FEDERAL RESERVE BANK
PHILADELPHIA

Working Papers
RESEARCH DEPARTMENT

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

Neil Bhutta
Federal Reserve Bank of Philadelphia
Consumer Finance Institute

Andreas Fuster
Swiss Finance Institute at EPFL

Aurel Hizmo
Federal Reserve Board

WP 24-11

PUBLISHED
June 2024

FRB CFI
PHILADELPHIA

ISSN: 1962-5361

DOI: https://doi.org/10.21799/frbp.wp.2024.11

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

NEIL BHUTTA ANDREAS FUSTER AUREL HIZMO†

June 5, 2024

Abstract

Comparing mortgage rates that borrowers obtain to rates that lenders could offer for the same loan, we find that many homeowners significantly overpay for their mortgage, with overpayment varying across borrower types and with market interest rates. Survey data reveal that borrowers' mortgage knowledge and shopping behavior strongly correlate with the rates they secure. We also document substantial variation in how expensive and profitable lenders are, without any evidence that expensive loans are associated with a better borrower experience. Despite many lenders operating in the US mortgage market, limited borrower sophistication may provide lenders with market power.

*This paper was previously circulated as "Paying Too Much? Price Dispersion in the US Mortgage Market." Adithya Raajkumar and Jacqueline Blair provided excellent research assistance. We thank Jason Allen, Robert Avery, John Campbell, Nick Embrey, Serafin Grundl, Katherine Guthrie, Hajime Hadeishi, 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, Bank of England, Baruch College (Zicklin), EPFL, Federal Reserve Bank of Atlanta, Federal Reserve Board, Freddie Mac, Norges Bank, NYU Stern, Oxford Saïd, University of Copenhagen, 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, and the NBER Summer Institute (Real Estate) for helpful comments. Special thanks to Jay Shultz, Saty Patrabansh, and Bob Avery 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.

†Bhutta: Federal Reserve Bank of Philadelphia; Fuster: EPFL, Swiss Finance Institute, and CEPR; Hizmo: Federal Reserve Board. Emails: neil.bhutta@phil.frb.org; andreas.fuster@epfl.ch; aurel.hizmo@frb.gov.

1 Introduction

Survey data indicate that half of the borrowers taking out a mortgage in the US only seriously considered one lender, and just 3 percent of the borrowers considered more than three lenders. Ninety-five percent of the respondents reported that they were satisfied that they received the lowest interest rate for which they could qualify.[1] 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 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. We argue that this overpayment reflects, at least in part, limited borrower sophistication, which provides lenders with market power.

To assess overpayment, we use unique 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. Importantly, these "offer" rates are lenders' private information, rather than publicly posted, and may not be automatically offered to prospective borrowers. 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 compare the rate they locked against the distribution of available rates for the same type of loan and borrower (same loan-to-value [LTV] ratio, credit score [FICO], points, etc.) on the same day in the same market. Motivated by a simple search model (e.g., Carlson and McAfee, 1983), we construct a metric we refer to as the *Expected Gain from Additional Search (EGain)*: the expected amount by which a borrower could reduce their rate if they were to obtain one additional rate quote.

We find that *EGain* differs substantially across borrower types. For example, borrowers

[1] Statistics are based on the National Survey of Mortgage Originations (NSMO), which covers mortgage originations from 2013 to 2020.

getting Federal Housing Administration (FHA) loans, who tend to have lower incomes and credit scores, have an average *EGain* of 28 basis points (bp), and, remarkably, for a quarter of these borrowers, *EGain* exceeds 40bp. In contrast, borrowers getting "jumbo" loans—who tend to have high incomes—have an average *EGain* of just 4bp, implying such borrowers are much more likely to obtain rates that come close to the best available. We interpret these results as consistent with borrowers across different market segments varying in their financial "sophistication," which we think of as encompassing factors such as the recognition that lenders may offer different rates, and the ability to search and negotiate for low rates.

We also find that the expected gain from search varies over time, inversely with interest rates. When market interest rates (and lenders' offer rates) are higher, the average *EGain* is lower, implying that borrowers tend to find better deals when rates are high. We argue that this relationship is in part driven by "behavioral" factors: When the level of rates is 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. Indeed, we also find explicit evidence from survey data on mortgage borrowers that shopping effort increases with market rates.

We provide supporting evidence of ineffective shopping and negotiation by analyzing price dispersion in the locks data alone, where we have a much larger sample and can further exploit unique features of the locks data. We find that the difference between the 90th and 10th percentile interest rate that observably identical borrowers lock in for the same loan in the same market, on the same day, and paying the same points, is 55bp; for a typical loan of $250,000, this corresponds to an upfront payment of about $6,250. Furthermore, the largest residual dispersion occurs for borrowers who may be the least financially sophisticated. For example, the 90-10 residual rate gap for FHA borrowers is about 72bp.[2]

We then unpack at which level this dispersion occurs—e.g., across lenders vs. within lenders, branches, or even loan officers. Estimated lender fixed effects exhibit quite substantial variation, but still only reduce the residual rate dispersion by a modest amount, suggesting limited explanatory power of factors such as (time-invariant) lender reputation or quality. Allowing lender fixed effects to vary by loan and borrower characteristics and over time improves explanatory power. However, even after further including branch and

[2]Notably, the lending platform providing our data is mostly used by monoline nonbank mortgage originators, and wide dispersion remains when we limit the sample exclusively to such lenders. Thus, we can rule out cross-selling or bundling of other services as explanations behind the observed patterns.

loan officer fixed effects, almost one-half of the residual dispersion remains, consistent with an important role for negotiation in this market.

In the final part of the paper, we present several findings that support the idea that mortgage lenders exert market power, which allows them to charge markups, and that a lack of borrower financial sophistication is a source of lender market power. First, we document that the largest lenders tend to be the most expensive (i.e., have high residualized rates), and exhibit the most dispersion in their (residualized) rates, suggestive of large lenders having market power. Next, we use detailed income and expense data for mortgage lenders from regulatory filings merged to our data on transacted rates to show that relatively expensive lenders have higher profit margins (net income per dollar originated). We also find that expensive lenders spend slightly more on occupancy and technology, but substantially more on salaries of loan officers and managers. This could imply a better borrower experience; yet our evidence from survey data indicates that, on the contrary, borrowers do not get better service when paying more for their mortgage. Finally, again using survey data, we show that the most sophisticated borrowers (those who shop the most and have high mortgage knowledge) get significantly cheaper mortgages than the least sophisticated. Moreover, this gap widens as local mortgage market concentration declines, suggesting that more intense lender competition is helpful primarily for sophisticated borrowers.

Overall, our empirical results indicate 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 who 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. Previous work in the US has assessed dispersion in offer rates (Alexandrov and Koulayev, 2017 and McManus et al., 2018); transacted mortgage rates (Agarwal et al., 2023 and Ambokar and Samaee, 2019); reset rates on adjustable-rate mortgages (Gurun et al., 2016); and mortgage broker fees (Woodward and

Hall, 2012). Our data contain key details not available in any other dataset (e.g., exact lock date, lock period, and points), which provides greater confidence that we are identifying true dispersion. Moreover, our data span multiple market segments—allowing us to show, for example, that dispersion is especially wide for FHA loans—and come from recent years after the implementation of post-crisis mortgage regulations (e.g., loan officer compensation rules) that could have reduced price dispersion. While we find considerable dispersion in mortgage rates, our estimates are substantively smaller than other estimates in the literature, as we discuss in a more detailed review of the literature in Appendix A.1.[3] Additionally, we document sizable dispersion within lender, branch, and even loan officer, suggestive of a role for negotiation, in contrast to assumptions in previous work that negotiation plays little role in the US mortgage market (Alexandrov and Koulayev, 2017).

Most notably, this paper is the first to compare transacted rates to offer rate distributions for identical loans, allowing us to measure the gains to additional search across all borrowers, including those at the bottom of the transaction rate distribution, and how gains vary by borrower type. We also document that borrower overpayment and shopping intensity change with market rates over time, consistent with behavioral factors affecting shopping effort.[4]

This paper is also the first to link data on firm-level mortgage pricing with revenues, costs, and profits. We confirm that our residualized lock rates are strongly correlated with firm revenues and also show that expensive lenders earn higher margins per dollar originated. Alongside these results, we provide novel evidence that borrowers who pay more are less satisfied with their lender, and that a lack of borrower sophistication may be undermining competition in this market.[5]

Our paper connects to the literature on (in)efficiency of consumer choice and price disper-

[3]Some work also exists outside the US, where the institutional details and the mortgage market structure are different. Allen et al. (2014b) study the Canadian market, where there is no dispersion in posted rates, but there is large dispersion in contracted rates, which they argue arises due to differences in bargaining leverage across consumers. Iscenko (2018) and Coen et al. (2023) document substantial heterogeneity in terms of how well UK borrowers fare when choosing mortgages from the options available at their lender; Liu (2019) shows that many UK borrowers appear to neglect nonsalient 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.

[4]The result that overpayment increases when market rates are low is distinct from and complements the finding of Fuster et al. (2023) that lender offers tend to feature higher markups when market rates are low.

[5]Our finding that lower market concentration may have limited benefits for less-sophisticated consumers is consistent with Allen et al. (2014a)'s evidence from mergers in the Canadian mortgage market. Malliaris et al. (2022) also use the NSMO data to document that indicators of sophistication are correlated with rates and study how the benefits of knowledge and shopping interact.

sion in various consumer finance markets, including mutual funds (Hortaçsu and Syverson, 2004; Choi et al., 2010), auto loans (Argyle et al., 2022), and credit cards (Stango and Zinman, 2016).[6] 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, 2017; Keys et al., 2016; Andersen et al., 2020), our results are in line with the growing literature pointing to financial literacy/sophistication as a key driver of differential outcomes in household finance (e.g., Hastings et al., 2013; Gomes et al., 2021).

In recent work, Agarwal et al. (2023) 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. They 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 who 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 provide evidence from the NSMO data that variation in sophistication is important to understanding cross-sectional dispersion.[7]

The rest of the paper is organized as follows. In the next section, we provide some background on the institutional details of this market. Section 3 describes our data sources. Section 4 explores how locked rates compare to the offer distribution; here we explain our $EGain$ measure and study how it varies across borrowers with different characteristics and over time. Section 5 assesses mortgage price dispersion in the locks data alone. Section 6 explores the connections between borrower sophistication, mortgage rates, and market power. Finally, Section 7 concludes with some potential policy implications.

[6] More broadly, a large literature studies price dispersion in various markets empirically and theoretically (see e.g., Baye et al., 2006, and Wright et al., 2021, for surveys of relevant work). One important conclusion from this literature is that a decrease in information cost (making shopping easier) does not necessarily reduce equilibrium price dispersion—neither empirically nor theoretically.

[7] 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 who 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 suffice to induce a borrower to accept a relatively "bad" offer.

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.[8]

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 can be used toward origination costs.

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

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

loans are ultimately sold.[9] 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).[10] Furthermore, lenders will add a margin that may depend, among other things, on the level of demand for loans (Fuster et al., 2023).

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," Woodward and Hall, 2012). 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.[11] 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 a 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.[12]

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

[9] 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.

[10] 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.

[11] 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.

[12] 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/pricing-discretion-fair-lending-risk.

application with a lender, but it can also happen at a later time. Not all rate locks ultimately lead to originated mortgages, since the loan application can still be rejected afterward (e.g., because the appraisal of the home comes in lower than expected) or the borrower could renege. However, the lock is binding for 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.

3 Data

In this section, we describe our primary data sources: (i) data on mortgage rate locks and lender offers from Optimal Blue; (ii) data on nonbank income statements from Mortgage Call Reports (MCR) collected by state regulators; and (iii) survey data from the National Survey of Mortgage Originations (NSMO).

3.1 Optimal Blue Data

Our main data come 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. More than 40,000 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 JPMorgan or Wells Fargo.

For this study, we use two components of the data generated by the platform: (i)) data on mortgage products and mortgage prices actually accepted by consumers, and (ii)) data on mortgage products available and mortgage prices offered by lenders.

3.1.1 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 of about 280 metropolitan areas as well as rural areas. All of the standard loan characteristics used for underwriting are included: 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 processing each mortgage. For some lenders, we can also observe loan officer compensation, expressed as a percentage of the loan amount.[13]

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 Home Mortgage Disclosure Act (HMDA) data.[14] This will be useful later to assess whether lender type and cross-selling might be driving the patterns in the data that we document.

We restrict the sample in various ways to ensure we study a relatively uniform set of

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

[14]We use a loan-level merge between Optimal Blue, administrative FHA data, and HMDA used in Bhutta and Hizmo (2021), which covers years 2014-2015. We also match Optimal Blue directly to the expanded public HMDA data from 2018-2019 to further classify lender types for lenders entering the Optimal Blue system after 2015.

loans that represent the type of mortgages originated in recent years. 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 Section 4, 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 (loans typically securitized through the GSEs, Fannie Mae or Freddie Mac), super-conforming (with loan amounts above the national conforming limit but below the local limit; the GSEs can still securitize these loans but potentially at slightly worse prices), jumbo (loan amount above the local conforming limit, meaning the loan cannot be securitized through the GSEs), and FHA loans (which require mortgage insurance from the FHA and are typically pooled into securities guaranteed by 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 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.2, 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 the data on originated loans.[15]

[15]See Appendix Table A-2 and Figure A-4 for details. For jumbo mortgages, the locked interest rates in

3.1.2 *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 (e.g., loan officers) 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 expect to pay at a particular lender. The Insight interface is designed for lenders to compare their pricing against that of peers.[16]

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 LO who has locked the most loans for that lender in that market.[17]

If a lender represents multiple different investors, the offer we observe is based on the most competitive investor offer. Thus, a borrower locking in 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.[18] Third, as noted earlier, borrowers may be able to negotiate and get an "exception" or a lower rate from the lender.

We conduct daily searches in one local market (Los Angeles), twice-weekly searches in four markets, and weekly searches for 15 additional markets.[19] We collect offer distributions for

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.

[16] The Pricing Insight data are different from earlier Optimal Blue data used by Fuster et al. (2023). 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").

[17] As explained further in Appendix A.3, we observe a distribution of prices (points) for a given note rate, which we transform into a distribution of rates for zero points.

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

[19] 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,

100 different loan types, differing across the following dimensions: FICO score, LTV ratio, loan program, loan purpose (purchase or cashout 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 for 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.

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 offers comparable, we subtract the median offer rate across lenders for the same product, day, and metropolitan area. 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) also document. In Appendix A.3, we further show that the degree of offer rate dispersion is quite similar across different types of loans and borrowers, and across all 20 cities in our sample.

3.2 Mortgage Call Report Data

We use firm-level data on mortgage-origination-related income and expenses of nonbank mortgage lenders from Mortgage Call Reports (MCR).[20] Nonbank financial institutions (aka shadow banks) that hold a state license through the Nationwide Multistate Licensing System (NMLS) must regularly complete an MCR, which consists of data on mortgage loan activity (e.g., total mortgage origination volume) and data on firm financial conditions, including detailed information on income and expenses related to mortgage lending.

In order to study how variation in mortgage pricing across firms relates to firm income, expenses, and profitability, we merge the MCR data with the Optimal Blue locks data. The MCR data cover about 90 percent of nonbank mortgage originations in 2018-2019; our MCR-Optimal Blue matched data cover 197 firms operating as independent nonbank mortgage lenders in 2018 or 2019 that accounted for about 40 percent of nonbank mortgage originations

Phoenix, Portland, San Diego, San Francisco, Seattle, Tampa, and Washington, DC.

[20]These data are available to Federal Reserve researchers through an agreement with the Conference of State Bank Supervisors, which owns and operates the system that collects MCR data on behalf of state regulators. In addition to income and expense data, firms also report data on assets and liabilities. Similar data are used by Jiang et al. (2020) to study the capital structure of nonbanks.

in 2018 and 2019. This lower coverage in our matched dataset relative to the MCR data mainly reflects the fact that not all nonbank lenders use the Optimal Blue platform.[21]

3.3 NSMO Data

The National Survey of Mortgage Originations (NSMO) is part of the National Mortgage Database® program, a joint initiative of the Federal Housing Finance Agency (FHFA) and the Consumer Financial Protection Bureau (CFPB). It surveys a nationally representative sample of borrowers with newly originated closed-end first-lien residential mortgages, focusing in particular on borrowers' experiences getting a mortgage, their perceptions and knowledge of the mortgage market, and their future expectations. We focus on the first 26 waves, including borrowers who took out mortgages from 2013 through 2019.[22] In addition to its unique focus on new mortgage borrowers, a key feature of the NSMO is that survey respondents are matched with extensive administrative data from mortgage servicing records and consumer credit files. Thus, we observe precise data on the credit risk of respondents (e.g., their credit score, DTI ratio) and the terms and conditions of the mortgages that respondents obtained.

For our analyses in this paper, we impose several sample restrictions. 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 focus on fixed-rate loans with a term of 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. These restrictions leave us with 22,567 mortgages for the analysis.[23] In all our NSMO analyses, we use the provided analysis weights, which are based on sampling weights and non-response adjustments.

[21]These coverage ratios use Home Mortgage Disclosure Act (HMDA) data to measure total market size (see Bhutta et al., 2017 for more on the HMDA data coverage and nonbank market share). We merge the MCR and Optimal Blue datasets on lender name, after first mapping Optimal Blue lender IDs to lender names by conducting a loan-level match to HMDA data. We match about 450 mortgage lenders operating in 2018 or 2019 from Optimal Blue to HMDA, of which about 240 are independent nonbanks.

[22]The public version of the data is available on FHFA's website. We are able to use a confidential version of the data that includes geographic information.

[23]The full NSMO dataset contains 39,615 loans originated before 2020. The restriction to 30-year mortgages drops just over 10,000 observations; most of the non-30-year mortgages have a term of 15 years. These sample sizes are as of July 2023, but the data may be revised in the future.

4 Comparing Locked Rates to Offer Rates

In this section, we study whether different types of borrowers get good or bad deals, relative to what is available in the market at the time they lock their mortgage. This will allow us to assess 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 distribution of lenders' offer rates (described in Section 3.1.2) by day, MSA, FICO, LTV, loan amount, and loan program (i.e., conforming, super-conforming, jumbo, and FHA). We match locked loans to the offers available for loans with nearly identical characteristics—see Appendix A.4 for additional detail.[24] We focus the analysis on purchase mortgages from the 20 MSAs for which offer data are collected and end up with 67,537 matched loans.

