortgage rates, controlling for an array of credit risk variables and other individual characteristics.[6] We then construct a composite measure of the rate component attributable to shopping/knowledge. Despite the coarseness of the survey questions, we find a sizeable 26bp gap between borrowers at the 90th percentile of this measure and those at the 10th percentile—novel evidence that a considerable amount of price dispersion stems from heterogeneity in financial knowledge and shopping. FHA, low-income, and low-FICO borrowers tend to do particularly poorly in terms of our composite measure.

In addition to these cross-sectional results, we also use the NSMO data to directly test our conjecture that a rise in rates encourages people to shop more. Indeed, we find that several measures of shopping activity rise with interest rates.

Overall, our empirical results suggest that a large fraction of the borrower population in the US overpays for mortgages, and a key reason for this seems to be a lack of financial sophistication. The borrowers that fare the worst often get government-guaranteed loans through the FHA program, which is aimed at lowering the cost of homeownership for lower-income households. Our results suggest that government entities such as the FHA might consider ways to reduce price dispersion and excessive markups to help fulfill their policy objectives. Our findings also suggest that the lack of consumer shopping is important for the pass-through of monetary policy to the mortgage market: reduced search effort appears to prevent borrowers' rates from falling as much as they could when market rates decrease, thereby weakening the pass-through of expansive policy.

This paper makes several contributions to the literature on overpayment and dispersion in the price of mortgages and consumer credit more broadly. This is the first paper to compare transacted mortgage rates to borrower-specific real-time available rates. This comparison is key to understanding *who* overpays, allowing the separation of overpayment from credit risk premia. To our knowledge this is also the first paper documenting that borrower overpayment and shopping intensity change with market rates over time, consistent with behavioral factors affecting shopping rather than standard search costs being the sole determinant.[7] We further support this conclusion by exploiting new survey data to provide novel direct evidence relating obtained rates to shopping

[6] We further provide complementary evidence using the 2016 Survey of Consumer Finances (SCF). We find that higher financial literacy, as gauged by the Lusardi-Mitchell financial literacy "test", is associated with significantly lower interest rates. We also find in the SCF data that borrowers who report shopping intensely for credit end up with substantially lower rates. In independent work made public after our initial draft, Malliaris et al. (2020) also use the NSMO data to document that indicators of sophistication are correlated with rates.

[7] Our finding of higher average locked-offer gaps when market rates are low is distinct from and complements the finding of Fuster et al. (2017) that lender offers tend to feature higher markups when market rates are low.

behavior and market knowledge—variables that are generally unobserved in other settings where price dispersion has been studied.

Our paper connects to the literature on (in)efficiency of consumer choice and price dispersion in various consumer finance markets, including mutual funds (Hortaçsu and Syverson, 2004; Choi et al., 2010), auto loans (Argyle et al., 2017), and credit cards (Stango and Zinman, 2016). In addition to mortgages being the largest household liability, the composition of mortgage borrowers spans the income, wealth, and financial sophistication spectrums, providing much cross-sectional variation to help shed light on the factors driving dispersion and overpayment in consumer financial markets more broadly. Along with other work documenting costly "mistakes" in the mortgage market (e.g., Agarwal et al., 2015, 2017b; Keys et al., 2016; Andersen et al., 2020), our results are in line with the growing literature pointing at financial literacy/sophistication as a key driver of differential outcomes in household finance (e.g., Hastings et al., 2013; Gomes et al., 2020).

In recent work, Agarwal et al. (2019) argue that overpayment by certain groups need not imply that they are unsophisticated (or have high search costs), but could be a rational response of relatively risky borrowers who fear being rejected. These authors document that the relationship between contracted mortgage rate and the number of "inquiries" recorded by credit bureaus—their proxy for borrower search—is U-shaped. This suggests that borrowers that search a lot may do so because their application gets rejected, which in turn may lead these borrowers to accept relatively worse offers. This channel may contribute to some of the overpayment we document. At the same time, however, we find considerable overpayment even among many well-qualified borrowers, and also provide evidence from NSMO and SCF data that variation in sophistication is important to understand cross-sectional dispersion.[8]

Related work by Alexandrov and Koulayev (2017) and McManus et al. (2018) shows dispersion in lenders' offer rates, while Gurun et al. (2016) and Agarwal et al. (2019) study transacted rates.[9] Our results on dispersion in offers are similar to those in the earlier work. However, when studying transacted rates, as noted earlier, our dataset allows much-improved identification. Indeed, while we find wide dispersion in locked rates, it is considerably narrower than the dispersion found in earlier work. Furthermore, we unpack the dispersion in transacted rates by assessing the relative explanatory power of (time-varying) controls for lenders, branches, and loan officers, as well as borrower characteristics.

Finally, Woodward and Hall (2012) avoid identification issues arising from differences in bor-

[8] Over the period we study, underwriting standards in the GSE and FHA segments of the market are largely dictated directly by these agencies. Thus, for the vast majority of borrowers that get approved for a loan, it should also be easy to get a loan from a different lender. That said, the *perception* that other lenders are unlikely to accept one's application may be sufficient to induce a borrower to accept a relatively "bad" offer.

[9] Some work also exists outside the US, where the institutional details and the mortgage market structure are different. 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 in bargaining leverage across consumers. In the UK market, Iscenko (2018) finds that many borrowers choose products that are dominated in cost terms by other available alternatives, while Liu (2019) shows that many borrowers appear to neglect non-salient fees and that lenders exploit this in their price setting. Damen and Buyst (2017) provide evidence that mortgage borrowers in Belgium who shop more achieve substantial savings.

rower characteristics by focusing on dispersion in the fees borrowers pay to mortgage brokers—who arrange loans between borrowers and lenders—which should be independent of credit risk. That said, broker fees are only one component of mortgage pricing, and many borrowers do not use mortgage brokers. Additionally, Woodward and Hall were limited to a small dataset of FHA loans from 2001—prior to regulations governing broker and loan officer compensation. Our much larger dataset covering 2015-19 enables us to examine overpayment and dispersion across market segments, origination channels, and borrower types, as well as within lender and even loan officer.

The rest of the paper is organized as follows. In the next section, we provide some institutional detail that will be important for the rest of the paper. Section 3 describes the Optimal Blue data on rate locks and mortgage offers. Section 4 measures and unpacks price dispersion in the offer data and the lock data. Section 5 explores how locked rates on average compare to the offer distribution, and how this varies across borrowers with different characteristics. Section 6 studies how these patterns evolve over time as market rates change. Section 7 introduces survey data from the NSMO and presents direct evidence on the connection between shopping, mortgage knowledge, and interest rate outcomes. Finally, Section 8 concludes with some potential policy implications.

2 Mortgage Pricing and Originations in the US

In this section, we provide a brief overview of some of the institutional details that will be important for the rest of the paper.[10]

In the US, there are multiple channels through which a borrower can obtain a loan. One of them is to go directly to a bank or credit union. An alternative is to obtain a loan through a specialized mortgage originator, a so-called "mortgage bank" (or "independent mortgage company"). These lenders, contrary to what the name suggests, are not depository institutions, and typically do not keep any of the mortgages on their own balance sheet. Finally, it is also possible to go through a mortgage broker, who may have relationships with both bank and nonbank originators, and acts as an intermediary connecting borrowers to those institutions. When a loan is originated directly by a lender who will either retain the loan in portfolio or sell it directly in the secondary (mortgage-backed securities, or MBS) market, this is called a "retail loan"; if a loan is originated via a nonbank entity that originates the loan for another lender, this is called "wholesale."

Regardless of the channel, a borrower will generally interact (in person or just by phone/online) with a loan officer or broker (henceforth LO) who will have access to various "rate sheets" that provide the detailed pricing available at a given point in time (generally updated at least once a day). Importantly, for any loan type and combination of characteristics, there is no single interest rate—instead, the rate sheet shows a combination of note rates and "(discount) points". To obtain a low note rate, a borrower can pay points (where 1 point = 1 percent of the loan amount). If the borrower is willing to take a higher rate, they can receive points (often called rebates or credits) which in turn can be used toward the origination costs. Whether a borrower "should" pay or receive

[10] For additional discussion, see e.g. Fuster et al. (2013) or `https://files.consumerfinance.gov/f/201301_cfpb_final-rule_loan-originator-compensation.pdf`.

points depends on their liquidity situation and on the amount of time they expect to stay in the mortgage. For a borrower with a high likelihood of prepaying the loan within a few years—either due to a move or an anticipated refinancing—it is likely not worth paying points.[11]

In the case of a retail loan, the available pricing will come directly from the lender's pricing desk; in the case of wholesale lending, the rate sheets can come from several different lenders (often referred to as "investors"). Each rate sheet will provide pricing for different loan programs (e.g. GSE loans, FHA, or jumbos) with adjustments depending on a few loan and borrower characteristics, typically FICO, LTV, loan amount, geographic region, loan purpose and property type. Pricing depends on the value that a lender assigns to the loan—often based on the current value of such a loan in the MBS market, where most loans are ultimately sold.[12] Prices also take into account required "guarantee fees" set by the agencies that securitize the loans and insure the credit risk, namely the GSEs and Ginnie Mae (for FHA/VA loans).[13] Furthermore, lenders will add a margin that may depend, among other things, on the level of demand for loans (Fuster et al., 2017).

On top of the prices from the rate sheet, the costs to the borrower include compensation of the LO and/or their employer (e.g., the mortgage bank). This compensation may be explicit (via upfront origination fees) or implicit (via lender profit margins on rate sheets). Historically, LOs had strong incentives to sell loans with higher interest rates, all else equal, and thereby generate more compensation not only for the lender but also for themselves (often called the "yield spread premium"). However, in the wake of the financial crisis, new regulations were imposed so that LO compensation may no longer vary with the interest rate and other terms of the loan. But lenders, of course, still profit when borrowers take higher interest rates.[14] Importantly, this does not imply that all LOs in a firm simply get paid an identical, fixed amount for each loan they originate. In fact, LOs are frequently given a choice between different possible compensation plans, for example trading off fixed salary for higher commission rates per dollar of originated loans.

Finally, it is not the case that the combination of rate sheets and a specific LO's compensation plan in all cases determine the final rate and points/fees that given borrower is offered: there may be "exceptions" granted, for instance to meet a competitive outside offer. Lenders generally have specific procedures for these exceptions, since they want to avoid violating fair lending laws.[15]

[11] The fact that the value of points depends on how long a mortgage remains active also means that comparing the cost of different mortgages based on their "annual percentage rate" (APR)—a mandated disclosure under Truth in Lending law—is not necessarily meaningful. The APR calculation distributes upfront points and fees over the life of the loan, assuming the loan runs to term (e.g., 30 years); however, most mortgages terminate much sooner. See e.g. https://www.mtgprofessor.com/tutorial_on_annual_percentage_rate_(apr).htm.

[12] Generally, prices in the MBS market depend on the yields on alternative investments (especially Treasuries) as well as investors' projections of future prepayments of the underlying mortgages (since mortgage borrowers have a free prepayment option). Prepayments are in turn affected by factors such as the volatility of interest rates, home price growth, or relevant policies by the GSEs and FHA (e.g. streamlined refinance programs).

[13] In addition to the guarantee fee, which is a flow insurance premium over the life of a mortgage, the GSEs charge upfront "loan-level price adjustments" that depend on borrower and loan characteristics—see e.g. https://www.fanniemae.com/content/pricing/llpa-matrix.pdf.

[14] These rules were first changed in 2011 as part of the Truth in Lending Act; the Consumer Financial Protection Bureau published its final rule on LO compensation requirements in January 2013.

[15] See e.g. https://www.crai.com/sites/default/files/publications/Managing-the-Fair-Lending-Risk-of%2DPricing-Discretion-Whitepaper-Oct-2014.pdf or https://www.mortech.com/mortechblog/

An important step in the origination process is the mortgage rate lock. A 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. However, the lock is binding on the lender, as long as the characteristics of the loan and borrower (such as the loan amount or the credit score) remain as specified at the time of the lock. Lenders typically do not charge an explicit fee for a rate lock, though there are generally loan application fees. Also, if a loan does not close by the time the lock period expires, extending the lock typically requires a fee.[16]

3 Optimal Blue Data

Our main data comes from an industry 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. More than 2.4 million mortgage locks were processed through this system in 2019, thus accounting for about 30% of loan originations nationally.

The lenders using the platform tend to be nonbank monoline mortgage lenders. These lenders have gained substantial market share in the post-crisis period (see e.g. Buchak et al., 2018); in 2019, they originated 56% of all purchase loans and 58% of refinance loans (CFPB, 2020). Optimal Blue is also used by smaller community banks or credit unions. That said, many institutions on this platform act as correspondent lenders, meaning that they originate loans intended to be sold to other financial institutions such as a large bank like JP Morgan or Wells Fargo.

For this study 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.

3.1 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. 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; we end in December 2019. The data have wide geographical coverage

pricing-discretion-fair-lending-risk.

[16] For more information on rate locks, see e.g. https://www.bankrate.com/finance/mortgages/questions-rate-lock-answered.aspx.

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, FICO score, debt-to-income (DTI) ratio, loan amount, loan program, loan purpose (purchase or refinancing), asset documentation, income documentation, employment status, occupancy status, property type, ZIP code location etc.

There are a number of unique features of the data relative to servicing data that are 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 observe the exact time-stamp of when the lock occurred, while in most other datasets only the closing date is recorded, which generally differs from the pricing-relevant lock date by several weeks or even months. Finally, we have unique identifiers for the lender, branch, and loan officer that processes each mortgage. For some lenders we can also observe loan officer compensation, expressed as a percentage of the loan amount.[17]

While the lock data features numeric lender identifiers, it does not directly provide us with information on the lenders. However, we are able to classify a subset of lenders into whether they are an independent nonbank or not by relying on a match between Optimal Blue locks and administrative FHA data used in Bhutta and Hizmo (2020). This will be useful later to assess whether lender type and cross-selling might be driving the patterns in the data that we observe.

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 owner-occupied single-unit properties, with full documentation of assets and income, and drop self-employed borrowers. We also drop loans for amounts under $100,000, and those with implausible values for LTV, DTI, or points/credits. Finally, we drop VA loans and streamline refinances (which are a small part of the sample). This leaves us with 3.6 million observations. For the analysis in Sections 5 and 6 we will further restrict the sample in order to match the locked mortgages to offers for identical characteristics, as will be described there.

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 (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 potentially 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 mortgage insurance from the FHA and are securitized through the government entity Ginnie Mae).

The table shows that FHA loans are most likely to go to first-time homebuyers with low FICO scores and high LTV and DTI. Jumbo loans, the only loan type where the credit risk is not guaranteed by the government, tend to go to the most creditworthy borrowers and feature relatively

[17]Some lenders process compensation outside of the Optimal Blue system, or do not compensate loan officers directly on a per-loan basis.

low LTVs. They only constitute about 2% of our sample. The table also shows that FHA borrowers on average pay fewer discount points than borrowers in the other programs; Appendix Figure A-1 displays the cumulative distribution of points paid (or received) by program.

As noted above, not all lenders use the Optimal Blue platform, and not all rate locks necessarily result in an originated mortgage. Thus, there is a concern that the distribution of interest rates recorded in our rate lock data may not accurately represent the rates that borrowers ultimately end up with. However, in Appendix A.1, we show that the interest rates observed in the rate lock data mirror the interest rates observed in the well-known McDash mortgage servicing dataset on originated mortgage loans, both in terms of averages and dispersion. Furthermore, loan/borrower characteristics in Optimal Blue locks also look very similar to those in data on originated loans.[18]

3.2 Mortgage Offers Data

As our second source of data, we collect data on the menu of mortgage products 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). Importantly, the offers we observe are 'customer facing', i.e. rates inclusive of margins and fees that borrowers could pay if they chose a particular lender. The Insight interface is designed for lenders to compare their pricing against that of peers.

For any combination of day, MSA, and loan/borrower characteristics, we measure an "offer" rate for each lender on the platform. This offer rate reflects the interest rate (with zero points) that the lender could offer a prospective borrower, including fees under the assumption that the loan is originated by the loan officer (LO) that has locked the most loans for that lender in that market.[19]

If a lender represents multiple different investors, the offer we observe is based on the most competitive investor offer. Thus, a borrower locking a loan with this lender would not necessarily get exactly the observed offer rate for three reasons. First, the locked rate can vary depending on which LO the borrower goes through, since different LOs can charge different markups. Second, the LO may offer a loan that is not based on the rate sheet of the most competitive investor, but on one from a different investor.[20] Third, as noted earlier, borrowers may be able to negotiate and get an "exception" or a lower rate from the lender.[21]

We conduct daily searches in one local market (Los Angeles), twice-weekly searches in four

[18] See Appendix Table A-1 and Figure A-4 for details. For jumbo mortgages, the locked interest rates in Optimal Blue tend to be higher than those in McDash, which could reflect that the relatively smaller lenders that use the Optimal Blue platform may not be as competitive for these types of loans as for FHA and conforming loans. It is also the case that average jumbo loan amounts are somewhat smaller in Optimal Blue locks than in McDash originations, which could reflect some differential selection of borrowers. The dispersion of rates is still very similar, however.

