Do Mortgage Brokers Help Borrowers Get Better Rates?[*]

Zirui Wang
McCombs School of Business, University of Texas at Austin

Job Market Paper

November 6, 2026
Link to the latest version

Abstract

Mortgage brokers claim to find lower rates, yet similar borrowers using brokers in the same local market on the same day pay widely different rates. Using U.S. mortgage data linking borrowers, brokers, and lenders, I show that broker firms deliver persistently different rates to similar borrowers. An IV exploiting broker proximity estimates that using a broker in the top rather than bottom tercile of past pricing raises mortgage rate spreads by 23–29 basis points. Lender choice helps explain these differences: some firms systematically choose relatively expensive lenders, even within their existing lender networks. These choices are associated with higher lender-paid compensation, consistent with steering. Steering practices vary across firms and persist within firms, and borrowers using firms with stronger past steering practices pay higher rates. Overall, the findings show how conflicts of interest can undermine the benefits of delegated search and raise household borrowing costs.

JEL classification: G14, G21, G24, G51
Keywords: Mortgage Brokers, Price Dispersion, Steering, Conflicts of Interest

[*]I am deeply grateful to my dissertation committee, John M. Griffin, Samuel Kruger, and Caitlin Gorback, for their invaluable guidance and relentless support. I also thank Andres Almazan, Aydogan Alti, Susan Cherry, Jonathan Cohn, Darrell Duffie, Andrew Fieldhouse, Cesare Fracassi, William Fuchs, John Hatfield, Zhiguo He, Erik Heitfield, Minjoo Kim, Eric Lee, Richard Lowery, Prateek Mahajan, Kevin Mei, Daniel Neuhann, Andrey Ordin, Aaron Pancost, Amiyatosh Purnanandam, Marius Ring, Nikolai Roussanov, Max Sacher, Lukas Schmid, Clemens Sialm, Shikhar Singla, Michael Sockin, Laura Starks, Luke Taylor, Sheridan Titman, Nancy Wallace, Toni Whited, Mindy Xiaolan, Liyan Yang, David Zhang, Xinran Zhang, Yiyuan Zhang (discussant), William Zheng (discussant), Qifei Zhu, and seminar and conference participants at the University of Texas at Austin and the Texas Finance Job Candidate Conference at Texas A&M University for their helpful comments. All errors are my own. Email: Zirui.Wang@mccombs.utexas.edu.

Mortgages are the largest liability on household balance sheets, yet borrowers with similar characteristics often pay very different interest rates. Search frictions help explain these differences: half of mortgage borrowers seriously consider only one lender, and those who know less about mortgages and shop less pay more (Stigler, 1961; Burdett and Judd, 1983; Bhutta, Fuster, and Hizmo, 2026). The use of mortgage brokers holds out hope of addressing these search frictions, as brokers claim to shop across lenders for lower rates. Yet rates on brokered loans are no less dispersed than those on retail loans.[1] Even among similar borrowers using brokers in the same local market on the same day, rates can differ by more than 50 basis points. Why do such differences persist when a mortgage market expert searches on the borrower's behalf? Does the choice of broker matter, and if so, why?

In this paper, I study why brokered mortgage rates remain dispersed and establish a causal link between broker identity and the mortgage rates borrowers pay. If brokers are all the same and equally effective at finding low rates, similar borrowers should pay similar rates regardless of which broker they use, reducing rate dispersion. Alternatively, broker firms may differ in pricing, with some consistently delivering higher rates because they choose more expensive lenders, particularly when those lenders pay brokers more. Understanding brokers' role in this dispersion is challenging because existing data do not jointly identify borrowers, brokers, and lenders, along with borrower characteristics, mortgage prices, and broker compensation. I address this challenge by constructing a new dataset that combines Optimal Blue rate locks, CoreLogic mortgage records, and Home Mortgage Disclosure Act (HMDA) records, allowing me to compare rates across broker firms and examine the incentives behind their lender choices.

My analysis delivers four primary findings. First, similar borrowers pay widely different rates across broker firms, and these differences persist across periods and loan types. An IV exploiting relative proximity to broker branches estimates that using a broker in the top rather than bottom tercile of past pricing raises rate spreads by 23–29 basis points. Second, some firms repeatedly choose lenders that are expensive for their borrowers, and these choices are associated with higher lender-paid broker compensation, consistent with steering. Third, steering practices vary across firms and

[1] Figure 1 compares residual rate spreads in the brokered and retail channels and shows similar dispersion.

persist within firms: brokers that chose expensive lenders in the past continue to do so, and borrowers who use firms with stronger past steering practices pay higher rates. Finally, the relationship between past steering and subsequent borrower rates mainly reflects the selection of expensive lenders within brokers' existing networks, rather than poor borrower-lender fit or differences in lender access.

I begin by documenting large and persistent differences in broker pricing. Among the 170 largest firms, estimated pricing effects range from about 41 basis points below the retail benchmark to 32 basis points above it, after accounting for borrower characteristics, county, and the date the rate was locked. A firm's pricing in earlier years predicts its pricing in later years, and firms that are expensive for purchase borrowers are also expensive for refinancing borrowers and across credit-score groups. In separate specifications, broker firm effects explain 12.1% of the remaining variation in brokered rates, compared with 9.4% for lender effects.

A natural concern is that these differences reflect the borrowers rather than the brokers. High-cost firms may attract borrowers with greater unobserved risk or screen applications more effectively to select safer borrowers. To address this sorting concern, I exploit borrowers' tendency to use nearby broker branches and instrument for using a high-cost broker with the relative distance to the nearest high-cost and low-cost branches. Firms are classified using their pricing in the previous year, so the current borrower's loan does not determine the classification. For purchase borrowers, I use their previous home addresses to measure distance from where they searched for a mortgage, rather than from the newly purchased property. Consistent with this distance-based design, borrowers are more likely to use high-cost firms when a high-cost branch is relatively closer, providing a strong first stage.

The exclusion restriction is that, after controlling for borrower characteristics and local market conditions, relative proximity affects mortgage rates only through broker choice. Living closer to a broker may make that broker more convenient or familiar, but should not directly affect the mortgage rates available to an otherwise similar borrower. A potential violation is that high-cost brokers may locate near borrowers who would pay higher rates regardless of which broker they use. Although the exclusion restriction cannot be tested directly, I use balance tests to gauge the extent of such geographic sorting. I examine a broad range of borrower and neighborhood characteristics, including

proxies for recent mortgage application denials that may indicate harder-to-place borrowers. I also examine the chosen brokers' website claims as proxies for advertised service quality. The differences are generally statistically insignificant or economically small. Although these tests cannot assess unobserved differences, the small imbalances across a broad range of characteristics reduce concerns that geographic sorting drives the results. The IV specifications also control for an extensive list of borrower risk, neighborhood characteristics, general access to broker branches, and market conditions at both the search origin and the mortgaged property.

The IV estimates imply that using a top-tercile rather than bottom-tercile broker raises rate spreads by 23–29 basis points. These estimates are 102–134% of the average rate spread in the IV sample, an economically large effect. Broker identity therefore plays an economically important role in mortgage pricing.

Why do some broker firms deliver higher rates? One explanation may lie in their lender choices. Brokers often advertise access to dozens of lenders on their websites, yet among firms making such claims, the median firm used only five lenders in the previous year. This motivates examining which lenders brokers regularly use and which lender they select for each borrower. For each loan, I use lender pricing models to predict the rate that each lender in a statewide comparison set would charge the borrower, including the chosen lender and alternative lenders. These models account for borrower characteristics, broker-lender relationships, and local market conditions. I then compare the chosen lender's predicted rate with those of alternative lenders. On average, the chosen lender's predicted rate is 27.7 basis points above the lowest predicted rate.

A concern is that lenders may observe borrower characteristics that my pricing model does not capture. The rates quoted to brokers may therefore differ from the predicted rates. A diligent broker who selects the lowest quote can therefore generate a positive observed gap if the selected lender does not have the lowest predicted price. To account for this, I simulate a diligent broker who chooses the lowest quote across lenders. I construct these quotes by adding random pricing errors to the predicted rates, drawn from a distribution calibrated to out-of-sample prediction errors. For each loan, I calculate the average gap that this simulated diligent broker would generate and subtract it

from the observed gap. Estimated overpayment remains positive after this adjustment, averaging 11.9 basis points; I call this difference the *steering gap* and use it as a proxy for steering practices. Larger gaps are associated with both higher realized mortgage rates and higher lender-paid broker compensation. These findings are consistent with brokers steering borrowers toward lenders that are expensive and pay them more.

These practices differ markedly across broker firms. Some of the largest firms have average steering gaps close to zero, while others have gaps of 15–20 basis points. Firms with larger steering gaps also have higher pricing effects relative to retail: a 10 basis point larger firm-level gap is associated with a 6.9 basis point higher pricing effect. This relationship helps explain the persistent differences in broker pricing documented earlier.

The differences also persist over time. Firms that used relatively expensive lenders in the previous year tend to continue doing so in subsequent loans. Those differences matter for a borrower choosing a broker. Borrowers who use a firm with a higher past steering gap pay higher rates themselves, even after accounting for their characteristics and local market conditions. The higher rates associated with past steering largely reflect the use of generally expensive lenders, rather than poor borrower-lender fit. They also mainly reflect expensive lender choices within brokers' existing networks, rather than differences in those networks.

Overall, the findings have three implications. First, choosing a broker matters as well as choosing a lender: broker identity explains at least as much variation in brokered rates as lender identity, and brokers' past lender choices predict the rates their future borrowers pay. Second, the costs to borrowers associated with steering can be much larger than the additional compensation earned by brokers. Comparing loans in the highest and lowest steering-gap deciles, brokers receive approximately $290 more in lender-paid compensation on a $330,000 mortgage, while borrowers pay about $7,500 more in present-value borrowing costs over an 11-year holding period. The estimated borrower cost is roughly 25 times the additional broker compensation, illustrating how relatively small financial incentives for intermediaries can be associated with substantial costs to consumers. Third, current regulation leaves room for steering across lenders. Regulation Z restricts compensation tied to mortgage terms

within a lender but allows compensation to differ across unaffiliated lenders, while its anti-steering safe harbor permits qualifying comparisons among a subset of the lenders with which a broker regularly does business. Greater transparency about lender-paid compensation and brokers' past lender choices could help borrowers and regulators identify costly practices.

Related Literature. This paper contributes to three strands of literature. First, I contribute to research on mortgage intermediaries and borrowing costs. Prior studies compare brokered and retail loans, finding higher costs for some borrowers, particularly those with lower incomes and credit scores (LaCour-Little, 2009). Other studies show that borrowers who compare too few brokers overpay (Woodward and Hall, 2012), and examine broker competition, profits, and origination fees (Berndt, Hollifield, and Sandås, 2010; Ambrose and Conklin, 2014). Related work examines brokered refinances, licensing, and racial differences in fees (Spader and Quercia, 2011; Agarwal, Ang, Deng, and Wang, 2021; Ambrose, Conklin, and Lopez, 2021). Outside the United States, brokers can increase lender competition while commissions distort advice, and borrower preferences help explain product choices (Robles-Garcia, 2020; Allen, Clark, Houde, Li, and Trubnikova, 2025).[2] I shift the focus from whether borrowers use a broker to which broker they use. I document persistent pricing differences across U.S. broker firms and use a proximity-based IV to show that broker choice has a substantial effect on borrowing costs.

Second, I contribute to research on conflicts of interest and steering by financial intermediaries. Seller-paid compensation can influence the advice intermediaries give their customers (Inderst and Ottaviani, 2012). Studies of mutual funds, convertible bonds, annuities, and life insurance link compensation to product recommendations and customer outcomes (Christoffersen, Evans, and Musto, 2013; Bergstresser, Chalmers, and Tufano, 2009; Egan, 2019; Egan, Ge, and Tang, 2022; Anagol, Cole, and Sarkar, 2017). Similar incentives shape insurer selection by health insurance brokers (Li, Liu, LoSasso, and Tilipman, 2026) and property recommendations by real estate agents (Barwick, Pathak, and Wong, 2017; Barry, Fried, and Hatfield, 2025), while brokered markets can sustain high commis-

[2] Related research examines lender screening and markups, specialization, and the role of local loan officers in credit provision (Bosshardt, Kakhbod, and Kermani, 2025; Paravisini, Rappoport, and Schnabl, 2023; Huang, Mayer, Titman, and Xu, 2026).

sions even with low concentration and easy entry (Hatfield, Kominers, and Lowery, 2025). In lending, prior work studies steering across loan products (Guiso, Pozzi, Tsoy, Gambacorta, and Mistrulli, 2022; Agarwal, Amromin, Ben-David, and Evanoff, 2025; Agarwal, Ambrose, and Yao, 2020; Benetton, Buchak, and Robles-Garcia, 2022) and real estate agents' influence on lender choice (Lopez, McCoy, and Sah, 2022; Zhang, Barwick, Han, and Kroah, 2026; Jorgensen, 2025). I provide first evidence of steering by mortgage brokers across wholesale lenders within the same mortgage product, rather than across loan products or through real estate agent referrals. By comparing the prices alternative lenders would charge the same borrower, I link costly lender choices to higher borrower rates and broker compensation, and show that these practices persist within broker firms.