To compare the locked rate to the offers, the main metric we consider is the expected gain from additional search, which we refer to as *EGain*. This metric is motivated by a simple search model (e.g., Carlson and McAfee, 1983).[25] Suppose that there are n mortgage lenders that are posting mortgage rate offers for a particular borrower type. Rates are ordered from lowest to highest: $r_1 \leq r_2 \leq ... \leq r_n$. Borrowers see only the mortgage rates available at the lenders they meet with, and we assume that each borrower has an equal chance of meeting any one of the lenders. Suppose a borrower has already found a rate r_k and is considering searching to obtain one additional rate quote from a new lender. Given that the probability of meeting a lender (different from lender k) that offers the rate r_i is $1/(n-1)$, the expected gain from an additional search is given by:

$$EGain_k = \sum_{i=1}^{k-1} (r_k - r_i) \frac{1}{n-1} = \left[r_k - \sum_{i=1}^{k-1} \frac{r_i}{k-1} \right] \frac{k-1}{n-1}. \tag{1}$$

Intuitively, the term in brackets is the locked rate minus the expected rate available at 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 remaining population, which is $(k-1)/(n-1)$. Therefore, this

[24]The offer rates are for a loan with zero points and fees. To compare locked rates to these offers, we adjust the locked rate for points paid or received by the borrower based on the empirical relationship between points and interest rates. See Appendix A.5 for more detail on how we estimate this empirical relationship.

[25]We are grateful to Gregor Matvos for suggesting this measure to us.

is a measure of how much money the borrower is leaving on the table, in expectation, by not conducting one more search. The measure takes into account not just where the borrower's rate is relative to the center of the distribution, but it also depends on the width of the offer distribution—the expected gain from search is higher the more widely dispersed the offers are. Furthermore, in our implementation, we take into account that some lenders do not offer a given loan type at all, which would correspond to $r_i = \infty$.[26] However, it is important to note that the measure takes the distribution of offers at a point as given—while, if all borrowers were to search more, the equilibrium distribution of offers might shift. Thus, it is best to consider it as a measure of overpayment from the point of view of a single borrower.

We also consider an alternative simpler measure where the locked rate is directly compared to the median offer rate for matched loans—we call the difference between the two the *Locked-Offer Rate Gap*. Positive values indicate that a borrower overpays relative to what a typical lender (as measured by the median) could offer for the same loan characteristics on the same day.

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 available rates, this fraction is 3.5%. Thus, even the best offers are indeed available to borrowers.[27]

4.1 Cross-sectional Differences in Expected Gain from Additional Search

Panel A of Figure 2 shows the distribution of *EGain* for all mortgages in our data. The dashed vertical line denotes the mean of the distribution. *EGain* is by definition bounded below at zero. The figure shows that about a quarter of locked loans would in expectation gain less than 5bp from an additional search. However, the distribution is highly skewed, and there are borrowers who could expect to lower their rate by 50bp or more by randomly contacting another lender for an offer on the exact same loan type.

[26]Thus, we take as n the number of unique lenders in a given MSA on a given day that offer at least one loan type across all programs. However, setting n equal to the number of lenders that do post offers for a given loan type instead does not qualitatively change the results.

[27]In addition, in Appendix Figure A-3, we validate that offer rates derived from Optimal Blue Insights closely track offer rates for comparable loans published by Mortgage News Daily, an industry website, as well as by Zillow and Freddie Mac's Primary Mortgage Market Survey.

Panel B of Figure 2 and the summary statistics in Table 2 show that the distributions of *EGain* vary substantially across different loan programs. *EGain* is largest for FHA borrowers, with an average of 28bp. Moreover, one-quarter of FHA borrowers would in expectation gain 40bp or more from an additional search. In contrast, the distributions for super-conforming and jumbo borrowers look very different: Even at the 75th percentile, the expected gain from an additional search is only 12bp (super-conforming) and 3bp (jumbo), indicating that many borrowers in these segments come close to the best offers posted on Optimal Blue.

Table 2 further shows that borrowers with lower FICOs and higher LTVs typically have higher remaining gains from search and overpay the most relative to the median available offer. Importantly, this overpayment is relative to offered rates that already incorporate risk-based pricing for low FICOs and high LTVs. First-time homebuyers and those who pay discount points also tend to fare worse.[28] Finally, borrowers at independent nonbanks appear to overpay by more, on average, than borrowers at other lender types; we return to discussing potential differences across lender types below.

It is worth noting that within each of the groups in Table 2, there is substantial dispersion in *EGain* and the locked-offer rate gap, as shown by the usually large gaps between the 75th and 25th percentile. Thus, even for high-FICO or low-LTV borrowers, a nontrivial fraction of borrowers lock rates well above what other lenders could offer them. However, dispersion tends to be largest for the groups that on average fare the worst.

Appendix Table A-12 provides analogous summary statistics based on the median income, college education share, minority share, and mortgage market concentration in a borrower's location.[29] Average *EGain* is largest in areas with low household income, low shares of college-educated residents, and high minority shares; these are also the areas where the dispersion in this measure is larger. In contrast, there is little covariation of the *EGain* distribution with local mortgage market concentration.

[28] Note that since we adjusted the mortgage note rate for points paid, this relationship is not "mechanical."

[29] Income, education, and minority shares are measured at the zip code level based on 2017 American Community Survey data; mortgage market concentration is the Herfindahl-Hirschman Index at the county level averaged over 2016-2019 using the HMDA data.

4.1.1 Regression Analysis

The summary statistics above indicate that borrowers who are financially less well-off, such as lower-FICO and higher-LTV borrowers, overpay the most. As noted above, this overpayment is relative to offered rates that already incorporate risk-based pricing for low FICOs and high LTVs. Instead of reflecting risk-based pricing, overpayment could be driven by borrowers who are less financially sophisticated, meaning that they are less effective at searching and negotiating for low rates. However, an alternative potential explanation for these patterns unrelated to sophistication 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 3, we regress *EGain* 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 have lower *EGain* (not shown in table), i.e., they get better rates relative to what is available in the market. However, conditional on loan amount, lower-FICO borrowers and higher-LTV borrowers continue to overpay more, to a similar degree as we observed in Table 2. Thus, such borrowers appear to obtain more expensive loans (relative to available offers) for reasons beyond the differential 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. Borrowers might choose expensive lenders because they offer better service or because they spend more on marketing and are more visible. To investigate this explanation, we add lender-branch fixed effects in columns (2) and (5) of Table 3. In these columns, the R^2 jumps sharply to about 40% from less than 20%, meaning that lender-specific pricing differences explain a fair amount of variation in *EGain*. Furthermore, the coefficients on FICO and LTV become slightly smaller, implying that sorting into lenders does explain some of the overpayment by low-FICO and high-LTV borrowers. Still, even within the same branch, these borrowers tend to pay differentially more relative to what is available in the market. This suggests that they may be less effective at negotiating a good rate.

Along similar lines, low-FICO and high-LTV borrowers may sort into LOs who charge more and who may be more experienced and offer more assistance in the loan application process. In order to test for this possibility, in columns (3) and (6), we directly control for LO compensation, which typically ranges from 1-2% of the loan amount, for the subset of lenders that report it. The coefficients on this variable are strongly significant; however, the coefficients on FICO and LTV change little, implying that low-FICO and high-LTV borrowers do not pay higher rates simply because they match with expensive LOs. Rather, these borrowers appear to pay more than other borrowers even when working with LOs with the same compensation—suggesting, again, that they may be less effective in shopping and negotiating for a good rate.

Robustness. Table A-13 in the Appendix reproduces these regressions with the Locked-Offer Rate Gap as dependent variable; the patterns are similar. In another robustness check, reported in Table A-14, we restrict the sample to lenders that we can identify as independent nonbanks. Doing so leaves the coefficients from Table 3 essentially unchanged. As noted earlier, the nonbank lenders that constitute the majority of our sample are only in the business of originating mortgages. Thus, differences in *EGain* 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.[30]

Finally, it may be that many of the lenders making offers in our dataset are small and hard to find. To ensure our results are unaffected by this possibility, we replicate our analysis using only offers from high-volume lenders, as designated on the Optimal Blue platform. Our results, shown in Appendix Table A-15, again remain qualitatively unchanged.

4.2 Time-Series Movements in Expected Gain from Additional Search

The last section explored the cross-sectional patterns in the expected gain from additional search. In this section, we instead study how *EGain* 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 what is offered in the market) when market rates are low, and more likely to get a good deal as rates increase? If so, what might explain this relationship?

Figure 3 plots the average *EGain* against market interest rates, here measured by the 10-year Treasury yield.[31] In the summer of 2016, the level of market interest rates as shown by

[30] Table 2 showed that borrowers who obtain loans from nonbanks tend to have higher *EGain* and 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.

[31] For the average *EGain* series, we use the estimated month fixed effects from a regression similar to

Treasury yields was very low. *EGain* 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, *EGain* 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 average *EGain* almost mirror movements in Treasury yields.

We confirm the statistical significance of the relationship between *EGain* and market rates in Table 4.[32] The first column regresses *EGain* on the 10-year Treasury yield, controlling for all borrower characteristics jointly, MSA fixed effects, and MSA-level house price growth over the past 12 months to account for housing market "hotness," which could affect borrowers' willingness to spend time shopping around. The coefficient implies that as the 10-year Treasury yield increases by 1 percentage point, the average *EGain* falls by about 5bp. This is sizable, given that over our sample as a whole, *EGain* averages 18bp.

In column (2), we add month fixed effects (interacted with MSA) and see that the relationship between Treasury yields and *EGain* is similarly strong within-month. Column (3) further adds lender-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 2.6bp), suggesting that some of the overall relationship may be due to borrowers selecting cheaper lenders when rates are higher (consistent with additional shopping).

In the remaining three columns, we test to what extent this relationship may be driven by affordability constraints: When market rates rise, constrained borrowers may be forced to shop more in order to find a relatively good rate that will allow them to qualify for a mortgage (Bhutta and Ringo, 2021). To test this mechanism, we interact the Treasury yield with an indicator for borrowers with a DTI no higher than 36% (who are likely unconstrained by the payment burden from higher rates).[33] The interaction term in columns (4)-(6) is positive and significant, meaning that the relationship between *EGain* and market rates is indeed weaker for unconstrained borrowers. However, the magnitude of the estimated coefficients

those in Table 3 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.

[32] Appendix Table A-16 repeats the analysis with locked-offer rate gap as dependent variable.

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

is relatively small, meaning that even for unconstrained borrowers, *EGain* is higher when market rates are lower.

This suggests that the relationship may be driven at least partly by "behavioral" factors: 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 Appendix Section A.7.2, we use the NSMO data to provide direct evidence that shopping effort increases with market rates. For example, we find that when market rates are higher, recent mortgage borrowers are more likely to report having considered or applied to more than one lender. These patterns hold even after controlling for local house price growth to account for variation in market "hotness," which might be correlated with interest rates and affect borrowers' ability to spend time shopping for a cheaper mortgage.

Importantly, the higher *EGain* when market interest rates are low is in addition to the higher "price of intermediation" when rates are low, identified by Fuster et al. (2023). That paper shows that *offers* feature higher lender markups (relative to loan values in the MBS market) at times of high demand and provides evidence that this is at least in part driven by lender capacity constraints.[34] 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.

5 Understanding Variation in Locked Rates

In the previous section, we observed that many borrowers overpay relative to the rates available for the same mortgage in the same location on the same day. Furthermore, this overpayment is amplified when market interest rates are low. In this section, we take a more comprehensive look at the dispersion in rates locked in by borrowers that are identical on observables (including the day of the lock and the location). This analysis is closer to existing literature on this topic (reviewed in Appendix A.1), but the additional detail available in

[34] Capacity constraints could also affect relative bargaining power and contribute to the patterns documented in this section. When market rates are low and mortgage demand is very high, a borrower potentially has less bargaining power because a loan officer may not care as much about losing a customer, since there are many others waiting to refinance (and the loan officer is already at or close to capacity).

our locks data allows us to more precisely measure and unpack price dispersion than what earlier work was able to do.

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 and as the gap between 75th-25th or 90th-10th percentiles. Comparing the residual dispersion along with the adjusted R^2 across specifications allows us to assess the relative importance of different drivers of price dispersion in this market.

Table 5 shows the results from various specifications, estimated on the same set of nearly 3 million loans locked over the five-year period 2015-2019.[35] 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. Given the importance of pure time-variation to explain variation in rates, in the remaining columns, we also report an adjusted R^2 within day and MSA (i.e., after absorbing the first set of fixed effects).

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.[36] 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).[37] This specification is similar to regressions one could typically run with a mortgage servicing dataset.[38] Within day and MSA, the adjusted R^2 from these variables is 0.36, and substantial dispersion remains: The standard deviation in residuals is

[35] The estimation drops "singleton" observations that are completely determined by the set of fixed effects. 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 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.

[36] 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 cashout refinance). The choice of FICO and LTV bins is motivated by the loan-level price adjustments set by the GSEs.

[37] 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).

[38] 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. In Appendix A.1, we report an alternative, more realistic comparison.

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).[39] 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 55bp. The adj. R^2 of 0.42 indicates that standard underwriting variables and upfront payments explain less than half of the variation in interest rates paid across borrowers within the same day/MSA.

Based on the regression coefficient on discount points (not shown in the table), we can translate interest rates to upfront points. This coefficient implies that 1 discount point changes the interest rate by about 21.8bp on average (see Appendix A.5 for details). Therefore, 55bp in rate is approximately equivalent to 2.5 upfront discount points or 2.5% 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 \$6,250 relative to someone borrowing at the 10^{th} percentile interest rate. Alternatively, if one prefers to think in terms of mortgage payments, 55bp correspond to about \$80/month for a \$250k loan at the average level of rates over our sample period.

Before discussing the remaining columns of Table 5, we note that Table 6 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 is highly 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.[40] Finally, we also see wide dispersion even when we focus

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

[40] 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. 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

just on low-risk borrowers: those with prime FICO scores in excess of 680, and those with LTVs of less than 75%.

Jumping back to Table 5, in column (4) we add lender fixed effects to allow for the possibility that some of the rate dispersion may reflect differences in lender characteristics such as service quality or advertising costs; we study correlates of these estimated fixed effects in Section 6. We find that adding these lender effects decreases the 90th-10th percentile difference by 7bp and increases the within-day-and-MSA adjusted R^2 by 0.1. In columns (5) and (6), we additionally 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. This reduces the 90-10 gap by a further 10bp from column (4), and adds another 0.1 to the adjusted R^2. These results suggest that in addition to time-invariant differences across lenders, 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 vary across different branches of a lender. 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.[41]

The branch fixed effects in column (7) have noticeable incremental explanatory power, increasing the adjusted R^2 and reducing the residual dispersion. Adding branch-by-month fixed effects in column (8) has more modest effects. Column (8) should come close to looking at observably identical borrowers getting a loan from the same branch at the same time, yet 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 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

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.

[41]We discuss correlates of estimated branch fixed effects (within lender) in Appendix A.6; see also the discussion in Section 6.1 below.

up with (within a branch) does appear to matter somewhat for the rate they end up with, since the adjusted R^2 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 90-10 gap remains at 26bp.

The last column of Table 6 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 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 3.1.1). This suggests that the large dispersion is not driven by unobservable pricing adjustments that banks or credit unions might make for customers who already have accounts or other business with them.

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 and potentially less sophisticated borrowers.

6 Borrower Sophistication and Lender Market Power

Thus far, we have presented evidence that many borrowers lock in mortgage rates that are high relative to available offers in the market. We have also seen that there is considerable dispersion in locked rates within lender and within location (e.g., zip code). These results suggest that price differences across borrowers cannot be fully explained by lender-specific attributes (e.g., service quality) or local market area characteristics (e.g., lender concentration). Moreover, we find that rate dispersion and the expected gains from additional search are highest among borrowers who may be the least financially sophisticated, such as low-FICO and FHA borrowers. These facts together suggest an imperfect market where lenders are able to charge markups to many borrowers, especially those who may lack the ability to shop and negotiate effectively.

As noted earlier, in most US cities, there are many active lenders, and market concentration is low (Amel et al., 2018). With many active firms selling a highly standardized product, it may be surprising to observe significant markups and price dispersion. However, theoretically, less concentration may not always improve competition and reduce markups. For example, the model of Armstrong and Vickers (2022), where consumers only consider subsets of firms when choosing from which one to buy, has the feature that entry does not always lower prices (see also Gabaix et al., 2016). As Syverson (2019) discusses, firms have "market power" when they can influence the price they charge as they do not face a perfectly elastic demand curve, and concentration can be a poor proxy for market power.

In this section, we provide further evidence, using additional data sources, that price dispersion reflects an imperfect market where lenders exert market power, as borrowers lack the ability to shop effectively. This section proceeds in four subsections. In the first, we measure lender "expensiveness" and document which firm characteristics (e.g., size) are correlated with expensiveness. In the second, we use the matched MCR-Optimal Blue data to test whether expensive lenders earn higher profits or have higher costs. In the last two subsections, we turn to the NSMO data. We first use these data to help assess whether more expensive lenders differentiate themselves by providing better service to borrowers. Then, we test how borrowers' sophistication relates to the mortgage rates they get and whether competition may be undermined by a lack of borrower sophistication.

6.1 Which Lenders Are the Most Expensive?

We begin by studying whether variation in lender "expensiveness" is related to other lender attributes such as size and type of firm (i.e., bank or nonbank). We measure expensiveness as the estimated lender fixed effects from specification (4) of Table 5. There are 678 unique lenders in the data, but for this analysis we restrict to the 452 lenders that locked at least 100 loans in the sample used for the regressions in Table 5. Figure 4 displays the distribution of fixed effects. The standard deviation of these fixed effects is about 0.19, or 19bp. Thus, there is considerable variation across lenders in their average mortgage rate expensiveness, in addition to considerable *within*-lender dispersion as we showed earlier.

Next, we run lender-level regressions of the lender fixed effects on lender size, proxied here by the number of locks a lender makes in total, and lender type (nonbanks vs. banks). To facilitate interpretation, we simply divide lenders into size quartiles. Column (1) of Table 7 shows that larger lenders tend to be significantly more expensive, and that nonbanks are more expensive than banks in the sample (for some lenders, type cannot be determined, but their fixed effects are not significantly different from the banks', which are the omitted category). In column (2), we add the share of locks by a lender that are FHA, jumbo and superconforming loans (with conforming loans as the omitted category). Lenders with higher shares of FHA loans are significantly more expensive, and adding this information on loan composition increases the explanatory power of the regression substantially.[42]

In the third column, we find that lenders with higher within-lender rate dispersion also tend to be more expensive.[43] Furthermore, column (4) shows that once we control for the lender size, loan type, and the variation in the residual, FHA share and within-lender dispersion remain highly statistically significant, while the effect of lender size is reduced. This reflects in part that larger lenders exhibit more variation in the rate residual (see last column of Table 7).