[19] As explained further in Appendix A.2, we observe a distribution of prices (points) for a given note rate, which we transform into a distribution of rates for zero points.

[20] One reason why an LO might want to do this is to maintain active relationships with multiple investors.

[21] The Pricing Insight data are different from earlier Optimal Blue data used by Fuster et al. (2017). Those data were based on rate sheets that did not include LO compensation and origination charges, unlike the Pricing Insight data we use here (where offers are "all included").

markets, and weekly searches for 15 additional markets.[22] We collect offer distributions for 100 different loan types, differing across the following dimensions: FICO score, LTV ratio, loan program, 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 documentation of income, assets and employment, and are used to finance single-unit homes.

An important limitation of the offers data is that 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. The time series is also slightly shorter than for the locks data, as we started systematically tracking offers in April 2016.

4 Dispersion in Mortgage Rates

4.1 Dispersion in Offer Rates

We begin by briefly presenting some findings from the Optimal Blue Insight data on offer rate dispersion. Our analysis here, along with additional findings presented in Appendix A.2, adds to recent work looking at offer rate dispersion using other sources of data in Alexandrov and Koulayev (2017) and McManus et al. (2018).

Figure 1 shows the dispersion in mortgage rates available from different lenders, pooling data over time and across all of the 20 metropolitan areas for which we obtained data. To make distributions comparable across time and locations, we demean the offer rates for each mortgage type in each market and day. Figure 1 indicates wide dispersion in offer rates. There is a 53bp difference between the 10th and 90th percentile offers, which is similar to what Alexandrov and Koulayev (2017) and McManus et al. (2018) have documented.

In Appendix A.2, we additionally show that the degree of offer rate dispersion is quite similar across different types of loans, different types of borrowers, and across all 20 cities in our sample. Finally, it worth noting that it is not necessarily the case that a given lender occupies the same spot in the offer distribution over time. Lenders could move around in the distribution if pricing does not simply reflect time-invariant cost factors. Unfortunately, since we cannot follow lenders over time in the Insights data, we cannot assess this directly in the offer data. However, the analysis in the next subsection will shed some light on whether lenders' relative pricing changes over time.

4.2 Dispersion in Locked Rates

In the previous subsection we observed wide variation in mortgage rates available from different lenders for identical borrowers on the same day and in the same market. In this section we aim to investigate whether identical borrowers who choose the same mortgage product, in the same

[22] 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.

market, and at the same time, actually *lock in* different interest rates. If many borrowers shop around, we may observe less dispersion in locked rates than we observe in offer rates.

To investigate dispersion in locked mortgage rates, we regress locked 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. Comparing the residual dispersion across specifications allows us to "unpack" the relative importance of different drivers of price dispersion in this market.

Table 2 shows the results from various specifications, estimated on the same set of 2.96 million loans locked over the five-year period 2015-2019.[23] In the first column, as a benchmark, we include only lock date-by-MSA fixed effects, in order to document the amount of overall interest rate dispersion within the same MSA on the same day. These day-by-MSA fixed effects explain just under 60 percent of the total variation in rates, and the standard deviation of the residual is 33bp.

In column (2), we add our baseline set of controls: an extensive set of underwriting variables, which consist of fully interacted bins of values for FICO, LTV, and loan program, interacted with lock month to allow for time-variation in risk pricing.[24] We also include borrower ZIP code fixed effects, lock period fixed effects, property type fixed effects, cubic functions of loan amount and DTI, as well as linear controls for FICO and LTV (to allow for within-bin variation).[25] This specification is similar to regressions one could typically run with a mortgage servicing dataset.[26] We see that the controls explain a sizable share of the raw variation in interest rates—the adjusted R-squared is 0.75—but that substantial dispersion remains: the standard deviation in residuals is 0.26, and the borrower at the 90th percentile of the residual distribution pays 58bp more than the borrower at the 10th percentile.

Column (3) adds bins for the points paid or received by the borrower (interacted with program by lock month).[27] This (usually unobserved) variable indeed explains some of the rate differences across borrowers, but substantial dispersion remains—e.g. the 90th-10th percentile difference is still 54bp, which is almost identical to the corresponding dispersion in offer rates shown in the previous subsection.

Based on the regression coefficient on discount points (not shown in the table), we can translate interest rates to upfront points.[28] This coefficient implies that 1 discount point changes the interest

[23] 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 (10) in all specifications. However, using the largest possible sample for each specification instead does not materially affect the results.

[24] We include 13 FICO bins, 9 LTV bins, and 12 dummies for the four loan programs interacted with three loan purposes (purchase, rate refinance, and cash-out refinance). The choice of FICO and LTV bins is motivated by the loan-level price adjustments set by the GSEs.

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

[26] It is already somewhat more precise, since here we control for the date 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.

[27] We include 8 point bins, as well as a linear function in points to allow for within-bin variation.

[28] We estimate the relationship between discount points and interest rates in a regression specification identical to

rate by about 21bp on average. Therefore, 54bp in rate is approximately equivalent to 2.6 upfront discount points or 2.6% of the mortgage balance. In other words, our results imply that a borrower with a $250k mortgage borrowing at the 90^{th} percentile interest rate should be getting—but in fact is not getting—a lender credit of about $6,500 relative to someone borrowing at the 10^{th} percentile interest rate. Alternatively, if one prefers to think in terms of mortgage payments, 54bp correspond to about $80/month for a $250k loan at the average level of rates over our sample period.[29]

Thus, observably identical borrowers within the same market, on the same day, getting the same loan can pay dramatically different prices. Table 3 shows how the residual dispersion in interest rates varies across different loan programs and characteristics. The middle column of the table uses the residuals from specification (3). We see an extreme amount of dispersion for the two lowest FICO groups. We also see substantial dispersion for FHA-insured loans, despite the fact that these loans are fully insured by the government and thus lenders and investors take very little, if any, credit risk. In other words, it seems unlikely that unobserved risk factors could explain the wide dispersion in FHA interest rates. Along the same lines, we also find fairly wide dispersion for conforming and super-conforming loans, which meet the credit standards of the GSEs and will likely be purchased and fully guaranteed by these institutions.[30] Finally, we also see wide dispersion even when we focus just on low-risk borrowers: those with prime FICO scores in excess of 680, and those with LTVs of less than 75 percent.

Jumping back to Table 2, in column (4) we add lender fixed effects to allow for the possibility that some of the price differences may reflect differences in lender characteristics such as service quality or advertising costs. We find that the 90th-10th percentile difference decreases only slightly, by 6bp. In columns (5) and (6), we further interact the lender fixed effects with lock day fixed effects and other controls, to allow for the possibility that lenders' (relative) pricing may change over time, or may differ across loan types. Here the 90-10 gap drops more substantially, by 10bp (or over 20 percent) from column (4). Overall, the results in columns (4)-(6) suggest that more so than time-invariant differences in lender quality, price dispersion may reflect lender pricing strategies that vary over time and across programs. Such variation would make it difficult for borrowers to find low rates simply by following the recommendations of family, friends or real estate agents—yet this is a common approach borrowers take to finding a mortgage.

In columns (7) and (8), we further allow for pricing to differ across different branches of a lender.

column (10) of Table 2, with the only exception that discount points are allowed to only enter linearly. Appendix Figure A-2 shows in a binned scatter plot that the relation between points and rate is indeed close to linear.

[29]Furthermore, a fixed-rate mortgage with a lower rate amortizes more quickly, so that the *interest* savings to the borrower with the lower rate are larger: e.g., a rate of 4.06% instead of 4.60% reduces the interest expense over the first year of the loan by $1,347.

[30]One caveat here is that lenders may be worried about so-called "put-back" risk where loans in default must be repurchased by the lender due to some defect in the underwriting found by the FHA or GSEs. However, at least in the case of the GSEs, Goodman (2017) documents that put-back risk has been negligible since lenders have stopped issuing low-documentation and other non-traditional loans. For FHA loans, perhaps the biggest concern for lenders has been litigation risk under the False Claims Act, which allows the federal government to sue lenders that knowingly submit false or fraudulent claims to the FHA. Under the Obama Administration, some of the largest lenders settled with the government, paying fines close to $5 billion. That said, this risk is most salient for large banks with significant capital at risk, unlike the nonbanks that dominate our data. Also, this risk has eased in recent years.

As discussed earlier, the lenders in our dataset tend to be nonbank monoline mortgage lenders and community banks. For a typical lender in our data, in a given MSA, most loans are originated through just 2 or 3 branches located within that MSA. Differential branch pricing could reflect differences in convenience of the office location and/or costs (e.g. office rent). In addition, as noted earlier, different branches can have different markups and pricing strategies.

The branch fixed effects in column (7) have noticeable incremental explanatory power, increasing the adjusted R-squared from 0.85 to 0.88 and reducing the residual dispersion. Adding branch-by-month fixed effects in column (8) further reduces residual dispersion—consistent again with time-varying price strategies playing a role in the rates borrowers obtain, but in this case at the branch level. Nevertheless, even in column (8), which should come close to looking at nearly-identical borrowers getting a loan from the same branch at the same time, the 90-10 gap remains at 31bp, and the interquartile range at 14bp.

Lastly, in columns (9) and (10), we further allow for pricing to vary across different loan officers (LOs) in the same branch, which could reflect for instance differences across LOs in terms of experience, compensation, or willingness/ability to negotiate. Which LO a borrower matches up with (within a branch) does appear to matter somewhat for the rate they end up with, since the adjusted R-squared further increases and the residual dispersion decreases in the last two columns. Nevertheless, even after including LO fixed effects that are allowed to vary across time and programs, the 90th-10th percentile difference remains at 26bp, and the interquartile range at 11bp.

The last column of Table 3 shows that the cross-sectional patterns in residual dispersion, already discussed above, remain similar in the most restrictive specification (10): the dispersion is substantially larger for loan types and borrower characteristics that are associated with being more financially constrained and potentially less sophisticated, such as FHA loans, low-FICO borrowers, or first-time homebuyers.

The final rows of the table show that the residual dispersion is identical if we only consider loans that were locked with lenders that we are able to classify as independent nonbanks (as discussed in Section 3.1).[31] This suggests that the large dispersion is not driven by unobservable pricing adjustments that banks or credit unions might make for customers that already have accounts or other business with them. The nonbank lenders are only in the business of originating mortgages.

To sum up the findings from this analysis, there is a large amount of dispersion in the rates that observably identical mortgage borrowers pay, even after controlling for the exact timing and upfront payments. Adding lender, branch and LO controls reduces the residual rate dispersion by about half. However, substantial dispersion remains, implying that two observably identical borrowers may get quite different deals from the same lender branch or even the same loan officer 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.

[31] The residual dispersion results are also essentially unchanged if we restrict the estimation sample to these independent nonbanks only.

5 Comparing the Locked Rates to Offer Rates

The analysis so far has focused on dispersion, or "second moments." We now 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. This will allow us to assess more directly which types of borrowers tend to "overpay" for their loans, and test different hypotheses for what is driving differences across borrowers and over time.

To do so, we use the data on lenders' offer rates (described in Section 3.2) to compute median offer rates by day, MSA, FICO, LTV, loan amount, and loan program (i.e. conforming, super-conforming, jumbo, and FHA). We then match these benchmark median offer rates to observations in the rate locks data with identical characteristics, and study the difference between the rate obtained by consumers and the median rate available—the *locked-offer rate gap*.[32]

We have fewer observations than in the previous analysis based on lock data only, since offer rates are only available for a subset of loan types/characteristics, 20 MSAs, and a shorter time period. In particular, this analysis is restricted to purchase mortgages, since we have the most granular offers for them. Appendix A.3 provides additional detail on the matching.

In our main analysis, we focus on the distance between the rate locked by a borrower and the rate available at the median lender, since we believe that this is a simple and transparent benchmark. However, in Appendix A.4 we consider an alternative measure that is more directly motivated based on search theory, namely the expected gain from obtaining one additional rate quote from a different lender. As we show there, the main results from this section are qualitatively identical when using this alternative measure.

Finally, one might be concerned that the distribution of offers could be a flawed benchmark for locked rates, if the best offers are not "achievable" for some reason. However, we note that overall, 5.3% of all borrowers obtain a rate in the best (bottom) 5% of their offer distribution. Even for FHA borrowers, which we will find below to do relatively poorly on average compared to the median available rate, this fraction is 3.5%. Thus, even the best offers are indeed available to borrowers.

5.1 Locked-Offer Rate Gaps by Borrower Type

The top panel of Figure 2 shows the distribution of the locked-offer rate gap for all mortgages in our data. The dashed vertical line denotes the mean of the distribution. The locked-offer rate gap is positive on average (dashed line just to the right of the thick black line that denotes zero), meaning that borrowers end up with mortgage rates that are more expensive than what the median lender

[32] We use the rate at which the median lender offers a loan with zero points and fees. To compare to this offer, we adjust the locked rate for points paid or received by the borrower based on the empirical relationship between points and interest rates. We estimate this relationship in a regression specification identical to column (10) of Table 2, with the only exception that points are allowed to only enter linearly (instead of entering in a binned fashion as in Table 2). In Figure A-3 in the appendix, we validate that our median offer rates derived from Optimal Blue Insights closely track offer rates for comparable loans published by Mortgage News Daily, an industry website.

could offer for identical mortgages.[33]

The bottom four panels of Figure 2 show the distribution of the locked-offer rate gap for various sub-segments of the market. Summary statistics for these distributions are given in Table 4. The locked-offer rate gap is largest for FHA loans, with an average of +25bp. This amounts to about 1.2% of the mortgage balance in upfront points/fees, or $2,400 for a typical FHA loan of $200k. Moreover, one-quarter of FHA borrowers overpay by 45bp or more relative to the median offer. In contrast, the distributions for super-conforming and jumbo mortgages look very different: the locked-offer rate gap is on average slightly negative at -4bp for super-conforming mortgages, and even more negative at -21bp for jumbo mortgages. Thus, in these two market segments, borrowers on average obtain relatively good deals, suggesting that they may be more sophisticated at shopping and negotiating. Note that the differences in average locked-offer gaps across market segments generally follow a similar pattern as the differences in dispersion seen in Table 3, but in some cases tell a more nuanced story: for instance, residual rate dispersion is identical in the conforming and jumbo segments, but jumbo borrowers fare significantly better relative to offers. This illustrates the value of studying the locked-offer rate gap, rather than relying on dispersion alone.

Table 4 further shows how the locked-offer rate gap distribution varies by FICO scores, LTV ratios, whether the borrower is a first-time homebuyer, whether the borrower paid or received points when taking out the loan, and whether we can classify the lender as independent nonbank or not. 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 660 on average pay 23bp more than what the median lender would offer for identical mortgages.

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 15bp more than what the median lender could offer them, while repeat homebuyers pay only 7bp more.

Borrowers that pay discount points (positive values in the table) tend to end up with a higher locked-offer rate gap than those who receive points (known as a rebate or credit) from the lender. Note that since we adjusted the mortgage note rate for points paid, this relationship is not "mechanical." Finally, the average rate gap is slightly higher when focusing on independent nonbanks only; we return to discussing potential differences across lender types below.

It is worth noting that within each of the groups in Table 4, there is substantial dispersion in the locked-offer rate gap, as shown in the table's final three columns. Thus, even for high-FICO or low-LTV borrowers, which on average have a gap close to zero, a non-trivial fraction of borrowers lock rates well above what the median lender could offer them. However, dispersion tends to be largest for the groups that on average fare the worst.

[33] Because the most popular lenders may be relatively expensive, and because there is substantial within-lender dispersion in rates as noted above, it is not necessarily surprising that the locked-offer gap would be positive.

Table A-5 provides analogous summary statistics based on the median income, college education share, minority share, and mortgage market concentration in a borrower's location.[34] In the first three cases, the observed differences between highest and lowest terciles are large: for instance, the tercile of borrowers in ZIP codes with the lowest fraction of college educated residents on average has a locked-offer rate gap of 16bp, while for the tercile with the highest fraction the average gap is only 5bp. Similarly, average gaps are larger in areas with lower median incomes and higher minority shares. In contrast, we do not find much evidence that average gaps increase in local mortgage market concentration, although the dispersion of gaps is larger in more concentrated markets.

5.2 Regression Analysis

Next, we turn to a regression analysis to investigate whether the differences across FICO and LTV groups in the locked-offer rate gap hold after controlling for certain loan characteristics, as well as fixed effects for the particular lender and branch to which borrowers went. For a subsample of loans, we can further control for loan officer compensation, which helps us assess whether differences in locked-offer rate gaps may be driven by low-FICO or high-LTV borrowers being "more work" for loan officers. Finally, we also test whether paying or receiving points is associated with getting a worse deal on the loan.