Finally, the paper contributes to research on price dispersion and search in financial markets. Search costs can sustain price differences across sellers (Stigler, 1961; Burdett and Judd, 1983). In mortgage markets, limited borrower shopping and lender advertising help explain why similar borrowers pay different rates (Allen, Clark, and Houde, 2014; Alexandrov and Koulayev, 2017; Bhutta, Fuster, and Hizmo, 2026; Woodward and Hall, 2010, 2012; Gurun, Matvos, and Seru, 2016). Related studies document search frictions in auto lending and price dispersion in commercial real estate (Argyle, Nadauld, and Palmer, 2023; Priest, 2023). Substantial price differences also arise for identical or nearly identical financial products, including municipal and corporate bonds, convertible bonds, and index funds (Green, Hollifield, and Schürhoff, 2007; Griffin, Hirschey, and Kruger, 2023, 2026; Egan, 2019; Hortaçsu and Syverson, 2004). I extend this literature by studying price dispersion when borrowers hire market experts to shop on their behalf. I show that substantial dispersion persists even with professional intermediaries, and that persistent differences in broker pricing and lender choices help explain why. Thus, price dispersion reflects not only borrowers' search frictions but also the choices and incentives of the intermediaries searching for them.

1 Data, Background, and Hypothesis

I first introduce the data, then describe the institutional setting and develop the hypotheses that guide the analysis.

1.1 Data

Studying broker pricing requires linking the borrower, broker, and lender on each loan while observing the borrower's risk characteristics and mortgage price. No single source provides all these elements. Mortgage records identify the participants but lack detailed borrower risk measures; rate locks contain risk and pricing information but do not identify the borrower; and HMDA reports prices and demographics but identifies the lender rather than the broker. I combine these sources at the loan level.

1.1.1 Data Sources

CoreLogic mortgage records identify borrowers and properties and report the mortgage amount, dates, and company and loan-officer identifiers. These records connect the borrower to the broker and the lender on a completed mortgage. Optimal Blue rate-lock data provide credit scores, loan-to-value and debt-to-income ratios, the date and rate of the lock, the requested lock period, and broker compensation. They measure the terms agreed upon before closing. HMDA data provide rate spreads, origination charges, income, and demographics. Linking the three sources allows me to follow the same mortgage from its rate lock to its recorded origination and compare pricing across the firms involved.

I supplement these data with NMLS licensing and branch records, Verisk and DataAxle consumer address histories, and American Community Survey neighborhood characteristics. The address histories locate purchase borrowers before they buy the mortgaged property, allowing me to measure their proximity to broker branches when searching. I also collect broker website claims to study how firms describe their services.[3]

1.1.2 Sample Construction

The three sources share no universal identifier for a loan or a borrower. I therefore match them on loan characteristics that they record in common, such as the property's location, the loan amount, and the loan's terms, and I do so in two steps. First, I link Optimal Blue locks to CoreLogic mortgages using property location, loan amount, dates, and loan characteristics. Second, I link these pairs to

[3]Websites come from NMLS records and supplementary searches by company name. They were collected in June and September 2026 and classified by claims about rates, lender access, approval, and service.

HMDA using geography, lender identity, loan amount, occupancy, and note rate. The match produces 15.8 million loans that closed from 2018 to 2025. Internet Appendix Section A describes the matching procedure in greater detail, and Figure IA.2 visualizes the matching procedure.

As a diagnostic for accidental matches, matching the 2022 locks against the 2021 HMDA file produces matches for about 1% of locks.

I classify a loan as brokered when CoreLogic identifies a non-funding broker firm and either Optimal Blue or HMDA identifies an intermediated application; the remaining matched loans are classified as retail. Distinguishing the funder from the broker is important because CoreLogic's company fields do not have a consistent lender-broker ordering. I resolve the roles using the lender of record and NMLS company identities. Each brokered loan is then assigned to the branch where its officer worked in the lock month using NMLS employment and branch records.

The main analysis focuses on conventional conforming fixed-rate mortgages, which makes the comparison less dependent on differences across mortgage products. It includes 8,150,052 loans, of which 335,997 are brokered, spanning 15,409 broker firms and 31,377 loan officers. The lender-choice analysis uses the 308,810 brokered loans for which lender prices can be predicted. The distance-based IV analysis additionally requires a located search origin and a previous-year price classification for the broker. I explain these requirements with the corresponding tests; minor filters are reported in the exhibit notes.

1.1.3 Main Variables

Mortgage prices. The main outcome is the HMDA rate spread: the loan's annual percentage rate (APR) minus the average prime offer rate (APOR) for a comparable loan when the rate is set, measured in basis points. Because APR includes points and other finance charges, the spread captures more of the borrower's cost than the note rate alone. Optimal Blue lock dates also allow me to compare borrowers who face the same market rates. Borrowers closing on the same day may have locked weeks apart; comparisons by lock date avoid this timing difference.

Broker compensation. Broker compensation is measured in basis points of the loan amount. This unit differs from the rate spread: compensation records a payment on the transaction, whereas the

rate spread is an annualized measure of borrowing cost. Keeping the two units distinct is important when interpreting whether a more expensive lender choice pays the broker more.

Lender networks. A broker's lender network consists of lenders that funded its loans during the previous 12 months. I construct these networks from all brokered CoreLogic mortgages, rather than only matched loans, so a lender relationship is not lost merely because a mortgage cannot be linked to the other sources. I measure lender concentration by the sum of squared lender shares (HHI). A high HHI means that the firm places a large share of its loans with a small number of lenders; it does not by itself indicate whether those lenders are cheap or expensive. The baseline pricing controls include income, credit score, loan amount, property value, combined loan-to-value ratio, debt-to-income ratio, origination charges, race, and sex.

1.1.4 Summary Statistics and Representativeness

Table 1 reports brokered-loan prices, borrower characteristics, loan characteristics, and lender networks. Rate spreads vary substantially, with a mean of 31.3 basis points and a standard deviation of 48.2 basis points. The 10th and 90th percentiles are -17.8 and 92.0 basis points. The average borrower has a credit score of 752, an income of \$130,000, and a loan of \$353,000. Average debt-to-income and combined loan-to-value ratios are 37.4% and 73.0%, respectively. These borrowers have substantial observable credit quality, yet face considerable variation in mortgage prices.

Retail borrowers have similar credit scores and debt-to-income ratios, although brokered loans are larger and more often cash-out refinances. These composition differences motivate the borrower and loan controls used in comparisons across channels. The matched brokered sample covers 4.8% of comparable HMDA brokered loans. It has a similar average rate spread under common sample definitions, but overrepresents 2020–2021 originations and principal residences. The data therefore provide detailed links for a subset of the broker market.[4]

[4] Table IA.1 compares brokered and retail loans. Table IA.2 compares the matched brokered sample with HMDA using common definitions.

1.2 Institutional Background

The wholesale mortgage market separates the intermediary that advises the borrower from the lender that funds the loan.

1.2.1 Brokers and Wholesale Lenders

A borrower can obtain a mortgage directly from a lender or through a mortgage broker. In the wholesale channel, the broker takes the application, collects documents, and submits the loan to a lender, which underwrites and funds it. Brokers work with lenders that have approved them as wholesale partners. Each broker-lender agreement sets the terms of the relationship, including compensation. Brokers therefore draw on a limited lender network: for the median brokered loan in the sample, the firm used five lenders in the previous year and the loan officer used three (Table 1, Panel D).

Brokers obtain pricing through lenders' rate sheets or online systems and lock the borrower's rate with the selected lender. A lender can offer different pricing to different broker accounts, and brokers may negotiate pricing on individual loans. The price available from a lender can therefore depend on which broker submits the loan. The firm and its loan officers enter the transaction at different levels. The firm maintains the wholesale relationships, while an individual loan officer works with the borrower and submits the application. A firm may therefore have access to more lenders than an individual officer regularly uses.

Figure 6, Panel A, presents the phrases used on mortgage broker websites, with larger phrases appearing on more firms' websites. Expertise and guidance are prominent, as are access to multiple lenders and finding a lower or better rate. Brokers thus advertise both help with the mortgage process and shopping on the borrower's behalf.[5]

Licensing and disclosure. Mortgage originators are licensed or registered through the Nationwide Multistate Licensing System (NMLS). Broker firms also obtain state licenses, and loan documents report originator and company identifiers.

[5] Figure IA.1 provides an example of a broker website.

1.2.2 Broker Compensation and Its Regulation

The broker is paid either by the borrower through fees at closing or by the lender, typically as a percentage of the loan amount; it cannot receive compensation from both on the same transaction. Among brokered loans whose Optimal Blue record states who pays the broker, the lender pays on 63.3% and the borrower on 36.7% (Table 1, Panel A). Recorded broker compensation averages 207.6 basis points of the loan amount. Lender-paid compensation is a cost of originating the mortgage and can be reflected in the borrower's price.

Federal rules restrict how brokers can be paid. Before 2011, lenders could pay brokers more for higher-rate loans through yield spread premiums (Woodward and Hall, 2012). Regulation Z now prohibits compensation based on the interest rate or other transaction terms, or proxies for those terms, while permitting compensation based on the amount of credit extended. Compensation can nevertheless differ across agreements with unaffiliated lenders. A broker can therefore earn more by placing a loan with one lender than with another, even though a given lender cannot pay more because the borrower receives a higher-rate loan.[6]

Anti-steering rule. The anti-steering provision seeks to prevent brokers from directing borrowers to loans to increase their own compensation. Under 12 CFR 1026.36(e)(1), a broker may not direct a borrower to a particular loan because it provides greater compensation unless the loan is in the borrower's interest. The rule does not require choosing the lender that pays the broker the least. A loan from a higher-paying lender may satisfy the rule when its terms serve the borrower's interest, such as when a higher rate comes with lower upfront charges that the borrower prefers.[7]

Safe harbor. The regulation also provides a safe harbor based on a limited comparison of loan options. A broker must obtain and review options from at least three lenders with which it regularly does

[6] See 12 CFR 1026.36(d) and its official commentary. For example, a lender cannot pay 1% on one loan and 2% on another because their interest rates or other transaction terms differ. Separate agreements under which unaffiliated Lender A pays 1% and Lender B pays 2% are permissible, subject to the compensation and anti-steering rules. Affiliated lenders are treated as one person under paragraph (d)(3). The CFPB also discusses brokers' ability to choose among lenders paying different compensation in its June 2013 final rule, 78 FR 35430.

[7] See official comments 36(e)(1)-2 and -3. Outside the safe harbor described below, the borrower-interest comparison considers other possible offers available through lenders with which the broker regularly does business and for which the borrower likely qualifies. Merely describing a loan as being in the borrower's interest does not establish compliance.

business, or all such lenders if it works with fewer than three, and present options meeting specified rate, fee, and loan-feature criteria for which the borrower likely qualifies.[8] It need not review every lender in its network or search the entire market. For example, a broker working with ten lenders can satisfy the comparison requirement by reviewing three, provided it meets the other safe-harbor conditions. Those three need not be the least expensive or lowest-paying lenders in its network.

Fiduciary duties. These federal compensation and anti-steering restrictions do not impose a general fiduciary duty on mortgage brokers. State laws can impose additional obligations. California, Nevada, and Washington expressly impose fiduciary duties on covered mortgage brokers or licensees, and New Mexico imposes such duties on mortgage loan originators.[9] Compliance with the federal safe harbor does not remove applicable state duties.

These rules seek to prevent compensation-driven steering, but they leave differences in compensation across lenders and discretion over which lenders a broker compares. If the comparison is limited to relatively expensive or high-paying lenders, its lowest-rate option can still be more expensive than alternatives elsewhere in the broker's network or in the market. The lender that pays the broker more need not offer the borrower a lower price. These features leave room for compensation to influence lender choice, even with federal anti-steering restrictions and applicable state duties.

1.3 Research Question and Hypotheses

If brokers shop across lenders for borrowers, they may reduce the search frictions associated with mortgage rate dispersion. I first compare the dispersion of brokered and retail rates after accounting for borrower characteristics and local market conditions:

$$RateSpread_i = \beta \, \mathbb{1}(Brokered)_i + X_i'\gamma + \delta_{c(i),d(i)} + \varepsilon_i, \tag{1}$$

[8] Under 12 CFR 1026.36(e)(2)–(4), for each loan type in which the borrower expresses interest, the required options include the lowest interest rate, the lowest total dollar amount of discount points and origination points or fees, and the lowest interest rate without specified features such as negative amortization, a prepayment penalty, or a balloon payment in the first seven years. The broker must have a good-faith belief that the borrower likely qualifies. One loan can satisfy multiple criteria; the options presented need not come from three different lenders. If more than three options are presented for a loan type, the qualifying options must be highlighted.

[9] See California Civil Code Section 2923.1; Nevada Revised Statutes Section 645B.0147; Washington RCW 19.146.095; and New Mexico Statutes Section 58-21B-20. The provisions differ in scope. For example, California requires covered brokers to place the borrower's economic interest ahead of their own, while New Mexico specifies duties for individual mortgage loan originators. These are examples, not an exhaustive list of state-law obligations.

where $RateSpread_i$ denotes the rate spread of loan i, in basis points, and $\mathbb{1}(Brokered)_i$ equals one for a brokered loan and zero for a retail loan. X_i denotes the baseline borrower and loan controls. The functions $c(i)$ and $d(i)$ identify the property's county and the loan's lock date, respectively. The fixed effect $\delta_{c,d}$ absorbs market conditions shared by borrowers in county c locking on date d, and ε_i denotes the remaining loan-level pricing variation. I estimate the model jointly on retail and brokered loans.

The coefficient β captures the average conditional difference between channels. The residuals measure how much more or less each borrower pays than the model predicts. Estimating a common pricing function and including the brokered indicator separates the comparison of dispersion from differences in average rates. In particular, a narrower brokered residual distribution would indicate that the rates of similar brokered borrowers vary less around their channel's conditional mean.