In sum, we find that larger lenders, and those that originate a larger share of FHA mortgages, are more expensive on average, and also exhibit more variation in the mortgage rates their borrowers lock, after finely accounting for observable characteristics. One possible interpretation of these findings is that larger lenders may have market power that enables them to charge more on average, while still offering competitive rates to some borrowers (likely those who shop around and/or negotiate).[44] Alternatively, some lenders may be able to have high market share and charge higher prices if they can differentiate themselves by providing better service to borrowers. Sections 6.2 and 6.3 look at whether loan pricing is related to lender costs and customer satisfaction.

6.1.1 Branch expensiveness

In Appendix A.6, we describe a similar analysis to understand pricing variation across branches *within* lenders. Similar to the finding for lender expensiveness, we document that

[42]Recall that the lender fixed effects are estimated from regressions where the loan type, as well as many other variables, are controlled. Thus, this regression is not reflecting that FHA loans are more expensive; rather, it indicates that lenders that make many FHA loans are on average more expensive across all loan types.

[43]Dispersion is measured as the standard deviation of residuals from specification (6) of Table 5—i.e., after accounting for lender fixed effects that are further allowed to vary across loan types and over time. The 10th percentile of this variable is 0.07, the 90th percentile is 0.20.

[44]The Armstrong and Vickers (2022) model predicts that larger firms choose higher maximum prices, in line with our evidence. In their model (where firms set a range of prices and play mixed strategies), this is because some consumers only consider one firm they are aware of; so large firms have less to gain from setting lower prices to attract consumers who consider multiple firms.

branches that originate more FHA loans are more expensive. We also find that the local characteristics of the areas where branches are originating loans correlate with their pricing. For example, branches in areas with a lower share of college-educated population are more expensive. This finding is reminiscent of findings in Drechsler et al. (2017) that deposit spreads vary with local college-education shares, which suggests a lack of depositor sophistication may be a source of market power in deposit markets. In Section 6.4 below, we use the NMSO data to measure borrower sophistication more directly and at the individual level, and we find that sophistication strongly predicts the rates borrowers get.

6.2 Do Expensive Lenders Earn Higher Profits?

In this subsection, we ask whether the large differences in loan expensiveness across lenders documented above are reflected in higher costs or in higher profitability reported by these lenders. To do this, we merge our measure of lender expensiveness from Optimal Blue with quarterly financial filings of these lenders (the MCR data; see Section 3.2). For our sample period of 2015:Q1-2019:Q4, there are 1897 lender-quarter filings, from 162 unique lenders.

The first three columns of Table 8 provide summary statistics on the main line items from lenders' financial statements.[45] The median of reported gross income is \$4.75 per \$100 originated, and most of it comes from secondary market income (or "gain-on-sale"), which is the income lenders earn from selling the loans they originate in the secondary market. The median of gross expenses for lenders is \$4.16 per \$100 originated, with the majority of expenses going toward personnel expenses. The second-to-last line shows after-tax net income for only residential origination before corporate allocations. The median of this measure is \$0.49 per \$100 originated. The last line of the table shows net income across all lines of business (including servicing) after taxes and corporate allocations, with a median of \$0.32 per \$100 originated. Note that there is also substantial heterogeneity in all of these variables across lenders within the same quarter. The cross-sectional standard deviation of gross and net income is \$1.97 and \$1.87 per \$100 originated, respectively.

Next, we investigate whether variation in lender "expensiveness" is related to these measures of lender income and costs. As in the previous subsection, we measure expensiveness as the lender fixed effects from specification (4) of Table 5. The last column of Table 8 displays

[45]Both income and expenses are only for origination, warehousing, and secondary marketing of mortgages for 1-4 unit residential and do not include income and expenses from other lines of business such as servicing, multifamily/commercial, or residential property portfolio management.

results from separate median regressions of each line item on our measure of lender expensiveness (expressed in percentage points) and year-quarter fixed effects.[46] A 1 percentage point increase in interest rates charged by lenders is associated with an extra gross income of \$4.05 per \$100 originated. The size of the effect is in line with expectations, as it implies that the value of 1 percentage point of interest rates to lenders is about 4 upfront points, which implies a reasonable point-rate tradeoff of about 25bp. Of the different income line items, lender expensiveness has the largest effect on secondary market income, although it does also have a moderate effect on origination income (i.e., fees). While these results are not surprising, they serve to validate that our measure of lender expensiveness is related to income exactly as we would expect.

Turning to lender costs, a 1 percentage point increase in interest rates charged is associated with \$3.50 of extra gross expenses per \$100 originated. Most of these extra expenses reflect increased personnel expenses, particularly payments to loan officers and managers. More expensive lenders do spend a bit more on technology, occupancy, and equipment; however these effects are fairly small.

Lastly, we find a positive and significant effect on net income: A 1 percentage point increase in interest rates charged is associated with \$0.45 of extra net income for \$100 of residential loans originated, and \$0.24 of extra net income across all lines of business after taxes and corporate allocations.

Overall, we find that the substantial variation in our estimated fixed effects (shown in Figure 4) is also reflected in the gross and net income that nonbank lenders report. Thus, it is not the case that the variation in interest rates across borrowers "nets out" within a lender such that some borrowers would cross-subsidize others but without an overall impact on the lenders; rather, some lenders are indeed more expensive and profitable than others.[47] Although more expensive lenders also have higher costs, these higher costs are not sufficient to offset their higher income. Perhaps even more importantly, the higher costs mostly reflect payments to loan officers and managers; only to a minor extent do more expensive lenders spend more on technology or occupancy (e.g., office rental) that would arguably improve

[46] We focus on median regressions in order to minimize the effect of outliers and possible data errors in the filings data. In Appendix Table A-17, we show results from OLS (mean) regressions, which are qualitatively similar, although less precise.

[47] This point is further illustrated in Appendix Figure A-2, which shows that the shares of borrowers who get particularly good deals or particularly bad deals are negatively related within lenders—not positively as in a world with pure cross-subsidization.

borrowers' experience. Of course, it is still possible that the more highly paid employees at expensive lenders provide superior service to borrowers; we test this hypothesis in the next subsection.

6.3 Are Expensive Lenders Better?

Above, we saw that lenders that charge more tend to have higher personnel costs. This leaves open the possibility that the loan officers at expensive lenders are of "better quality" and provide a better service to customers, which in turn might justify their higher costs.

We cannot directly test this hypothesis within our Optimal Blue and MCR data, but we can do so using the NSMO survey data. This survey asks respondents a battery of questions about the quality of their experience in obtaining a mortgage. For example, the survey asks "yes/no" questions about whether borrowers experienced delays in their closing date and in paperwork processing. It also asks borrowers about their level of satisfaction ("very," "somewhat," or "not at all" satisfied) with their lender, with various aspects of the lending process, and with the interest rate that they got.

To test whether borrowers who pay more receive better service, we run a series of regressions of lender service quality indicators on the interest rate borrowers obtained, conditional on a detailed set of controls:

$$Y_{ijtw} = \beta Rate_i + \Gamma Z_{ij} + \alpha_t + \delta_w + \epsilon_{ijtw}. \tag{2}$$

In equation (2), Y refers to a service-quality outcome for borrower i with mortgage characteristics j who originated a loan in month t and was surveyed in wave w. $Rate_i$ is the contract rate on the mortgage for borrower i, and Z_{ij} is a rich set of borrower and mortgage characteristics. The full list of controls is provided in the note to Table 9; it contains, for instance, flexible controls for credit score and LTV, fixed effects for county, program (e.g., GSE or FHA) and purpose (purchase or refinance), as well as borrower characteristics capturing income, employment, wealth, race, and more.[48] We further include origination month fixed effects α_t, which will absorb any economy-wide changes in service quality (e.g., when lenders hit capacity constraints), and survey wave fixed effects δ_w.

[48] One limitation of the NSMO data is that they do 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 differences in rates due to variation in points.

We emphasize here that most of the right-hand-side variables in equation (2) are drawn from administrative data rather than being self-reported. Because of the large number of precisely measured loan and borrower controls, we interpret β as the relationship between service quality and mortgage expensiveness, as the controls account for variation in mortgage rates that are due to borrower risk, loan type, and aggregate time series fluctuations.

The results in Table 9 fail to support the notion that more expensive mortgages are associated with better service quality. In fact, more expensive mortgages appear to be associated with more delays (though not statistically significant) and less satisfaction. Borrowers with a mortgage that is 100bp more expensive are 13 percentage points less likely to report being very satisfied with their mortgage rate and about 2 percentage points less likely to be very satisfied with their lender, the application process, and the closing process. The other coefficients, though smaller in magnitude, also have a negative sign. Overall, rather than getting better service quality in return for a more expensive mortgage, these results point in the direction of borrowers getting worse service despite paying more.

6.4 Sophistication, Concentration, and Mortgage Rates

If consumers lack knowledge about mortgages or do not shop around effectively, this could generate some degree of market power for lenders, even in low-concentration markets where many lenders offer mortgages. Furthermore, sophisticated borrowers may be better able to take advantage of more intense competition (lower market concentration). In this section, we examine these hypotheses based on the NSMO data.

We start by studying the relationship between a borrower's mortgage contract rate and their shopping and knowledge. We estimate OLS regressions of the form

$$Rate_{ijtw} = \beta X_i + \Gamma Z_{ij} + \alpha_t + \delta_w + \epsilon_{ijtw}, \tag{3}$$

where $Rate_{ijtw}$ is the contract rate on the mortgage for borrower i with loan characteristics j, loan origination month t, and responding to survey wave w. X_i are different measures of borrower i's shopping effort or knowledge about the mortgage market based on several questions in the NSMO. Z_{ij} is a rich set of borrower and mortgage characteristics that could influence the pricing of the loan and is similar to the set of controls described above in equation (2) (see the note to Table 10 for the full list of controls; recall that these are mostly drawn from loan-level administrative data, i.e., not self-reported).

The first column of Table 10 shows how mortgage rates obtained by borrowers correlate with 8 shopping and knowledge variables simultaneously. We think of the first four items as capturing shopping effort, while the other four capture knowledge about the mortgage market and their own mortgage. All are individually significant, suggesting that there are different dimensions to shopping and knowledge that can contribute to a borrower obtaining a low rate.[49] For instance, borrowers who are 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.[50] Conversely, shopping alone does not guarantee a good rate if a borrower's knowledge is low (see also Malliaris et al., 2022).[51]

In the second column, we regress $Rate_{ijctw}$ on a composite "sophistication index" measure, which we construct as the sum of six of the shopping and knowledge dummy variables, and then divide by six so that it ranges from 0 to 1.[52] The coefficient on the sophistication index of -0.226 implies that the difference in rates between the most and least sophisticated borrowers is nearly 23bp. This result suggests that variation across consumers in shopping and knowledge is an important contributor to mortgage rate dispersion, especially considering that our sophistication index is based on coarse responses to qualitative survey questions, likely leading to individual-specific noise and attenuation of the resulting coefficients.[53]

[49] In Appendix A.7, we provide a more complete description and summary statistics for each shopping and knowledge variable, and regression results for each of these variables individually. The coefficients are only slightly larger relative to the simultaneous regression in Table 10. Borrowers may apply to 2+ lenders for reasons other than "better terms" (e.g., because they got turned down at one lender), which we condition on. In the appendix, we show that applying to 2+ lenders for other reasons is *positively* related to rates, in line with the findings of Agarwal et al. (2023).

[50] One may be concerned that borrowers who use a mortgage broker report only considering one lender even though the broker may be shopping across many lenders on the borrower's behalf. When we control for whether a borrower used a mortgage broker, our results remain virtually unchanged, and using a broker is associated with getting a slightly higher mortgage rate.

[51] In an earlier version of this paper (available at `https://doi.org/10.17016/FEDS.2020.062`), we provide a complementary analysis using data from the 2016 Survey of Consumer Finances. 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.

[52] We include the dummy "considered 3+ lenders" but not the dummy for considering exactly 2 lenders; we also exclude the dummy for whether "most lenders offer the same rate" since this question was not asked in early waves. The median borrower has a sum of 3 (i.e., an index value of 0.5); about 14 percent of borrowers have a sum of 1 or 0, and about 8 percent of borrowers have a sum of 5 or 6.

[53] 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. In Appendix A.7.1, we provide evidence that mortgage knowledge and shopping is higher for jumbo borrowers relative to FHA borrowers, and high-FICO borrowers relative to low-FICO borrowers, consistent with our earlier results that overpayment is correlated

In column (3) of Table 10, we interact the sophistication index with market concentration, measured as the county-level HHI of lender concentration in the year before origination and standardized to have mean zero and standard deviation of one. Bear in mind that most of the US population live in counties with a low HHI, limiting the variation and range of HHI over which we can estimate the role of concentration. Nonetheless, we estimate a statistically significant and positive coefficient on the interaction term $HHI \times Sophistication$. This means that sophisticated borrowers tend to pay lower rates in less concentrated markets (HHI low) than in more concentrated ones (HHI high), while for less sophisticated borrowers, the differences across market types are smaller. Thus, this finding is consistent with the idea that a reduction in market concentration may be less effective in promoting competition and limiting markups when borrowers are less sophisticated.

Further support for this story is provided in Appendix A.8. We use the Optimal Blue locks data to assess whether the propensity of borrowers to get "good deals" (a rate residual less than -20bp) or "bad deals" (a rate residual larger than +20bp) varies with county HHI. We find that with higher market concentration, the likelihood of getting a good deal is significantly lower (while the average rate is not significantly related with HHI).

7 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 attempt to 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.

Given our findings, future research might focus on the design of policies that would help borrowers search and negotiate more effectively. Alternatively, future research could study whether the problem can be alleviated if the guaranteeing agencies were to impose requirements on the maximum locked-offer rate spreads 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 Agarwal et al. 2023, Alexandrov and Koulayev 2017, or Guiso et al. 2022).

with loan program and credit score.

The negative relationship between overpayment and the level of market rates that we document in Section 4.2 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.[54] 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 policymakers, depending on whether slowing the housing market through higher mortgage rates is seen as desirable in a given situation or not.

[54] See Amromin et al. (2020) for a review of related work.

References

Agarwal, Sumit, 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," *Econometrica*, 2023, *forthcoming.*

__ **, 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-François Houde, "The Effect of Mergers in Search Markets: Evidence from the Canadian Mortgage Industry," *American Economic Review*, October 2014a, *104* (10), 3365–96.

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

Ambokar, Sumedh and Kian Samaee, "Mortgage Search Heterogeneity, Statistical Discrimination and Monetary Policy Transmission to Consumption," Working Paper, University of Pennsylvania 2019.

Amel, Dean, Elliot Anenberg, and Rebecca Jorgensen, "On the geographic scope of retail mortgage markets," *FEDS Notes*, 2018.

Amromin, Gene, Neil Bhutta, and Benjamin J. Keys, "Refinancing, Monetary Policy, and the Credit Cycle," *Annual Review of Financial Economics*, 2020, *12* (1), 67–93.

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," *Review of Financial Studies*, 2022, *36* (7), 2685–2720.

Armstrong, Mark and John Vickers, "Patterns of Competitive Interaction," *Econometrica*, 2022, *90* (1), 153–191.

Baye, Michael R., John Morgan, and Patrick Scholten, "Information, Search, and Price Dispersion," in T. Hendershott, ed., *Economics and Information Systems*, Elsevier, 2006, pp. 323–375.

Bhutta, Neil and Aurel Hizmo, "Do Minorities Pay More for Mortgages?," *Review of Financial Studies*, 2021, *34* (2), 763–789.

— **and Daniel Ringo**, "The effect of interest rates on home buying: Evidence from a shock to mortgage insurance premiums," *Journal of Monetary Economics*, 2021, *118*, 195–211.

— **, Steven Laufer, and Daniel Ringo**, "Residential Mortgage Lending in 2016: Evidence from the Home Mortgage Disclosure Act Data," *Federal Reserve Bulletin*, 2017, *103* (6).

Buchak, Greg and Adam Jørring, "Do Mortgage Lenders Compete Locally? Implications for Credit Access," Working Paper, Stanford University and Boston College 2021.

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

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.

Coen, Jamie, Anil Kashyap, and May Rostom, "Price Discrimination and Mortgage Choice," Working Paper 31652, National Bureau of Economic Research 2023.

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

d'Avernas, Adrien, Andrea L. Eisfeldt, Can Huang, Richard Stanton, and Nancy Wallace, "The Deposit Business at Large vs. Small Banks," Working Paper 31865, National Bureau of Economic Research 2023.

Drechsler, Itamar, Alexi Savov, and Philipp Schnabl, "The Deposits Channel of Monetary Policy," *Quarterly Journal of Economics*, 2017, *132* (4), 1819–1876.

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," *Journal of Finance*, 2023, *forthcoming.*

Gabaix, Xavier, David Laibson, Deyuan Li, Hongyi Li, Sidney Resnick, and Casper G. de Vries, "The impact of competition on prices with numerous firms," *Journal of Economic Theory*, 2016, *165*, 1–24.

Gomes, Francisco, Michael Haliassos, and Tarun Ramadorai, "Household Finance," *Journal of Economic Literature*, 2021, *59* (3), 919–1000.

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," *Journal of Financial Economics*, 2022, *143* (3), 1209–1226.

Gurun, Umit G., Gregor Matvos, and Amit Seru, "Advertising Expensive Mortgages," *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.

Hurst, Erik, Benjamin Keys, Amit Seru, and Joseph Vavra, "Regional Redistribution through the US Mortgage Market," *American Economic Review*, 2016, *106* (10), 2982–3028.

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

Jiang, Erica, Gregor Matvos, Tomasz Piskorski, and Amit Seru, "Banking without Deposits: Evidence from Shadow Bank Call Reports," Working Paper 26903, National Bureau of Economic Research 2020.

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.

Malliaris, Steven, Daniel A. Rettl, and Ruchi Singh, "Is competition a cure for confusion? Evidence from the residential mortgage market," *Real Estate Economics*, 2022, *50* (1), 206–246.

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.

Syverson, Chad, "Macroeconomics and Market Power: Context, Implications, and Open Questions," *Journal of Economic Perspectives*, 2019, *33* (3), 23–43.

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.