One potential explanation for the results in Table 4 is that lower-FICO borrowers and higher-LTV borrowers tend to have smaller loans and thus less of an incentive (in dollar terms) to shop around. In columns (1) and (4) of Table 5, we regress locked-offer gaps on bins for different FICO scores and LTV ratios, respectively, as well as fine loan amount bins and MSA-by-month fixed effects. It is indeed the case that borrowers with the largest loan amounts pay substantially less relative to their median offer rate (not shown in table). However, conditional on loan amount, lower-FICO borrowers and higher-LTV borrowers continue to pay more, to a similar degree as we observed in Table 4. Thus, such borrowers appear to obtain more expensive loans for reasons beyond the differential monetary incentive to shop stemming from loan size variation.

Another potential explanation for why low-FICO and high-LTV borrowers are more likely to pay too much is that they sort into more expensive lenders or branches. Borrowers might choose expensive lenders because they offer better service or simply because they spend more on marketing and are more visible. To investigate this explanation, we include branch fixed effects in columns (2) and (5) of Table 5. In these columns, the R-squared jumps sharply to about 50 percent from less than 20 percent, meaning that branch-specific pricing differences explain a fair amount of variation in the locked-offer gap. Furthermore, the coefficients on FICO and LTV become slightly smaller in magnitude, implying that sorting into lenders does explain some of the "overpayment" by low-FICO and high-LTV borrowers, but the coefficients remain large.

Thus, it does not appear that, for example, lower-FICO borrowers end up with higher locked-

[34] Income, education, and minority shares are measured at the ZIP code level based on 2017 American Community Survey data; mortgage market concentration is measured at the county level as the share of the largest four lenders (following Scharfstein and Sunderam 2016) in the 2016 HMDA data.

offer gaps just because they get their loans from more expensive lenders or branches. Even within the same branch, low-FICO and high-LTV borrowers tend to pay more relative to their benchmark median offer. One reason why this might occur is that (some of) these borrowers could be "more work" for loan officers, who therefore require additional compensation through a higher rate (or equivalently, more points upfront). Since by law the LO compensation can no longer depend on the interest rate paid by the borrower, the postulated channel would have to work through low-FICO and high-LTV borrowers being matched with LOs that "specialize in difficult cases" and charge more. In order to test for this possibility, in columns (3) and (6) we directly control for LO compensation (in % of the loan amount) for the subset of loans for which we observe it.[35] The coefficients on this variable are strongly significant, and their magnitude of about +0.15 suggests that higher LO compensation is reflected almost one-for-one in the rate the borrower pays (since we earlier noted that one percent of the loan amount—one point—corresponds to about 0.2% in rate terms). However, the coefficients on FICO and LTV remain similar, implying that low-FICO and high-LTV borrowers do not pay higher rates simply because they match with expensive LOs.

The final two columns of the table test whether borrowers who pay or receive points get a worse deal relative to the omitted category (those with points between -0.2 and +0.2).[36] Column (7) reproduces the result seen in Table 4 that borrowers who pay (receive) points tend to pay high (low) rates relative to what is available in the market. Column (8) shows that once we control for lender/branch, the coefficients on the dummies for having paid or received points are close to zero; this means that the overall relationship is driven by sorting into cheap/expensive lenders.

The main takeaway from this analysis is that low-FICO and high-LTV borrowers on average tend to pay substantially higher rates not just due to credit 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. This is illustrated in Figure 3, where the magnitude of the coefficients on FICO and LTV bins from columns (1) and (4) of Table 5 are compared to coefficients from a similar regression where we use the offer rates as dependent variable. We see that for FICO, the locked-offer rate gap is about one-third as large as the offer differences.[37] For LTV, it is in fact not the case that lender offers for high-LTV loans on average feature worse rates; if anything, the reverse is true. This may be surprising, but is mostly due to the fact that in the conforming segment, borrowers with LTVs above 80 are required to get private mortgage insurance, which effectively reduces the risk to the lender/GSE (at least in terms of loss-given-default).[38] Thus, what this analysis implies is that high-LTV borrowers only pay higher interest rates due to their less effective search/negotiation process, rather than due to differences in offer rates.

[35] We only observe loan officer compensation for a subset of lenders. LO compensation typically amounts to 1-2% of the loan amount originated.

[36] One reason why borrowers who pay/receive points might pay higher rates could be that lender offers become more difficult to compare than at zero points, so that suboptimal decisions become more likely.

[37] One reason for the higher offer rates for low-FICO borrowers is that the GSEs charge additional loan-level price adjustments for such loans.

[38] Reflecting this, the GSEs' loan-level price adjustments tend to be lower at LTVs above 80 than at 80.

Robustness. Table A-6 in the Appendix reproduces the same regressions for FHA loans only, and obtains similar results. Thus, the previous findings are not due simply to sorting into different loan programs, or driven by the different benchmark offer rates across programs.

In another robustness check, reported in Table A-7, we restrict the sample to lenders that we can identify as independent nonbanks. Doing so leaves the coefficients from Table 5 essentially unchanged. As noted earlier, the nonbank lenders that constitute the majority of our sample are only in the business of originating mortgages. Thus, the results on differential locked-offer gaps cannot be explained by potential price advantages that bank lenders might grant to financially well-off (high FICO, low LTV) customers, for instance because they also have significant account balances or other business with the bank.[39]

As mentioned at the beginning of this section, in Appendix A.4 we reproduce the two tables from this section using a borrower's expected gain from additional search instead of the locked-offer gap, and obtain qualitatively identical results.

Finally, it may be that most of the lenders making offers in our dataset are 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 our analysis using only offers from high-volume lenders, as designated on the Optimal Blue platform. Our results remain qualitatively unchanged.

6 Time-series Movements in the Locked-Offer Rate Gap

The last section explored the cross-sectional patterns in the locked-offer rate gap. In this section, we instead study how this gap moves over time, with a particular focus on how it responds to changes in market interest rates. Are borrowers more likely to end up with worse rates (relative to the median lender offer) when market rates are low, and more likely to get a good deal as rates increase? If so, what might explain this relationship?

Figure 4 plots the average locked-offer rate gap against market interest rates, here measured by the 10-year Treasury yield.[40] In the summer of 2016, the level of market interest rates as shown by Treasury yields was very low. The locked-offer rate gap during this time was high, meaning that borrowers were locking rates from the higher end of the offer rate distribution. As Treasury yields increased, and as a result lenders increased their offer rates, the locked-offer gap shrunk, indicating that borrowers moved toward the cheaper end of the offer distribution. When rates fell again starting in late 2018, the inverse happened. Overall, the movements in the locked-offer gap almost mirror movements in the Treasury yields.

[39] Table 4 showed that borrowers who obtain loans from nonbanks tend to have higher locked-offer gaps; this could either reflect overall advantageous pricing by banks/credit unions, or selection by borrowers. What we emphasize here is that such differential pricing, if it exists, does not appear to vary with borrower creditworthiness.

[40] For the average locked-offer gap, we use the estimated month fixed effects from a regression similar to those in Table 5 but controlling simultaneously for FICO and LTV. We use the 10-year Treasury yield as our measure of market rates 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.

We confirm the statistical significance of the relationship between the locked-offer rate gap and market rates in Table 6.[41] The first column adds the 10-year Treasury yield as a control to a regression similar to the ones estimated in Table 5 (controlling for all borrower characteristics jointly, along with MSA fixed effects). The coefficient implies that as the 10-year Treasury yield increases by 1 percentage point, the average locked-offer gap falls by about 6bp. 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.

In column (2) we add month fixed effects (interacted with MSA) and see that the relationship between Treasury yields and the locked-offer rate gap is even stronger within-month—the magnitude of the coefficient increases to 8.3bp. Column (3) further adds branch fixed effects, to see to what extent the estimated relationship gets weaker once we control for potentially time-varying selection of borrowers into expensive or cheap lenders/branches. The coefficient on the Treasury yield is reduced (to 5.7bp), suggesting that some of the overall relationship may be due to borrowers selecting cheaper lenders when rates are higher (consistent with additional shopping).

Next, we test the hypothesis 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.[42] To study whether this is likely to be an important factor behind the relationship, we separate borrowers into those with a DTI up to 36 percent (who are likely unconstrained by the payment burden) and those with a higher DTI (for whom a higher rate may mean they run up against underwriting constraints).[43]

We thus repeat the same regressions, allowing for separate coefficients on the Treasury yield depending on whether a borrower's DTI is above 36 or not. Across columns (4) to (6), we see that the estimated coefficient on the Treasury yield is indeed slightly more negative for the high-DTI borrowers, suggesting that affordability constraints play some role in the relationship. However, the coefficient on the Treasury yield remains sizeable even for those borrowers that are most likely not constrained by the payment burden.

This suggests that the relationship may be driven at least partly 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. In Section 7.3, we will show that according to survey data, shopping effort does indeed increase when market rates are higher.

Importantly, the higher locked-offer gap when market interest rates are low is in addition to the

[41] Appendix Table A-10 repeats the analysis in this section with the expected gain from search as the dependent variable, which leaves the qualitative results unchanged. This implies that the results cannot be explained by time variation in the width of the offer distribution.

[42] 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., 2020). 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.

[43] Using alternative DTI cutoffs to separate borrowers, e.g. 43 percent, leaves the results qualitatively unchanged.

higher "price of intermediation" when rates are low, identified by Fuster et al. (2017). That paper shows that *offers* feature higher lender markups (relative to loan values in the MBS market) at times of high demand, and provide evidence that this is at least in part driven by lender capacity constraints. Thus, there are two complementary reasons why after a drop in market rates, borrowers obtain worse mortgage rates than they could in a frictionless world: lenders make worse offers relative to an MBS-market benchmark, *and* borrowers fare worse relative to those offers.

The hypothesis that borrowers anchor their beliefs about mortgage rates to a reference rate also has implications for the cross-section of the locked-offer rate gap. If borrowers use a heavily advertised rate, such as the prime conforming rate, as a reference rate, they should be willing to search/negotiate more when the offer rates in their program are high relative to this reference rate. Therefore, the locked-offer rate gap should be low when the gap between the offer rates in other programs and the advertised prime conforming rate is high. We test this hypothesis in the last two columns of Table 6.

We first compute a daily median offer rate for a typical borrower in each program.[44] Then, we calculate the spread between that program-specific offer rate and the prime conforming offer rate from Optimal Blue; by construction this spread is zero for borrowers in the conventional conforming market.[45] Specification (7) regresses the locked-offer rate gap on the offer spread to the prime conforming rate and shows that the coefficient of interest is negative and statistically significant. Specification (8) also includes Treasury yields to control for the overall level of the interest rates and the coefficient is unchanged. This mechanism may also partly explain why FHA borrowers overpay more and jumbo borrowers underpay less than standard conforming borrowers: the offer rates are typically lower for FHA than conforming (by 33bp on average over our sample) but higher for jumbo than conforming (by 7bp on average).[46] Overall, these results support the hypothesis the variation in the locked-offer rate gap is likely to be driven by behavioral phenomena, such as anchoring to the level of a salient and observable mortgage rate.

7 Survey Evidence on Shopping, Knowledge, and Mortgage Rates

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 a 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

[44] For conforming, super-conforming, and jumbo loans, we compute the daily median offer rate for a borrower with LTV=80, FICO=750, DTI=36. For FHA loans we compute it for a borrower with LTV=96, FICO=680, DTI=36.

[45] Our results are robust to different choices for reference rate such as the Freddie PMMS rate, the Bankrate prime rate, Mortgage News Daily 30-year fixed rate, or MBS yields.

[46] FHA rates are lower because they do not include the insurance premium, which borrowers need to pay separately.

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.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 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 borrowers 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.

7.1 The Relationship between Shopping, Knowledge, and Contract Rates

We estimate OLS regressions of the form

$$RateSpread_{ijtw} = \beta X_i + \Gamma Z_{ij} + \alpha_t + \delta_w + \epsilon_{ijtw} \tag{1}$$

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.[47] 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 7; it contains for instance flexible controls for FICO and LTV, fixed effects for MSA, loan term, program (e.g. GSE or FHA) and purpose (purchase or refinance), as well as borrower income, education, age, and race.[48] We further include origination month fixed effects α_t and survey wave

[47] 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. In the public dataset, the gap is truncated at -1.5 and +1.5 percentage points; however, we were able to run the analyses described in this section at FHFA on a version of the data without truncation (and containing MSA indicators, which we are able to use as fixed effects). In earlier drafts, we reported results based on the public version of the data (version as of February 12, 2019), with little qualitative difference.

[48] One limitation of the NSMO data is that it does not contain a direct measure of points paid or received by the borrower. However, the controls for borrower wealth and expected time in the mortgage should help absorb

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?" 49.0% of respondents (weighted) answer 1, 35.3% 2, 13.0% 3, 1.7% 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, 76.7% answer 1, 18.7% 2, 3.6% 3, 0.7% 4, and 0.3% 5 or more. We combine the last four groups into "2+".

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 (81.4% of those that apply to more than one lender, or 18.6% of the sample overall).[49]

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" (32.7% a little, 9.2% a lot); "Websites that provide information on getting a mortgage" (32.1% a little, 22.2% a lot); and "Friends/relatives/co-workers" (32.0% a little, 15.1% 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?" 61.7% respond "Very", 32.9% "Somewhat", and 5.5% "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.

differences in rates due to variation in points.

[49] Overall, 4.6% of respondents stated that they applied to more than one lender because they got "turned down on an earlier application"; 6.7% because of "concern over qualifying for a loan"; and 7.4% because of "information learned from the 'loan estimate'," with overlap across these categories.

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, 68.2% 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 7. 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 9.5bp lower rate, while applying to more than one lender in search of better loan terms is associated with a 7bp 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 20bp 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 a 6bp 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 low rate.[50] 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. Conversely, shopping alone does not guarantee a good rate if a borrower's knowledge is low (see also Malliaris et al., 2020). 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.

In Appendix A.5, we provide a complementary analysis using data from the 2016 Survey of Consumer Finances (SCF). Consistent with the NSMO results, we find that borrowers who report shopping more, and borrowers with high financial literacy—based on their answers to the Lusardi-Mitchell financial literacy questions—get significantly lower interest rates, even after controlling for loan characteristics, borrower credit risk, and borrower demographics.

7.2 Who Pays More Because of a Lack of Shopping or Knowledge?

The previous subsection provides evidence that more intense mortgage shopping and more knowledge about the mortgage market is associated with lower contracted rates. We next ask which

[50]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), in line with the findings of Agarwal et al. (2019).

observable borrower and loan characteristics are associated with stronger reported shopping intensity and mortgage knowledge, resulting in lower interest rates. To do this, we first isolate the part of the interest rate spread that can be attributed solely 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 using a regression almost identical to the one in specification (10) of Table 7. 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 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 spread relative to a hypothetical borrower that indicates that they did not engage in any shopping-related activities and have the poorest possible understanding of the mortgage market.

We summarize this predicted rate spread in the top row of Table 8. Due to shopping and mortgage knowledge, the average borrower pays 35bp 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 26bp. This implies that there are substantive differences across borrowers in shopping behavior and mortgage knowledge amounting to a 26bp difference in rates paid.

If shopping and mortgage knowledge are correlated with borrower and loan characteristics, then interest savings will differ by group. In Table 8 we also show group-specific predicted interest rate spreads that can be attributed to shopping and mortgage knowledge. The differences across groups are most pronounced at the lower end of the group-specific shopping/knowledge distribution. For example, at the 10th percentile, borrowers in the jumbo market pay about 12bp less than FHA borrowers due to shopping and mortgage knowledge, whereas the jumbo-FHA difference is about 3bp at the 90th percentile.[51] In other words, the gap in knowledge and shopping is not as big between the most savvy FHA and jumbo borrowers as the gap between the least savvy FHA and jumbo borrowers.

The table further shows that the predicted rate spread decreases in a borrower's FICO score and increases in the LTV, meaning that low-FICO and high-LTV borrowers pay higher rates due to shopping and knowledge. The same is true for borrowers with low loan amounts.

Turning to other borrower characteristics, borrowers with incomes of $175k or higher pay less due to shopping and knowledge than borrowers with incomes of less than $35k, with a 13bp difference at the 10th percentile. In addition, 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 across groups in Table 8 may appear relatively small. However, it bears remembering that the right-hand-side variables of the underlying regression are coarse

[51] To be clear, the percentiles are calculated within group, so across-group differences are driven only by differences in shopping/knowledge.

responses to qualitative survey questions, likely leading to substantial individual-specific noise and attenuation of the resulting coefficients.[52]

With that caveat in mind, we believe the findings here lend considerable support to the mechanism we postulated in our earlier analysis using the rate locks and offers data. Namely, at least some of the overpayment by many borrowers is likely due to ineffective shopping and negotiation, reflecting a lack of financial sophistication and knowledge of the market. Such knowledge is particularly important in this setting where, as documented earlier, there is considerable dispersion in prices across lenders, and even across branches and loan officers of the same lender. Furthermore, comparing offers is complicated due to the multi-dimensional pricing with upfront points.