Figure 1 presents the residual distributions, with gray bars for retail loans and a red line for brokered loans. The distributions overlap closely. The residual standard deviation is 38.2 basis points for retail loans and 39.8 basis points for brokered loans. Both distributions have long right tails: some borrowers pay at least 50 basis points more than similar borrowers in the same market. Average rates are also similar across the two channels.[10]

The comparison motivates the paper's question: why do similar borrowers pay different rates even when brokers shop on their behalf? I examine whether systematic differences across broker firms help explain this dispersion.

Similar dispersion in the two channels does not establish the role of brokers. Even if all brokers delivered the same expected price to a given borrower, unobserved borrower characteristics and loan-specific pricing could generate dispersion. Alternatively, some brokers might deliver lower prices and others higher prices, with their differences obscured by the channel average. Distinguishing these explanations requires following broker identities across loans, rather than comparing channels alone.

Hypothesis 0. *Observably similar borrowers in the same market pay the same expected rate regardless of which broker firm they use.*

[10] Table IA.3 compares average rates under alternative fixed effects and finds no statistically significant difference between the two channels. Figures IA.4 and IA.3 compare the channels across credit scores and with finer geographic controls and find similar patterns.

Under this null, broker firms should not have persistent differences in conditional price levels. Their estimated effects may differ in a finite sample, but a firm's effect in one group of loans should not systematically predict its effect in another. The null does not require the brokered and retail channels to have the same average price. The alternative concerns differences among broker firms themselves.

Hypothesis 1. *Observably similar borrowers in the same market pay different expected rates depending on which broker firm they use.*

If some brokers deliver lower rates and others higher rates, the brokered average can resemble the retail average while substantial dispersion remains. One potential explanation is lender choice. Brokers may select expensive lenders when those lenders offer higher compensation.

Hypothesis 1A. *Some broker firms systematically choose lenders that are expensive for their borrowers. These choices are associated with higher borrower rates and higher lender-paid compensation.*

I first test for persistent differences in broker pricing and estimate the effect of using a high-cost firm. I then examine lender choice and compensation as a potential explanation for those differences.

2 Are All Broker Firms the Same?

Brokered rates remain widely dispersed after accounting for borrower characteristics and local market conditions. I first ask whether the broker firm a borrower uses helps explain this dispersion. I estimate firm-specific price levels and examine whether the differences persist across periods and loan types.

2.1 Broker Firms Differ in Pricing

If broker firms deliver similar rates to similar borrowers, their estimated price levels should differ only because of sampling variation. I test this prediction by estimating a separate effect for each broker firm:

$$RateSpread_i = \sum_f \beta_f \, \mathbb{1}(BrokerFirm_i = f) + X_i'\gamma + \delta_{c(i),d(i)} + \varepsilon_i, \tag{2}$$

where $RateSpread_i$ denotes loan i's rate spread, and $\mathbb{1}(BrokerFirm_i = f)$ equals one if the loan was originated through broker firm f and zero otherwise. The sum runs over broker firms. X_i denotes borrower and loan controls, while $c(i)$ and $d(i)$ identify the county and lock date. The fixed effect $\delta_{c,d}$ absorbs common local pricing conditions, and ε_i denotes the residual. Retail loans are the omitted group, so β_f measures how much more or less firm f's borrowers pay than similar retail borrowers in the same local market on the same day.

Figure 2 presents these estimates for the 170 largest broker firms. Each dot shows a firm's estimated price level relative to retail, together with its confidence interval. Firms are ordered from lowest to highest price level, with zero indicating the retail benchmark. Blue and red dots identify firms significantly below and above retail, respectively, after adjusting for multiple comparisons. The estimates range from about -41 to $+32$ basis points. Fifty-nine firms have significantly lower estimated rates than retail, while 30 have significantly higher rates. These differences are large despite comparing borrowers with similar observable characteristics in the same market. The pattern extends beyond the largest firms: among all 10,104 firms in the regression, 947 are significantly below retail and 1,190 above it after correcting for multiple comparisons. Thus, the near-zero average price difference between brokered and retail loans masks substantial differences in pricing across broker firms.

2.2 Broker Pricing Persists Across Periods and Loan Types

The estimated pricing differences could reflect a particular mix of borrowers or loans rather than persistent differences across firms. If broker pricing is persistent, firms that deliver higher rates in one group of loans should also deliver higher rates in another. I test this by estimating equation (2) separately across periods and borrower and loan groups.

Figure 3 compares firm price levels across six splits: 2018–2021 versus 2022–2025, purchases versus refinances, conventional versus government-backed loans, and groups defined by credit score, loan term, and loan-to-value ratio. Each panel plots firm price levels estimated separately for the two groups, using retail loans in each group as the benchmark. The relationships are positive throughout: firms that deliver higher rates in one group also tend to deliver higher rates in the other. Correlations adjusted for estimation noise range from 0.65 to 0.95 (Table IA.4). These results show that broker

pricing differences persist across time, loan products, and borrower groups.

How much do these persistent differences contribute to mortgage rate dispersion? To quantify their importance, I compare the share of residual rate-spread variation explained by broker firm and lender identities. Table 2 starts with borrower and loan controls and county × lock-date fixed effects, then adds lender effects, broker firm effects, or both. Broker firm effects explain 12.1% of the baseline residual variance, compared with 9.4% explained by lender effects. Including both sets of effects explains 18.2%. Thus, broker identity explains a substantial share of the remaining variation in mortgage rates, even more than lender identity alone.

The importance of broker pricing differences is also reflected in local market outcomes. Figure IA.5 shows counties with a greater share of loans from high-price firms have higher residual brokered-rate dispersion and higher brokered rates relative to retail.[11] These findings establish that broker pricing differences are persistent and economically important. The next question is whether these differences reflect the brokers themselves or the borrowers who select into them. I address this question using an IV based on borrowers' proximity to broker branches.

3 Do High-Cost Broker Firms Raise Borrowers' Rates?

A natural concern is that the persistent pricing differences reflect the borrowers rather than the brokers. High-cost firms may serve borrowers with risks that lenders observe but I do not, or borrowers with different preferences for mortgage services. To address this sorting concern, I exploit borrowers' tendency to use nearby broker branches. I first document this relationship and then use borrowers' relative proximity to high-cost and low-cost branches as an instrument for broker choice.

3.1 Borrowers Use Nearby Broker Branches

I begin by examining how distance affects borrowers' choice of broker branches. For each borrower, I form a choice set of branches active during the previous 12 months whose firms are licensed in the property's state. The choice set includes both the selected branch and alternatives the borrower could have used. Distance is measured from the borrower's home when searching, which I call the search

[11] Figure IA.5 measures each firm's price level using its loans in other counties, so the local loans do not directly determine that classification.

origin. For a purchase borrower, this is the previous address traced through consumer address histories; for a refinance borrower, it is the mortgaged property. I estimate

$$\mathbb{1}(Chosen)_{i,j} = \sum_k \beta_k \, \mathbb{1}(Distance_{i,j} \in Bin_k) + \alpha_i + \eta_{j,t(i)} + u_{i,j}, \tag{3}$$

where each observation pairs loan i with a potential branch j. $\mathbb{1}(Chosen)_{i,j}$ indicates whether the borrower chose that branch, and $Distance_{i,j}$ denotes the distance from the search origin to the branch. The coefficients β_k measure differences in branch choice probabilities across distance bins, relative to branches at least 20 miles away. Loan fixed effects α_i hold the borrower fixed, while branch $\times$ lock-month effects $\eta_{j,t}$ absorb a branch's overall popularity at the time of the loan. The specification therefore compares branches at different distances from the same borrower, accounting for differences in their overall popularity.

Figure 4, Panel A, presents the estimated choice probabilities relative to branches at least 20 miles away. Dots show the coefficients in percentage points. The probability of choosing a branch falls steadily with distance. A branch within half a mile is 2.67 percentage points more likely to be used, compared with 1.30 percentage points for a branch one to two miles away and about 0.7 percentage points for a branch three to five miles away. These results show that proximity is an important predictor of broker choice. When a high-cost branch is relatively closer than a low-cost branch, borrowers should therefore be more likely to use a high-cost broker. This relationship motivates the differential-distance instrument developed in the next subsection.[12]

3.2 Differential Distance as an Instrument

The tendency of borrowers to use nearby branches motivates an instrument based on their relative proximity to high-cost and low-cost brokers. I classify firms using their estimated price levels in the previous calendar year, so the borrower's current loan does not enter the classification. To reduce noise, I shrink these estimates toward the monthly mean, with greater shrinkage for less precisely estimated firms. A firm is high-cost if its lagged price level is in the top tercile across branches in the lock month

[12] The average choice set contains about 1,700 branches. A conditional logit specification also finds a strong distance relationship (Table IA.5).

and low-cost if it is in the bottom tercile. The instrument is the difference between the distances to the nearest low-cost and high-cost branches:

$$dd_i = \textit{Distance}(\textit{Home}_i,\ \text{nearest low-cost branch}) - \textit{Distance}(\textit{Home}_i,\ \text{nearest high-cost branch}),$$

where $\textit{Home}_i$ denotes the search origin of the borrower on loan i, and dd_i denotes differential distance. A positive value means that the nearest high-cost branch is closer than the nearest low-cost branch.[13] I estimate the effect of using a high-cost broker by two-stage least squares:

$$\mathbb{1}(\textit{HighCost})_i = \pi\, dd_i + X_i'\gamma_1 + W_i'\lambda_1 + \alpha_{o(i),t(i)} + \mu_{c(i),t(i)} + u_i, \tag{4}$$

$$\textit{RateSpread}_i = \beta\, \widehat{\mathbb{1}(\textit{HighCost})}_i + X_i'\gamma_2 + W_i'\lambda_2 + \alpha_{o(i),t(i)} + \mu_{c(i),t(i)} + \varepsilon_i, \tag{5}$$

where $\mathbb{1}(\textit{HighCost})_i$ equals one if loan i uses a high-cost broker firm and zero otherwise. The first stage predicts broker choice using differential distance, and the second stage estimates its effect on mortgage rate spreads. Borrower controls X_i include income, loan amount, property value, origination charges, race, sex, and deciles of credit score, combined loan-to-value ratio, and debt-to-income ratio. Origin controls W_i include neighborhood characteristics and distance to the nearest branch of either type. The fixed effects $\alpha_{o(i),t(i)}$ and $\mu_{c(i),t(i)}$ absorb county $\times$ lock-month conditions at the borrower's search origin and the mortgaged property, respectively. Controlling for distance to the nearest branch separates relative proximity from general access to brokers, while neighborhood controls account for differences in population density, income, and education. The coefficient β estimates the effect of using a high-cost rather than low-cost broker for borrowers whose choice responds to relative proximity, under the IV assumptions.

3.3 Relevance and the Exclusion Restriction

The instrument must predict broker choice and affect mortgage rates only through that choice. I first test relevance. Figure 4, Panel B, plots the probability of using a high-cost broker against differential distance, after accounting for the controls and fixed effects. As the nearest high-cost branch becomes

[13] Figure IA.7 illustrates the instrument for a borrower in Travis County, Texas.

relatively closer, borrowers are more likely to use a high-cost firm. A 10-mile increase in differential distance is associated with a 4.2–4.8 percentage point increase in high-cost broker use, compared with a sample mean of about 40%. First-stage F-statistics range from 66.3 to 77.7 (Table 3).

The exclusion restriction requires that relative proximity affect the borrower's rate only through the type of broker used. The intuition is that proximity makes a broker more convenient or familiar, but should not directly affect the mortgage rate available to an otherwise similar borrower. A concern is that high-cost brokers may locate near borrowers who would pay higher rates regardless of the broker they use. For example, these borrowers may have greater unobserved credit risk or be harder to place. High-cost brokers may also provide better services that borrowers are willing to pay for. Although the exclusion restriction cannot be tested directly, I examine balance across a broad range of observable characteristics to gauge the extent of these concerns.

Figure 5 presents these comparisons. Each point is the coefficient from a separate regression of a standardized characteristic on the standardized instrument, with county $\times$ month fixed effects but no other controls. The tests cover borrower and loan characteristics, neighborhood characteristics, and the chosen brokers' website claims as proxies for advertised service quality. Almost all differences are statistically insignificant at the 1% level, with the exception of the neighborhood share of residents with a bachelor's degree. Borrowers living relatively closer to high-cost branches tend to come from neighborhoods with a lower college-educated share, although the difference is economically small. A one-standard-deviation increase in differential distance is associated with changes of no more than 0.02 standard deviations in borrower and loan characteristics and 0.06 standard deviations in neighborhood characteristics. This includes proxies for harder-to-place borrowers, measured using matched denied HMDA applications in the same or previous year (-0.010 and -0.014 standard deviations). Differences in advertised broker services are also small and statistically insignificant at the 1% level. Although balance tests cannot directly assess unobserved characteristics, the broad balance across observables alleviates concerns that geographic sorting, borrower risk, or differences in broker services drive the results. The IV regressions also control for borrower risk and neighborhood characteristics.[14]

[14]Tables IA.12 and IA.13 and Figure IA.6 provide complementary comparisons by the sign of the instrument. Table IA.14 reports a placebo test on retail refinance borrowers, who do not use a broker, and finds that differential distance

3.4 High-Cost Brokers Raise Borrowers' Rates

I next use the variation in broker choice induced by relative proximity to estimate its effect on pricing. Table 3 presents OLS estimates in Panel A and 2SLS estimates in Panel B. Across columns, I add fixed effects for the mortgage county × month, lock date, and lender × month.