Wright, Randall, Philipp Kircher, Benoît Julien, and Veronica Guerrieri, "Directed Search and Competitive Search Equilibrium: A Guided Tour," *Journal of Economic Literature*, March 2021, *59* (1), 90–148.

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

Notes: Sample includes 30-year fixed-rate mortgages on owner-occupied single-unit properties, with full documentation of assets and income. Self-employed borrowers, loans for amounts under $100,000, VA loans, and streamline refinances are excluded.

Table 2: Summary Statistics of the Expected Gain from Search and Locked-Offer Rate Gap

		Expected Gain from Search			Locked-Offer Rate Gap		
	Observations	Mean	$25^{th} pct.$	$75^{th} pct.$	Mean	$25^{th} pct.$	$75^{th} pct.$
All Mortgages	67,637	0.18	0.04	0.25	0.10	-0.08	0.25
Program							
FHA	14,715	0.28	0.08	0.40	0.25	0.01	0.46
Conforming	46,535	0.17	0.04	0.22	0.09	-0.06	0.22
Super-Conforming	4,448	0.10	0.01	0.12	-0.05	-0.21	0.08
Jumbo	1,939	0.04	0.00	0.03	-0.23	-0.35	-0.08
FICO							
[640, 660)	7,629	0.27	0.06	0.40	0.23	-0.02	0.46
[680, 700)	9,617	0.22	0.05	0.31	0.16	-0.06	0.35
[720, 740)	10,666	0.19	0.04	0.26	0.11	-0.06	0.26
740+	39,725	0.15	0.04	0.21	0.07	-0.09	0.20
LTV							
(75, 80]	22,270	0.13	0.03	0.17	0.02	-0.12	0.16
(85, 90]	7,295	0.15	0.04	0.21	0.05	-0.09	0.20
(90, 95]	16,573	0.16	0.04	0.22	0.08	-0.08	0.21
(95, 97]	21,499	0.27	0.08	0.38	0.23	0.01	0.42
First-Time Homebuyer							
No	33,378	0.15	0.03	0.20	0.06	-0.09	0.20
Yes	34,253	0.21	0.05	0.30	0.15	-0.05	0.31
Discount Points							
[-5, -0.2)	14,133	0.14	0.02	0.19	0.00	-0.17	0.19
[-0.2, 0.2]	23,430	0.16	0.03	0.21	0.08	-0.10	0.21
(0.2, 5]	30,074	0.22	0.06	0.30	0.18	-0.01	0.31
Lender Type							
Independent Non-bank	56,483	0.19	0.05	0.26	0.12	-0.05	0.27
Other/Unclassified	11,154	0.13	0.02	0.17	0.00	-0.15	0.16

Data Source: Optimal Blue

Notes: The expected gain from an additional search, given by equation (1). The locked-offer rate gap is the difference between each locked rate and the median offer rate in the same market on the same day for an identical mortgage. 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 3: Regressions of the Expected Gain from Search on Observables

	(1)	(2)	(3)	(4)	(5)	(6)
FICO (omitted cat.: [640,660))						
$I_{680 \leq FICO < 700}$	-0.039***	-0.029***	-0.024***			
	(0.006)	(0.005)	(0.008)			
$I_{720 \leq FICO < 740}$	-0.069***	-0.047***	-0.045***			
	(0.008)	(0.006)	(0.009)			
$I_{FICO \geq 740}$	-0.097***	-0.067***	-0.056***			
	(0.008)	(0.006)	(0.008)			
LTV (omitted cat.: (60,80])						
$I_{85 < LTV \leq 90}$				0.015***	0.011***	0.011***
				(0.003)	(0.003)	(0.004)
$I_{90 < LTV \leq 95}$				0.033***	0.025***	0.021***
				(0.004)	(0.004)	(0.004)
$I_{LTV > 95}$				0.128***	0.102***	0.074***
				(0.009)	(0.008)	(0.011)
Loan Officer Comp (%)			0.107***			0.096***
			(0.025)			(0.027)
Loan amount F.E. ($10k bins)	Yes	Yes	Yes	Yes	Yes	Yes
MSA x Month F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Lender-Branch F.E.		Yes	Yes		Yes	Yes
Adj. R-Squared	0.122	0.403	0.403	0.160	0.429	0.417
Observations	67637	65757	15444	67637	65757	15444

Data Source: Optimal Blue

Notes: The dependent variable is the expected gain from an additional search, given by equation (1). The data cover mortgage rates for 20 metropolitan areas during the period 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 4: The Relationship Between the Expected Gain from Search and Treasury Yields

	(1)	(2)	(3)	(4)	(5)	(6)
Treasury Yield	-0.051***	-0.044**	-0.026**	-0.057***	-0.051***	-0.031**
	(0.007)	(0.018)	(0.012)	(0.007)	(0.018)	(0.012)
Treasury Yield x DTI $\leq$ 36				0.013**	0.015***	0.012***
				(0.005)	(0.005)	(0.004)
Borrower and Loan Controls	Yes	Yes	Yes	Yes	Yes	Yes
MSA F.E	Yes	Yes	Yes	Yes	Yes	Yes
MSA x Month F.E.		Yes	Yes		Yes	Yes
Lender-Branch F.E.			Yes			Yes
Adj. R-Squared	0.145	0.163	0.430	0.146	0.165	0.430
Observations	67241	67241	65349	67241	67241	65349

Data Source: Optimal Blue

Notes: The dependent variable is the expected gain from an additional search, given by equation (1). Treasury Yield is the daily 10-year yield on the day of the mortgage rate lock. All specifications include controls for FICO, LTV, and loan amount, as well as for MSA house price growth (from Zillow) over the past 12 months. The data cover mortgage rates for 20 metropolitan areas during the period 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 5: 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	0.21	0.18	0.16	0.15	0.15	0.13
75-25 Percentile	0.36	0.28	0.26	0.22	0.21	0.17	0.15	0.14	0.13	0.12
90-10 Percentile	0.79	0.58	0.55	0.48	0.44	0.38	0.33	0.31	0.30	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	0.58	0.75	0.77	0.81	0.82	0.85	0.88	0.88	0.89	0.90
Adj. R-Squared (Within Date x MSA)		0.36	0.42	0.52	0.55	0.62	0.69	0.70	0.72	0.74
Observations	2996149	2996149	2996149	2996149	2996149	2996149	2996149	2996149	2996149	2996149

Data Source: Optimal Blue

Notes: The dependent variable is the mortgage interest rate locked. The data cover mortgage rates locked for 277 metropolitan areas during the period 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 6: Summary Statistics of the Residualized Locked Rate

	Observations	$90^{th} - 10^{th}$ Percentile Gap	
		Spec. (3) of Table 5	Spec. (10) of Table 5
All Mortgages	2,996,149	0.55	0.26
Program			
FHA	884,681	0.72	0.31
Conforming	2,001,083	0.48	0.25
Super-Conforming	72,419	0.43	0.21
Jumbo	37,966	0.48	0.24
FICO			
< 600	42,369	0.93	0.47
$[600, 640)$	221,431	0.82	0.37
$[640, 680)$	490,343	0.68	0.30
$[680, 740)$	933,110	0.55	0.27
≥ 740	1,308,896	0.44	0.23
LTV			
≤ 75	506,400	0.47	0.23
$(75, 80]$	624,315	0.46	0.23
$(80, 95]$	968,492	0.50	0.26
>95	846,179	0.71	0.33
First-Time Homebuyer			
No	1,991,677	0.50	0.24
Yes	1,004,130	0.64	0.31
Loan Purpose			
Purchase	2,294,776	0.55	0.27
Cashout	351,269	0.50	0.23
Rate Refi	350,104	0.57	0.25
Lender Type			
Independent Non-bank	1,897,661	0.55	0.27
Other/Unclassified	1,098,488	0.54	0.26

Data Source: Optimal Blue

Notes: This table summarizes the residualized locked mortgage rate from specifications (3) and (10) of Table 5.

Table 7: Which Lenders Are the Most Expensive?

	(1)	(2)	(3)	(4)	(5)
Dep. Var.:	Lender FE	Lender FE	Lender FE	Lender FE	Within-lender dispersion
Size Quartile 2	0.0533**	0.0488*		0.0253	0.0216***
	(0.0258)	(0.0256)		(0.0238)	(0.00672)
Size Quartile 3	0.0740***	0.0812***		0.0282	0.0487***
	(0.0256)	(0.0243)		(0.0234)	(0.00577)
Size Quartile 4	0.127***	0.131***		0.0465*	0.0777***
	(0.0244)	(0.0235)		(0.0257)	(0.00573)
Nonbank	0.108***	0.0622***		0.0579***	0.00395
	(0.0220)	(0.0215)		(0.0202)	(0.00499)
Lender type unknown	0.00916	0.0151		-0.0147	0.0273***
	(0.0215)	(0.0210)		(0.0208)	(0.00550)
FHA share		0.385***		0.267***	0.108***
		(0.0763)		(0.0767)	(0.0169)
Jumbo share		-0.154		-0.253**	0.0910
		(0.202)		(0.105)	(0.110)
Superconf. share		0.0405		0.0460	-0.00502
		(0.214)		(0.206)	(0.0449)
Within-lender dispersion			1.637***	1.091***	
			(0.177)	(0.203)	
Constant	-0.205***	-0.296***	-0.317***	-0.355***	0.0540***
	(0.0234)	(0.0266)	(0.0251)	(0.0267)	(0.00687)
Adj. R2	0.16	0.28	0.22	0.34	0.37
Obs.	452	452	452	452	452

Data Source: Optimal Blue

Notes: Table displays lender-level regression results using data for 452 lenders that use the Optimal Blue platform and have at least 100 mortgage locks. The outcome variable in columns (1)-(4) are the lender fixed effects from specification (4) in Table 5. Within-lender dispersion is measured as the standard deviation of residuals from specification (6) of Table 5. Robust standard errors in parentheses. * p<.1, ** p<0.05, *** p<0.01.

Table 8: Nonbank Income and Expenses (from MCR Data) and Their Relationship with Estimated Lender Expensiveness

	Mean	Median	St. Deviation	Regression on Lender Expensiveness β	St. Err.
Income					
Origination Income	1.33	0.80	1.53	0.78***	(0.19)
Interest Income	0.28	0.23	0.32	0.03	(0.04)
Secondary Market Income (Gain-on-Sale)	3.35	3.52	1.76	3.37***	(0.43)
Other Income	0.35	0.02	2.11	0.05**	(0.02)
Gross Income	5.03	4.75	1.97	4.05***	(0.41)
Expenses					
Loan Production Officers (Sales Employees)	1.37	1.36	0.58	1.06**	(0.47)
Loan Origination (Fulfillment/Non-Sales)	0.65	0.58	0.42	0.45***	(0.13)
Origination-Related Management and Directors	0.28	0.15	0.36	0.25***	(0.07)
Employee Benefits	0.27	0.25	0.19	0.33***	(0.05)
Other Personnel Expenses	0.44	0.31	0.54	0.24*	(0.13)
Interest Expenses	0.25	0.21	0.37	0.06	(0.04)
Occupancy and Equipment	0.25	0.23	0.15	0.30***	(0.03)
Technology	0.10	0.08	0.09	0.12***	(0.03)
Outsourcing, Professional, and Subservicing Fees	0.15	0.09	0.22	0.10***	(0.02)
Other non-interest expenses	0.65	0.51	0.64	0.26	(0.23)
Gross Expenses	4.33	4.16	1.76	3.50***	(0.41)
Corporate overheads					
Total Corporate Expenses	0.35	0.11	1.22	0.28***	(0.10)
Net Income					
Net Income (residential originations, pre-corporate allocations)	0.67	0.49	1.21	0.45**	(0.18)
Net Income (all lines, after corporate allocations and taxes)	0.31	0.32	1.87	0.24**	(0.12)

Data Source: Conference of State Bank Supervisors Mortgage Call Reports; Optimal Blue

Notes: This table is based on 1897 quarterly filings from 162 unique lenders between 2015:Q1 and 2019:Q4. All variables are shown as dollars per $100 originated. Some categories for income and expenses are combined or not shown since they only have zeros for all firms. The standard deviation reported in the third column is the cross-sectional standard deviation, computed after taking out year-quarter fixed effects. The reported coefficients in the last two columns are from separate median regressions of each line item on lender expensiveness (in percentage points) as measured from lender fixed effects (from specification (4) of Table 5), and year-quarter fixed effects. The standard errors determining statistical significance are clustered at the lender level. * p<.1, ** p<0.05, *** p<0.01.

Table 9: Are Higher Mortgage Rates Associated with Better Service Quality?

	β	SE	N	Outcome mean
Did you experience delays in...				
...Your closing date	.008	(.008)	22,567	0.17
...Paperwork processing	.009	(.009)	22,567	0.23
Were you very satisfied with...				
...Your interest rate	-.13***	(.01)	22,567	0.69
...Your lender	-.023***	(.009)	22,567	0.77
...The application process	-.019**	(.01)	22,567	0.67
...The documentation process	-.008	(.011)	16,188	0.62
...The loan closing process	-.019*	(.01)	22,567	0.68
...The timeliness of disclosures	-.007	(.01)	22,567	0.68
Overall satisfaction	-.035***	(.009)	16,188	0.69

Source: Authors' calculations based on National Survey of Mortgage Originations and the National Mortgage Database

Notes: Each row represents a separate regression with the dependent variables listed in column (1). The displayed coefficients are those on the borrower's mortgage rate, in percentage points. All outcomes are dummy variables, except "Overall satisfaction," which is the sum of the previous six satisfaction questions divided by six (thus, its range is 0 to 1). Sample restricted to first-lien 30-year FRM loans for single-family principal residence properties, with no more than two borrowers, originated from 2013 through 2019 (the documentation question was not asked in every wave and thus has fewer observations). All regressions control for origination month fixed effects, survey wave fixed effects, county fixed effects, credit score (linear term plus dummies for 11 credit score bins), LTV (linear term plus dummies for 6 LTV bins), indicators for loan purpose (purchase, refinance, or cashout refinance), 9 loan amount categories, loan program (Freddie, Fannie, FHA, VA, FSA/RHS, other), first-time homebuyer status, single borrowers, log borrower income, self-employment status of borrower(s), respondent gender, race, ethnicity, whether the household owns 4 different types of financial assets, whether the household could pay their bills for 3 months without borrowing, metropolitan CRA low-to-moderate income tract status, self-assessed creditworthiness, and the likelihood of moving, selling, or refinancing within a couple years. Observations weighted by NSMO sample weights. Robust standard errors in parentheses. * $p<.1$, ** $p<0.05$, *** $p<0.01$.

Table 10: Sophistication, Lender Concentration, and Mortgage Rates

	(1)	(2)	(3)
Considered 2 lenders	-0.022**		
	(0.009)		
Considered 3+ lenders	-0.044***		
	(0.012)		
Applied to 2+ lenders for better loan terms	-0.075***		
	(0.019)		
Used web/broker/friends to get info? A lot	-0.021***		
	(0.008)		
Familiar with mortgage rates? Very	-0.042***		
	(0.015)		
Most lenders offer same rate? No	-0.022**		
	(0.010)		
Knows their interest rate	-0.060***		
	(0.007)		
Answered whether rate is fixed or variable	-0.056***		
	(0.019)		
Sophistication Index		-0.226***	-0.219***
		(0.019)	(0.019)
Sophistication Index × County HHI (last year)			0.050**
			(0.020)
County HHI (last year)			-0.025*
			(0.014)
Adj. R2	0.53	0.53	0.53
Obs.	22567	22567	22563

Source: Authors' calculations based on National Survey of Mortgage Originations and the National Mortgage Database

Notes: Dependent variable is the borrower's mortgage interest rate in percentage points. Sample restricted to first-lien 30-year FRM loans for single-family principal residence properties, with no more than two borrowers, originated from 2013 through 2019 (waves 1-24; "most lenders offer same rate" was not asked in the first 6 waves of the survey so we include a separate dummy for these missing observations). All regressions control for origination month fixed effects, survey wave fixed effects, county fixed effects, credit score (linear term plus dummies for 11 credit score bins), LTV (linear term plus dummies for 6 LTV bins), indicators for loan purpose (purchase, refinance, or cashout refinance), 9 loan amount categories, loan program (Freddie, Fannie, FHA, VA, FSA/RHS, other), first-time homebuyer status, single borrowers, log borrower income, self-employment status of borrower(s), respondent gender, race and ethnicity, whether the household owns 4 different types of financial assets, whether the household could pay their bills for 3 months without borrowing, CRA low-to-moderate income tract status, self-assessed creditworthiness, applying to 2+ lenders for reasons other than "better loan terms," and the likelihood of moving, selling, or refinancing within a couple years. Observations weighted by NSMO sample weights. Sophistication index is a composite of shopping/knowledge questions and ranges from zero to one (see text for details). HHI is market concentration at the county-year level, as of the year prior to origination, standardized to have mean 0 and standard deviation 1. Robust standard errors in parentheses. * $p<.1$, ** $p<0.05$, *** $p<0.01$.