7.3 Time-series Variation in Shopping Intensity

Earlier, we saw that the locked-offer 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:

$$Shopping_{ijtw} = \beta \cdot PMMS_{it} + \Gamma Z_{ij} + \delta_w + \epsilon_{ijtw} \tag{2}$$

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 7: (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 7, except for some variables that seem likely endogenous to the shopping effort itself.[53] Furthermore, we add the knowledge variables used in Table 7 as well.

Panel A of Table 9 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 a statistically significant way. For instance, column (1) implies that a 1 percentage point increase in market mortgage rates increases the probability that a borrower

[52] 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.

[53] These variables are whether a borrower obtained their mortgage through a broker, the term of the loan, and whether it has an adjustable rate. We also do not include MSA fixed effects, though adding them has minimal effects.

considered more than one lender by 4.5 percentage points, relative to a sample average of 51 percent.[54] Column (2) shows that this coefficient is unaffected by the addition of fine borrower- and loan-level controls, 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 loans; this alleviates 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 affordability 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 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 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 Appendix A.6, 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.

8 Conclusion and 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. The involved dollar amounts in this scenario are large not just at the individual level but also in aggregate: for instance, if the average locked-offer rate gap of FHA borrowers moved to zero (assuming nothing else changes in the market structure), this would amount to savings of

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

about \$2.75 billion/year for these borrowers.[55]

Given our findings, future research should consider the effects of 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, future research could study whether the problem can be alleviated if the guaranteeing agencies were to impose requirements on the maximum locked-offer gap they allow for loans to be securitized. Of course, to understand the effectiveness of such policies one would need to consider general equilibrium effects on the offers that lenders make (as in Alexandrov and Koulayev 2017, Agarwal et al. 2019, or Guiso et al. 2020).

The negative relationship between the average locked-offer rate gap and the level of market rates that we document in Section 6 also matters 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 rates available in the market, likely at least in part 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 expansive monetary policy to the mortgage market.[56] On the other hand, the pass-through of increases in policy rates to rates on new mortgages may be dampened by more intense borrower shopping. This could be good or bad news for monetary policy makers, depending on whether slowing the housing market through higher mortgage rates is seen as desirable in a given situation or not.

[55] This calculation is based on average FHA originations over 2015-2019 of about \$230bn/year (see `https://www.hud.gov/sites/dfiles/Housing/documents/FHA_SF_MarketShare_2019Q3.pdf`) multiplied by 1.2 points, which is the upfront equivalent of the average locked-offer rate gap of +25 bp that we documented.

[56] 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., 2017a; 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., 2020; Keys et al., 2016).

References

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," NBER Working Paper 21512 2017.

_ , Itzhak Ben-David, and Vincent Yao, "Systematic mistakes in the mortgage market and lack of financial sophistication," *Journal of Financial Economics*, 2017, *123* (1), 42 – 58.

_ , John Grigsby, Ali Hortaçsu, Gregor Matvos, Amit Seru, and Vincent Yao, "Searching for Approval," Working Paper 2019.

_ , Richard J. Rosen, and Vincent Yao, "Why Do Borrowers Make Mortgage Refinancing Mistakes?," *Management Science*, 2015, *62* (12), 3494–3509.

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, "Sources of Inaction in Household Finance: Evidence from the Danish Mortgage Market," *American Economic Review*, 2020, *110* (10), 3184–3230.

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

Bhutta, Neil and Aurel Hizmo, "Do Minorities Pay More for Mortgages?," *Review of Financial Studies*, 2020, *forthcoming.*

Buchak, Greg, Gregor Matvos, Tomasz Piskorski, and Amit Seru, "Fintech, Regulatory Arbitrage, and the Rise of Shadow Banks," *Journal of Financial Economics*, 2018, *130*, 453–483.

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

Carlson, John A. and R. Preston McAfee, "Discrete Equilibrium Price Dispersion," *Journal of Political Economy*, 1983, *91* (3), 480–493.

CFPB, "Data Point: 2019 Mortgage Market Activity and Trends," Technical Report, Consumer Financial Protection Bureau 2020.

Choi, James J., David Laibson, and Brigitte C. Madrian, "Why Does the Law of One Price Fail? An Experiment on Index Mutual Funds," *Review of Financial Studies*, 2010, *23* (4), 1405–1432.

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," *Review of Economic Studies*, 2020, *87* (2), 914–958.

Duncan, Doug, "Shopping Around for a Mortgage Pays Off for Consumers," Fannie Mae Perspectives, https://www.fanniemae.com/portal/research-insights/perspectives/shopping-mortgage-consumers-duncan-080519.html August 2019.

Fuster, Andreas, Laurie Goodman, David Lucca, Laurel Madar, Linsey Molloy, and Paul Willen, "The Rising Gap between Primary and Secondary Mortgage Rates," *Federal Reserve Bank of New York Economic Policy Review*, 2013, *19* (2), 17–39.

_ , 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.

Gathergood, John and Jörg Weber, "Financial literacy, present bias and alternative mortgage products," *Journal of Banking & Finance*, 2017, *78*, 58–83.

Gomes, Francisco, Michael Haliassos, and Tarun Ramadorai, "Household Finance," *Journal of Economic Literature*, 2020, *forthcoming.*

Goodman, Laurie S, "Quantifying the tightness of mortgage credit and assessing policy actions," *BCJL & Soc. Just.*, 2017, *37*, 235.

Guiso, Luigi, Andrea Pozzi, Anton Tsoy, Leonardo Gambacorta, and Paolo Emilio Mistrulli, "The Cost of Steering in Financial Markets: Evidence from the Mortgage Market," Working Paper 2020.

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

Hastings, Justine S., Brigitte C. Madrian, and William L. Skimmyhorn, "Financial Literacy, Financial Education, and Economic Outcomes," *Annual Review of Economics*, 2013, *5* (1), 347–373.

Hortaçsu, Ali and Chad Syverson, "Product Differentiation, Search Costs, and Competition in the Mutual Fund Industry: A Case Study of S&P 500 Index Funds," *Quarterly Journal of Economics*, 2004, *119* (2), 403–456.

Huston, Sandra J., "Financial literacy and the cost of borrowing," *International Journal of Consumer Studies*, 2012, *36* (5), 566–572.

Iscenko, Zanna, "Choices of dominated mortgage products by UK consumers," Occasional Paper 33, Financil Conduct Authority 2018.

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

Liu, Lu, "Non-Salient Fees in the Mortgage Market," Working Paper, Imperial College London 2019.

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.

Malliaris, Steven G., Daniel A. Rettl, and Ruchi Singh, "Is Competition a Cure for Confusion? Evidence from the Residential Mortgage Market," Working Paper, University of Georgia 2020.

McManus, Doug, Liyi Liu, and Mingzhe Yi, "Why Are Consumers Leaving Money On The Table?," *Freddie Mac Economic & Housing Research Insight,* *http://www.freddiemac.com/research/insight/20180417_consumers_leaving_money.page*, 2018.

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," *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)	255	94	544	71	720	262	222	92
Interest Rate	4.33	0.51	4.31	0.47	4.21	0.50	4.30	0.61
Discount Points Paid	0.15	0.95	0.28	0.97	0.19	0.74	0.06	1.14
FICO	742	47	750	41	763	33	669	47
LTV	81	14	80	12	77	10	93	8
DTI	35	9	36	9	31	9	42	10
First-time Homebuyer %	24		23		11		49	
Refinance Share %	31		33		33		17	
N. observations	2,316,400		119,894		76,941		1,092,535	

Data Source: Optimal Blue

Table 2: Unpacking the Dispersion in Locked Interest Rates

	Underwriting Grid			Add Lender Controls			Add Branch Controls		Add LO Controls	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Standard Deviation	0.33	0.26	0.24	0.22	0.20	0.18	0.16	0.15	0.14	0.13
75-25 Percentile	0.36	0.28	0.26	0.22	0.20	0.17	0.15	0.14	0.13	0.12
90-10 Percentile	0.78	0.58	0.54	0.48	0.44	0.38	0.33	0.31	0.29	0.26
Underwriting Variables Grid										
Lock Date x MSA F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FICO x LTV x Program x Lock Month F.E.		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
ZIP Code F.E.		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Discount Points x Program x Lock Month F.E.			Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Add Lender Controls										
Lender F.E.				Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lender x Lock Date F.E.					Yes	Yes	Yes	Yes	Yes	Yes
Lender x FICO x LTV x Program x Lock Month F.E.						Yes	Yes	Yes	Yes	Yes
Lender x Points x Lock Month F.E.						Yes	Yes	Yes	Yes	Yes
Add Branch Controls										
Branch F.E.							Yes	Yes	Yes	Yes
Branch x Lock Month F.E.								Yes	Yes	Yes
Add Loan Officer Controls										
Loan Officer F.E.									Yes	Yes
Loan Officer x Program F.E.										Yes
Loan Officer x Lock Year F.E.										Yes
Adj. R-Squared	0.59	0.75	0.78	0.81	0.83	0.85	0.88	0.89	0.89	0.90
Observations	2959539	2959539	2959539	2959539	2959539	2959539	2959539	2959539	2959539	2959539

Data Source: Optimal Blue

Notes: The dependent variable is the mortgage interest rate locked. The data covers mortgage rates locked for 277 metropolitan areas during the period between 2015-2019. We focus on 30 year, fixed rate, fully documented mortgages. “Program” refers to 12 dummy variables representing four loan programs interacted with three loan purposes. Specifications (2)-(10) also include lock period f.e., property type f.e., cubic functions of loan amount and DTI, as well as linear functions of FICO, LTV, and (from specification (3) onward) discount points. For MSAs that span across multiple states we include MSA x State fixed effects.

Table 3: Summary Statistics of the Residualized Locked Rate

		$90^{th} - 10^{th}$ Percentile Gap	
	Observations	Spec. (3) of Table 2	Spec. (10) of Table 2
All Mortgages	2,959,539	0.54	0.26
Program			
FHA	876,640	0.71	0.31
Conforming	1,972,913	0.47	0.24
Super-Conforming	72,038	0.43	0.21
Jumbo	37,948	0.47	0.24
FICO			
< 600	41,836	0.92	0.46
$[600, 640)$	219,492	0.81	0.36
$[640, 680)$	487,004	0.67	0.30
$[680, 740)$	921,992	0.55	0.27
≥ 740	1,289,215	0.44	0.22
LTV			
≤ 75	500,902	0.46	0.22
$(75, 80]$	619,803	0.46	0.23
$(80, 95]$	951,705	0.50	0.25
>95	837,151	0.71	0.32
First-Time Homebuyer			
No	1,981,043	0.50	0.24
Yes	978,020	0.64	0.31
Loan Purpose			
Purchase	2,269,213	0.55	0.27
Cashout	344,735	0.56	0.25
Rate Refi	345,591	0.50	0.23
Lender Type			
Independent Non-bank	1,878,780	0.54	0.26
Other/Unclassified	1,080,759	0.54	0.26

Data Source: Optimal Blue

Notes: This table summarizes the residualized locked mortgage rate from specifications (3) and (10) of Table 2.

Table 4: Summary Statistics of the Locked-Offer Rate Gap

	Observations	Mean	St. Deviation	Percentiles	
				25^{th}	75^{th}
All Mortgages	64,788	0.11	0.31	-0.07	0.26
Program					
FHA	14,441	0.25	0.38	0.02	0.45
Conforming	44,040	0.09	0.27	-0.06	0.22
Super-Conforming	4,478	-0.04	0.26	-0.20	0.09
Jumbo	1,829	-0.21	0.32	-0.34	-0.06
FICO					
$[640, 660)$	7,406	0.23	0.40	-0.02	0.45
$[680, 700)$	9,390	0.16	0.36	-0.05	0.35
$[720, 740)$	10,207	0.12	0.30	-0.06	0.26
740+	37,785	0.07	0.27	-0.08	0.21
LTV					
$(75, 80]$	21,334	0.02	0.26	-0.12	0.16
$(85, 90]$	6,882	0.06	0.28	-0.08	0.20
$(90, 95]$	15,782	0.08	0.27	-0.08	0.22
$(95, 97]$	20,790	0.24	0.36	0.01	0.42
First-Time Homebuyer					
No	32,437	0.07	0.28	-0.09	0.21
Yes	32,345	0.15	0.34	-0.05	0.32
Discount Points					
$[-5, -0.2)$	14,015	0.01	0.33	-0.16	0.19
$[-0.2, 0.2]$	22,735	0.08	0.29	-0.09	0.22
$(0.2, 5]$	28,038	0.18	0.30	-0.00	0.31
Lender Type					
Independent Nonbank	45,618	0.13	0.30	-0.04	0.27
Other/Unclassified	19,170	0.05	0.32	-0.13	0.23

Data Source: Optimal Blue

Notes: 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 offer rate (the "locked-offer rate gap"). In the discount points category, negative values mean that the borrower receives points (also known as a rebate or credit) while positive values mean that the borrower pays points.

Table 5: Regressions of the Locked-Offer Rate Gap on Borrower/Loan Characteristics, Lender-Branch Fixed Effects, and Loan Officer Compensation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
FICO (omitted cat.: [640,660))								
$I_{680 \leq FICO < 700}$	-0.056***	-0.044***	-0.043***					
	(0.007)	(0.006)	(0.011)					
$I_{720 \leq FICO < 740}$	-0.088***	-0.059***	-0.058***					
	(0.010)	(0.008)	(0.013)					
$I_{FICO \geq 740}$	-0.123***	-0.080***	-0.071***					
	(0.011)	(0.009)	(0.013)					
LTV (omitted cat.: (60,80])								
$I_{85 < LTV \leq 90}$				0.017***	0.009	0.018**		
				(0.005)	(0.006)	(0.008)		
$I_{90 < LTV \leq 95}$				0.049***	0.033***	0.031***		
				(0.006)	(0.006)	(0.008)		
$I_{LTV > 95}$				0.178***	0.138***	0.101***		
				(0.012)	(0.011)	(0.019)		
Discount Points								
$I_{-5 < Points < -0.2}$							-0.094***	0.001
							(0.020)	(0.006)
$I_{0.2 < Points \leq 5}$							0.108***	0.034***
							(0.013)	(0.012)
Loan Officer Comp (%)			0.158***			0.140***		
			(0.036)			(0.039)		
Loan amount f.e. ($10k bins)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
MSA x Month f.e.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Branch f.e.		Yes	Yes		Yes	Yes		Yes
Adj. R-Squared	0.116	0.480	0.442	0.154	0.505	0.456	0.157	0.476
Observations	64693	62783	14659	64693	62783	14659	64693	62783

Data Source: Optimal Blue

Notes: 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 2016-2019. We focus on 30 year, fixed rate, fully documented purchase mortgages. Standard errors shown in parentheses are two-way clustered at the month and lender level. Significance: * p<0.1, ** p<0.05, *** p<0.01.

Table 6: The Relationship Between the Locked-Offer Gap and Treasury Yields

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treasury Yield	-0.059***	-0.083***	-0.057***					-0.058***
	(0.007)	(0.015)	(0.014)					(0.014)
Offer Spread to Prime Conforming Rate							-0.176***	-0.177***
							(0.026)	(0.026)
Treasury Yield ×								
DTI > 36				-0.066***	-0.090***	-0.064***		
				(0.008)	(0.016)	(0.015)		
DTI ≤ 36				-0.049***	-0.074***	-0.046***		
				(0.009)	(0.015)	(0.012)		
Borrower and Loan Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
MSA F.E	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
MSA x Month F.E.		Yes	Yes		Yes	Yes	Yes	Yes
Branch F.E.			Yes			Yes	Yes	Yes
Adj. R-Squared	0.145	0.156	0.505	0.147	0.157	0.505	0.509	0.508
Observations	64397	64317	62398	64397	64317	62398	62783	62398
P-val. for equality of DTI coefficients				0.024	0.031	0.010		

Data Source: Optimal Blue

Notes: 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 offer spread to conforming rate is defined as the average offer rate for a typical borrower in the same program in the same day minus the average offer rate for a typical prime conforming borrower. All specifications include controls for FICO, LTV, and loan amount. 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. Standard errors shown in parentheses are two-way clustered at the month and lender level. Significance: * p<0.1, ** p<0.05, *** p<0.01.