The OLS estimates range from 12.8 to 15.8 basis points. The IV estimates are larger, ranging from 23.5 to 29.1 basis points and significant at the 1% level throughout. With both origin and mortgage county × month fixed effects, the estimate is 24.0 basis points.[15] These results indicate that using a broker in the top rather than bottom tercile of past pricing raises mortgage rate spreads by 23–29 basis points for borrowers whose broker choice responds to relative proximity. The results also hold when comparing loans funded by the same lender in the same month. Thus, differences in lender identity do not fully account for the higher rates delivered by high-cost brokers.

The larger IV estimates are consistent with downward selection bias in OLS. For example, high-cost firms may screen borrowers more effectively and serve borrowers with lower unobserved credit risk, making their pricing differences appear smaller in OLS. The IV instead exploits variation in broker choice driven by relative proximity, and the balance tests show little systematic variation in borrower risk or advertised broker services with the instrument.

The increase is economically meaningful: it is about the size of the mean rate spread in the IV analysis. An illustrative 24 basis point increase in the note rate on a $330,000, 30-year mortgage at 3.38% would increase payments by about $531 per year.[16]

One alternative explanation is that high-cost brokers provide better service by closing loans faster. Borrowers facing closing deadlines may be willing to pay higher rates for faster processing. The linked

is much more weakly related to their rates than to the rates of brokered borrowers.

[15] Tables IA.7 and IA.8 explore alternative definitions of high-cost firms, such as quintile and median splits, a continuous measure of the previous-year price effect, and price effects estimated on all other years, and find qualitatively similar results. Table IA.9 explores alternative measures of relative proximity, such as log distances, the mean distance to the three nearest branches of each type, and the high-cost share of nearby branches, and shows that the results are robust across these measures.

[16] Discounted at 3.38%, the additional payments have a present value of about $9,994 if the mortgage is held for the full 30 years. This illustration treats the estimated APR-spread difference as a note-rate change; realized costs depend on fees, prepayment, and the holding period. Table IA.10 shows that high-cost broker use also raises the note rate. The estimated effects on origination charges, discount points, and total loan costs are positive but imprecise, so the higher rate does not appear to be offset by lower upfront fees.

rate-lock and mortgage records provide lock dates, closing dates, and requested lock periods. Using the same instrument, I find no statistically significant differences in either the time from lock to closing or the requested lock period between high-cost and low-cost firms (Table IA.11). These results do not support faster loan processing as an explanation for the higher rates.

4 Why Are Some Broker Firms High-Cost?

The previous section shows that using a high-cost broker raises borrowers' rates. Why do some firms deliver higher rates when brokers claim to shop across multiple lenders for lower rates? One explanation may lie in their lender choices.

Figure 6 examines how brokers advertise their services and which lenders they use. Panel A shows that expertise, finding lower rates, and access to multiple lenders are common claims on broker websites. Panel B compares the number of lenders firms advertise with the number that actually fund their loans. Among 637 firms advertising a number of lenders, the median advertised number is 25, while the median number used in the previous year is only five. Nearly 95% advertise at least 20 lenders, yet a similar share used fewer than 20, and 97% used fewer lenders than they advertised. Most firms therefore place their loans with a small subset of the lenders they advertise. This motivates a closer look at which lenders brokers regularly use and which lender they choose for each borrower. I examine whether these choices are costly, whether they persist within firms, and whether they predict the rates subsequent borrowers pay.

4.1 Measuring Lender Choice

To assess whether brokers choose expensive lenders, I compare the lender selected for each loan with alternative lenders that could fund the same borrower. The challenge is that I observe the rate only at the lender that funds the loan. Comparing lenders' average rates would not solve the problem: a lender may have high average rates because it serves riskier borrowers, yet be inexpensive for a particular borrower. I therefore estimate a pricing model for each lender and use it to predict the rate that lender would charge the same borrower. This allows me to compare the selected lender with alternatives while holding the borrower's observed characteristics fixed.

The comparison set S_i consists of lenders active in borrower i's state, together with the lenders

used by the broker there in the previous year. This provides a common benchmark for all brokers in the same state and allows comparisons with lenders outside each broker's existing network. If the comparison included only lenders a broker already uses, a firm with a small network of expensive lenders could have a small gap between its chosen lender and other lenders in its network, even when cheaper alternatives exist elsewhere in the state.[17]

The lender pricing model is

$$RateSpread_i = X_i'\gamma_\ell + \phi_{\ell,f(i)} + \psi_{\ell,s(i),y(i)} + \tau_{\ell,t(i)} + \varepsilon_i, \tag{6}$$

where $RateSpread_i$ denotes loan i's rate spread and X_i denotes its borrower and loan characteristics.[18] The subscript ℓ identifies the lender, and the coefficients γ_ℓ allow each lender to price those characteristics differently. The functions $f(i)$, $s(i)$, $y(i)$, and $t(i)$ identify the broker firm, state, year, and lock month. Lender $\times$ broker effects $\phi_{\ell,f}$ capture relationship-specific pricing; lender $\times$ state $\times$ year effects $\psi_{\ell,s,y}$ capture local pricing differences; lender $\times$ lock-month effects $\tau_{\ell,t}$ capture changes in the lender's prices over time; and ε_i denotes the residual.

The 100 lenders with the most brokered loans have separate pricing models, estimated on their retail and brokered loans; the remaining lenders share a pooled model. For each loan, I evaluate the models using the borrower's characteristics to obtain a predicted rate spread $\hat{r}_{i\ell}$ for every lender in S_i. I also predict the chosen lender's price, leaving out the focal loan to avoid using its realized rate in the prediction. Thus, all lenders are compared using predicted prices rather than mixing observed and predicted rates.

Predicting a price outside the broker's existing lender network requires accounting for relationships that have not been observed. I decompose the lender $\times$ broker effect into a broker component, a lender component, and a relationship-specific component. For every lender, I include the broker's general pricing component. I include the relationship-specific component only when the broker has

[17] The state set includes lenders that funded brokered loans for at least 10 broker firms in the state over the previous 12 months, together with the broker's own lenders in that state.

[18] The characteristics are income, credit score, loan amount, property value, combined loan-to-value ratio, debt-to-income ratio, occupancy, property type, and number of units.

an observed relationship with that lender. This allows existing broker-lender relationships to have different prices without assuming that unobserved relationships offer the same negotiated terms.

The models imply substantial price differences for the same borrower. The mean predicted spread across lenders in the state set is 41.0 basis points, compared with 5.7 basis points at the lender with the lowest predicted spread. Lenders in the broker's own network are cheaper than the state average, with a mean predicted spread of 33.9 basis points. The median lender-model R^2 is 0.49, and the out-of-sample root mean squared prediction error on brokered loans is 35.1 basis points.[19]

I first calculate the *observed gap*, the chosen lender's predicted price relative to a benchmark price in the state set:

$$ObservedGap_i^b = \hat{r}_{i\ell^*(i)} - B_i^b, \tag{7}$$

where

$$B_i^b \in \left\{ \min_{\ell \in S_i} \hat{r}_{i\ell},\ \frac{1}{|S_i|} \sum_{\ell \in S_i} \hat{r}_{i\ell} \right\}$$

is the benchmark, with $b \in \{\min, \text{mean}\}$ selecting the lowest or the mean predicted price in the state lender set S_i. Here $\ell^*(i)$ denotes the lender that funded loan i, and $\hat{r}_{i\ell}$ denotes lender ℓ's predicted spread for that loan. I use the lowest predicted price as the baseline benchmark. Under this benchmark, the observed gap measures how much more the selected lender is predicted to charge than the cheapest lender in the comparison set.

A positive observed gap does not necessarily indicate that a broker failed to find the cheapest available loan. Lenders may observe borrower characteristics that my pricing model does not capture, so their actual quotes can differ from the predicted rates. Even a diligent broker who selects the lowest quote may therefore choose a lender that does not have the lowest predicted price.

To account for this, I simulate a *diligent broker* who shops across every lender in S_i and selects the lowest quote. For each borrower-lender pair, I generate a quote by adding a random pricing error to the predicted rate. These errors are independently drawn from a mean-zero normal distribution with

[19] Figure IA.10 reports model fit, and Figure IA.11 compares lender-specific prices of borrower characteristics. Internet Appendix Section B records additional implementation details.

a standard deviation of 35 basis points, calibrated to the 35.1 basis point out-of-sample prediction error. The prediction error may include borrower risk that raises or lowers all lenders' quotes by the same amount, leaving their ranking unchanged. Assigning the full prediction-error variation to lender-specific quotes therefore allows substantial scope for unobserved differences to affect which lender offers the lowest rate.

The *simulated gap* is the observed gap a diligent broker would generate on average across the possible quotes:

$$SimulatedGap_i^b = \sum_{\ell \in S_i} p_{i\ell} \hat{r}_{i\ell} - B_i^b, \tag{8}$$

where $p_{i\ell}$ denotes the probability that a diligent broker selects lender ℓ for loan i, and B_i^b is the same benchmark as in equation (7). The simulated broker chooses a lender based on its randomly generated quote, but I calculate the resulting gap using that lender's predicted rate, just as for the actual broker. Averaging across possible quotes gives the expected gap under diligent shopping.

I subtract the simulated gap from the observed gap to obtain the *steering gap*, which measures how much the actual lender choice exceeds the gap expected from diligent shopping:

$$\begin{aligned} SteeringGap_i &= ObservedGap_i^b - SimulatedGap_i^b \\ &= \left(\hat{r}_{i\ell^*(i)} - \cancel{B_i^b}\right) - \left(\sum_{\ell \in S_i} p_{i\ell} \hat{r}_{i\ell} - \cancel{B_i^b}\right) \\ &= \hat{r}_{i\ell^*(i)} - \sum_{\ell \in S_i} p_{i\ell} \hat{r}_{i\ell}, \end{aligned} \tag{9}$$

A positive steering gap means that the chosen lender's predicted price is higher than the price expected from diligent shopping after allowing for unobserved pricing differences. The expected gap for a diligent broker is zero, although individual loans may have positive or negative gaps. I therefore use the measure to identify systematic lender-choice practices rather than classify every individual loan as steered.

The benchmark B_i^b cancels from the difference, so the steering gap is the same whether the observed and simulated gaps are measured against the minimum or the mean. The measure compares the

chosen lender directly with a probability-weighted alternative, rather than measuring how far the borrower is from a single unusually cheap prediction. Figure IA.8 illustrates the observed gap, simulated gap, and steering gap for an example loan.

The steering gap captures lender choices that remain expensive after accounting for the pricing differences allowed by the simulation. Because competing quotes and the advice given to borrowers are not observed, I use it as a proxy for steering practices. In the following sections, I examine whether these practices vary across firms, persist over time, predict borrowers' mortgage rates, and are associated with higher broker compensation.

I also test whether the findings depend on the comparison set and pricing-error assumptions. A lender active in a state need not be accessible through every broker or willing to make every loan. I therefore repeat the analysis using only lenders in the broker's existing network. I also allow for larger and lender-specific pricing errors, including a common error standard deviation of 111 basis points. The main relationships remain positive and significant under these alternatives (Table IA.15). Thus, the relationship between costly lender choices and borrower rates is not driven solely by comparisons with lenders outside the broker's network or by the baseline pricing-error assumptions.

4.2 The Steering Gap Predicts Realized Rates

If the steering gap captures costly lender choices, borrowers with larger gaps should pay higher realized rates. I test this relationship using

$$RateSpread_i = \beta\, SteeringGap_i + X_i'\gamma + \delta_{c(i),t(i)} + \zeta_{f(i)} + \varepsilon_i, \tag{10}$$

where $RateSpread_i$ denotes the realized rate spread on loan i, and $SteeringGap_i$ denotes its steering gap, both in basis points. X_i denotes borrower and loan controls.[20] The functions $c(i)$, $t(i)$, and $f(i)$ identify the county, lock month, and broker firm. County × lock-month effects $\delta_{c,t}$ absorb local pricing conditions, broker effects ζ_f absorb persistent firm characteristics, and ε_i denotes the residual. With firm effects, β compares the rates of a firm's borrowers whose loans have different steering gaps.

[20] The controls are income, credit score, loan amount, property value, combined loan-to-value ratio, debt-to-income ratio, race, ethnicity, sex, occupancy, property type, number of units, loan purpose, and loan term.

Figure 7 introduces the gaps and their relationship with realized rates. Panel A compares the observed-gap distribution with the distribution implied by diligent-broker choices. Panel B shows the steering-gap distribution in gray bars on the left axis and a binned scatter plot of realized spreads against the gap on the right axis. Each dot averages loans after accounting for borrower characteristics and local market conditions.

The observed gap exceeds the gap generated by diligent shopping after accounting for the assumed pricing errors. Their means are 27.7 and 15.8 basis points, leaving an average steering gap of 11.9 basis points. Measured against the mean predicted price in the state set, the two means are -7.7 and -19.6 basis points, so the average steering gap is again 11.9 basis points (Figure IA.12). Thus, diligent selection of the lowest quote does not explain the observed lender choices under the calibrated benchmark. This excess supports a steering interpretation. The loan-level steering gap is positive for 71.9% of loans. Panel B also shows that realized spreads increase with the gap throughout its distribution.

Table 4 presents the regression estimates.[21] The first column compares borrowers within local markets; later columns add broker firm effects, lender concentration, and lock-date effects. A one basis point higher gap is associated with a 0.79 basis point higher realized spread. Within the same broker firm, the coefficient is about 0.73 and changes little with the additional controls. A 10 basis point larger gap therefore corresponds to a roughly 7.3 basis point higher realized spread. The predicted lender-price differences are reflected in what borrowers pay.