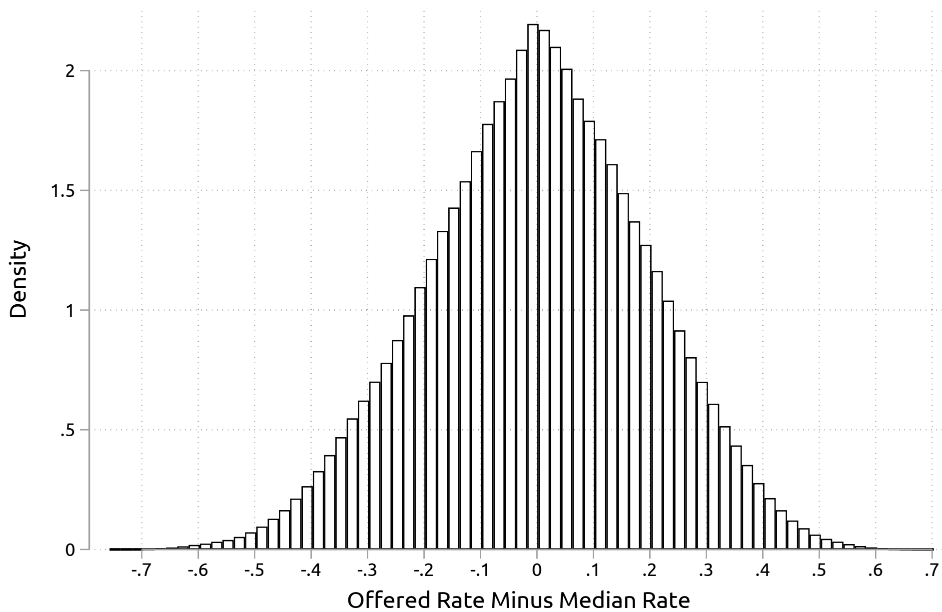
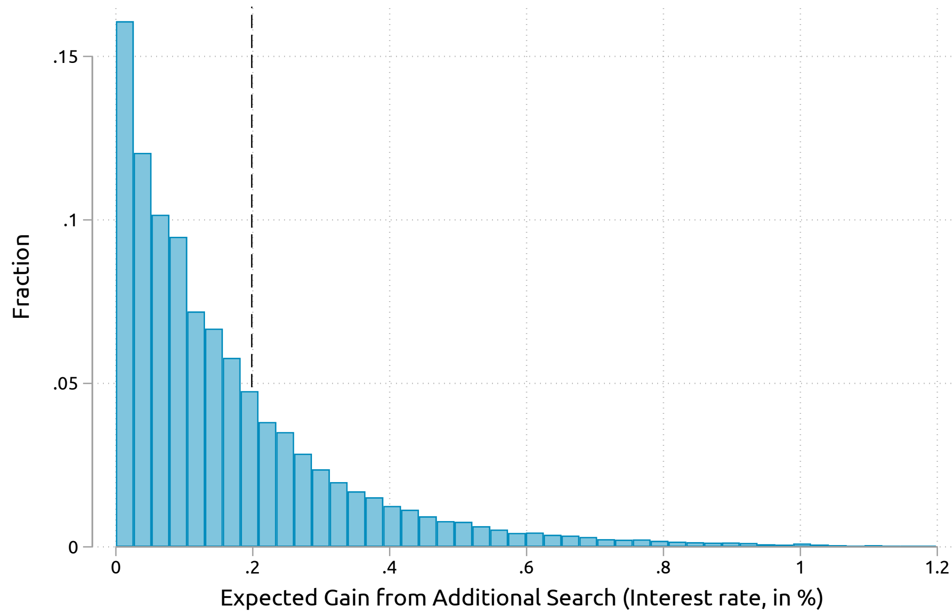


Figure 1: Offer Dispersion for Identical Mortgages

Data Source: Optimal Blue
Note: Figure shows the distribution of real-time offered interest rates, in percentage points, 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).

A. All Programs



B. Differences Across Programs

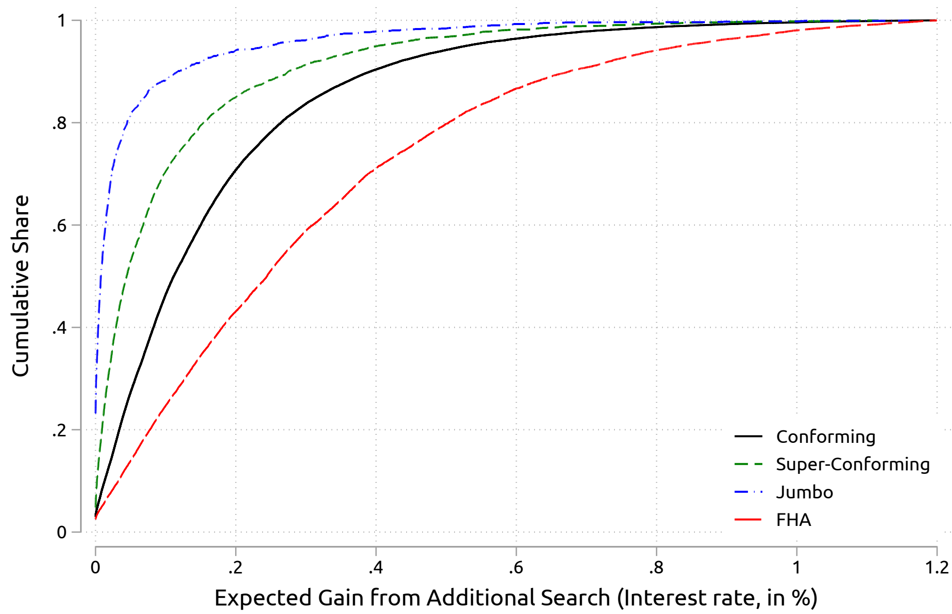


Figure 2: Distribution of the Expected Gain from Additional Search for Identical Mortgages

Data Source: Optimal Blue
Note: Panel A shows the distribution of the expected gain from additional search, defined in equation (1), for all borrowers in the sample. Being at (or close to) zero means that the rate locked by a borrower is near the lowest available rate, so that the borrower would have little to gain from additional search. Larger values mean that a borrower would have more to gain from additional search. The dashed line denotes the mean of the distribution. Panel B shows the cumulative share of loans below a certain expected gain across the four loan types in the sample.

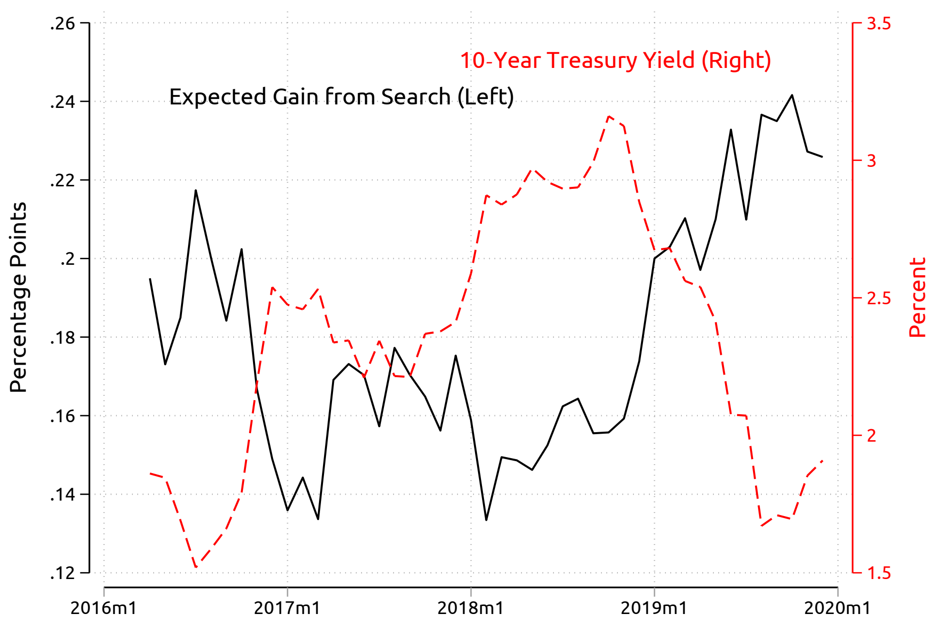


Figure 3: The Evolution of Average Expected Gain from Additional Search and Treasury Yields

Data Source: Optimal Blue
Note: The dashed red line is the 10-year Treasury yield. The solid black line is the monthly average expected gain from additional search (*EGain*) after controlling for borrower and loan characteristics (FICO, LTV, loan amount, and MSA fixed effects).

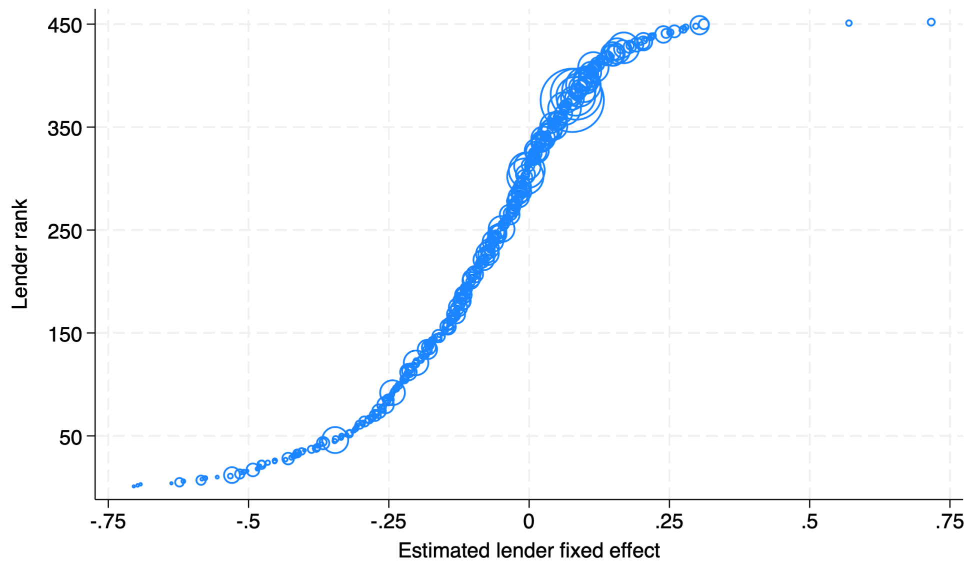


Figure 4: Distribution of Estimated Lender Fixed Effects from Interest Rate Regressions

Data Source: Optimal Blue
Note: Lender fixed effects, in percentage points, come from specification (4) in Table 5. Only lenders with at least 100 locked loans in Optimal Blue are included in the figure; size of circles is proportional to total number of locks.

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

A.1 Relationship to the Literature

Table A-1 contains a summary of other papers that have studied price dispersion in the US mortgage market.[1] These other papers look either only at offers or only at originations. Among those that look at originations, arguably the results in Agarwal et al. (2023) and Ambokar and Samaee (2019) are most closely related to ours.[2]

There are at least five major distinctions between our work and these papers. First, we have data from all segments of the mortgage market (conforming, FHA, jumbo)—this is important as it allows us to document differences in dispersion across segments. Second, our data come (only) from the post-financial-crisis period, when new regulations as well as the increased popularity of online lending (and online mortgage shopping) could have been expected to reduce or even eliminate price dispersion. We show that substantial dispersion remains.

Third, we are able to observe identifiers for branch and loan officer within lender. This allows us to unpack at which "level" price dispersion occurs (across vs. within lenders, and then further across branches and across loan officers within lender). We are thereby able to quantify remaining dispersion for observably identical borrowers who get loans not only from the same lender at the same time, but even in the same branch and from the same loan officer.

Fourth, we are able to observe discount points and the exact lock date, variables that are both highly pricing relevant. Without these variables, it is not possible to know how much of the estimated price dispersion is simply an artifact of missing controls. We can compare results for the conforming segment, which these other papers also study. We find a lower dispersion, which may reflect either the different time period or the additional controls we are able to include.[3] The standard deviation of rate residuals in our conforming mortgage subsample is 19bp, after controlling for lender fixed effects and 12bp after controlling for a full set of observables including (time-varying) branch and loan officer fixed effects.[4]

To illustrate to what extent our finer controls matter for the amount of residual price dispersion,

[1]There also exist comparable studies from non-US markets; cf. footnote 3 in the main text. Since institutional arrangements differ quite substantially across countries, we do not include the quantitative findings of these papers in our discussion in this section.

[2]Gurun et al. (2016) look at "reset rates" on privately securitized adjustable-rate mortgages prior to the financial crisis; Woodward and Hall (2012) study broker fees on a sample of FHA mortgages in 2001.

[3]Note that Agarwal et al. (2023) could control more finely for loan characteristics and timing in their data, but measuring price dispersion is not the main focus of their paper.

[4]Papers, including ours, usually also report R-squared values from the regressions, but these are not easily comparable across studies because they are to a large extent driven by the amount of time-series variation in overall rate levels in the sample period considered.

we consider a version of our analysis in Table 5 where we do not directly control for points and only control for an approximate closing month.[5] In this case, the standard deviation of the rate residual from a specification similar to (4) in Table 5 is 24bp, which is about 26% higher than the 19bp when we control for points and exact lock date. Similarly the 90th-10th percentile gap in the rate residual is 53bp when we do not include the additional controls versus the 42bp when we do.

Fifth, and finally, we are able to match the prices (rates) that borrowers obtain with the distribution of available rates at the same time for loans with identical characteristics. This allows us to cleanly measure expected (unrealized) gains from additional search, and how those vary with borrower and loan characteristics. Note that one could consider doing something similar based on transacted rates only, by taking the best accepted rate by a borrower with certain characteristics as a benchmark, and determine overpayment relative to that. However, this strategy would require a large number of borrowers with certain characteristics getting a loan in a given location on a given date. In our data, if we group loans by MSA, day, FICO bin, LTV bin, and loan program, there are on average 2.4 loans per bin, and 90% of the bins have 5 or fewer loans in them. Clearly, using the lowest rate obtained by one of these borrowers would not provide a robust benchmark. Moreover, without matching to the offer data, it would not be possible to detect if every borrower of a certain type is overpaying relative to what is available in the market.

[5]To be exact, we generate a closing date to be a random number between (lock date + lock period)/2 and (lock date + lock period), and then convert that date to a month.

Table A-1: Other Estimates of Price Dispersion in US Mortgage Market

Paper	Data	Main Findings	Comments
Agarwal et al. (2023)	Mortgages originated between 2001 and 2011 and insured by one of the GSEs. 85% of originations are from 2001-2009.	Residual standard deviation of 39bp (after controlling for origination quarter, state, and various loan and borrower characteristics). 90th-10th percentile gap in residual is 90bp.	Interest rates are adjusted for points and fees. Adding lender-by-quarter fixed effects appears to have little effect on residual dispersion (see their Figure A1.C).
Alexandrov and Koulayev (2017)	Rate sheet data (offers data) for 31 lenders from Informa. Calculate dispersion in offered rates for characteristics of 1.3M loans originated in 2014.	Focus is on the spread between highest and lowest offer across lenders. The average of this spread is 50bp, though with variation across loan types (10th percentile is around 33bp, 90th percentile around 69bp).	No data on transactions. Argue that wide dispersion in offers implies borrowers don't shop much, and that search costs, incorrect beliefs about dispersion, and non-price lender preferences matter (with some evidence from NSMO survey data).
Ambokar and Samaee (2019)	Fannie Mae and Freddie Mac public loan-level data on originations from 1999 to 2016.	Residual standard deviation of 27bp (after controlling for origination month, FICO bins, LTV bins, 3-digit zip code and other loan characteristics available in these data, as well as originator fixed effects).	Data do not contain information lock date or fees/points. Only relatively large originators are individually identified.
Gurun et al. (2016)	Privately securitized adjustable-rate mortgages through 2006 (data source: LoanPerformance)	Find 95th-5th percentile gap in residual *reset* rate of 310bp; also construct "lender expensiveness" measure from residual; 95th-5th gap there averages 280bp.	Reset rate typically applies 2+ years after origination, if borrower does not refinance sooner. Main focus is on relationship between lender expensiveness and advertising.
McManus et al. (2018)	PMMS survey data (offer rates at lender level for a first-lien, prime, conventional, conforming home purchase mortgages with a loan-to-value of 80%) from 1995 through 2017.	For most periods, the standard deviation of offer rates ranges from 15 to 25 bp (mostly below 20). During the global financial crisis, the rate dispersion peaked at about 45 bp. On a specific date in 2018, show that range of offers is about 90bp.	No data on transactions. Calculate potential gains from shopping simply based on calculating the minimum rate obtained from X random searches (for $X = 1$ to 5).
Woodward and Hall (2012)	Sample of 1,500 FHA loans from six weeks in 2001; brokered mortgages only.	90th-10th percentile gap in fees paid to brokers (upfront fee + yield spread premium) is about 300bp.	300bp in upfront costs would correspond to about 75bp in rates just from broker costs; there might be additional rate variation across borrowers.

A.2 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-2 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-2 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 (2021). 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.[6] 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-2 also indicates that the distribution of FICO scores and LTVs in Optimal Blue almost mirrors the HUD data, whereas the McDash data are skewed

[6]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.

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.

A.3 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 3.1.2 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 1% 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 part of the price of the mortgage, they are part of 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 1% to originate a mortgage at 4% interest is equivalent to a lender that charges no fees but requires the borrower to pay 1 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 1 point, while a price of 99 means the borrower would have to pay 1 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 detail in Section A.5, we allow for this conversion factor vary by loan program $\times$ lock month. The estimated conversion factor averages about 22bp 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 4.14%.

A.3.1 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.

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 differs across these mortgage types, 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-3 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 Mae or Freddie Mac 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-4 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.3.2 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 in the screenshot in Figure A-5 appears representative of the universe of offer distributions.

Table A-5 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 $6,000 difference in upfront costs between these lenders. Even going from the 75^{th} percentile lender to the 25^{th} percentile lender would save about $3,000 for a typical borrower with a $250k loan.

A.4 Matching Offers and Locks

As described in Section 3.1.2, 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%. 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. Similarly, restricting more conservatively to matches where FICO and LTV are within 1 point between datasets does not affect the results shown in Table 2 but considerably reduces the sample size from 67,637 to 4,156.

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.5 Estimating the Discount Point to Interest Rate Tradeoff

In some parts of the paper, we need to know how much 1 discount point (or credit) buys down (or up) the interest rate borrowers pay. For example, we use this tradeoff to match each rate locked to the zero point offer rate, or for back-of-the envelope calculations of how much money borrowers leave on the table. We estimate the relationship between discount points and interest rates in a regression specification identical to column (10) of Table 5, with the only exception that discount points are allowed to only enter linearly and are only allowed to vary by loan program $\times$ lock month. Appendix Figure A-8 shows in a binned scatter plot that the relation between points and rate is indeed close to linear. The slope of the line in the chart is the average point-rate tradeoff, which implies that 1 discount point changes the interest rate by about 21.8bp on average.

In Table A-6, we show how the point-rate tradeoff varies with FICO, DTI, and LTV by adding several interactions of discount points to the same regression we run in the paper. As a baseline, the average coefficient on discount points over all of the data in the paper is 0.218, so 1 discount point paid buys down the interest rate by 21.8bp. Relative to this magnitude, the coefficients shown in Table A-6 are small and economically insignificant, all of them being 0.1 to 0.8bp. It is safe to say that the point-rate tradeoff does not vary in a meaningful way with these loan characteristics.

Another possibility is that the point-rate tradeoff may vary across lenders. In Table 5, we do interact binned discount points with lender fixed effects. However, in the parts of the paper where we consider whether borrowers are overpaying, we think it would be overcontrolling for us to soak up the variation across lenders in the point-rate tradeoff. The fact that some lenders might offer worse tradeoffs than others reflects the fact that those lenders are just more expensive to do business with. In other words, we should not control for lender specific point-rate tradeoffs since that is part of how expensive a lender is relative to others, which is what we are trying to capture when comparing locked and offered rates.