Table 7: Relationship Between Mortgage Rates and Measures of Shopping and Knowledge

Dep. var.: Interest rate spread (%)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Seriously considered 2 lenders	-0.047***									-0.021
	(0.012)									(0.013)
Seriously considered 3+ lenders	-0.095***									-0.049***
	(0.015)									(0.018)
Applied to 2+ lenders		-0.052***								0.052*
		(0.013)								(0.029)
Applied to 2+ lenders in search of better loan terms			-0.073***							-0.092***
			(0.014)							(0.031)
Used other lenders/brokers to get info? A little				-0.030**						-0.001
				(0.012)						(0.013)
Used other lenders/brokers to get info? A lot				-0.071***						-0.022
				(0.019)						(0.020)
Used web to get info? A little					-0.053***					-0.042***
					(0.012)					(0.013)
Used web to get info? A lot					-0.078***					-0.043***
					(0.015)					(0.015)
Used friends/relatives to get info? A little						0.001				0.004
						(0.013)				(0.013)
Used friends/relatives to get info? A lot						0.009				0.012
						(0.018)				(0.018)
Familiar with mortgage rates? Somewhat							-0.099***			-0.078**
							(0.033)			(0.033)
Familiar with mortgage rates? Very							-0.197***			-0.145***
							(0.033)			(0.033)
Index of mortgage knowledge (Std)								-0.061***		-0.044***
								(0.006)		(0.006)
Most lenders offer same rate? Yes									0.033**	0.024
									(0.016)	(0.016)
Adj. R2	0.14	0.13	0.13	0.13	0.14	0.13	0.14	0.14	0.13	0.14
Obs.	19824	19824	19824	19824	19824	19824	19824	19824	19824	19824

Data Source: National Survey of Mortgage Originations

Dependent variable: spread between a borrower's mortgage interest rate and the market mortgage rate prior to origination (as measured by PMMS), in percentage points. 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, MSA 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. N =19,824 instead of 19,906 because singleton observations are dropped. Robust standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table 8: Summary Statistics of the Interest Rate Spread that Can Be Attributed to Shopping and Mortgage Knowledge

	Observations	10^{th} Percentile	Mean	90^{th} Percentile
All Mortgages	19,906	-0.21	-0.35	-0.47
Program				
Conforming	11,103	-0.22	-0.36	-0.47
Jumbo	679	-0.29	-0.39	-0.48
FHA	2,734	-0.17	-0.32	-0.45
FICO				
$\leq$ 600	411	-0.17	-0.31	-0.45
601-640	1,089	-0.18	-0.32	-0.45
641-680	2,195	-0.20	-0.34	-0.46
681-740	4,784	-0.20	-0.34	-0.46
$>$ 740	11,427	-0.24	-0.36	-0.47
LTV				
$\leq$ 75	8,216	-0.23	-0.36	-0.47
76-80	3,551	-0.23	-0.36	-0.47
81-95	4,551	-0.21	-0.35	-0.47
96-97	1,805	-0.15	-0.31	-0.44
Loan Amount				
<100k	3,011	-0.17	-0.32	-0.45
[100k, 200k)	7,736	-0.21	-0.34	-0.46
[200k, 300k)	4,656	-0.22	-0.36	-0.47
[300k, 400k)	2,405	-0.24	-0.37	-0.48
$\geq$ 400k	2,098	-0.27	-0.38	-0.48
First-Time Homebuyer				
No	16,717	-0.23	-0.36	-0.47
Yes	3,189	-0.15	-0.31	-0.46
Income				
<35k	1,189	-0.14	-0.30	-0.43
[35k, 75k)	6,014	-0.18	-0.32	-0.45
[75k, 175k)	9,752	-0.23	-0.36	-0.47
$\geq$ 175k	2,951	-0.27	-0.39	-0.48
Education				
Less than college	3,322	-0.17	-0.31	-0.44
Some college	3,975	-0.20	-0.34	-0.46
College grad	7,017	-0.22	-0.36	-0.47
Postgrad	5,592	-0.24	-0.37	-0.48

Data Source: National Survey of Mortgage Originations
Notes: The variable we are summarizing here is the interest rate spread that is only due to shopping and knowledge about the mortgage market (so a more negative value is better from the perspective of a borrower). 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 7 (see text for details).

Table 9: Relationship Between Various Binary Measures of Mortgage Shopping and Mortgage 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 $\leq$ *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

Data Source: National Survey of Mortgage Originations

Notes: 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.

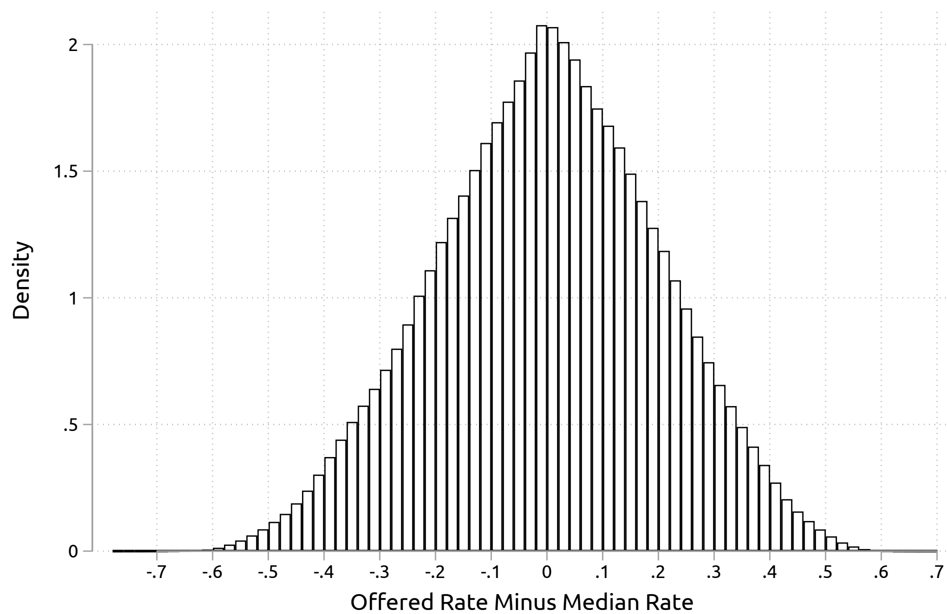


Figure 1: Offer Dispersion for Identical Mortgages

Data Source: Optimal Blue
Note: Figure shows the distribution of real-time offered interest rates, where for each offer rate we subtract the median offered rate across lenders for an identical mortgage in the same metropolitan area. The histogram includes data between April 2016 and December 2019 from 20 metropolitan areas for 52 combinations of loan characteristics (FICO, LTV, program, loan amount).

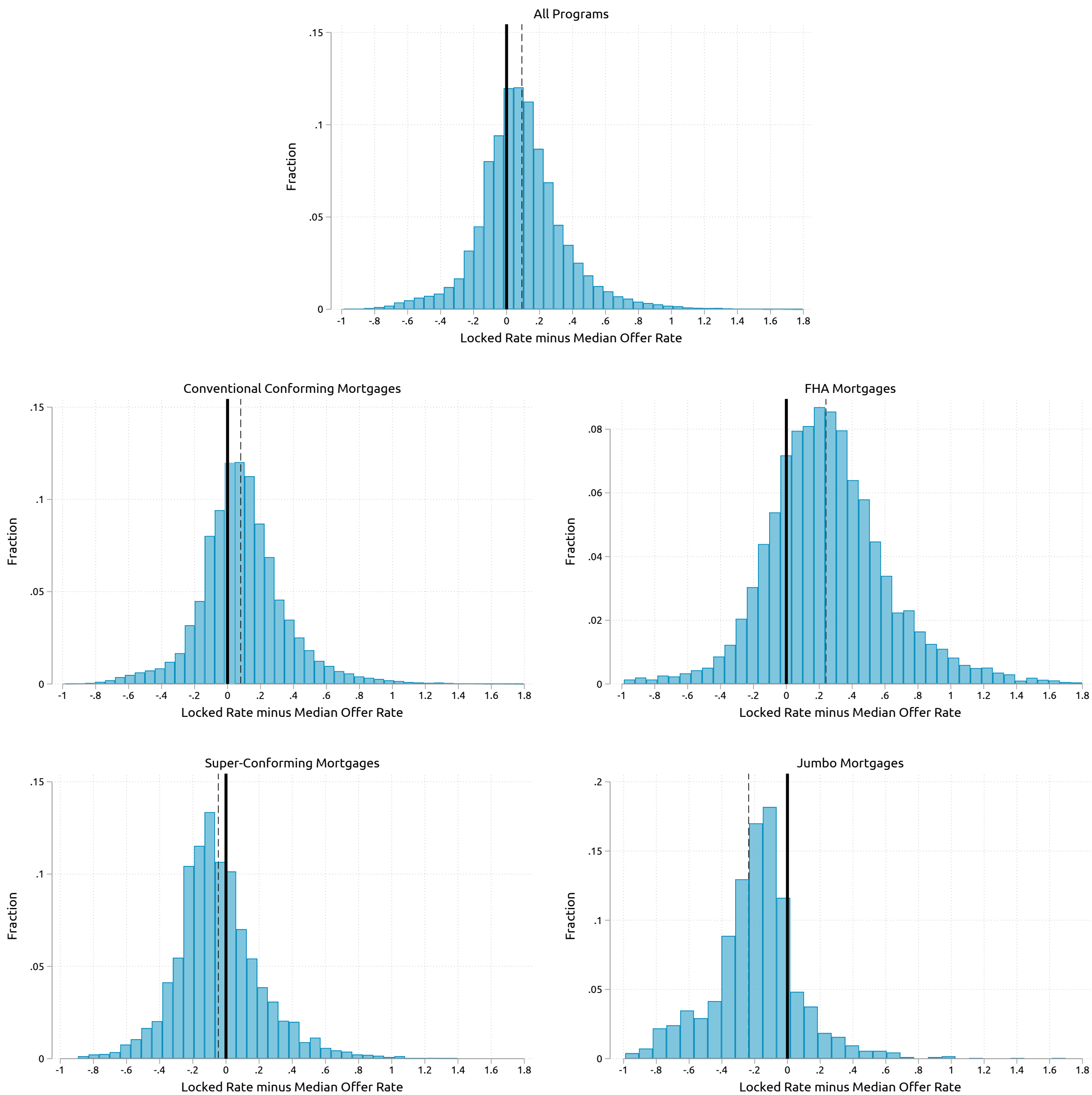
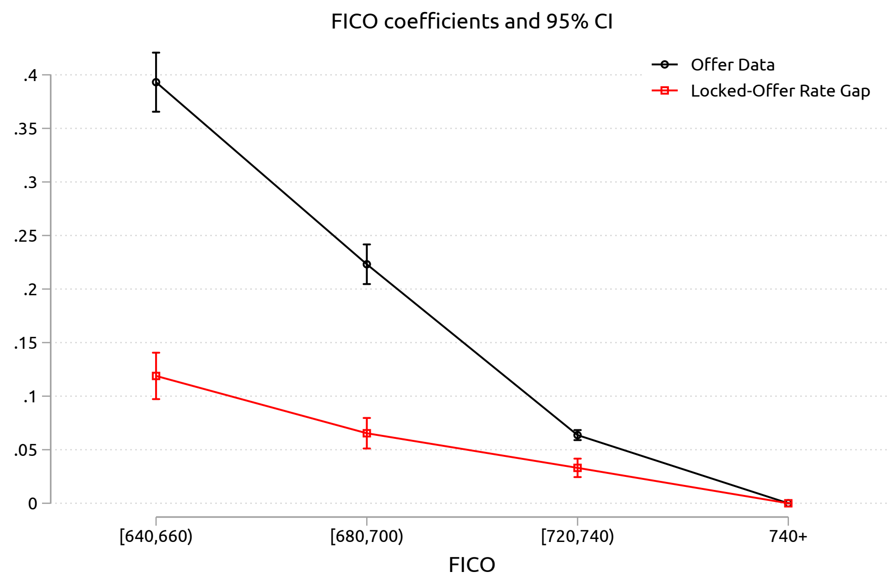


Figure 2: Distribution of Rate Locked Minus the Median Offer Rate for Identical Mortgages

Data Source: Optimal Blue
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 figure shows the distribution of the difference between each locked rate and the median offer rate. The dashed line denotes the mean of the distribution.



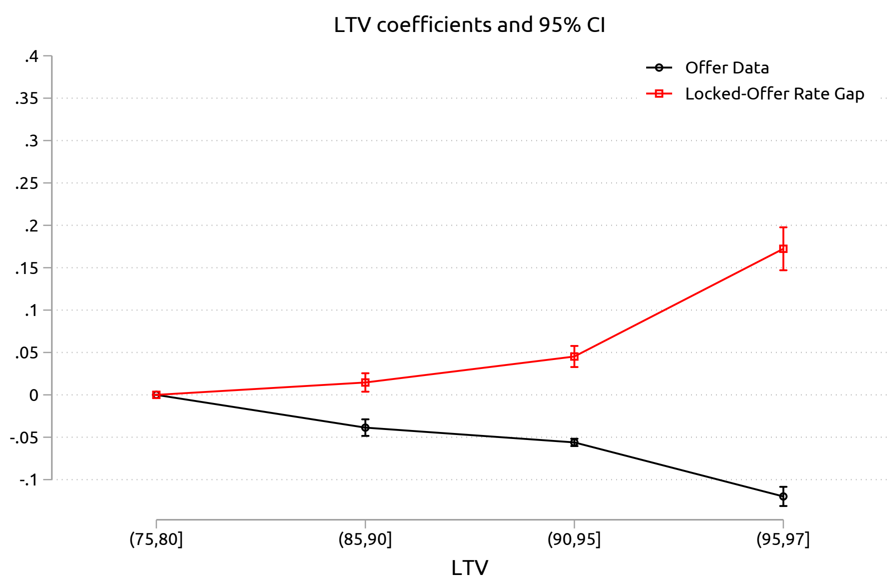


Figure 3: Comparing Locked-Offer Rate Gap to Offer Differences for Different FICO and LTV Levels

Data Source: Optimal Blue
Note: The red squares plot the coefficients on FICO bins and LTV bins from columns (1) and (4) of Table 5, where the dependent variable is the locked-offer rate gap. The black circles are the corresponding coefficients from a regression where the dependent variable is the offer rate (and where program fixed effects are also included).

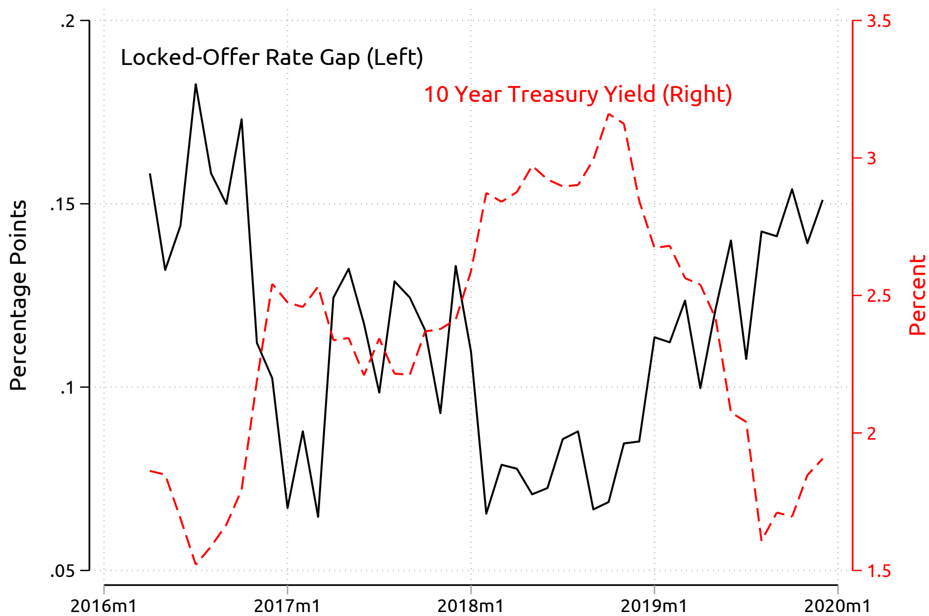


Figure 4: The Evolution of Rate Locked Minus the Median Offer Rate and Treasury Yields

Data Source: Optimal Blue
Note: The dashed red line is the 10 year Treasury yield. The solid black line is the monthly average locked-offer gap after controlling for borrower and loan characteristics. For the average locked-offer gap, we use the estimated month fixed effects from a regression similar to those in Table 5 but controlling simultaneously for FICO, LTV, loan amount, and MSA f.e..

Internet Appendix for "Paying Too Much? Borrower Sophistication and Overpayment in the US Mortgage Market"

A.1 Comparing Offer and Locked Interest Rates in Optimal Blue to Other Data Sources

In this section we assess whether the interest rates we observe in the Optimal Blue data align with other data sources. To begin, we compare median offer rates from Optimal Blue to offer rates from Mortgage News Daily (MND) for various 30-year fixed-rate loan programs. MND uses several sources of information to estimate typical offer rates, including directly obtaining rate sheets from the largest lenders. The three panels in Figure A-3 plot median offer rates from Optimal Blue against MND's offer rates for conforming, FHA, and jumbo mortgages, respectively. In the top two panels, the Optimal Blue median offer rates for conforming and FHA loans—which are the bulk of our data—move almost in lockstep with the MND offer rates. For jumbo loans, the Optimal Blue median offer rate exhibits a little more variation from trough to peak, but on average the level is quite similar. Overall, these results help establish that our median offer rates from Optimal Blue are representative of the overall market.