4.3 Lender Choices Differ Across Broker Firms

The loan-level proxy can still reflect an individual pricing error or an unusual lender choice. If it captures broker practices, however, those differences should also appear when loans are aggregated within firms. I therefore calculate each broker firm's mean steering gap. A firm following the simulated diligent-shopping rule has an expected mean gap of zero; persistent positive averages indicate choices that are expensive relative to that rule.

Figure 8, Panel A, compares the 50 largest broker firms. It reports each firm's mean steering gap

[21] Internet Appendix Tables IA.16 to IA.19 repeat Tables 4 to 7 with the observed gap, before the diligent-broker adjustment, and find qualitatively similar results. Tables IA.20 to IA.22 explore alternative fixed effects and clustering, exclude loans funded by the largest lenders, and split the sample by loan purpose, lock year, credit score, and firm size. The main results are robust across these specifications.

and confidence interval, ordered from lowest to highest.

Mean steering gaps range from about zero to 40 basis points. Some firms choose lenders close to the diligent-shopping benchmark on average, while others have gaps of 15–20 basis points, and two exceed 30 basis points, even after the adjustment for pricing errors. The simulated gaps vary less across these firms (Figure IA.13), so the differences do not simply reflect different simulated allowances. These practices differ markedly across broker firms: the variation remains when lender choices are averaged across each firm's borrowers.

These lender choices are related to the firm price levels in Section 2. Panel B relates each firm's estimated effect β_f from equation (2) to its mean steering gap. Firms with larger gaps have higher price levels relative to retail lenders: across 4,414 broker firms, a 10 basis point larger mean steering gap is associated with a 6.9 basis point higher price level.

4.4 Lender Choices Persist Within Firms

A firm's average gap could reflect a temporary group of borrowers or lending conditions. If it reflects persistent lender-choice practices, the firm's past gap should predict the gap on subsequent loans. I estimate

$$SteeringGap_i = \beta\, PastSteeringGap_{f(i),t(i)} + X_i'\gamma + \kappa\, HHI_{f(i),t(i)} + \delta_{c(i),t(i)} + \zeta_{f(i)} + \varepsilon_i, \qquad (11)$$

where $SteeringGap_i$ denotes the current loan's gap, and $PastSteeringGap_{f,t}$ denotes the mean steering gap across firm f's loans in months $t-12$ through $t-1$, measured in basis points and updated each month. The focal loan and other loans in lock month t do not enter this average. $HHI_{f,t}$ denotes the firm's lender concentration over the same prior period, standardized by its sample standard deviation. The functions $f(i)$, $t(i)$, and $c(i)$ identify the firm, lock month, and county of loan i. The controls X_i, local market effects $\delta_{c,t}$, firm effects ζ_f, and residual ε_i are defined as in equation (10).

Without firm effects, the coefficient compares firms with different recent lender-choice histories. With firm effects, it asks whether a firm's current choices become more expensive when its own recent choices have been more expensive. The concentration control distinguishes the prices of the lenders

used from the extent to which the firm relies on a small number of lenders. A concentrated network can be inexpensive, and a broad network can still produce expensive choices.

Figure 9 presents two binned scatter plots. Panel A compares the focal loan's gap with its firm's past gap. Panel B compares firm averages in two successive, non-overlapping 12-month windows. Both relationships are positive and approximately linear.

The second comparison uses a firm's mean in months $t-12$ to $t-1$ and its mean in months $t-24$ to $t-13$. The two means contain different loans, so their relationship does not mechanically arise from a loan appearing on both sides of the regression. Together, the loan-level and annual-window tests distinguish a persistent firm pattern from an isolated expensive lender choice.

Table 5 reports the corresponding estimates. In Panel A, the coefficient is 0.71 across firms and 0.46 within firms. Thus, a 10 basis point increase in a firm's past gap is associated with a 4.6 basis point higher gap on its current loans, holding persistent firm characteristics fixed. In Panel B, coefficients across non-overlapping annual windows range from 0.60 to 0.66. Firms that used expensive lenders in the past tend to continue doing so.

4.5 Past Lender Choices Predict Borrowers' Rates

The persistence of the steering proxy makes it possible to characterize a firm's lender-choice practices before a new borrower arrives. I next ask whether borrowers who use firms with higher past steering gaps pay higher rates themselves. I relate the current loan's rate spread to the same measure of past firm steering practices:

$$RateSpread_i = \beta\, PastSteeringGap_{f(i),t(i)} + X_i'\gamma + \kappa\, HHI_{f(i),t(i)} + \delta_{c(i),t(i)} + \zeta_{f(i)} + \varepsilon_i, \qquad (12)$$

where $RateSpread_i$ denotes loan i's realized spread and $PastSteeringGap_{f,t}$ denotes the mean steering gap across firm f's loans in months $t-12$ through $t-1$. In this regression, the measure is standardized by its standard deviation of 18.6 basis points. The functions $f(i)$, $t(i)$, and $c(i)$ identify the broker firm, lock month, and county. $HHI_{f,t}$ denotes standardized lagged lender concentration; X_i denotes borrower and loan controls; $\delta_{c,t}$ and ζ_f denote local market and firm effects; and ε_i denotes the residual. The

coefficient β measures the rate difference associated with a one-standard-deviation higher past gap.

This specification uses the broker's history rather than the lender choice on the current loan. The historical measure can be constructed before the focal loan is locked, without observing that loan's price. Firm effects then compare borrowers at the same firm during periods with different recent lender-choice histories. The test asks whether the persistence documented above has a corresponding pattern in subsequent borrowing costs.

Figure 10 presents this within-firm relationship as a binned scatter plot. The horizontal axis is the firm's past steering gap and the vertical axis the realized rate spread, adjusted for controls and fixed effects. Rates rise with the past gap. Table 6 reports the estimates with and without firm effects and lender concentration controls.

A one-standard-deviation higher past steering gap is associated with a 10.0 basis point higher rate spread across firms. Within the same firm, the estimate is 7.1 basis points, or 6.8 basis points with lock-date effects. Borrowers therefore pay more when they use a firm with a recent history of larger steering gaps, even after accounting for borrower characteristics, local market conditions, and persistent firm differences. Because the proxy uses earlier loans, this relationship extends beyond the lender choice and price on the focal loan. The firm's past lender-choice practices predict the borrowing costs of its subsequent customers.

4.6 Expensive Lenders or Poor Borrower Fit?

A lender can be expensive for a borrower for two different reasons. It may be expensive in general, charging more than other lenders for most borrowers or for most of the broker's borrowers. Or it may be inexpensive in general but price this borrower's characteristics less favorably than other lenders do. A low credit score, for example, may raise the rate more at one lender than at another. The two explanations describe different broker behavior. In the first, the broker uses lenders that are expensive in general. In the second, the broker uses lenders that can be competitive but does not place each borrower with the lender that prices that borrower best.

The pricing model in equation (6) distinguishes the two. As noted above, the lender $\times$ broker effect is the sum of a broker component, a lender component, and a relationship-specific component,

$\phi_{\ell,f} = \phi^B_f + \phi^L_\ell + \phi^R_{\ell,f}$. Let X^{ref} denote a reference borrower with the average characteristics of brokered loans. Each predicted price then splits into four components:

$$\hat{r}_{i\ell} = \underbrace{X^{\text{ref}\prime}\hat{\gamma}_\ell + \hat{\phi}^L_\ell + \hat{\psi}_{\ell,s(i),y(i)} + \hat{\tau}_{\ell,t(i)}}_{\text{lender level } \hat{r}^{\text{Lev}}_{i\ell}} + \underbrace{\hat{\phi}^R_{\ell,f(i)}}_{\text{relationship } \hat{r}^{\text{Rel}}_{i\ell}} + \underbrace{(X_i - X^{\text{ref}})'\hat{\gamma}_\ell}_{\text{borrower fit } \hat{r}^{\text{Fit}}_{i\ell}} + \underbrace{\hat{\phi}^B_{f(i)}}_{\text{broker}}. \tag{13}$$

The lender level $\hat{r}^{\text{Lev}}_{i\ell}$ is lender ℓ's price for the reference borrower in the loan's state, year, and lock month. The relationship component $\hat{r}^{\text{Rel}}_{i\ell}$ is the part of the lender's price specific to this broker, and it is zero for lenders outside the broker's observed menu. The borrower-fit component $\hat{r}^{\text{Fit}}_{i\ell}$ is the price difference that the borrower's own characteristics make at lender ℓ, relative to the reference borrower. The broker component is the same for every lender. The lender level and the relationship component therefore measure how expensive a lender is in general, for any borrower or for any borrower of this broker, while the borrower-fit component measures how the lender prices this borrower's characteristics.

Substituting equation (13) into the steering gap in equation (9) splits the gap into the same parts:

$$\begin{aligned} \textit{SteeringGap}_i = &\underbrace{\left(\hat{r}^{\text{Lev}}_{i\ell^*(i)} - \sum_{\ell\in S_i} p_{i\ell}\,\hat{r}^{\text{Lev}}_{i\ell}\right)}_{\text{lender level}} + \underbrace{\left(\hat{r}^{\text{Rel}}_{i\ell^*(i)} - \sum_{\ell\in S_i} p_{i\ell}\,\hat{r}^{\text{Rel}}_{i\ell}\right)}_{\text{relationship}} \\ &+ \underbrace{\left(\hat{r}^{\text{Fit}}_{i\ell^*(i)} - \sum_{\ell\in S_i} p_{i\ell}\,\hat{r}^{\text{Fit}}_{i\ell}\right)}_{\text{borrower fit}}. \end{aligned} \tag{14}$$

Each part compares the chosen lender's component with the component that the diligent broker is expected to select. Because the selection probabilities $p_{i\ell}$ sum to one, the broker component cancels. The probabilities depend on the full predicted prices, so the decomposition is an accounting split at the benchmark's probabilities rather than a set of separate counterfactuals.

On the average loan, most of the steering gap reflects lenders that are expensive in general. Lender levels account for 62.1% of the 11.9 basis point average gap and relationship components for 16.3%. Borrower fit accounts for the remaining 21.6%, so poor fit is part of the gap but the smaller part.

Which of these components helps explain why borrowers at firms with higher past gaps pay more? To examine this, I split the past gap in equation (12) into its three parts, each a trailing 12-month firm average divided by the standard deviation of the past gap. I regress the rate spread on the three parts together, with the controls and fixed effects of Table 6, Column (4). A part's contribution is its coefficient multiplied by its share of the past gap, the coefficient from regressing the part on the past gap with the same controls and fixed effects. The contributions add up to the coefficient on the past gap.

Figure 11 reports the contributions. Of the 7.1 basis point within-firm estimate, lender levels account for 58.5%, relationship components for 38.6%, and borrower fit for 2.9%. The observed gap against the lowest predicted spread splits similarly, with borrower fit accounting for 5.3% of its estimate. The relationship between past steering gaps and subsequent rates is concentrated in firms' use of lenders that are expensive in general; differences in how well lenders price particular borrowers account for little of that relationship. Borrower fit accounts for about a fifth of the average gap but for almost none of its relationship with subsequent rates.

4.7 Lender Choice or Menu Choice?

A broker's menu is the set of lenders it worked with in the borrower's state during the previous year. An expensive choice can arise within this menu or from the menu itself. The broker may choose a lender with a higher predicted price than others on its menu. Or it may choose well among its own lenders while the menu as a whole is expensive relative to the other lenders active in the state. This distinction matters because access to more lenders would not by itself improve choices among existing partners.

To separate the two, I compare three predicted rates for the same loan. The first is the rate at the chosen lender, $\hat{r}_{i\ell^*(i)}$. The second is the expected predicted price of the lender selected by the diligent broker in equation (8), which shops the full state set, $\bar{r}_i^S = \sum_{\ell \in S_i} p_{i\ell} \hat{r}_{i\ell}$. The third is the corresponding expected predicted price when the diligent broker shops only the broker's menu M_i, $\bar{r}_i^M = \sum_{\ell \in M_i} p_{i\ell}^M \hat{r}_{i\ell}$, where $p_{i\ell}^M$ denotes the probability that lender ℓ quotes the lowest price among the menu lenders, with the same pricing-error distribution as in the state-set benchmark. The chosen

lender's rate can then be written as

$$\hat{r}_{i\ell^*(i)} = \underbrace{\bar{r}_i^S}_{\text{state-set benchmark}} + \underbrace{\left(\bar{r}_i^M - \bar{r}_i^S\right)}_{\text{menu choice}} + \underbrace{\left(\hat{r}_{i\ell^*(i)} - \bar{r}_i^M\right)}_{\text{lender choice}}. \tag{15}$$

The first term is the expected predicted price of the lender selected by a diligent broker shopping the state set. The second measures how much higher that expected predicted price is when shopping is restricted to the broker's menu. The third compares the chosen lender's predicted price with the diligent-shopping benchmark for the same menu. All three terms use predicted prices; the pricing errors affect which lender the diligent broker selects.

Subtracting the state-set benchmark from both sides splits the steering gap into the two parts:

$$\textit{SteeringGap}_i = \hat{r}_{i\ell^*(i)} - \bar{r}_i^S = \underbrace{\left(\hat{r}_{i\ell^*(i)} - \bar{r}_i^M\right)}_{\text{lender choice}} + \underbrace{\left(\bar{r}_i^M - \bar{r}_i^S\right)}_{\text{menu choice}}. \tag{16}$$

The observed gap splits in the same way, with the lowest predicted spreads on the menu and in the state set in place of the two diligent brokers' expected predicted prices: the chosen lender's predicted spread minus the lowest on the menu, plus the lowest on the menu minus the lowest in the state set. Figure IA.9 illustrates both splits for one loan.