A.6 Variation in Pricing Across Branches Within Lender

This section complements Section 6.1 in the main text, where we study covariates of the estimated lender fixed effects from the regressions in Table 5. Here, we instead focus on variation in average fixed effects across branches within a given lender. The goal is to see whether the "expensiveness" of branches within a lender covaries with characteristics of the branch and its lending, or with local characteristics based on where the branch originates loans. This analysis is thus related to work that studies the pricing of deposits across branches of the same bank (e.g., Drechsler et al., 2017; d'Avernas et al., 2023).

We again restrict to lenders with at least 100 loans locked in the sample, as in Section 6.1. Those 452 lenders have a total of 15,881 branches.[7] In our baseline specification below, we restrict the sample to branches that made at least 50 loans. We are interested in correlates of variation in branch fixed effects (i.e., effectively the average rate residual per branch) within lenders.[8]

As in Table 7 in the main text, we start by considering the role of branch size and the share of different loan types originated by a branch. Column (1) of Table A-7 indicates that while size does not appear to correlate with the branch's expensiveness (as measured by its fixed effect), branches that originate a larger share of FHA loans are more expensive. The standard deviation (SD) of FHA shares across branches in the sample of column (1) is 0.2 (or 0.15 within lender) such

[7]We do not know whether all of these are truly "branches" with a separate physical location; in some cases, what is recorded as a "branch" in the data could also be small teams of loan officers. Conversely, a few lenders, including larger ones, have only one or two branches recorded in the data, suggesting that perhaps they do not use different branch IDs in Optimal Blue (instead simply relying on LO IDs). Below, we try different sample restrictions to ensure that our results are not sensitive to removing what appear to be "unusual" branches.

[8]The estimated fixed effects we take for this exercise come from a specification like in column (7) of Table 5 of the main text, except that we replace the lock date × MSA fixed effects and the zip code fixed effects by lock date × state fixed effects. This is because we want to use the local variation associated with differences across counties in the regressions in this appendix section.

that the strength of the economic association is meaningful, with a one-SD move in FHA share corresponding to a 5.5bp larger branch fixed effect (relative to a SD of fixed effects across branches of 14bp).

In the second specification, we relate the expensiveness of a branch with the characteristics of the counties where it lends (weighted by how many loans a given branch made in each county).[9] We observe that branches that lend in counties with lower mortgage market concentration (HHI), lower average income, lower shares of college-educated households, and higher minority shares tend to be more expensive. In terms of the economic magnitude of the coefficients in column (2), the implied effect per one-SD move in a variable is largest for the share of college-educated households (where a one-SD increase is associated with a 3.3bp decrease in the branch fixed effect), while the smallest effect is for the HHI (where a one-SD increase is associated with a 0.5bp decrease in the branch fixed effect). Thus, even though the direction of the estimated effect for local market concentration is surprising (as one might expect a positive relationship), the economic magnitude of the association is very small. This remains the case across the other columns of the table, where we control for FHA share and the demographic variables jointly. Not surprisingly, given that FHA share and socioeconomic characteristics are correlated, the associated coefficients are attenuated somewhat but remain significant. To study the robustness of these conclusions, in column (4), we remove the minimum-loan-count restriction, while in column (5) we retain only lenders where the largest branch makes no more than 20% of that lender's total loans (so we remove dominant branches). In either case, results remain qualitatively similar.

In sum, this evidence suggests that the pricing variation across branches within lender is meaningfully related to the branch's typical borrower clientele. Branches that make predominantly FHA mortgages or are active in locations with a less educated population (as well as lower incomes and higher minority share) tend to be more expensive. The latter finding is related to the finding of Drechsler et al. (2017) that bank branches in counties with a lower share of college-educated adults tend to increase deposit spreads by more (i.e., pay lower deposit rates to their customers) in response to a Fed funds rate increase. As a caveat to our analysis, we are not able to disentangle to what extent these within-lender differences are driven by variation across branches in their "ex-ante" (strategic) price setting, versus variation arising during the negotiation process (e.g., with borrowers in less economically well-off locations being less willing or able to negotiate for a better deal).

A.7 Additional Details on NSMO Analysis

In Section 6.4 of the main text, we consider the following mortgage shopping and mortgage knowledge variables in our analysis of how shopping and knowledge relates to the mortgage rates that

[9] We have also considered specifications where instead we defined for each branch the county where it made the most loans as its "home county," and used that county's characteristics. Results are similar.

borrowers get:

1. The answer to the question, "How many different lenders/mortgage brokers did you seriously consider before choosing where to apply for this mortgage?" 48% of respondents (weighted) answer 1; 36% 2; 13% 3; 2% 4; and 1%, 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, nearly 75% answer 1; 20% 2; 4% 3; 0.9% 4; and 0.4%, 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 (about 82% of those that apply to more than one lender, or 21% of the sample overall).

4. A series of questions are asked about possible sources of information on mortgages and mortgage lenders that the borrower could have used. For each source, a respondent can say they used it "a lot," "a little," or "not at all." As a proxy for search effort, we create a dummy variable that equals one if the respondent says they used one of "Other lenders or brokers," "Websites that provide information on getting a mortgage," or "Friends/relatives/co-workers"; about 38% of respondents said they used at least one of these information sources "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?" About 55% respond "Very," whereas 38% say "Somewhat," and 8% say "Not at all".

6. 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; of those who were asked, almost 70% agree with the statement.

7. We create an indicator variable for whether a borrower knows their interest rate by comparing the self-reported contract rate to the contract rate available from the linked loan-level administrative data. We set this variable equal to one if the borrower's self-reported rate is within 5bp of the actual rate. Just over 49% know their rate by this standard.

8. An indicator for whether the borrower said, "Don't know" in response to the question "Is this an adjustable-rate mortgage?" (About 5% answered "Don't know.")

In Table A-8, we regress mortgage rate against each of these measures one at a time, and then in a final specification jointly (the joint specification is also shown in the main text in Table 10). In all specifications, we control finely for other factors (e.g., credit risk, loan type) that likely influence loan pricing. We see that most proxies for intense shopping and better mortgage market knowledge are associated with lower mortgage rates. In the final column, where we include all X_i jointly, some of the coefficients are slightly attenuated relative to the earlier columns, but all remain individually significant, suggesting that there are different dimensions to shopping and knowledge that can contribute to a borrower obtaining a low rate.[10]

A.7.1 Who Pays More Because of a Lack of Shopping or Knowledge?

The previous subsection provides evidence that more intense mortgage shopping and more mortgage knowledge is associated with lower contracted rates. We next ask which observable borrower and loan characteristics are associated with stronger shopping intensity and mortgage knowledge.

In Table A-9, we test whether a composite "sophistication index" is correlated with loan type (FHA, GSE, jumbo, etc.), credit score, and first-time homebuyer status. The sophistication index is the same measure we use in Section 6.4, in which we construct as the sum of six of the shopping and knowledge dummy variables shown in Table A-8, divided by six so that the index ranges from 0 to 1.[11] Column (1) of Table A-9 tests for a correlation between sophistication and loan type, and indicates that FHA borrowers tend to be the least sophisticated (the omitted category is jumbo and other non-conforming conventional loans). In column (2), we find a monotonic relationship between credit score and sophistication. In column (3), we do not find a correlation between first-time homebuyer status and sophistication. In column (4), we include loan type, credit score, and first-time status simultaneously, and continue to find that FHA and low-score borrowers are the least sophisticated. Finally, in column (5), we control for loan amount bins to account for the possibility that borrowers getting smaller loans, which may correlated with FHA status and lower credit scores, may be have less incentive to shop around. Of course, loan amount also likely reflects income, education, etc., and therefore might be a proxy for sophistication. We find that the coefficients on loan type and credit score are only somewhat smaller and remain statistically significant.

Overall, we believe the findings here lend 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 and dispersion in observed locked rates is likely due to ineffective shopping and negotiation, reflecting a lack of financial sophistication and knowledge of the market.

[10]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. (2023).

[11]We include the dummy "considered 3+ lenders" but not the dummy for considering exactly 2 lenders; we also exclude the dummy for whether "most lenders offer the same rate" since this question was not asked in early waves. We include "applied to 2+ lenders for better loan terms" but not "applied to 2+ lenders" since the latter is associated with higher rates conditional on the former. The median borrower has a sum of 3 (i.e., an index value of 0.5); about 14% of borrowers have a sum of 1 or 0, about 8% of borrowers have a sum of 5 or 6, and the standard deviation of the index is 0.2.

A.7.2 Time-series Variation in Shopping Intensity

In Section 4.2 in the main text, we documented that our measures of borrower overpayment in the Optimal Blue data, namely the expected gain from additional search ($EGain$) and the locked-offer rate gap, decrease when market interest rates are higher, even for borrowers who do not appear constrained. We speculated that this may be driven in part by an increase in shopping intensity when interest rates are higher. The NSMO enables us to test this hypothesis directly. We estimate linear probability models of the form:

$$Shopping_{ijctw} = \beta \cdot PMMS_t + \Gamma Z_{ij} + \Theta W_{ct} + \delta_w + \epsilon_{ijctw}, \tag{A1}$$

where $Shopping_{ijctw}$ is a binary measure of shopping intensity (discussed below) by borrower i with loan characteristics j, located in county c, loan origination month t and responding to survey wave w. $PMMS_t$ is our main variable of interest, the market mortgage rate on average during the month of loan origination. Z_{ij} are borrower and mortgage controls, similar to the controls in equation (3). W_{ct} includes county fixed effects, and controls for county house price growth in the 12 and 24 months prior to the survey to account for "hotness" of the local housing market, which might influence mortgage shopping behavior. Finally, δ_w are survey wave fixed effects.

As dependent variables, we use binary versions of the shopping variables that were associated with lower contract interest rates in Table A-8: (i) whether a borrower seriously considered at least two lenders; (ii) whether a borrower applied to at least two lenders in search of better terms; (iii) whether a borrower used other lenders/brokers to get information "a little" or "a lot"; and (iv) whether a borrower used websites that provide information on getting a mortgage "a little" or "a lot." For each of these variables, we report regressions with and without other covariates (except for survey wave fixed effects).

Table A-10 reports the results of these regressions. Across the different measures in the full sample (panel A), a higher level of market mortgage rates is associated with more shopping effort, with statistical significance for three of four outcomes. The coefficients are largely unaffected by the addition of controls, which alleviates concerns that the relationship is driven by variation in the type of borrower who applies at different points in time (and at different levels of market rates).

Columns (1)-(2) imply that a 1 percentage point increase in market mortgage rates increases the probability that a borrower considered more than one lender by 4-5 percentage points, relative to a sample average of 52%.[12] Columns (3)-(4) indicate that the likelihood of applying to 2+ lenders for better terms rises by about 5-6 percentage points when rates rise by 100bp, relative to the sample average of 21%. Finally, the coefficients in the last four columns on the likelihood of using other lenders and the web to gain information are positive but not quite as large. In panel B, we restrict the sample to borrowers whose DTI ratio ends up below 36 percent, suggesting that

[12] Over our sample period (2013-2019), the market mortgage rate as measured by PMMS varied from about 3.3% to nearly 5%.

they had additional room to make larger payments. The coefficients are somewhat larger than in panel A, and thus it does not appear that the positive relationship between market interest rates and shopping is mainly driven by affordability constraints.

A.8 Is Getting a "Good Deal" Related to Market Concentration?

In Section 6.4, we found evidence using NSMO data that higher lender concentration in a county is associated with a relative increase in the rate that highly "sophisticated" borrowers get relative to less sophisticated borrowers. In other words, market concentration seems to matter more for borrowers who are more likely to shop around and have strong knowledge about the mortgage market, whereas for less sophisticated borrowers, lower concentration may be less effective in increasing borrower surplus.

In this section, we study the relationship between concentration and Optimal Blue mortgage lock rate residuals, and in particular whether the propensity of borrowers to get particularly "good" or "bad" deals varies with local (county-level) market concentration, as measured by the Herfindahl-Hirschman Index (HHI). We regress either a loan's rate residual, its absolute value, or indicators for good and bad deals (defined as rate residuals of less than -20bp or more than +20bp, respectively), on the county-level HHI from the previous calendar year, various other local controls (the log number of loans originated in the county in the previous year, as well as median household income, the share of college-educated individuals, and the minority share, all at the zip-code level), as well as loan type-by-month fixed effects. In some specifications, we further add lender-by-month fixed effects; this allows us to study effects of market concentration on rates within a lender at a point in time (relying on lenders that are active in multiple counties).

For the absolute rate residuals, as well as the good-deal and bad-deal dummies, we rely on the residuals from column (3) of Table 5 (i.e., the regression specification that finely controls for loan characteristics but not lender fixed effects). For the rate residual itself, we use residuals from a slightly different regression specification, since the one in column (3) of Table 5 uses zip-code fixed effects that would absorb most county-level variation in average residuals. Therefore, we generate residuals based on a specification with only state-by-month fixed effects, so we are effectively studying within-state variation in residuals as a function of county-level concentration. When interpreting the results below, keep in mind that the HHI in the mortgage market is generally low: Across the loans in our sample, the median lagged county-level HHI is 0.020, with a 90th percentile at 0.037 and a 10th percentile at 0.006.[13]

Table A-11 shows that local concentration is not significantly related to average residuals

[13]These HHI numbers may appear low, but that is partly because counties are weighted by the number of loans in our sample, and large counties tend to be less concentrated. If we keep only one observation per county-year in our sample, the median HHI increases to 0.035, with 90th and 10th percentiles at 0.076 and 0.016.

(columns 1-2).[14] However, columns (3) and (4) indicate a strong relationship between local concentration and the absolute value of the residual: In more concentrated markets (higher HHI), there is less dispersion in rate residuals (absolute values are lower).[15] The remaining columns show that what is particularly strongly affected is the likelihood of getting good deals: In more concentrated markets, borrowers are significantly less likely to obtain a rate residual of -20 bp or lower. For instance, the coefficient in column (6) implies that moving from the 90th to the 10th percentile of HHI changes the probability of a good deal by about 1.7 percentage points, or 10% of the mean. Columns (7) and (8) show that higher concentration also reduces the probability of bad deals, although this effect is only significant without lender fixed effects.

A potential rationalization of these findings is that when a market is less concentrated (which could be seen as more competitive), lenders may try harder to earn high rents from unsophisticated borrowers, because they know that they need to compete harder for sophisticated borrowers (e.g., it becomes more likely that a borrower obtains an outside offer). Therefore, it appears that sophisticated borrowers may indeed benefit from increased competition, while less sophisticated borrowers do not benefit or may even be better off in more concentrated markets.

[14] This finding is in line with Hurst et al. (2016) and Scharfstein and Sunderam (2016), who also do not find a relationship between local concentration and mortgage rates. Buchak and Jørring (2021) similarly find no relationship between concentration and rates, but do find that lenders charge higher fees (as reported in HMDA since 2018) in more concentrated markets.

[15] This finding is in line with Allen et al. (2014a), who find that after concentration-increasing mergers in the Canadian mortgage market, there is a decrease in price dispersion.

Table A-2: 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-3: 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-4: Interest Rate Dispersion for Offered Conforming Mortgages with No Points and Fees, Across MSAs

	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-5: Dispersion in Points and Fees Charged to Originate at the Median Interest Rate, from Lender Offers

	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-6: Variation of the Point-Rate Tradeoff with FICO, LTV, and DTI

	(1)	(2)	(3)	(4)
$DiscountPoints \times I_{FICO \geq 720}$	0.001			0.001
	(0.000)			(0.000)
$DiscountPoints \times I_{DTI>36}$		-0.002***		-0.002***
		(0.000)		(0.000)
$DiscountPoints \times I_{LTV>80}$			0.008***	0.008***
			(0.000)	(0.000)
Average Coefficient on Discount Points	0.218	0.218	0.218	0.218
Observations	3001321	3001321	3001321	3001321

Data Source: Optimal Blue

Notes: We estimate the relationship between discount points and interest rates in a regression specification identical to column (10) of Table 5, with the only exception that discount points are allowed to only enter linearly. Discount points are allowed to vary by loan program x lock month, and the average across all data is 0.218. Significance: * p<0.1, ** p<0.05, *** p<0.01.

Table A-7: Pricing Variation Across Branches Within Lender

	(1)	(2)	(3)	(4)	(5)
Size Quartile 2	0.00319		0.00264	0.00791	0.00739
	(0.00608)		(0.00589)	(0.00517)	(0.00627)
Size Quartile 3	0.00137		0.000757	0.00657	0.0116**
	(0.00504)		(0.00487)	(0.00445)	(0.00518)
Size Quartile 4	-0.000998		-0.000242	0.00896	0.0120*
	(0.00584)		(0.00578)	(0.00577)	(0.00703)
FHA share	0.278***		0.201***	0.115***	0.121***
	(0.0304)		(0.0346)	(0.0167)	(0.0205)
Superconf. share	-0.0273		-0.0176	-0.0639***	-0.0377*
	(0.0442)		(0.0393)	(0.0201)	(0.0202)
Jumbo share	-0.0194		0.143	0.00720	0.0108
	(0.112)		(0.106)	(0.0396)	(0.0441)
Avg. county HHI		-0.799*	-1.371***	-1.191***	-1.425***
		(0.466)	(0.470)	(0.436)	(0.512)
Avg. income (1000s)		-0.000890**	-0.000711*	-0.000381	-0.000626*
		(0.000384)	(0.000368)	(0.000314)	(0.000366)
Avg. BA+ share		-0.426***	-0.244***	-0.198***	-0.163**
		(0.0546)	(0.0592)	(0.0545)	(0.0667)
Avg. minority share		0.111***	0.0633***	0.0554***	0.0593***
		(0.0221)	(0.0228)	(0.0183)	(0.0209)
Lender FE?	Yes	Yes	Yes	Yes	Yes
Mean Dep. Var.	0.00	0.00	0.00	0.00	0.00
SD Dep. Var.	0.14	0.14	0.14	0.19	0.19
Adj. R2	0.11	0.10	0.13	0.10	0.10
Obs.	6815	6815	6815	15881	12059
Sample restriction?	Branch loans > 50	Branch loans > 50	Branch loans > 50	-	Share largest branch < 20%

Data Sources: Optimal Blue, HMDA, ACS

Notes: Standard errors in parentheses are clustered by lender. * p<.1, ** p<0.05, *** p<0.01.