Next, we compare *lock* rates to interest rates on *closed* mortgages. A concern with the locks data is that high and low lock rates may systemically be less likely to actually proceed all the way to origination. For example, borrowers who lock in a high rate at one lender may continue to shop around and ultimately find a better rate.

The top panel of Table A-1 compares unconditional distributions of interest rates from the Optimal Blue locks data with interest rate distributions from other administrative data sources on closed mortgages, by loan type. If the Optimal Blue locks are representative of closed loans, then the rate distributions across these datasets should be very similar.

The first four columns compare distributions for FHA loans locked or closed in 2014-15. For these years, we have access to administrative data from the Department of Housing and Urban Development (HUD) on the *universe* of originated FHA loans, which serves as an ideal benchmark. In addition, we compare the locks data to well-known and widely used Black Knight McDash servicing data, which contains loans serviced by the largest mortgage servicers in the US. We can see in the top left portion of Table A-1 that average and 90th percentile locked rates line up identically to both the HUD and McDash data. This remains true whether we look at all FHA locks in Optimal Blue (column 1) or only those that we were able to match to an originated loan in the HUD data (column 2), based on the procedure of Bhutta and Hizmo (2020). The fact that the distributions in columns (1) and (2) are almost identical (also for other characteristics) implies that there is little evidence of "selection" in terms of which locks end up in originated loans.[1] Moreover, the full HUD and McDash data are slightly lower at the 10th percentile, suggesting an even wider distribution than in Optimal Blue. Table A-1 also indicates that the distribution of FICO scores and LTVs in Optimal Blue almost mirrors the HUD data, whereas the McDash data are skewed slightly toward less risky borrowers.

The remaining columns compare Optimal Blue locks to McDash loans in 2016-18, separately for FHA, conforming, and jumbo loans. The most notable difference is for jumbo loans, where we observe higher interest rates in Optimal Blue by 30-40bp, although the amount of dispersion is similar to McDash. In Figure A-4, we plot the average, 10th and 90th percentile rates over time from Optimal Blue locks and McDash. Rates move closely together across the distribution, with McDash rates lagging locked rates a bit—as expected since mortgages typically do not get originated until a few weeks after the rate mortgage rate is locked in. Again, while the levels of rates are very similar across the two datasets for FHA and conforming mortgages, Optimal Blue rates tend to be higher than McDash for jumbo loans, although the amount of dispersion is similar.

[1] This remains true if we plot the distribution of rates in the Optimal Blue locks over time: the distribution of all locks and the matched (i.e. originated) locks are almost always nearly identical.

A.2 Price Dispersion in Mortgage Offers

In this appendix, we provide additional detail on our analysis of price dispersion in offer interest rates across lenders, already briefly discussed in Section 4.1 of the main text.

There are two things to consider when thinking about the "price" of a mortgage with certain characteristics. First, 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. Alternatively, they can choose negative points, known as lender credits or rebates, in return for a higher mortgage rate. In this case, borrowers receive cash from the lender which can be used toward closing costs. Either way, one point in upfront payments corresponds to about 20bp in mortgage rate (so a borrower could get e.g. a 4% mortgage rate with no points, a 4.2% rate but receive one point, or a 3.8% rate by paying one point).

Second, 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. We can think of lender fees and discount points as interchangeable: from the borrower's 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 the Optimal Blue Pricing Insights interface, we observe how lenders compare in terms of the sum of points and fees that they charge for a given mortgage rate, on a given day in a given location and for certain borrower and loan characteristics. The interface allows users to specify the key underwriting and loan characteristics, including location (MSA), FICO score, LTV, loan amount, DTI, loan type and term (e.g. 30-year fixed), loan purpose (e.g. cashout refinance), program (e.g. FHA or conforming), as well as details about the property (e.g. whether it is a single-family home or a condo) and whether it will be owner-occupied or not. Furthermore, the user specifies the desired lock period (e.g. 30 days). One could furthermore specify a given mortgage rate for which offers should be compared (e.g. 4%), but by default the system instead shows the comparison of points/fees for the mortgage rate at which the median lender that makes an offer does so at (as close as possible to) zero points and fees.

An example of the resulting output is shown in Figure A-5. Lenders are sorted based on the "price" they offer for a loan with the desired characteristics, where the price equals 100 minus the points/fees the borrower would be charged. Thus, a price of 101 means the borrower would receive one point, while a price of 99 means the borrower would have to pay one point to get this loan. As can be seen in the screenshot, the range of offers in this example spans almost 4 points, which for a typical loan of $250,000 would correspond to a difference between the cheapest and most expensive lenders of $10,000.

As noted in the main text, we conduct searches for 100 different combinations of FICO, LTV, program, loan amount, loan purpose, occupancy, and rate type, across 20 MSAs (at different frequencies). For each of these searches, we then receive the underlying individual price offers for the mortgage rate the system chooses (as explained above).

For our main analysis, we then transform these prices into the rate each lender would offer at zero points and fees, by converting points into rates using a conversion factor that we estimate based on the lock data. As explained in the main text, we allow for this conversion factor to be time-varying. The estimated conversion factor averages about 21bp in rate per 1 point upfront, which is also in line with what is typically observed in lender rate sheets. So for instance, a lender that is shown as offering a price of 100.5 for a 4.25% mortgage rate is assigned a rate of 4.145%.

A.2.1 Dispersion in Offer Rates

We start by documenting the dispersion in mortgage rates available from different lenders for identical mortgages in Los Angeles, since we have daily searches for this MSA. The first panel of Figure A-6 shows the distribution of rates offered by different lenders for conforming mortgages with an amount of $300k, FICO=750, LTV=80 and DTI=36. There are about 120 different lenders offering this mortgage in Los Angeles on any given day. The histogram shows the daily offer rates after subtracting the median (for the same day) over the period of April 2016 to December 2019.

Figure A-6 shows that the rate difference between the cheapest and the most expensive lender is about 100bp. 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-6, which plot the dispersion for a typical FHA mortgage and a jumbo mortgage. The exact shape of the distribution does look different across these different mortgages, but the amount of dispersion is similar.

Figure 1 in the main text shows the dispersion in mortgage rates available from different lenders in all of the 20 metropolitan areas. Table A-2 shows more detailed summary statistics of the rate dispersion in this pooled offer data, broken down by mortgage types. There are typically over 100 unique lenders on 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. Generally, the price dispersion is a bit higher for mortgages with low FICO scores, high LTVs and FHA mortgages. Overall, there is about a 50-55bp difference in mortgage rates between the 10^{th} percentile lender and the 90^{th} percentile lender, and a 90bp difference between the 1^{st} and the 99^{th} percentile lender.

Table A-3 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.2.2 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 interest rate at which the median lender offers a mortgage (with given characteristics) at zero points or fees. Figure A-7 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. This figure shows that the range of offers shown in the screenshot in Figure A-5 appears representative of the universe of offer distributions.

Table A-4 summarizes this dispersion for different mortgage types. The differences in the upfront costs of a mortgage with an identical rate across lenders are very large. The difference between the 90^{th} percentile and 10^{th} percentile lender is around 2.2 to 2.5% of the mortgage balance. For a typical conforming loan of $250K that amounts to roughly a $6000 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.3 Matching Offers and Locks

As described in Section 3.2, we collect data on mortgage offers for 20 MSAs (some daily, others twice or once per week) and for different loan programs (conforming, super-conforming, jumbo, and FHA) and borrower/loan characteristics. In particular, we collect rates for FICO scores of 640, 680, 720, and 750, and LTV ratios of 70, 80, 90, 95, and 96 percent. When matching locks to these offers, we allow for some variation in the characteristics around the values that we collect rates for, but do so in a conservative way. What this means is that (with two small exceptions noted below) we match locks with FICO scores slightly *above* the FICO value from the rate offer and with LTV ratios slightly *below* the LTV value from the offer, as follows:

- Offer FICO 640: Lock FICO range 640-659
- Offer FICO 680: Lock FICO range 680-699
- Offer FICO 720: Lock FICO range 720-739
- Offer FICO 750: Lock FICO range 740-850 (maximum FICO)
- Offer LTV 70: Lock LTV range 60.01-70
- Offer LTV 80: Lock LTV range 75.01-80

- Offer LTV 90: Lock LTV range 85.01-90
- Offer LTV 95: Lock LTV range 90.01-95
- Offer LTV 96: Lock LTV range 95.01-97

In choosing these ranges, we follow Fannie Mae's loan-level pricing adjustment (LLPA) grid (https://www.fanniemae.com/content/pricing/llpa-matrix.pdf). This grid is also why we decided to assign FICO scores of 740-749 the FICO 750 offer as well, and similarly for LTVs of 96.01-97 for the LTV 96 offer. (LTV values above 95 are uncommon for GSE loans, but are very common for FHA loans, where the modal LTV is 96.5.) We do not include some intermediate values (e.g. FICO 660-679, 700-719; LTV 80-85) since LLPAs can be different and do not always change linearly; however, matching less conservatively in that regard does not materially affect the results.

In addition to matching on date, FICO, LTV, MSA and loan program, we also only retain purchase mortgages with a 30 day lock period (since that is what the rate search is for). 30 days is also the most common lock period in the data.

A.4 An Alternative to the Locked-Offer Rate Gap: Expected Gains from Search

Our headline measure of the "locked-offer rate gap" captures how far the rate a particular borrower locked is from what the median lender could offer them for an identical loan on the same day. We construct this simple measure for each borrower to see how well they are doing relative to the median lender, and to uncover which groups of borrowers do particularly badly. An alternative approach to this is to construct a measure of expected benefits from one extra search for each borrower by making some assumptions on how borrowers shop and what rates they obtain when doing so.

We start with a simple search model similar to Carlson and McAfee (1983). Suppose that there are n mortgage lenders who are posting mortgage rate offers on Optimal Blue for a particular borrower type. Rates are ordered from lowest to highest:

$$r_1 \leq r_2 \leq ... \leq r_n$$

Borrowers only see the mortgage rates available at the lenders they meet with. Assuming each borrower has an equal chance of meeting any one of the lenders, the probability of finding a lender that offers the rate r is $f(r) = 1/n$. Suppose a borrower has already found a rate r_k and is considering searching one more time for a cheaper lender. The expected gain from doing so is given by:

$$\begin{aligned} x_k &= \sum_{i=1}^{k-1} (r_k - r_i) f(r_i) \\ &= \sum_{i=1}^{k-1} (r_k - r_i) \frac{1}{n} \\ &= \left[r_k - \sum_{i=1}^{k-1} \frac{r_i}{k-1} \right] \frac{k-1}{n} \end{aligned} \tag{A1}$$

Intuitively, the term in the brackets is the locked rate minus the expected rate from going to the $k-1$ lenders that are offering rates lower than r_k. Of course, the borrower does not know which lenders are offering rates lower than r_k, so we have to adjust the expectation by the share of these lenders in the population, which is $(k-1)/n$. Therefore, this is a measure of how much money the borrower is leaving on the table, in expectation, by not conducting one more search. Compared to the locked-offer rate gap we use in our main analysis, where only the median available rate matters for our assessment of "how well" a borrower did, here the width of the offer distribution also plays a role: for a given mean of the offer distribution, x_k will be higher when offers are more widely dispersed (as this leads $E(r|r < r_k)$ to be lower).

Table A-8 summarizes the expected gains from search for different cuts of the data similar to Table 4 in the main text. Not surprisingly, the overall level of expected gains from search measure is larger than the locked-offer rate gap, since the expected gain is by definition non-negative. Taking into account this difference in levels, however, all the cross-sectional patterns we are interested in are very similar to the ones in Table 4.

Table A-9 replicates the results of Table 5 using the alternative measure of expected gains from one more search. The results are identical in both of these tables, suggesting that the choice of using locked-offer rate gap or the alternative measure of expected gains from one more search is immaterial for our results. This is not very surprising, given the result in Table A-2 that the dispersion in offer rates does not vary much with borrower/loan characteristics.

A.5 Evidence from the SCF on the Effects of Financial Literacy and Shopping

As a complement to our analysis of the new NSMO data in Section 7, here 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 US 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 three questions designed by Annamaria Lusardi and Olivia Mitchell to gauge individuals' general financial literacy.[2] 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.[3]

For our analysis here, we focus on a subsample of SCF households that own their home and recently took out a fixed-rate 30-year or 15-year 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 A-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[4], loan

[2] 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). The only other paper examining the relationship between financial literacy and mortgage rates is Huston (2012). More recently, Gathergood and Weber (2017) study the relationship between financial literacy and mortgage product choice.

[3] 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.

[4] 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. Another caveat is that we do not observe

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?"[5]

Table A-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. Finally, column 6 regresses mortgages rates on financial literacy and shopping simultaneously. The estimated coefficients on both variables are almost unchanged, indicating that both shopping and financial literacy are independently important for the mortgage rates consumers obtain. 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.

A.6 Correlates of Shopping Intensity and Knowledge

Section 7 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-12, we report results from regressions of the four binary shopping measures already used in Section 7.3 on the three mortgage knowledge measures introduced in Section 7.1, 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-12 while the latter type is called "multivar."

In Table A-13, 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-12 indicate that borrowers that are more knowledgeable also shop more. Of course, in this case it is difficult to rule out reverse causality, namely that the additional shopping made them more knowledgeable (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.[6] 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 9.[7] 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 knowledgeable,

points or fees in the SCF, which might bias our estimates if less literate borrowers are actually paying fewer points in return for paying higher rates.

[5] Just over one-quarter of our sample of mortgage borrowers answered "10", while less than 3% answered "0"; the mean response was about 7.5, with a standard deviation of 2.5.

[6] This self-assessed creditworthiness was also used as a control variable in Table 7.

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

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 knowledgeable. Similarly, FHA borrowers do not appear to shop less, but tend to be significantly less knowledgeable 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-offer 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 knowledgeable 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 knowledgeable—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.

Table A-1: Comparing Mortgage Locks in Optimal Blue to Closed Mortgages

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	FHA Loans, 2014-15				FHA Loans, 2016-18		Conventional Conforming Loans, 2016-18		Conventional Jumbo Loans, 2016-18	
	Optimal Blue		HUD	McDash	Optimal Blue	McDash	Optimal Blue	McDash	Optimal Blue	McDash
	All	Matched								
Interest Rate										
10th	3.75	3.75	3.625	3.625	3.625	3.5	3.75	3.625	3.75	3.375
mean	4.14	4.14	4.11	4.09	4.40	4.26	4.44	4.29	4.33	3.95
90th	4.625	4.625	4.625	4.625	5.25	5.125	5.125	5	4.875	4.625
FICO Score										
10th	628	629	630	641	620	629	681	686	719	726
mean	679.4	679.3	680.5	688.9	672.3	684.0	745.2	750.0	766.1	771.3
90th	744	742	745	754	738	751	800	802	801	803
LTV										
10th	93.7	95	94.3	87.6	93.4	87.9	66.6	64.4	66.7	65.0
mean	95.3	95.5	95.8	93.7	95.4	93.7	83.6	82.0	77.6	82.7
90th	96.5	96.6	96.5	96.5	96.5	96.5	95.0	95.0	85.0	85.0
Loan Amount										
10th	89,745	92,640	84,000	81,987	100,360	97,697	116,000	113,715	482,000	485,100
mean	187,624.3	186,804.2	180,450.1	173,106.5	204,065.3	203,275.5	255,892.9	255,738.2	729,963.4	850,403.6
90th	300,000	294,325	293,250	276,892	321,985	325,004	417,000	418,125	1,060,000	1,260,000
N	282,933	162,244	1,318,700	777,763	860,579	1,468,968	1,547,776	2,695,218	61,430	190,993

Data Source: Optimal Blue, HUD, Black Knight McDash

Note: All statistics are for 30-year fixed rate home purchase mortgages for owner-occupied properties. Conventional conforming include super-conforming loans that have loan amounts under the higher loan limits in high-cost geographies. “McDash” refers to Black Knight McDash data. “Matched” in column (2) means Optimal Blue locks that matched to originated FHA loans in the HUD data.

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

	Median No. Offers	Median Rate	Standard Deviation	Percentile Differences $75^th - 25^{th}$	$90^{th} - 10^{th}$	$99^{th} - 1^{st}$
All Offers	118	4.67	0.20	0.27	0.53	0.90
Program						
FHA	117	4.08	0.22	0.32	0.59	0.93
Conforming	122	4.54	0.19	0.27	0.51	0.88
Super-Conforming	144	4.68	0.20	0.27	0.52	0.88
Jumbo	106	5.06	0.20	0.26	0.53	0.92
FICO						
640	107	5.23	0.21	0.29	0.54	0.92
680	118	4.64	0.20	0.28	0.53	0.90
720	122	4.48	0.20	0.27	0.52	0.90
750	122	4.44	0.20	0.27	0.52	0.90
LTV						
70	122	4.67	0.20	0.27	0.52	0.90
80	117	4.78	0.20	0.28	0.53	0.91
90	105	4.78	0.20	0.27	0.52	0.91
95	128	4.63	0.20	0.27	0.51	0.88
96	119	4.27	0.21	0.30	0.55	0.91

Data Source: Optimal Blue

Notes: 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 4-6 show the difference between various percentiles of the offer distribution.