On the average loan, lender choice accounts for 8.8 basis points of the 11.9 basis point steering gap, or 73.5%, and menu choice for 3.2 basis points, or 26.5%. Menu choice is a larger part of the observed gap, 12.3 of 27.7 basis points. The lowest predicted spread in the larger state set is low partly because it is the minimum over more noisy predictions; the diligent-broker comparison accounts for this.

Using the method of Section 4.6, I split the within-firm relationship between the past gap and borrowers' rates (Table 6, Column (4)) into the two parts. Figure 11, Panel B, reports the contributions. Of the 7.1 basis point estimate, lender choice accounts for 78.9% and menu choice for 21.1%. The observed gap splits almost identically, 80.3% and 19.7%. Per basis point, the menu part predicts higher rates than the lender-choice part, but it varies less across firms and over time, accounting for 15% of the variation in the past gap. The higher rates paid by subsequent borrowers at firms with larger past gaps

are therefore mainly associated with expensive choices from the firms' existing menus; expensive menus account for about a fifth of the relationship. Table IA.15, Column (5), reports the main estimates with the lender-choice part alone as the gap.

5 Why Do Some Broker Firms Use Expensive Lenders?

Why would a broker repeatedly choose lenders that are expensive for its borrowers? Compensation provides a possible explanation. A broker's compensation agreements can differ across lenders, so a lender that charges the borrower more may also pay the broker more. I test this explanation by examining whether the costly lender choices captured by the steering gap are associated with higher lender-paid compensation, and then compare the broker's gain with the borrower's cost in dollars.

5.1 Costly Lender Choices Pay Brokers More

This subsection asks whether brokers earn more on loans placed with lenders that are expensive for the borrower. The relevant comparison is borrower-specific. A lender that pays high compensation need not be expensive for every borrower because lenders price credit and loan characteristics differently. Relating compensation to the steering gap accounts for this distinction: the gap assesses the selected lender against alternatives for the same loan. Firm effects further separate the relationship from a tendency for some broker firms to earn more on all their loans.

Lenders pay the broker on about 63% of brokered loans (Section 1.2). For lender-paid loans, I estimate:

$$Comp_i = \beta\, SteeringGap_i + X_i'\gamma + \delta_{c(i),t(i)} + \zeta_{f(i)} + \varepsilon_i, \tag{17}$$

where $Comp_i$ denotes the compensation paid by the lender to the broker on loan i, in basis points of the loan amount, and $SteeringGap_i$ denotes the loan's gap in basis points of rate spread. X_i denotes borrower and loan controls. The functions $c(i)$, $t(i)$, and $f(i)$ identify the county, lock month, and broker firm. The fixed effects $\delta_{c,t}$ and ζ_f absorb local market conditions and persistent firm compensation levels, respectively, and ε_i denotes the residual. The coefficient β measures whether a firm receives more compensation on loans placed with relatively expensive lenders.

Table 7 reports the estimates. Column (1) includes borrower and loan controls without fixed ef-

fects, Column (2) adds county × lock-month and broker firm fixed effects, and Column (3) adds lock-date fixed effects. Compensation increases with the steering gap. Without fixed effects, a one basis point higher gap is associated with 0.28 basis points more compensation. Within firms, the coefficient is 0.11 and remains unchanged when lock-date effects are added. A 10 basis point larger steering gap is therefore associated with about 1.1 basis points more compensation, relative to a mean of 205.4 basis points. Because the estimate holds within the same firm, it does not reflect differences in average compensation across brokers.[22]

5.2 How Much Do Borrowers Pay for the Broker's Gain?

The previous subsection shows that brokers earn more on costly lender choices. This subsection compares that gain with the borrower's cost. Compensation is measured in basis points of the loan amount, whereas the steering gap and the rate spread are measured in basis points of the interest rate. Comparing the broker's gain with the borrower's cost therefore requires expressing both in dollars. Figure 12 does so by decile of the steering gap. I re-estimate the specification in Column (2) of Table 7 with indicators for the ten deciles in place of the linear gap, once with compensation and once with the rate spread as the dependent variable. Each dot plots a decile's adjusted mean rate spread against its adjusted mean compensation, in basis points on the inner axes. The steering gap averages -11.7 basis points in the lowest decile and 50.8 basis points in the highest. From the lowest to the highest decile, adjusted compensation rises from 200.7 to 209.5 basis points of the loan amount, and the adjusted rate spread rises from 18.6 to 46.5 basis points. Both increase across deciles. Compensation rises most from the first to the second decile and from the ninth to the tenth, and the rate spread rises most in the top decile.

The outer axes convert these basis points to dollars for a typical loan, using the sample medians in Table 1: a loan amount of $330,000 and a note rate of 3.38%. Compensation is a one-time payment at origination, so its dollar value is its share of the loan amount, and each basis point is worth $33. A higher rate spread instead raises the borrower's payment every month. I treat a difference in the rate spread as a difference in the note rate of a 30-year fixed-rate mortgage and assume that the borrower

[22] Using the differential-distance instrument of Section 3, Table IA.10 also shows that brokers receive more lender-paid compensation when borrowers use high-cost firms.

holds the mortgage for 11 years, a common horizon before a mortgage is refinanced or the home is sold, and then repays the remaining balance. The borrower's cost has two parts: the additional monthly payments over the 11 years, and the larger balance repaid at the end, because a higher rate also slows amortization. Both are discounted at the 3.38% note rate.

From the lowest to the highest steering-gap decile, broker compensation rises by 8.9 basis points, or about $290 per loan. Over the same deciles, the rate spread rises by 27.9 basis points, which raises the monthly payment on a $330,000 mortgage by about $52. Over 11 years, the additional payments have a present value of about $5,670, and the balance repaid at the end is about $2,720 larger, with a present value of about $1,880. The borrower's present-value borrowing costs therefore rise by about $7,550, roughly 25 times the broker's additional compensation. The two amounts are not a transfer from the borrower to the broker: the broker receives a one-time payment at origination, while the higher rate raises the borrower's payments to the lender for as long as the loan is outstanding. Lender choices that earn the broker a few hundred dollars more thus cost the borrower several thousand dollars more.

6 Conclusion

Mortgage brokers promise to shop across lenders for lower rates, yet mortgage rates remain widely dispersed even when borrowers use these market experts. Using novel data linking borrowers, brokers, and lenders, I study whether differences across broker firms help explain this dispersion. I document large and persistent differences in broker pricing. An IV exploiting borrowers' relative proximity to broker branches estimates that using a broker in the top rather than bottom tercile of past pricing raises rate spreads by 23–29 basis points.

I then examine brokers' lender choices to understand why some firms are high-cost. Comparing predicted lender prices for the same borrower and adjusting for unobserved pricing differences using a simulated diligent broker, I find that some firms systematically choose relatively expensive lenders. These choices are associated with higher realized mortgage rates and higher lender-paid broker compensation, consistent with steering. Steering practices differ across firms and persist over time, and borrowers using firms with stronger past steering practices pay higher rates.

These findings have three implications. First, choosing a broker matters as well as choosing a

lender. Broker identity explains more of the remaining variation in mortgage rates than lender identity, and brokers' past lender choices predict the rates their future borrowers pay. Second, relatively small financial incentives for brokers can be associated with substantial costs to borrowers. Comparing loans in the highest and lowest steering-gap deciles, brokers receive approximately $290 more in lender-paid compensation on a $330,000 mortgage, while borrowers pay about $7,500 more in present-value borrowing costs over an 11-year holding period, roughly 25 times the additional broker compensation. Third, expanding brokers' access to lenders alone may not ensure lower rates, as most costly lender choices occur within their existing networks. Current regulation also leaves room for compensation differences across lenders and discretion over which lenders brokers compare. Greater transparency about lender-paid compensation and brokers' past lender choices could help borrowers compare firms and regulators identify costly steering practices.

Overall, the findings show how conflicts of interest can raise household borrowing costs and undermine the integrity of financial intermediation, with broader implications for access to affordable credit.

References

Agarwal, Sumit, Brent W. Ambrose, and Vincent W. Yao, 2020, Lender Steering in Residential Mortgage Markets, *Real Estate Economics* 48, 446–475.

Agarwal, Sumit, Gene Amromin, Itzhak Ben-David, and Douglas D. Evanoff, 2025, Predatory Lending and Hidden Risks, *Journal of Financial and Quantitative Analysis* 60, 2526–2554.

Agarwal, Sumit, Swee Hoon Ang, Yongheng Deng, and Yonglin Wang, 2021, Mortgage Brokers and the Effectiveness of Regulatory Oversights, *Management Science* 67, 5278–5300.

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

Allen, Jason, Robert Clark, and Jean-François Houde, 2014, Price Dispersion in Mortgage Markets, *Journal of Industrial Economics* 62, 377–416.

Allen, Jason, Robert Clark, Jean-François Houde, Shaoteng Li, and Anna Trubnikova, 2025, The Role of Intermediaries in Selection Markets: Evidence from Mortgage Lending, *Review of Financial Studies* 38, 3284–3328.

Ambrose, Brent W., and James N. Conklin, 2014, Mortgage Brokers, Origination Fees, Price Transparency and Competition, *Real Estate Economics* 42, 363–421.

Ambrose, Brent W., James N. Conklin, and Luis A. Lopez, 2021, Does Borrower and Broker Race Affect the Cost of Mortgage Credit?, *Review of Financial Studies* 34, 790–826.

Anagol, Santosh, Shawn Cole, and Shayak Sarkar, 2017, Understanding the Advice of Commissions-Motivated Agents: Evidence from the Indian Life Insurance Market, *Review of Economics and Statistics* 99, 1–15.

Argyle, Bronson, Taylor Nadauld, and Christopher Palmer, 2023, Real Effects of Search Frictions in Consumer Credit Markets, *Review of Financial Studies* 36, 2685–2720.

Barry, Jordan M., Will Fried, and John William Hatfield, 2025, Et Tu, Agent? Commission-Based Steering in Residential Real Estate, *Iowa Law Review* 110, 1473.

Barwick, Panle Jia, Parag A. Pathak, and Maisy Wong, 2017, Conflicts of Interest and Steering in Residential Brokerage, *American Economic Journal: Applied Economics* 9, 191–222.

Benetton, Matteo, Greg Buchak, and Claudia Robles-Garcia, 2022, Wide or Narrow? Competition and Scope in Financial Intermediation, Working Paper.

Bergstresser, Daniel, John M. R. Chalmers, and Peter Tufano, 2009, Assessing the Costs and Benefits of Brokers in the Mutual Fund Industry, *Review of Financial Studies* 22, 4129–4156.

Berndt, Antje, Burton Hollifield, and Patrik Sandås, 2010, The Role of Mortgage Brokers in the Subprime Crisis, Working Paper.

Bhutta, Neil, Andreas Fuster, and Aurel Hizmo, 2026, Paying Too Much? Borrower Sophistication and Overpayment in the U.S. Mortgage Market, *Journal of Finance* 81, 49–90.

Bosshardt, Joshua, Ali Kakhbod, and Amir Kermani, 2025, Do Intermediaries Improve GSE Lending? Evidence from Proprietary GSE Data, *Journal of Financial Economics* 170, 104082.

Burdett, Kenneth, and Kenneth L. Judd, 1983, Equilibrium Price Dispersion, *Econometrica* 51, 955–969.

Christoffersen, Susan E. K., Richard Evans, and David K. Musto, 2013, What Do Consumers' Fund Flows Maximize? Evidence from Their Brokers' Incentives, *Journal of Finance* 68, 201–235.

Egan, Mark, 2019, Brokers versus Retail Investors: Conflicting Interests and Dominated Products, *Journal of Finance* 74, 1217–1260.

Egan, Mark, Shan Ge, and Johnny Tang, 2022, Conflicting Interests and the Effect of Fiduciary Duty: Evidence from Variable Annuities, *Review of Financial Studies* 35, 5334–5386.

Green, Richard C., Burton Hollifield, and Norman Schürhoff, 2007, Financial Intermediation and the Costs of Trading in an Opaque Market, *Review of Financial Studies* 20, 275–314.

Griffin, John M., Nicholas Hirschey, and Samuel Kruger, 2023, Do Municipal Bond Dealers Give Their Customers "Fair and Reasonable" Pricing?, *Journal of Finance* 78, 887–934.

Griffin, John M., Nicholas Hirschey, and Samuel Kruger, 2026, What Drives Large Corporate Bond Trading Costs?, Working Paper.

Guiso, Luigi, Andrea Pozzi, Anton Tsoy, Leonardo Gambacorta, and Paolo Emilio Mistrulli, 2022, The Cost of Steering in Financial Markets: Evidence from the Mortgage Market, *Journal of Financial Economics* 143, 1209–1226.

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

Hatfield, John William, Scott Duke Kominers, and Richard Lowery, 2025, Collusion in Brokered Markets, *Journal of Finance* 80, 1417–1462.

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

Huang, Ruidi, Erik J. Mayer, Sheridan Titman, and David Xu, 2026, Human Capital and Local Credit Supply: Evidence from the Mortgage Industry, Working Paper.

Inderst, Roman, and Marco Ottaviani, 2012, Competition through Commissions and Kickbacks, *American Economic Review* 102, 780–809.