Table A-8: Relationship Between Mortgage Rates and Measures of Shopping and Knowledge

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Considered 2 lenders	-0.028***								-0.022**
	(0.007)								(0.009)
Considered 3+ lenders	-0.062***								-0.044***
	(0.011)								(0.012)
Applied to 2+ lenders		-0.022**							0.065***
		(0.009)							(0.017)
Applied to 2+ lenders for better loan terms			-0.040***						-0.075***
			(0.010)						(0.019)
Used web/broker/friends to get info? A lot				-0.034***					-0.021***
				(0.007)					(0.008)
Familiar with mortgage rates? Very					-0.058***				-0.042***
					(0.015)				(0.015)
Most lenders offer same rate? No						-0.028***			-0.022**
						(0.010)			(0.010)
Knows their interest rate							-0.066***		-0.060***
							(0.007)		(0.007)
Answered whether rate is fixed or variable								-0.079***	-0.056***
								(0.019)	(0.019)
Adj. R2	0.52	0.52	0.52	0.52	0.52	0.52	0.53	0.52	0.53
Obs.	22567	22567	22567	22567	22567	22567	22567	22567	22567

Source: Authors' calculations based on National Survey of Mortgage Originations and the National Mortgage Database

Notes: Dependent variable is the borrower's mortgage interest rate in percentage points. Sample restricted to first-lien 30-year, fixed-rate loans for single-family principal residence properties, with no more than two borrowers, originated from 2013 through 2019 (waves 1-24; "most lenders offer same rate" was not asked in the first 6 waves of the survey so we include a separate dummy for these missing observations). All regressions control for origination month fixed effects, survey wave fixed effects, county fixed effects, credit score (linear term plus dummies for 11 credit score bins), LTV (linear term plus dummies for 6 LTV bins), indicators for loan purpose (purchase, refinance, or cashout refinance), 9 loan amount categories, loan program (Freddie, Fannie, FHA, VA, FSA/RHS, other), first-time homebuyer status, single borrowers, log borrower income, self-employment status of borrower(s), respondent gender, race and ethnicity, whether the household owns 4 different types of financial assets, whether the household could pay their bills for 3 months without borrowing, CRA low-to-moderate income tract status, self-assessed creditworthiness, and the likelihood of moving, selling, or refinancing within a couple years. Observations weighted by NSMO sample weights. Robust standard errors in parentheses. * p<.1, ** p<0.05, *** p<0.01.

Table A-9: Is Mortgage Knowledge and Shopping Correlated with Loan Type and Credit Score?

	(1)	(2)	(3)	(4)	(5)
Loan Program					
GSE	-0.038***			-0.036***	-0.017***
	(0.006)			(0.006)	(0.006)
FHA	-0.068***			-0.053***	-0.034***
	(0.007)			(0.007)	(0.007)
VA/RHS/FHLB	-0.044***			-0.034***	-0.024***
	(0.007)			(0.007)	(0.007)
Credit Score					
720-779		-0.013***		-0.011***	-0.011***
		(0.004)		(0.004)	(0.004)
620-719		-0.039***		-0.033***	-0.029***
		(0.004)		(0.004)	(0.004)
< 620		-0.055***		-0.045***	-0.038***
		(0.007)		(0.008)	(0.008)
First-time homebuyer					
Yes			0.000	0.006	0.012***
			(0.004)	(0.004)	(0.004)
Loan amount bins					Yes
Adj R-sq	0.05	0.05	0.04	0.05	0.06
N	22567	22567	22567	22567	22567

Data Source: Authors' calculations from the National Survey of Mortgage Originations and the National Mortgage Database

Notes: Dependent variable is the "sophistication index," which is a composite of six mortgage knowledge and shopping questions and ranges from zero to one (see text for more details), and has a mean of 0.46 and standard deviation of 0.2. On the right-hand side, the omitted loan type category is nonconforming (e.g., jumbo) loans; the omitted credit score category is 780-850. Sample is the same as described in Appendix Table A-8. All regressions control for origination month fixed effects, survey wave fixed effects, and county fixed effects. Robust standard errors in parentheses. * p<.1, ** p<0.05, *** p<0.01.

Table A-10: Do Borrowers Shop More When Mortgage Market Interest Rates Are Higher?

	Considered 2+ lenders		Applied to 2+ lenders for better terms		Used other lenders to get info		Used web to get info	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. All borrowers								
PMMS rate	0.050**	0.040*	0.058***	0.045***	0.033	0.028	0.036*	0.045**
	(0.020)	(0.021)	(0.016)	(0.017)	(0.020)	(0.021)	(0.020)	(0.021)
Controls?		Yes		Yes		Yes		Yes
Mean of Dep. Var.	0.52	0.52	0.21	0.21	0.42	0.42	0.54	0.54
Obs.	22567	22567	22567	22567	22567	22567	22567	22567
B. DTI < 36								
PMMS rate	0.074**	0.073**	0.077***	0.088***	0.061**	0.070**	0.032	0.053*
	(0.029)	(0.031)	(0.023)	(0.025)	(0.029)	(0.032)	(0.029)	(0.031)
Controls?		Yes		Yes		Yes		Yes
Mean of Dep Var	0.52	0.52	0.20	0.20	0.42	0.42	0.56	0.56
Obs.	10561	10561	10561	10561	10561	10561	10561	10561

Source: Authors' calculations based on National Survey of Mortgage Originations and the National Mortgage Database

Notes: Dependent variables are listed above each column; all are dummy variables. Sample restricted to first-lien, 30-year, fixed-rate mortgage for single-family principal residence properties, with no more than two borrowers, originated from 2013 through 2019 (the documentation question was not asked in every wave and thus has fewer observations). Controls include survey wave fixed effects, county fixed effects, house price growth in the 12 and 24 months before the survey, credit score (linear term plus dummies for 11 credit score bins), LTV (linear term plus dummies for 6 LTV bins), indicators for loan purpose (purchase, refinance, or cashout refinance), 9 loan amount categories, loan program (Freddie, Fannie, FHA, VA, FSA/RHS, other), first-time homebuyer status, single borrowers, log borrower income, self-employment status of borrower(s), respondent gender, race and ethnicity, whether the household owns 4 different types of financial assets, whether the household could pay their bills for 3 months without borrowing, metropolitan CRA low-to-moderate income tract status, self-assessed creditworthiness, and the likelihood of moving, selling, or refinancing within a couple years. Observations weighted by NSMO sample weights. Robust standard errors in parentheses. * p<.1, ** p<0.05, *** p<0.01

Table A-11: Market Concentration and Price Dispersion

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Rate residual		$\lvert$Rate residual$\rvert$		I(residual$< -0.2\%$)		I(residual$> 0.2\%$)	
County HHI (year-1)	-0.280	0.167	-0.307***	-0.380***	-0.419**	-0.794***	-0.363*	-0.147
	(0.246)	(0.218)	(0.104)	(0.0827)	(0.209)	(0.207)	(0.188)	(0.166)
Loan type × month FE?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Local controls?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lender × month FE?	No	Yes	No	Yes	No	Yes	No	Yes
Mean Dep. Var.	-0.00	-0.00	0.17	0.17	0.17	0.17	0.15	0.15
Adj. R2	0.00	0.16	0.08	0.13	0.03	0.13	0.02	0.08
Obs.	2995231	2995229	2995231	2995229	2995231	2995229	2995231	2995229

Standard errors in parentheses two-way clustered by lender and county.

* p<.1, ** p<0.05, *** p<0.01

Notes: Loan type × month FE are indicators for a loan being FHA, jumbo, or superconforming, interacted with the lock month. Local controls are the county's one-year lagged log number of loans originated (from HMDA), as well as the loan zip code's median household income, population share with at least a BA degree, and minority share. Rate residuals and HHI are winsorized at the 0.5th and 99.5th percentile.

Table A-12: Summary Statistics of Expected Gains from Additional Search by Zip Code Demographics

	Observations	Mean	St. Deviation	Percentiles	
				25^{th}	75^{th}
All Mortgages	67,637	0.18	0.21	0.04	0.25
Median Household Income					
First Tercile	22,585	0.21	0.23	0.06	0.29
Second Tercile	22,491	0.18	0.20	0.04	0.24
Third Tercile	22,536	0.16	0.20	0.03	0.21
Percent College Educated					
First Tercile	22,569	0.21	0.23	0.06	0.29
Second Tercile	22,539	0.19	0.21	0.04	0.25
Third Tercile	22,518	0.15	0.18	0.03	0.19
Minority Share					
First Tercile	22,560	0.16	0.19	0.03	0.21
Second Tercile	22,587	0.17	0.20	0.04	0.23
Third Tercile	22,479	0.22	0.23	0.05	0.30
County HHI					
First Tercile	23,011	0.18	0.21	0.04	0.25
Second Tercile	22,117	0.18	0.21	0.04	0.24
Third Tercile	22,509	0.19	0.21	0.04	0.26

Data Source: Optimal Blue, American Community Survey (ACS), HMDA

Notes: This table summarizes the expected gain from an additional search, given by equation (1) in the main text. The median household income, percent college educated, and minority share (share of Hispanic/Latino plus non-Hispanic Black) are based on ACS data and observed at the zip code level. The mortgage market HHI (Herfindahl-Hirschman Index) is computed at the county level, averaging over 2016-2019 using the HMDA data.

Table A-13: Regressions of the Locked-Offer Rate Gap on Observables

	(1)	(2)	(3)	(4)	(5)	(6)
FICO (omitted cat.: [640,660))						
$I_{680 \leq FICO < 700}$	-0.060***	-0.046***	-0.039***			
	(0.008)	(0.006)	(0.009)			
$I_{720 \leq FICO < 740}$	-0.092***	-0.059***	-0.056***			
	(0.010)	(0.008)	(0.013)			
$I_{FICO \geq 740}$	-0.128***	-0.081***	-0.064***			
	(0.011)	(0.009)	(0.013)			
LTV (omitted cat.: (60,80])						
$I_{85 < LTV \leq 90}$				0.018***	0.009*	0.018***
				(0.005)	(0.005)	(0.006)
$I_{90 < LTV \leq 95}$				0.050***	0.034***	0.029***
				(0.005)	(0.005)	(0.007)
$I_{LTV > 95}$				0.186***	0.143***	0.102***
				(0.012)	(0.011)	(0.019)
Loan Officer Comp (%)			0.172***			0.154***
			(0.037)			(0.039)
Loan amount f.e. ($10k bins)	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	0.111	0.481	0.446	0.151	0.508	0.461
Observations	67637	65757	15444	67637	65757	15444

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 cover mortgage rates for 20 metropolitan areas during the period 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-14: Regressions of the Expected Gain from Search on Observables, for Independent Nonbank Originators Only

	(1)	(2)	(3)	(4)	(5)	(6)
FICO (omitted cat.: [640,660))						
$I_{680 \leq FICO < 700}$	-0.037***	-0.028***	-0.025**			
	(0.007)	(0.005)	(0.010)			
$I_{720 \leq FICO < 740}$	-0.067***	-0.047***	-0.041***			
	(0.009)	(0.007)	(0.010)			
$I_{FICO \geq 740}$	-0.096***	-0.068***	-0.051***			
	(0.009)	(0.007)	(0.008)			
LTV (omitted cat.: (60,80])						
$I_{85 < LTV \leq 90}$				0.015***	0.013***	0.013***
				(0.003)	(0.003)	(0.005)
$I_{90 < LTV \leq 95}$				0.033***	0.026***	0.018***
				(0.004)	(0.004)	(0.005)
$I_{LTV > 95}$				0.131***	0.106***	0.065***
				(0.010)	(0.009)	(0.009)
Loan Officer Comp (%)			0.107***			0.097***
			(0.026)			(0.027)
Loan amount f.e. ($10k bins)	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	0.130	0.400	0.410	0.169	0.428	0.420
Observations	56483	55053	13123	56483	55053	13123

Data Source: Optimal Blue

Notes: The dependent variable is the expected gain from an additional search, given by equation (1). The data cover mortgage rates for 20 metropolitan areas during the period 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-15: Regressions of the Expected Gain from Search, Constructed Using Offers from High-Volume Lenders Only, on Observables

	(1)	(2)	(3)	(4)	(5)	(6)
FICO (omitted cat.: [640,660))						
$I_{680 \leq FICO < 700}$	-0.037***	-0.026***	-0.021**			
	(0.006)	(0.004)	(0.008)			
$I_{720 \leq FICO < 740}$	-0.069***	-0.047***	-0.044***			
	(0.008)	(0.006)	(0.009)			
$I_{FICO \geq 740}$	-0.096***	-0.067***	-0.052***			
	(0.008)	(0.006)	(0.008)			
LTV (omitted cat.: (60,80])						
$I_{85 < LTV \leq 90}$				0.014***	0.010***	0.010**
				(0.003)	(0.003)	(0.004)
$I_{90 < LTV \leq 95}$				0.033***	0.025***	0.020***
				(0.004)	(0.004)	(0.005)
$I_{LTV > 95}$				0.126***	0.100***	0.069***
				(0.009)	(0.009)	(0.012)
Loan Officer Comp (%)			0.106***			0.096***
			(0.024)			(0.026)
Loan amount f.e. ($10k bins)	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	0.107	0.377	0.374	0.142	0.402	0.385
Observations	67637	65757	15444	67637	65757	15444

Data Source: Optimal Blue

Notes: The dependent variable is the expected gain from an additional search, given by equation (1), and it is computed using offers only from high-volume lenders as identified in the Optimal Blue Pricing Insights data. The data cover mortgage rates for 20 metropolitan areas during the period 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-16: The Relationship Between the Lock-Offer Rate Gap and Treasury Yields

	(1)	(2)	(3)	(4)	(5)	(6)
Treasury Yield	-0.059***	-0.081***	-0.051***	-0.067***	-0.089***	-0.058***
	(0.008)	(0.015)	(0.015)	(0.008)	(0.016)	(0.016)
Treasury Yield x DTI $\leq$ 36				0.020***	0.020***	0.017***
				(0.006)	(0.006)	(0.006)
Borrower and Loan Controls	Yes	Yes	Yes	Yes	Yes	Yes
MSA F.E	Yes	Yes	Yes	Yes	Yes	Yes
MSA x Month F.E.		Yes	Yes		Yes	Yes
Branch F.E.			Yes			Yes
Adj. R-Squared	0.141	0.150	0.503	0.143	0.152	0.504
Observations	65938	65936	64025	65938	65936	64025

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 cover mortgage rates for 20 metropolitan areas during the period 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-17: The Relationship Between Lenders' Expensiveness and Their Income and Expenses — Mean (OLS) Regression vs. Median Regression

		Mean Regression		Median Regression	
	Observations	β	St. Err.	β	St. Err.
Income					
Origination Income	1,897	1.02	(0.69)	0.78***	(0.19)
Interest Income	1,897	−0.07	(0.10)	0.03	(0.04)
Secondary Market Income (Gain-on-Sale)	1,897	2.44***	(0.76)	3.37***	(0.43)
Other Income	1,897	0.71	(0.46)	0.05**	(0.02)
Gross Income	1,897	3.47***	(0.70)	4.05***	(0.41)
Expenses					
Loan Production Officers (Sales Employees)	1,897	0.86***	(0.29)	1.06**	(0.47)
Loan Origination (Fulfillment/Non-Sales)	1,897	0.36**	(0.15)	0.45***	(0.13)
Origination-Related Management and Directors	1,897	0.36**	(0.14)	0.25***	(0.07)
Employee Benefits	1,897	0.31***	(0.09)	0.33***	(0.05)
Other Personnel Expenses	1,897	0.37*	(0.22)	0.24	(0.13)
Interest Expenses	1,897	−0.08	(0.12)	0.06	(0.04)
Occupancy and Equipment	1,897	0.29***	(0.05)	0.30***	(0.03)
Technology	1,897	0.08*	(0.04)	0.12***	(0.03)
Outsourcing, Professional, and Subservicing Fees	1,897	0.07	(0.10)	0.10***	(0.02)
Other non-interest expenses	1,897	0.44	(0.32)	0.26	(0.23)
Gross Expenses	1,897	3.05***	(0.65)	3.50***	(0.41)
Corporate overheads					
Total corporate expenses	1,897	0.50	(0.31)	0.28***	(0.10)
Net Income					
Net Income (residential originations, pre-corporate allocations)	1,897	0.45*	(0.25)	0.45**	(0.18)
Net Income (all lines, after corporate allocations and taxes)	1,897	0.46	(0.57)	0.24**	(0.12)

Data Source: Conference of State Bank Supervisors Mortgage Call Reports; Optimal Blue

Notes: The table displays results from OLS and median regressions of lenders' financial filing line items on lender expensiveness (in percentage points), as measured from lender fixed effects column 4 of Table 5, and year-quarter fixed effects. All line items are shown as dollars per $100 originated. Some categories for income and expenses are combined or not shown since they only have zeros for all firms. The standard errors are clustered at the lender level. * p<.1, ** p<0.05, *** p<0.01.