Table A-3: The real-time interest rate dispersion for offered conforming mortgages with no points and fees

	Median No. Offers	Median Rate	Standard Deviation	Percentile Differences $75^th - 25^{th}$	$90^{th} - 10^{th}$	$99^{th} - 1^{st}$
Atlanta, GA	112	4.68	0.20	0.28	0.54	0.92
Boston-Worcester-Lawrence, MA-NH-ME-CT	77	4.49	0.21	0.30	0.56	0.93
Charlotte-Gastonia-Rock Hill, NC-SC	93	4.67	0.21	0.28	0.55	0.93
Chicago-Gary-Kenosha, IL-IN-WI	103	4.57	0.20	0.28	0.53	0.90
Cleveland-Akron, OH	61	4.71	0.21	0.30	0.57	0.92
Dallas-Fort Worth, TX	136	4.67	0.21	0.29	0.55	0.93
Denver-Boulder-Greeley, CO	119	4.69	0.19	0.25	0.49	0.88
Detroit-Ann Arbor-Flint, MI	76	4.68	0.21	0.29	0.56	0.94
Las Vegas, NV	87	4.88	0.21	0.28	0.55	0.92
Los Angeles-Riverside-Orange County, CA	147	4.69	0.20	0.27	0.52	0.89
Miami-Fort Lauderdale, FL	95	4.66	0.21	0.30	0.56	0.93
Minneapolis-St. Paul, MN	73	4.65	0.19	0.26	0.51	0.89
New York-Northern New Jersey-Long Island, NY-NJ	93	4.60	0.21	0.30	0.56	0.92
Phoenix-Mesa, AZ	117	4.80	0.21	0.29	0.54	0.91
Portland-Salem, OR	88	4.77	0.20	0.27	0.52	0.88
San Diego, CA	103	4.71	0.19	0.26	0.51	0.89
San Francisco-Oakland-San Jose, CA	112	4.75	0.19	0.26	0.51	0.88
Seattle-Tacoma-Bremerton, WA	101	4.79	0.19	0.26	0.51	0.88
Tampa-St. Petersburg-Clearwater, FL	124	4.80	0.20	0.27	0.53	0.92
Washington-Baltimore, DC-MD-VA	116	4.61	0.21	0.28	0.55	0.93

Data Source: Optimal Blue

Notes: 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-4: 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			
FHA	1.42	2.59	3.83
Conforming	1.19	2.22	3.69
Super-Conforming	1.23	2.35	3.79
Jumbo	1.13	2.31	3.84
FICO			
640	1.30	2.41	3.83
680	1.22	2.35	3.77
720	1.19	2.30	3.78
750	1.20	2.30	3.78
LTV			
70	1.19	2.28	3.77
80	1.24	2.37	3.81
90	1.19	2.29	3.81
95	1.20	2.26	3.72
96	1.32	2.44	3.80

Data Source: Optimal Blue

Notes: 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 chosen such that the median lender charges no points and fees at this interest rate.

Table A-5: Summary Statistics of the Rate Locked Minus the Median Offer Rate for Identical Mortgages by ZIP Code Demographics

	Observations	Mean	St. Deviation	Percentiles	
				25^{th}	75^{th}
All Mortgages	64,788	0.11	0.31	-0.07	0.26
Median Household Income					
First Tercile	21,673	0.16	0.32	-0.03	0.31
Second Tercile	21,517	0.10	0.30	-0.07	0.25
Third Tercile	21,585	0.07	0.31	-0.11	0.22
Percent College Educated					
First Tercile	21,610	0.16	0.32	-0.03	0.32
Second Tercile	21,602	0.12	0.31	-0.06	0.26
Third Tercile	21,576	0.05	0.29	-0.11	0.19
Minority Share					
First Tercile	21,619	0.07	0.30	-0.09	0.22
Second Tercile	21,574	0.09	0.30	-0.08	0.24
Third Tercile	21,595	0.16	0.33	-0.04	0.32
Market Share of Top 4 Lenders					
First Tercile	21,711	0.11	0.28	-0.04	0.24
Second Tercile	21,513	0.09	0.31	-0.09	0.25
Third Tercile	21,564	0.12	0.34	-0.08	0.29

Data Source: Optimal Blue, American Community Survey, HMDA

Notes: 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 offer rate. The median household income, percent college educated, and minority share (share of Hispanic/Latino plus non-Hispanic Black) are only observed at the ZIP code level. The market share of the top four lenders is observed at the county level.

Table A-6: Regressions of the Locked-Offer Rate Gap on Observables, for FHA Loans Only

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
FICO (omitted cat.: [640,660))								
$I_{680 \leq FICO < 700}$	-0.033***	-0.039***	-0.061***					
	(0.009)	(0.009)	(0.019)					
$I_{720 \leq FICO < 740}$	-0.069***	-0.065***	-0.086***					
	(0.010)	(0.010)	(0.019)					
$I_{FICO \geq 740}$	-0.073***	-0.067***	-0.082***					
	(0.013)	(0.010)	(0.016)					
$I_{LTV > 95}$				0.040	0.062**	0.074*		
				(0.026)	(0.024)	(0.038)		
Discount Points								
$I_{-5 < Points < -0.2}$							-0.137***	0.021
							(0.021)	(0.013)
$I_{0.2 < Points \leq 5}$							0.145***	0.027
							(0.023)	(0.023)
Loan Officer Comp (%)			0.182***			0.180***		
			(0.057)			(0.057)		
Loan amount f.e. ($10k bins)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
MSA x Month f.e.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Branch f.e.		Yes	Yes		Yes	Yes		Yes
Adj. R-Squared	0.128	0.564	0.627	0.122	0.559	0.615	0.199	0.560
Observations	14330	12857	2965	14330	12857	2965	14330	12857

Data Source: Optimal Blue

Notes: The dependent variable is the mortgage interest rate locked minus the median offer rate in the same market and day for an identical mortgage. Unlike in the corresponding table in the main text, here we only use two LTV bins (separated at 95) since the majority of FHA loans have very high LTVs. 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. Standard errors shown in parentheses are two-way clustered at the month and lender level. Significance: * p<0.1, ** p<0.05, *** p<0.01.

Table A-7: Regressions of the Locked-Offer Rate Gap on Observables, for Independent Nonbank Originators Only

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
FICO (omitted cat.: [640,660))								
$I_{680\leq FICO<700}$	-0.057***	-0.044***	-0.048***					
	(0.009)	(0.007)	(0.012)					
$I_{720\leq FICO<740}$	-0.093***	-0.066***	-0.057***					
	(0.011)	(0.009)	(0.013)					
$I_{FICO\geq 740}$	-0.126***	-0.087***	-0.069***					
	(0.012)	(0.010)	(0.013)					
LTV (omitted cat.: (60,80])								
$I_{85<LTV\leq 90}$				0.013**	0.010	0.018*		
				(0.006)	(0.006)	(0.010)		
$I_{90<LTV\leq 95}$				0.043***	0.030***	0.025**		
				(0.006)	(0.006)	(0.011)		
$I_{LTV>95}$				0.178***	0.140***	0.089***		
				(0.013)	(0.012)	(0.016)		
Discount Points								
$I_{-5<Points<-0.2}$							-0.078***	0.005
							(0.025)	(0.007)
$I_{0.2<Points\leq 5}$							0.089***	0.017**
							(0.014)	(0.008)
Loan Officer Comp (%)			0.155***			0.139***		
			(0.036)			(0.038)		
Loan amount f.e. ($10k bins)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
MSA x Month f.e.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Branch f.e.		Yes	Yes		Yes	Yes		Yes
Adj. R-Squared	0.134	0.469	0.429	0.174	0.496	0.440	0.155	0.462
Observations	45522	44319	11720	45522	44319	11720	45522	44319

Data Source: Optimal Blue

Notes: 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 2016-2019. We focus on 30 year, fixed rate, fully documented purchase mortgages. Standard errors shown in parentheses are two-way clustered at the month and lender level. Significance: * p<0.1, ** p<0.05, *** p<0.01.

Table A-8: Summary Statistics of the Expected Gain from Search

	Observations	Mean	St. Deviation	Percentiles	
				25^{th}	75^{th}
All Mortgages	64,788	0.20	0.23	0.05	0.27
Program					
FHA	14,441	0.32	0.30	0.10	0.45
Conforming	44,040	0.17	0.19	0.05	0.23
Super-Conforming	4,478	0.10	0.15	0.01	0.13
Jumbo	1,829	0.05	0.12	0.00	0.05
FICO					
[640, 660)	7,406	0.31	0.31	0.08	0.45
[680, 700)	9,390	0.25	0.27	0.06	0.35
[720, 740)	10,207	0.20	0.22	0.05	0.27
740+	37,785	0.16	0.18	0.04	0.22
LTV					
(75, 80]	21,334	0.13	0.15	0.03	0.18
(85, 90]	6,882	0.16	0.17	0.04	0.22
(90, 95]	15,782	0.17	0.19	0.04	0.23
(95, 97]	20,790	0.30	0.29	0.09	0.42
First-Time Homebuyer					
No	32,437	0.16	0.19	0.04	0.22
Yes	32,345	0.23	0.25	0.06	0.32
Discount Points					
[-5, -0.2)	14,015	0.15	0.19	0.02	0.21
[-0.2, 0.2]	22,735	0.18	0.21	0.04	0.23
(0.2, 5]	28,038	0.24	0.25	0.07	0.32
Lender Type					
Independent Non-bank	45,618	0.21	0.23	0.06	0.28
Other	19,170	0.17	0.21	0.03	0.24

Data Source: Optimal Blue

Note: For each mortgage rate locked by borrowers in our data, we compute the expected gain from search using equation (A1).

Table A-9: Regressions of the Expected Gains from Search on Observables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
FICO (omitted cat.: [640,660))								
$I_{680 \leq FICO < 700}$	-0.055***	-0.045***	-0.042***					
	(0.006)	(0.005)	(0.010)					
$I_{720 \leq FICO < 740}$	-0.095***	-0.073***	-0.069***					
	(0.009)	(0.007)	(0.012)					
$I_{FICO \geq 740}$	-0.125***	-0.094***	-0.084***					
	(0.009)	(0.007)	(0.011)					
LTV (omitted cat.: (60,80])								
$I_{85 < LTV \leq 90}$				0.017***	0.013***	0.016***		
				(0.003)	(0.003)	(0.006)		
$I_{90 < LTV \leq 95}$				0.034***	0.026***	0.024***		
				(0.004)	(0.004)	(0.006)		
$I_{LTV > 95}$				0.148***	0.121***	0.090***		
				(0.010)	(0.009)	(0.012)		
Discount Points								
$I_{-5 < Points < -0.2}$							-0.036***	0.004
							(0.009)	(0.005)
$I_{0.2 < Points \leq 5}$							0.065***	0.018*
							(0.010)	(0.010)
Loan Officer Comp (%)			0.112***			0.099***		
			(0.026)			(0.029)		
Loan amount f.e. ($10k bins)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
MSA x Month f.e.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Branch f.e.		Yes	Yes		Yes	Yes		Yes
Adj. R-Squared	0.135	0.407	0.408	0.175	0.435	0.421	0.132	0.390
Observations	64693	62783	14659	64693	62783	14659	64693	62783

Data Source: Optimal Blue

Notes: The dependent variable is the expected gain from an additional search, given by equation (A1). 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. Standard errors shown in parentheses are two-way clustered at the month and lender level. Significance: * p<0.1, ** p<0.05, *** p<0.01.

Table A-10: The Relationship Between the Expected Gains from Search and Treasury Yields

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treasury Yield	-0.053***	-0.056***	-0.036***					-0.037***
	(0.006)	(0.010)	(0.008)					(0.008)
Offer Spread to Prime Conforming Rate							-0.116***	-0.117***
							(0.020)	(0.020)
Treasury Yield ×								
DTI > 36				-0.058***	-0.062***	-0.042***		
				(0.007)	(0.011)	(0.010)		
DTI ≤ 36				-0.046***	-0.048***	-0.028***		
				(0.007)	(0.009)	(0.005)		
Borrower and Loan Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
MSA F.E	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
MSA x Month F.E.		Yes	Yes		Yes	Yes	Yes	Yes
Branch F.E.			Yes			Yes	Yes	Yes
Adj. R-Squared	0.165	0.181	0.438	0.166	0.182	0.438	0.441	0.441
Observations	64396	64316	62397	64396	64316	62397	62782	62397
P-val. for equality of DTI coefficients				0.037	0.028	0.012		

Data Source: Optimal Blue

Notes: The dependent variable is the expected gain from an additional search, given by equation (A1). The offer spread to conforming rate is defined as the average offer rate for a typical borrower in the same program in the same day minus the average offer rate for a typical prime conforming borrower. All specifications include controls for FICO, LTV, and loan amount. 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. Standard errors shown in parentheses are two-way clustered at the month and lender level. Significance: * p<0.1, ** p<0.05, *** p<0.01.

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Financial Literacy (Fraction Correct)	-0.247**	-0.202**	-0.246**				-0.245**
	(0.110)	(0.097)	(0.097)				(0.100)
Shops Around for Credit				-0.262***	-0.259***	-0.230***	-0.222**
				(0.090)	(0.084)	(0.085)	(0.087)
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, property type, 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. * p<0.1, ** p<0.05, *** p<0.01.

Table A-12: 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

Data Source: National Survey of Mortgage Originations

Note: 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-13: 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. (1)	Multivar. (2)	Univar. (3)	Multivar. (4)	Univar. (5)	Multivar. (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

Data Source: National Survey of Mortgage Originations
Note: 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.

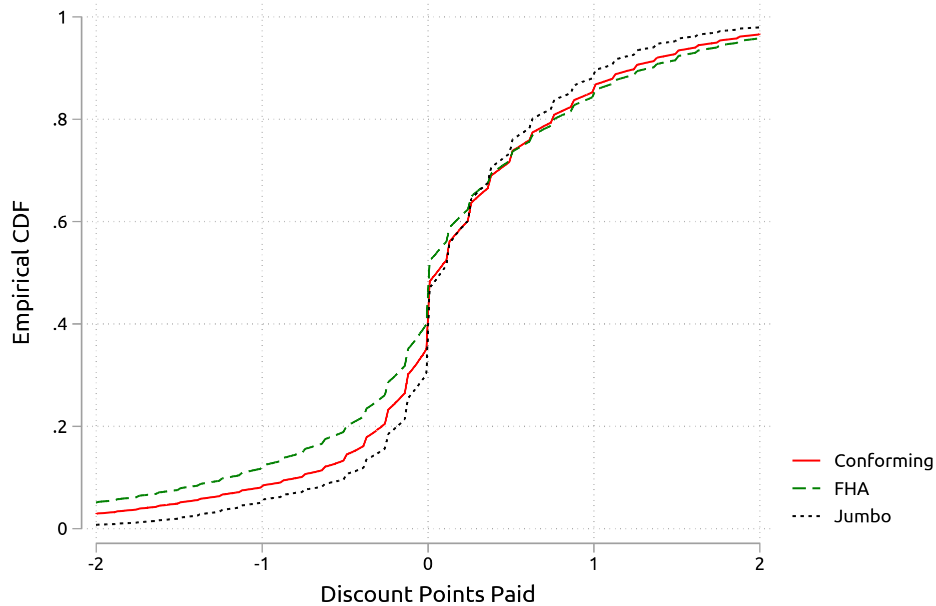


Figure A-1: The Empirical Cumulative Distribution of Discount Points Paid, by Program

Data Source: Optimal Blue
Note: Figure shows cumulative share of borrowers that paid up to a certain amount of discount points; negative values represent credits/rebates. Data includes purchase and refinance rate locks in 2015-2019.