Jorgensen, Rebecca A., 2025, The Economic Consequences of Mergers Between Real Estate Agencies and Mortgage Lenders, Working Paper.

LaCour-Little, Michael, 2009, The Pricing of Mortgages by Brokers: An Agency Problem?, *Journal of Real Estate Research* 31, 235–264.

Li, Anran, Tong Liu, Anthony LoSasso, and Nicholas Tilipman, 2026, Going for Broker? Intermediation in Health Insurance Markets, Working Paper.

Lopez, Luis Arturo, Shawn J. McCoy, and Vivek Sah, 2022, Steering Consumers to Lenders in Residential Real Estate Markets, *Real Estate Economics* 50, 1596–1641.

Paravisini, Daniel, Veronica Rappoport, and Philipp Schnabl, 2023, Specialization in Bank Lending: Evidence from Exporting Firms, *Journal of Finance* 78, 2049–2085.

Priest, Alex, 2023, Does Credit Supply Affect Real Estate Cap Rates?, Working Paper.

Robles-Garcia, Claudia, 2020, Competition and Incentives in Mortgage Markets: The Role of Brokers, Working Paper.

Spader, Jonathan S., and Roberto G. Quercia, 2011, Mortgage Brokers and the Refinancing Transaction: Evidence from CRA Borrowers, *Journal of Real Estate Finance and Economics* 42, 181–210.

Stigler, George J., 1961, The Economics of Information, *Journal of Political Economy* 69, 213–225.

Woodward, Susan E., and Robert E. Hall, 2010, Consumer Confusion in the Mortgage Market: Evidence of Less than a Perfectly Transparent and Competitive Market, *American Economic Review* 100, 511–515.

Woodward, Susan E., and Robert E. Hall, 2012, Diagnosing Consumer Confusion and Sub-Optimal Shopping Effort: Theory and Mortgage-Market Evidence, *American Economic Review* 102, 3249–3276.

Zhang, Dayin, Panle Jia Barwick, Lu Han, and Jonathan Kroah, 2026, Market Power in Mortgage Pricing: The Role of Referral Lending, Working Paper.

Figure 1: Residual Rate Dispersion in Retail and Brokered Loans

This figure compares mortgage rate dispersion between retail and brokered loans after accounting for borrower characteristics and local market conditions. The horizontal axis shows residual rate spreads in basis points, and the vertical axis shows density. Gray bars represent retail loans, and the red line represents brokered loans. Residuals come from a pooled regression of rate spreads on a brokered indicator, income, credit score, loan amount, property value, combined loan-to-value ratio, debt-to-income ratio, origination charges, race, sex, and county × lock-date fixed effects. The legend reports residual standard deviations adjusted for absorbed degrees of freedom. The sample consists of conventional conforming fixed-rate mortgages matched across Optimal Blue, CoreLogic, and HMDA from 2018–2025. The displayed range is restricted to the pooled 0.5th–99.5th percentiles.

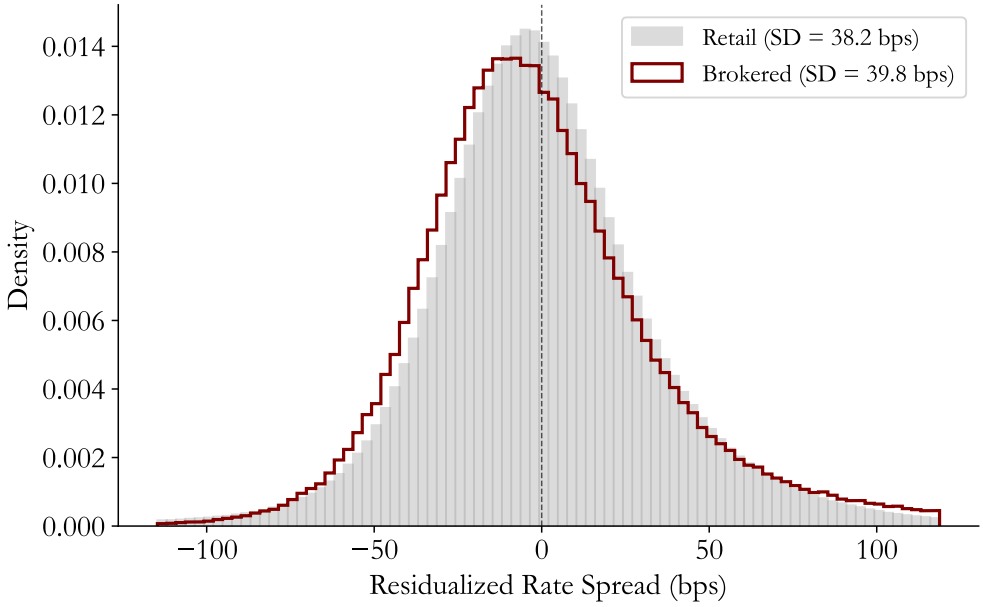


Figure 2: Broker Firm Price Levels Relative to Retail

This figure compares broker firm pricing relative to retail loans after controlling for borrower characteristics and local market conditions. The horizontal axis shows the 170 largest broker firms, ordered by their estimated effects, and the vertical axis shows rate differences relative to retail, in basis points. Dots represent firm effects, with vertical bars showing confidence intervals. Blue and red indicate firms significantly below and above retail, respectively; white indicates insignificant differences. The dashed line marks the retail benchmark. Confidence intervals and significance use a Bonferroni correction for 170 comparisons. Firm effects are estimated from a regression of rate spreads on broker firm indicators, income, credit score, loan amount, property value, combined loan-to-value ratio, debt-to-income ratio, origination charges, race, sex, and county × lock-date fixed effects, with retail loans omitted. Rate spread is APR minus APOR. The sample includes matched Optimal Blue, CoreLogic, and HMDA loans from 2018–2025 in county-days with brokered loans. Standard errors are clustered by broker firm (lender for retail loans), county, and lock date.

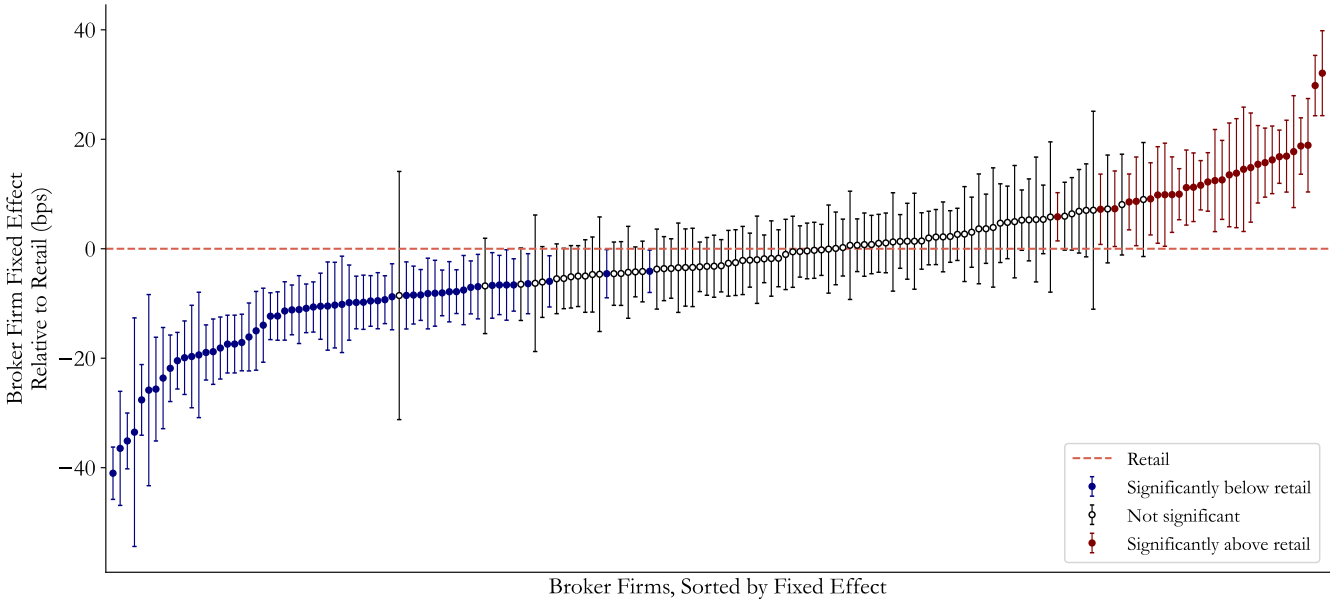


Figure 3: Broker Firm Price Levels Across Periods and Loan Types

This figure examines whether broker firm pricing persists across periods, loan products, and borrower groups. Each panel plots firm price levels in one group against those in another, with the horizontal and vertical axes showing the respective estimated firm effects. Dots represent averages across 20 bins, with a linear fit and 95% confidence band. The six comparisons are 2018–2021 versus 2022–2025, purchase versus refinance, conventional versus government-backed, credit scores at least 740 versus below 740, 30-year versus 10–25-year terms, and combined loan-to-value ratios at most 80% versus above 80%. Firm effects are estimated separately within each group from regressions of rate spreads on broker firm indicators, income, credit score, loan amount, property value, combined loan-to-value ratio, debt-to-income ratio, origination charges, race, sex, and county × lock-date fixed effects, with retail loans in each group as the benchmark. Firms must have at least 10 brokered loans in each group and are weighted by the harmonic mean of their loan counts. Effects are in basis points. The sample is matched Optimal Blue, CoreLogic, and HMDA loans, 2018–2025. Standard errors are clustered by broker firm.

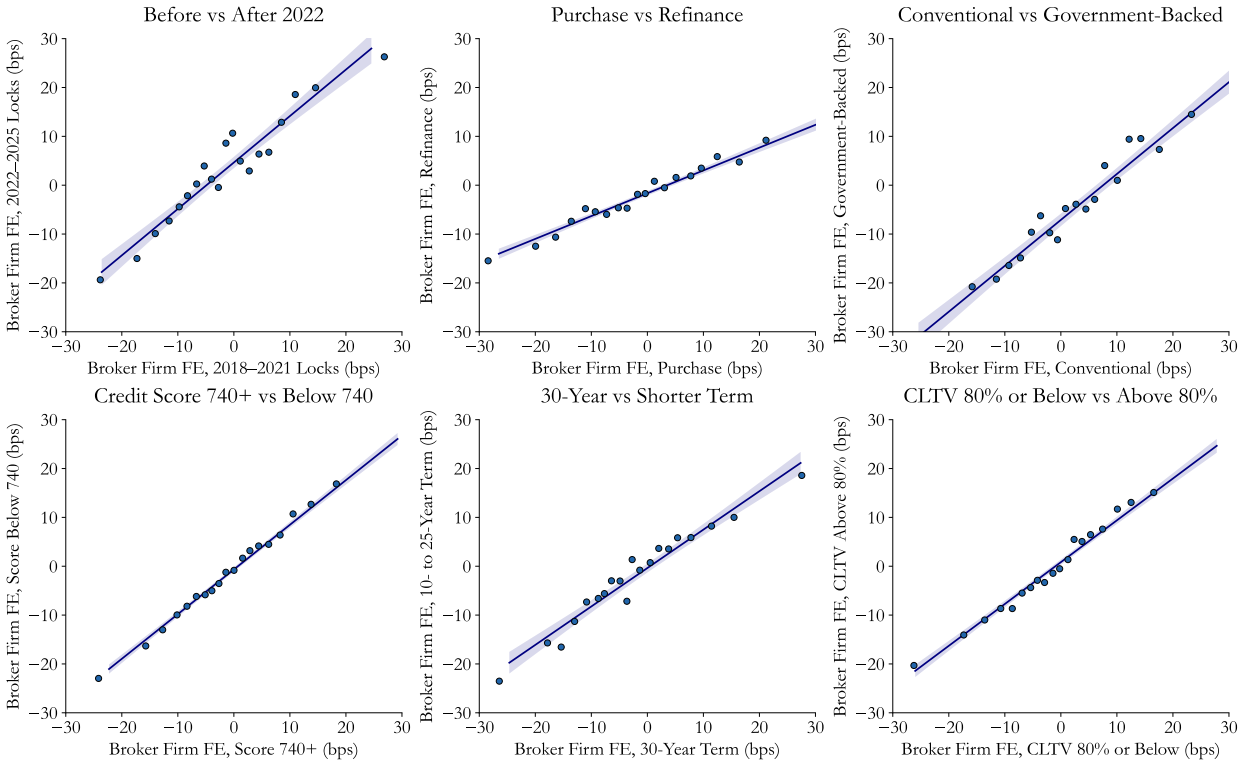
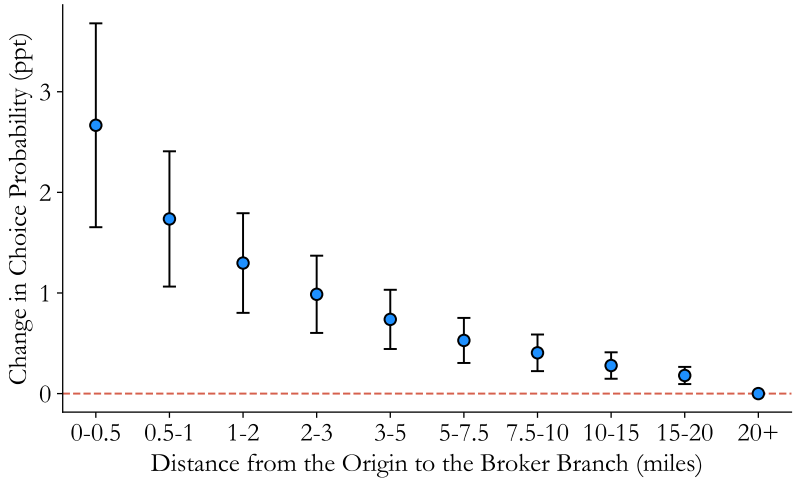


Figure 4: Proximity and Broker Choice

This figure examines how proximity to broker branches affects branch choice and the likelihood of using a high-cost broker. Panel A reports a linear probability regression of branch choice on distance-bin indicators, with loan and branch × lock-month fixed effects. Each observation pairs a loan with a potential branch. Dots show coefficients in percentage points relative to branches at least 20 miles away. Alternatives include the chosen branch, up to 100 branches within 20 miles, and 30 other branches, sampled and weighted by inverse selection probabilities. Panel B presents the first stage of the differential-distance IV. The horizontal axis measures distance to the nearest low-cost branch minus distance to the nearest high-cost branch, and the vertical axis shows high-cost broker use. High- and low-cost firms are classified by previous-year price effects in the top and bottom terciles across branches in the lock month. Dots represent 20 equal-count bins, adjusted for income, loan amount, property value, origination charges, race, sex, credit-score, combined loan-to-value, and debt-to-income deciles, origin characteristics, nearest-branch distance, and origin and mortgage county × lock-month effects. In both panels, vertical bars show 95% confidence intervals, and the choice set contains branches active in the previous year whose firms are licensed in the property's state. The borrower's search origin is the previous address for purchases (from Verisk and DataAxle) and the mortgaged property for refinances. The data link Optimal Blue, CoreLogic, HMDA, and NMLS. Panel B uses the 2019–2025 IV sample in Column (2) of Table 3. Standard errors are clustered by origin county in Panel A and mortgage county in Panel B.