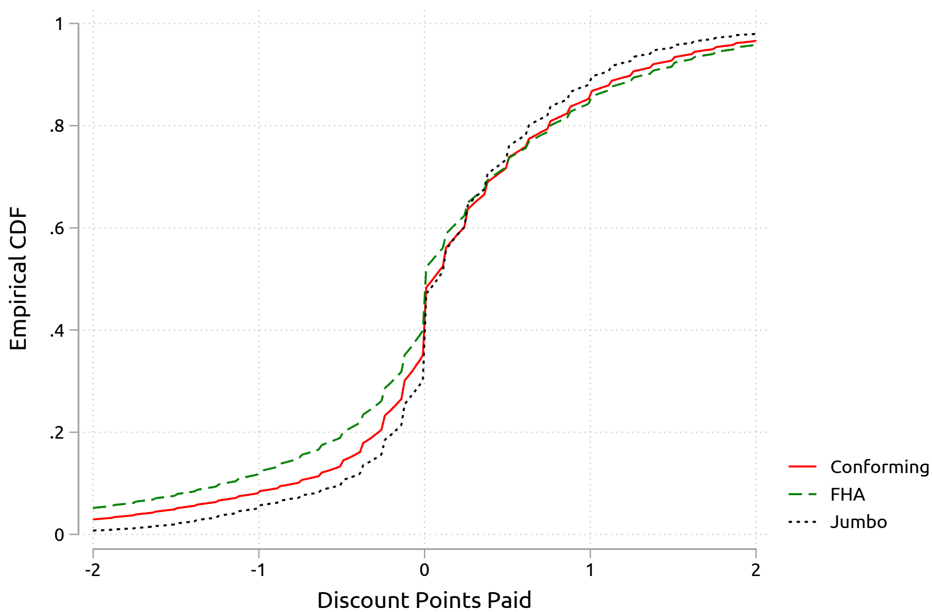


Figure A-1: The Empirical Cumulative Distribution of Discount Points Paid, by Program

Data Source: Optimal Blue

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

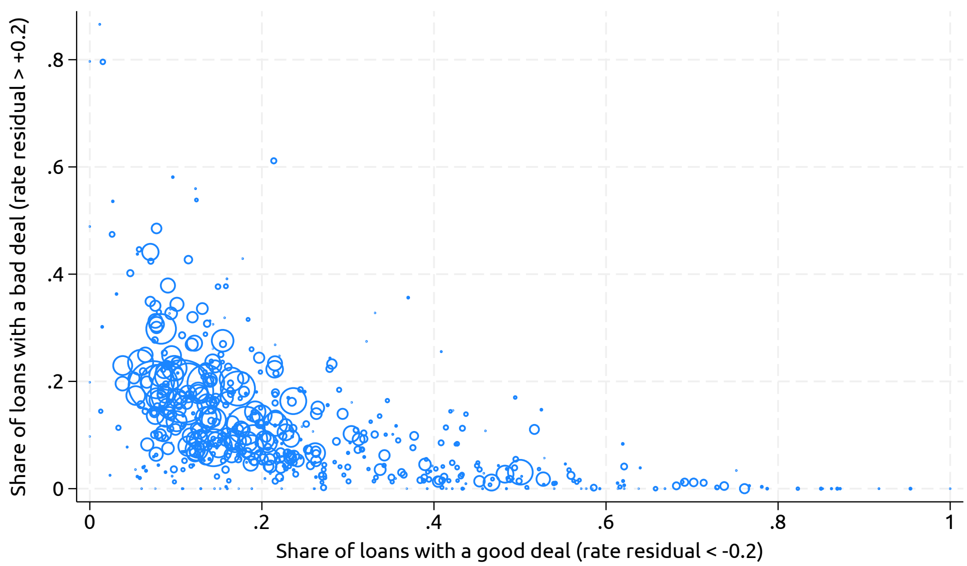


Figure A-2: Are Lenders That Give Good Deals More Likely to Also Give Bad Deals?

Data Source: Optimal Blue
Note: The chart shows lender level shares of locks considered to be good and bad deals. We define whether a lock is a good or bad deal by using the rate residual from specification (3) of Table 5. A good deal is defined as a lock with a rate residual more than 20bp below the average rate lock for that mortgage product in a particular date after controlling for borrower and loan characteristics. Conversely, a bad deal is a lock with a rate residual of more than 20bp. The size of the bubble represents the size of the lender. The Spearman correlation of the share of loans with good and bad deals is -0.64.

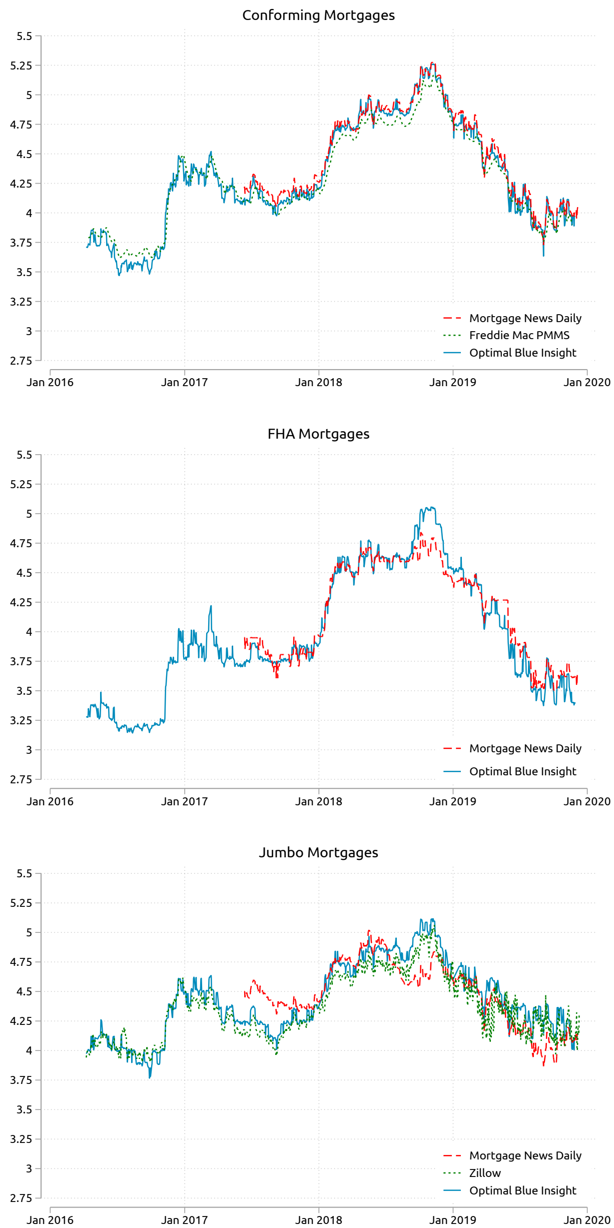


Figure A-3: Comparison of Average Offer Rates from Optimal Blue with Mortgage News Daily Data

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.

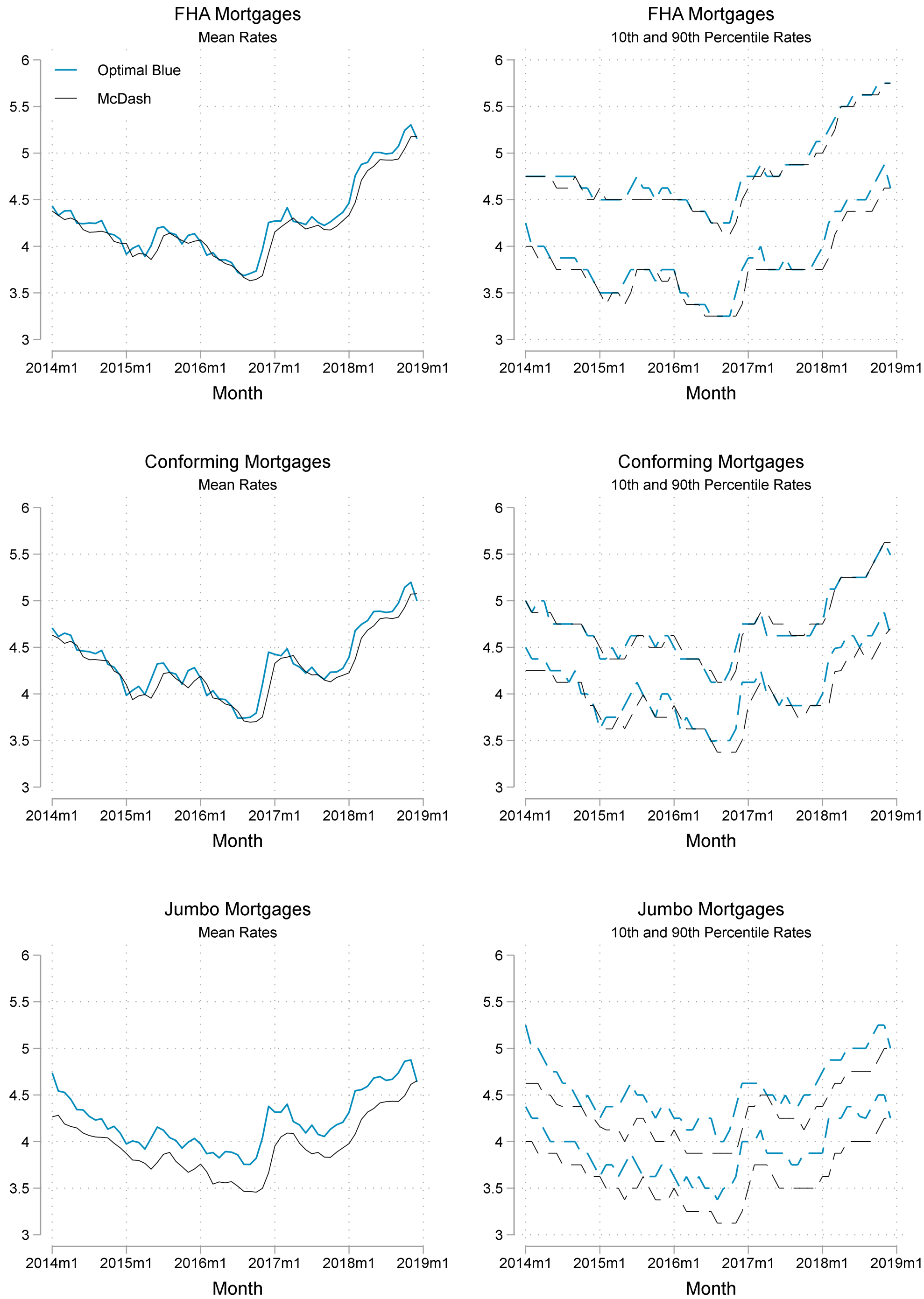


Figure A-4: Comparison of Locked Interest Rates from Optimal Blue with Interest Rates on Closed Originations in McDash

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.

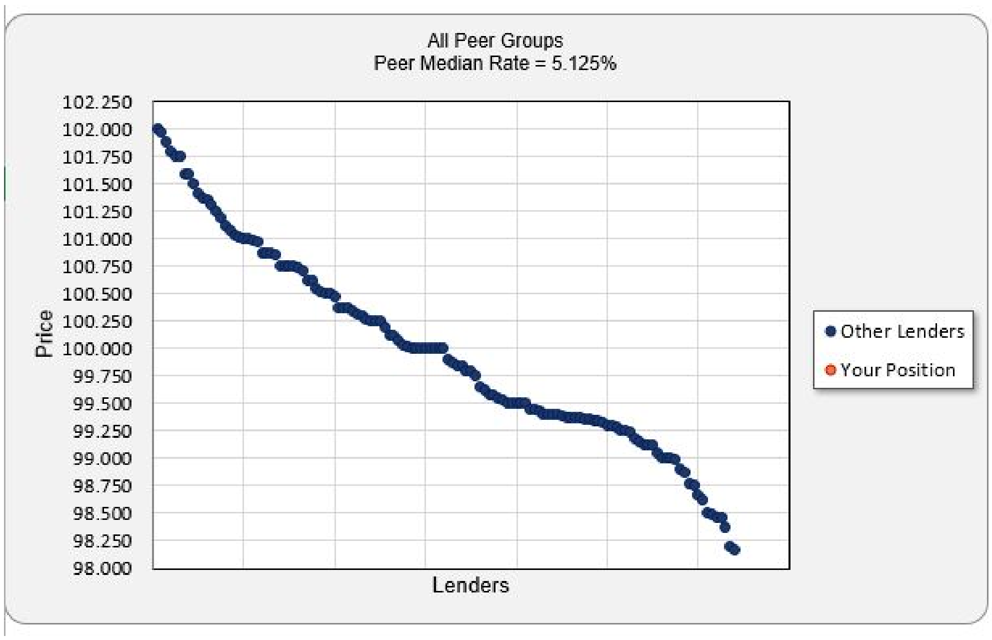


Figure A-5: Screenshot of Sample Offer Distribution from Optimal Blue Pricing Insights

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.

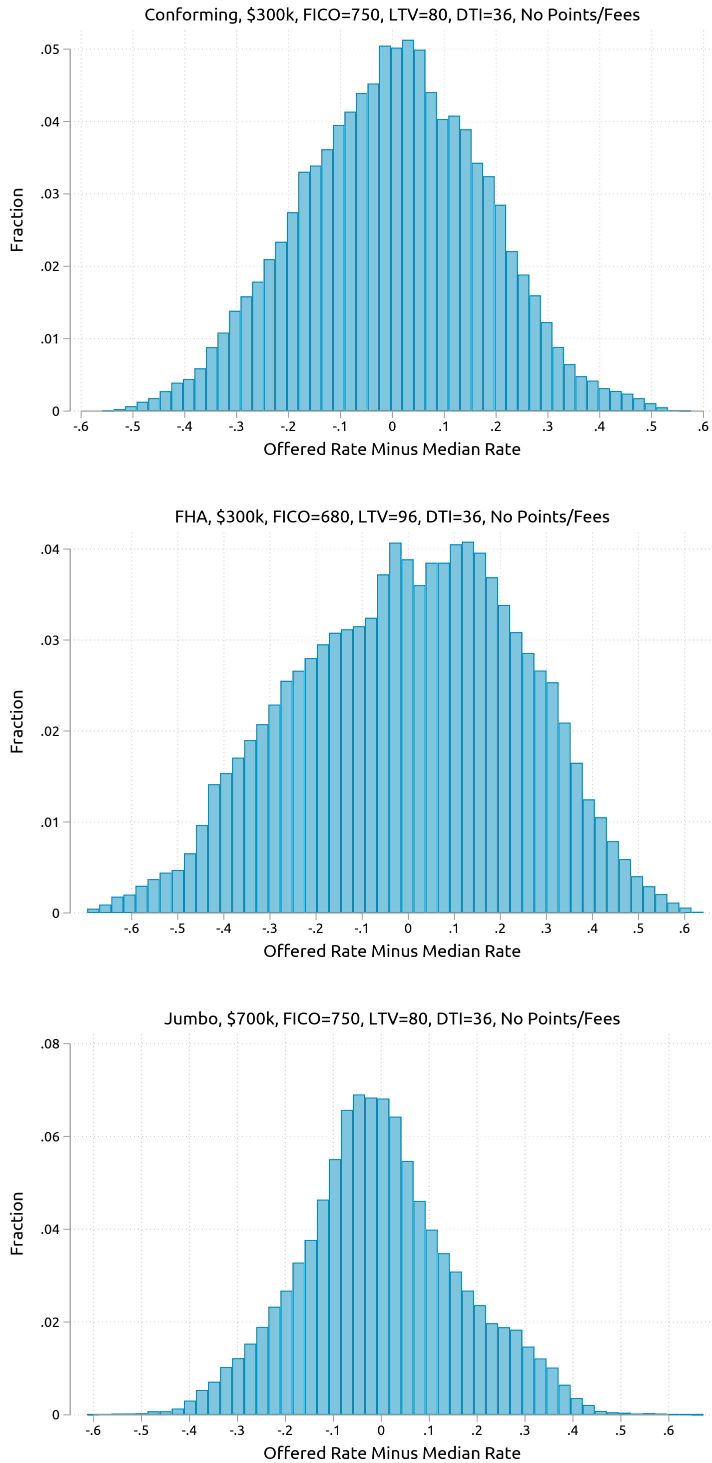


Figure A-6: Interest Rate Offer Dispersion for Identical Mortgages in Los Angeles

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.

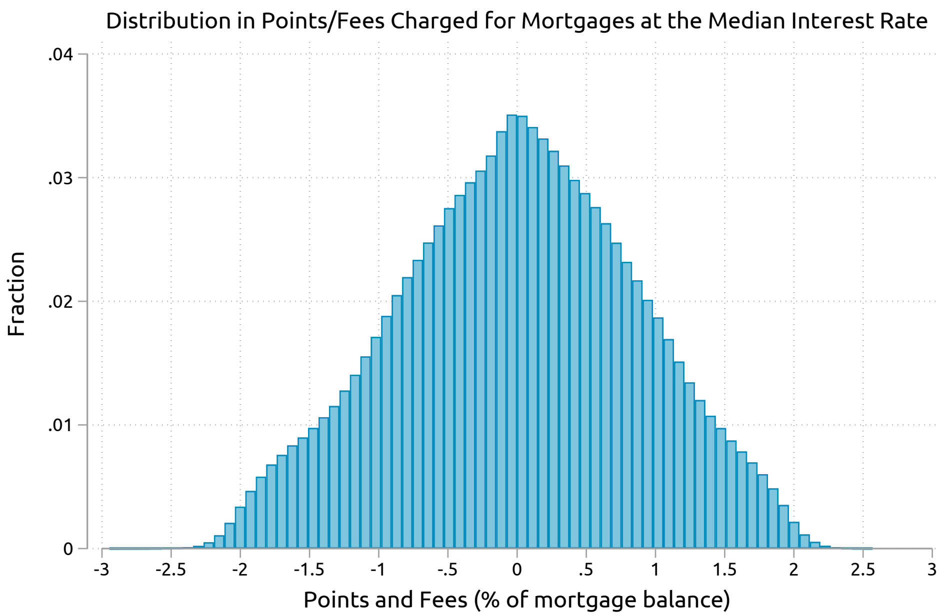


Figure A-7: Dispersion in Points and Fees Lenders Charge for Identical Mortgages at the Median Interest Rate

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.

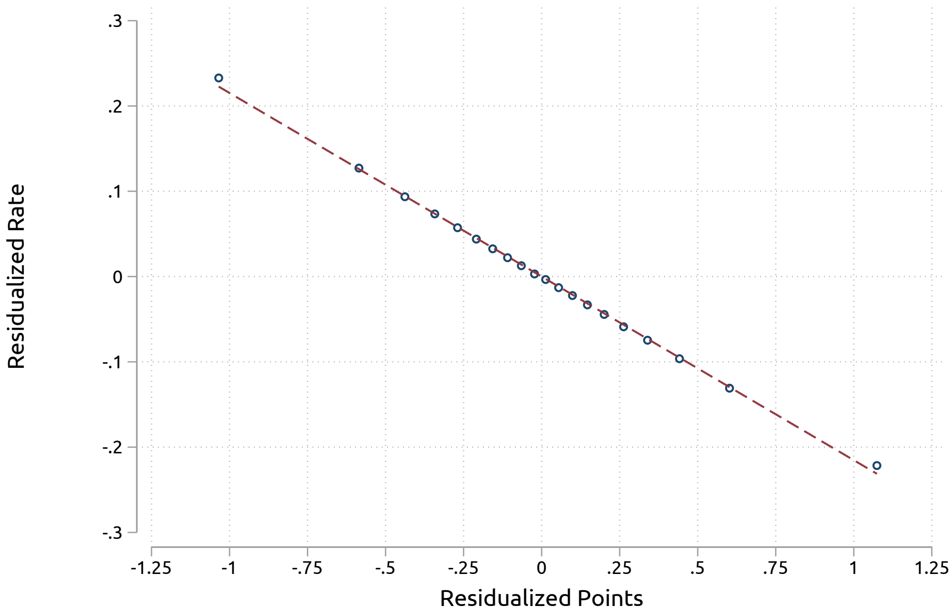


Figure A-8: The Relationship Between Discount Points Paid and Mortgage Rates

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 5, with the only exception that we are not controlling for discount points.