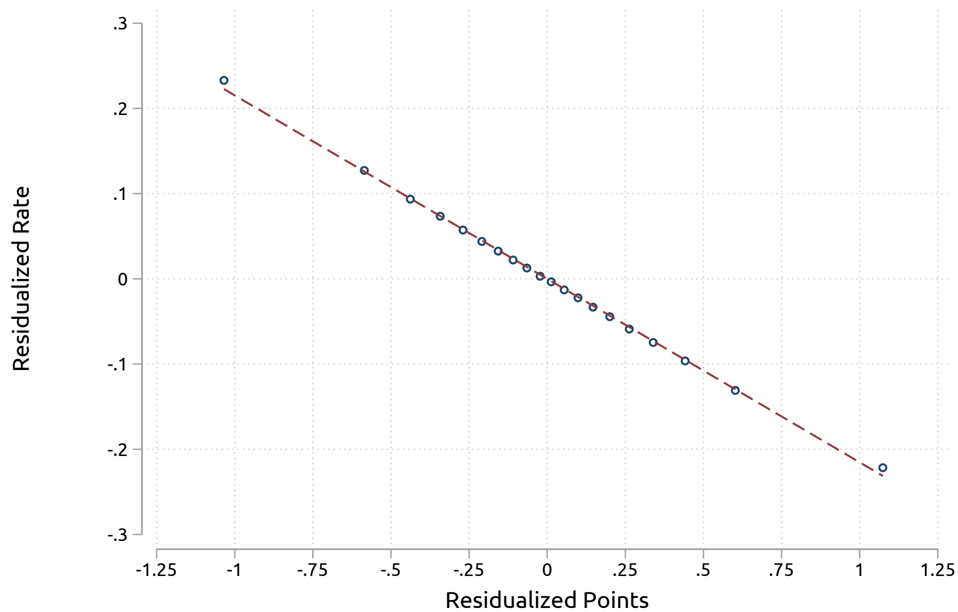


Figure A-2: The Relationship between Discount Points Paid and Mortgage Rates

Data Source: Optimal Blue
Note: Binned scatter plot. Discount points and mortgage rates are first residualized using a regression specification identical to column (10) of Table 2, with the only exception that we are not controlling for discount points.

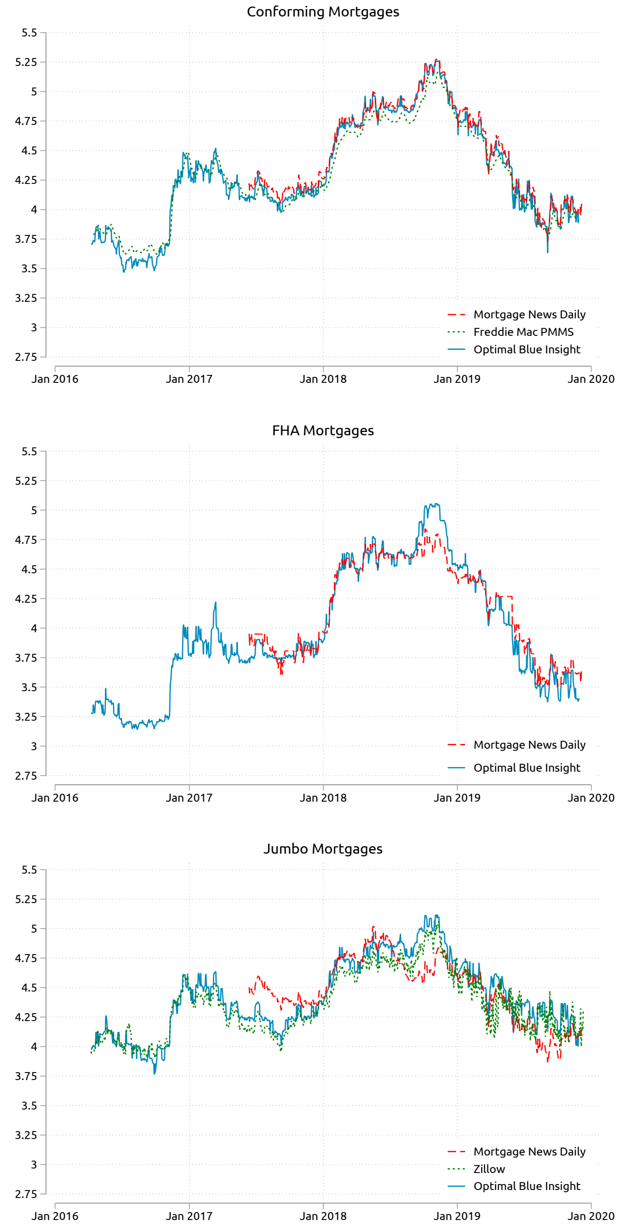


Figure A-3: Comparison of Average Offer Rates from Optimal Blue with Mortgage News Daily Data

Data Source: Optimal Blue, Mortgage News Daily, Freddie Mac, Zillow
Note: The Optimal Blue Data are for borrowers with FICO=750, DTI=36, with no points/fees, and LTV=80 for conforming and jumbo, and LTV=96.5 for FHA. The Mortgage News Daily (MND) data reflect rates for "top-tier" borrowers, and we adjust the MND rates assuming they include 0.5% points and fees.

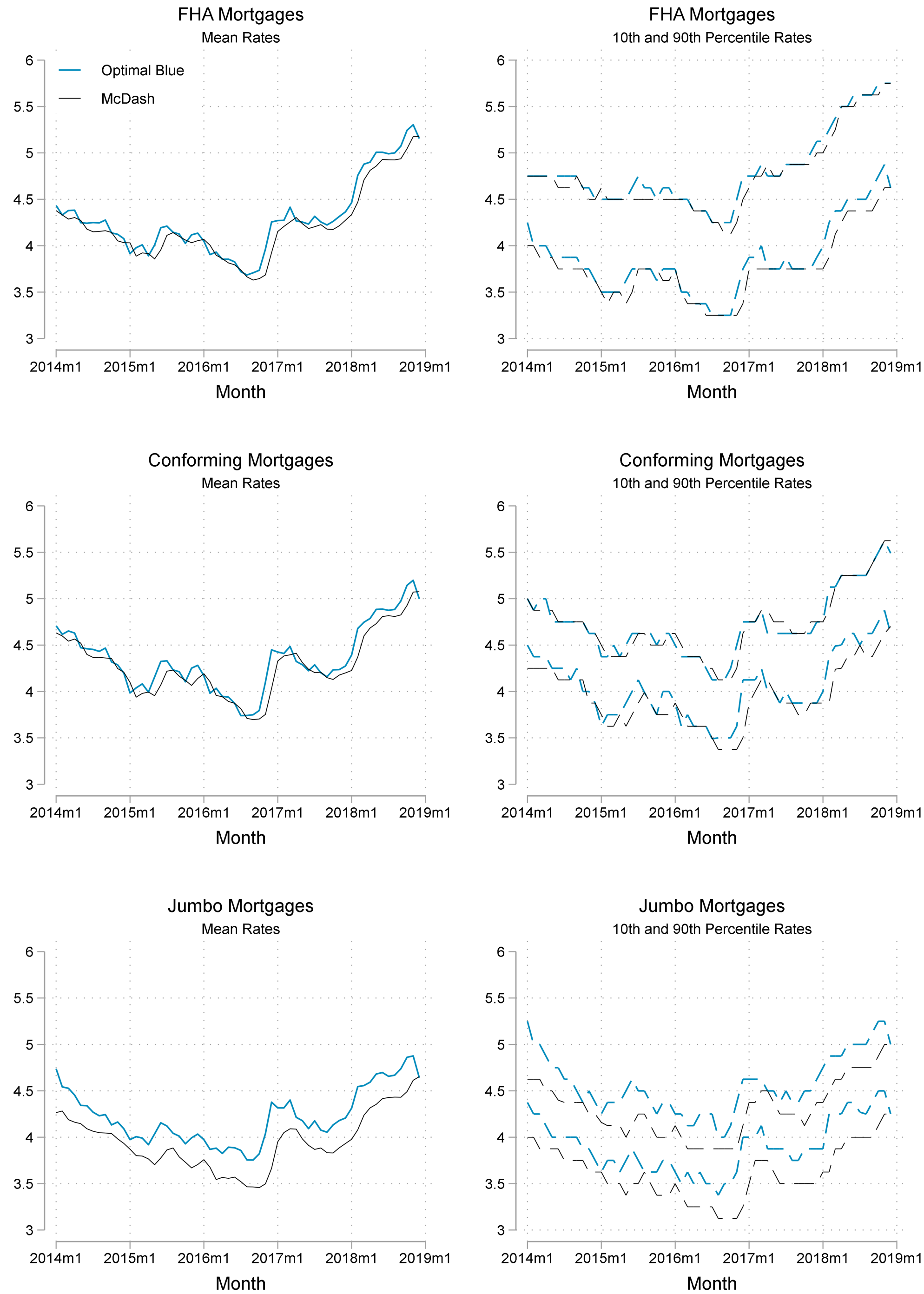


Figure A-4: Comparison of Locked Interest Rates from Optimal Blue with Interest Rates on Closed Originations in McDash

Data Source: Optimal Blue, Black Knight McDash
Note: The Optimal Blue series lead the McDash series because for Optimal Blue we observe the date when the loan terms are locked, while in McDash we observe when a loan is originated.

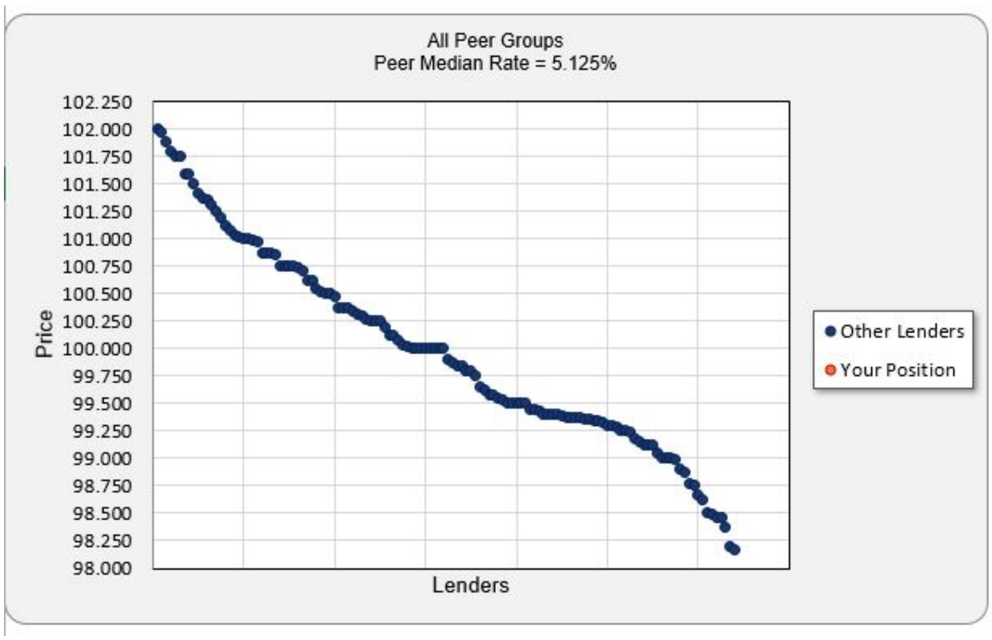


Figure A-5: Screenshot of Sample Offer Distribution from Optimal Blue Pricing Insights

Data Source: Optimal Blue
Note: Figure shows an example of the real-time distribution of offers across lenders in the same metropolitan area for a loan with given characteristics and at a note rate of 5.125%. Lenders are sorted by "price", which equals 100 + the points (rebate/credit) the lender pays to the borrower (so "102" means the borrower receives two points at closing, while "98" means they would have to pay two points). The mortgage note rate for which offers are shown is chosen such that the median lender offers a price as close as possible to 100. For actual lenders using the interface, an orange dot would show their position in the distribution.

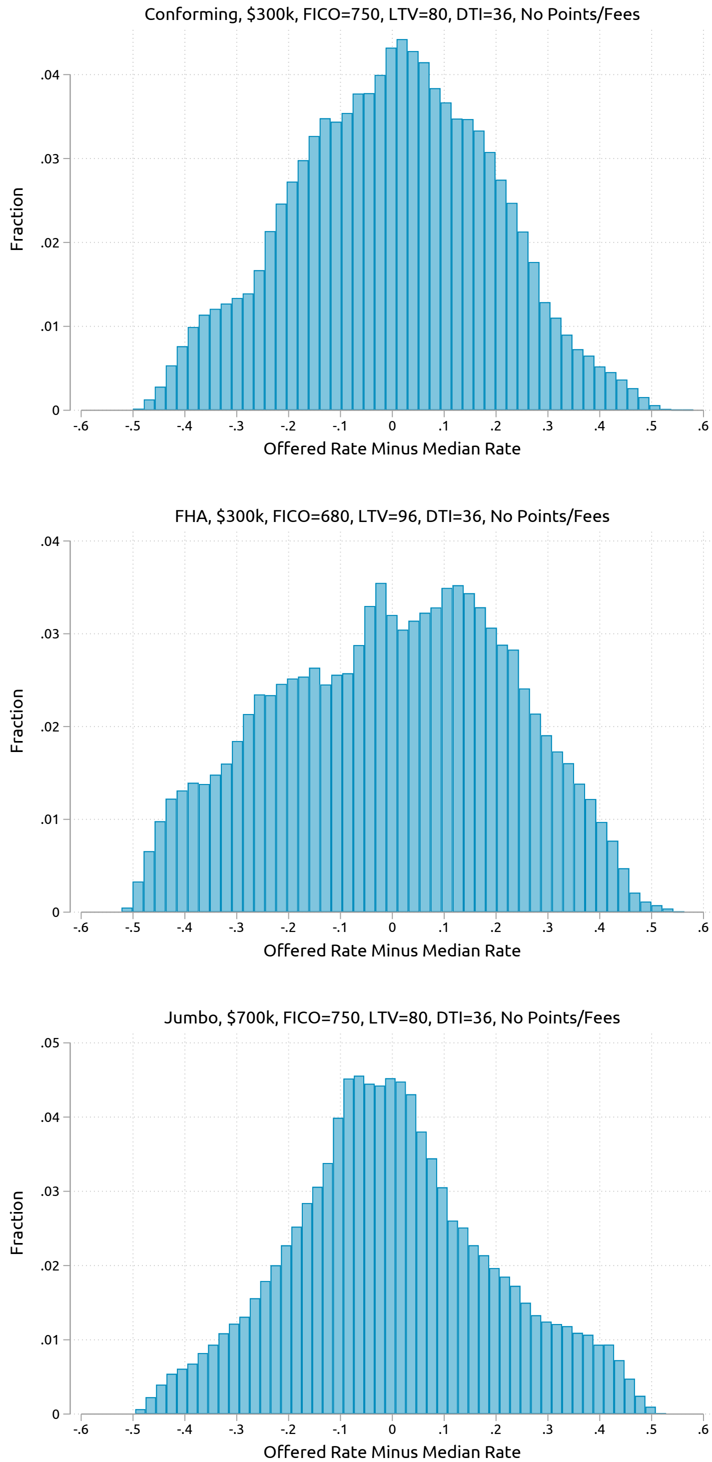


Figure A-6: Interest Rate Offer Dispersion for Identical Mortgages in Los Angeles

Data Source: Optimal Blue
Note: The spread is defined as the difference between real-time mortgage rate offers and the median offer rate for identical mortgage products. The histogram includes daily data between April 2016 and December 2019.

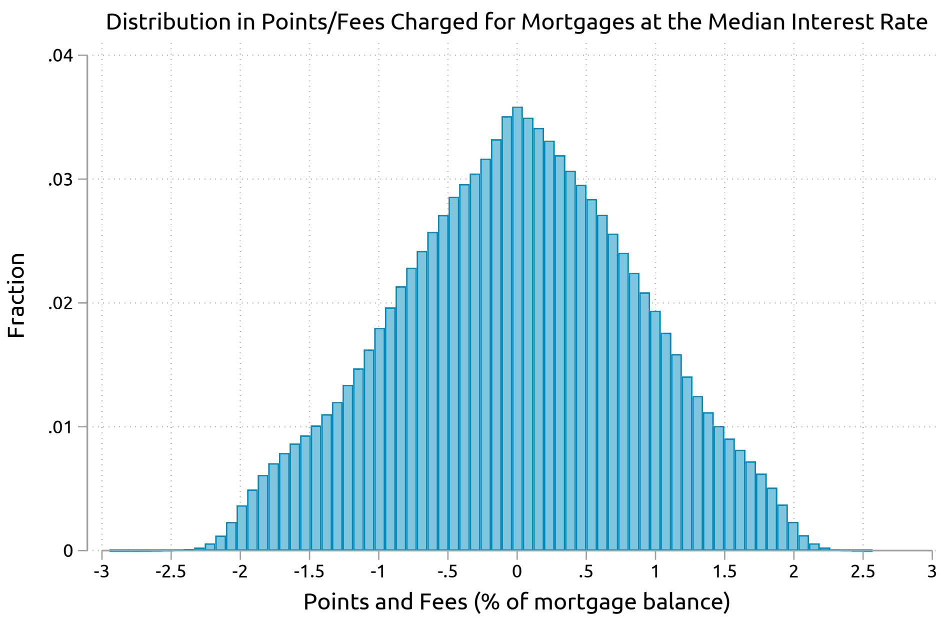


Figure A-7: Dispersion in Points and Fees Lenders Charge for Identical Mortgages at the Median Interest Rate

Data Source: Optimal Blue

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