Panel A: Distance and the Choice of Broker Branch



Panel B: Differential Distance and the Use of a High-Cost Broker

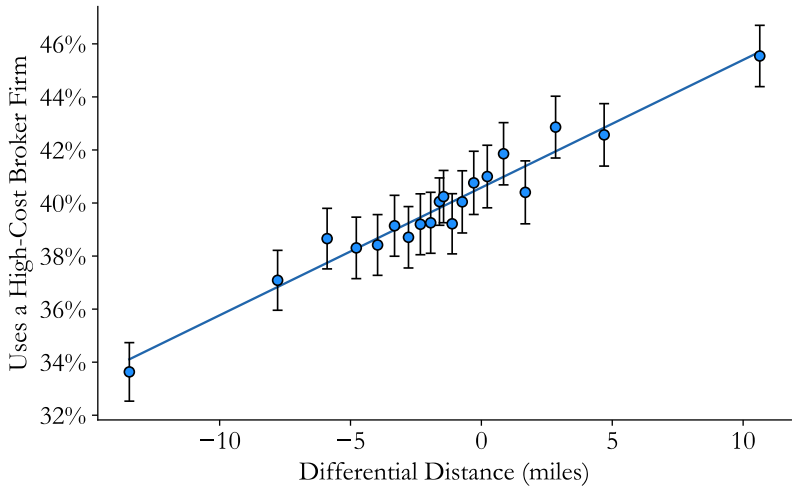


Figure 5: Balance Tests for the Differential Distance Instrument

This figure tests whether differential distance is associated with borrower and neighborhood characteristics or the services advertised by the chosen broker. Differential distance is the distance to the nearest low-cost branch minus the distance to the nearest high-cost branch, with firms classified by the top and bottom terciles of previous-year price effects. Each dot shows the coefficient from a separate regression of a characteristic on differential distance, with mortgage county × lock-month fixed effects and no other controls. Both variables are standardized, so coefficients measure standard-deviation changes in characteristics for a one-standard-deviation (19.9-mile) increase in differential distance. Black and slate-gray bars show 95% and 99% confidence intervals, respectively. Characteristics include borrower and loan variables, origin-neighborhood characteristics from the American Community Survey, and the chosen broker's website claims. Linked HMDA denials identify applications denied in the same or previous year, matched on census tract, income, age, sex, race, ethnicity, occupancy, loan purpose, and property value within 10%. Website claims are indicators, except advertised lender count, measured as log one plus the stated number. Collected in 2026, these claims describe the selected broker rather than predetermined borrower characteristics. The sample includes owner-occupied brokered loans matched across Optimal Blue, CoreLogic, and HMDA, with traced search origins and chosen firms in the top or bottom tercile, 2019–2025. Standard errors are clustered by mortgage county.

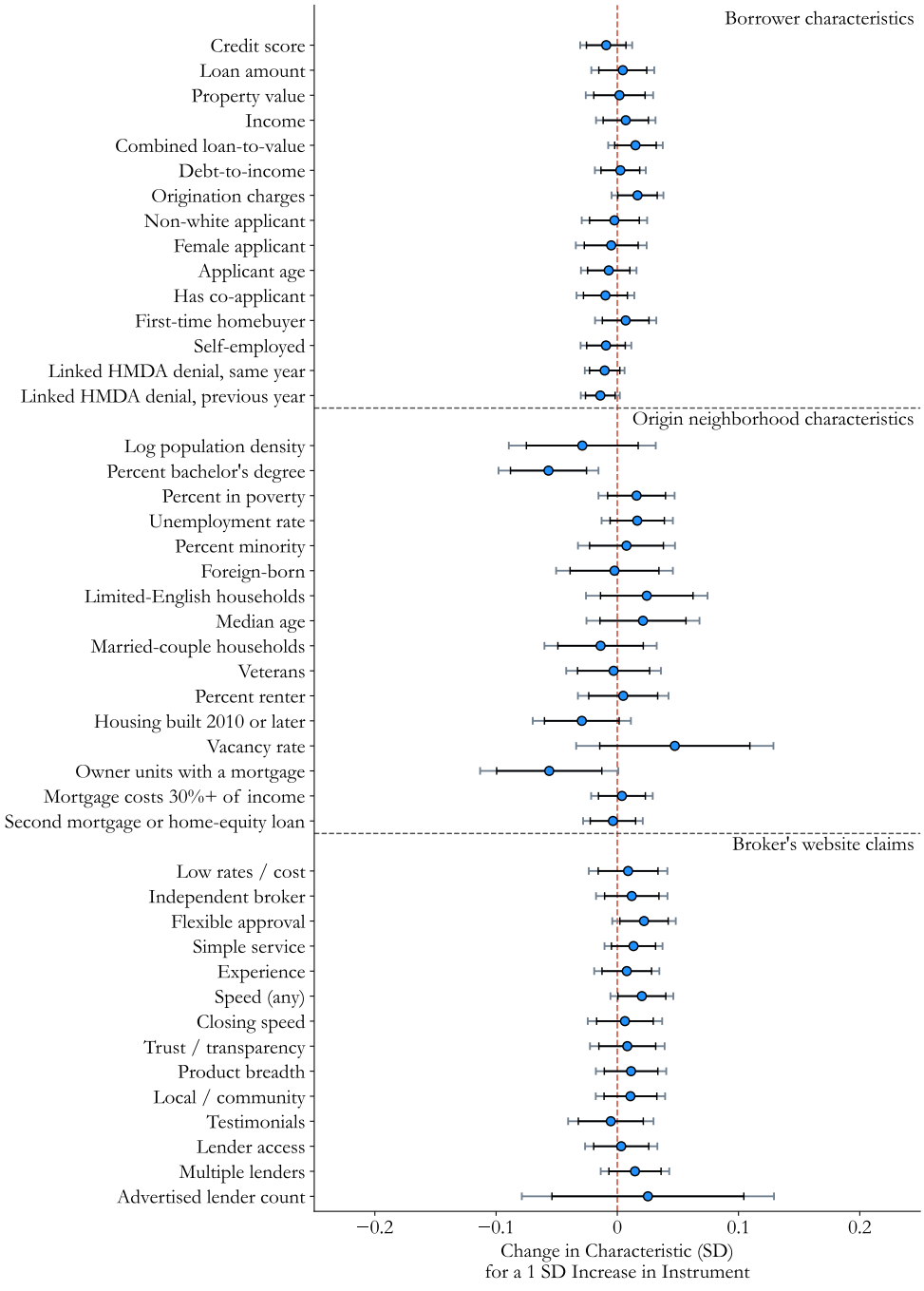
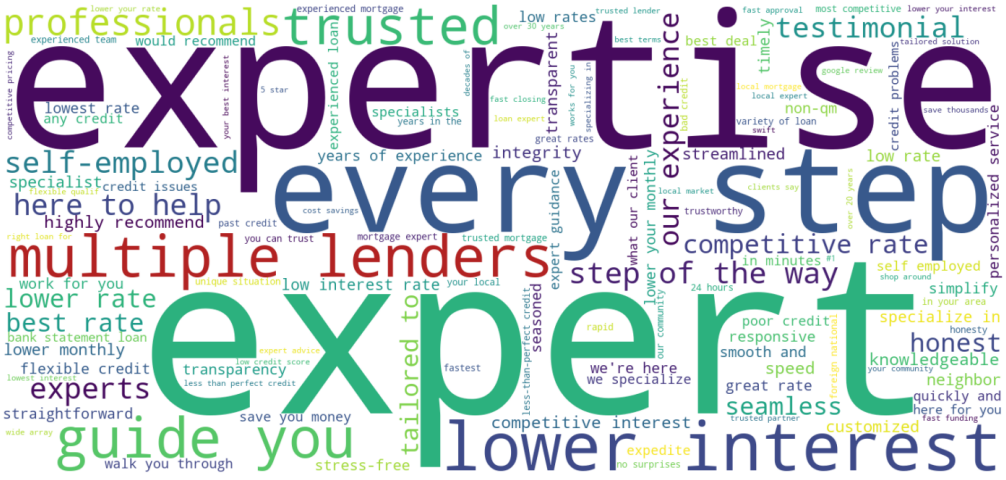


Figure 6: Broker Advertising and Lender Use

This figure compares mortgage brokers' advertised services and lender access with their actual lender use. Panel A shows common phrases on broker websites; phrase size is proportional to the number of firms using each phrase. Panel B compares the number of lenders advertised on each firm's website (horizontal axis) with the number of distinct lenders funding its brokered CoreLogic mortgages in 2025 (vertical axis). Bubble size is proportional to the firm's number of brokered loans in 2025. The gray dashed line marks equal advertised and actual counts, with the share of firms below the line labeled. The dark red dashed lines mark 20 lenders; the braces show the share of firms advertising at least 20 lenders (top) and the share using fewer than 20 (right). The marginal histograms show the share of firms (filled bars) and their loans (hollow bars) in bins of five lenders. Axes are truncated at 100 lenders, but the histograms and reported shares include all firms. Websites were collected from NMLS records and supplementary company-name searches in June and September 2026. Panel B includes 637 firms reporting a lender count on their websites and having at least one brokered CoreLogic mortgage in 2025.

Panel A: Phrases on Mortgage Broker Websites



Panel B: Advertised and Actual Number of Lenders

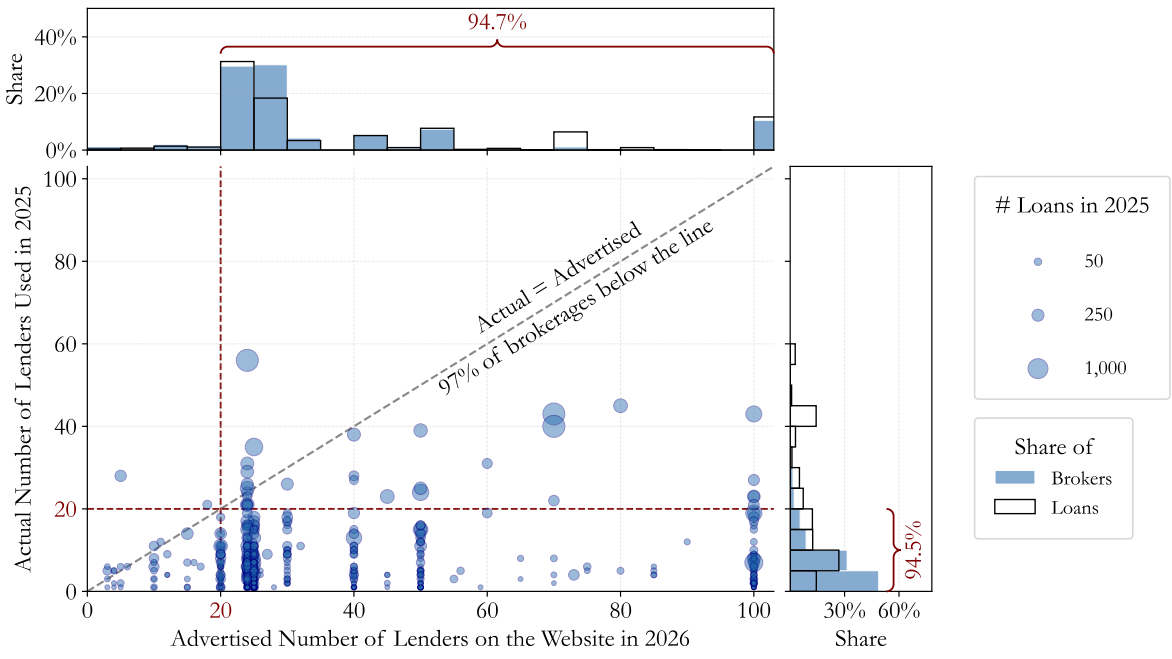
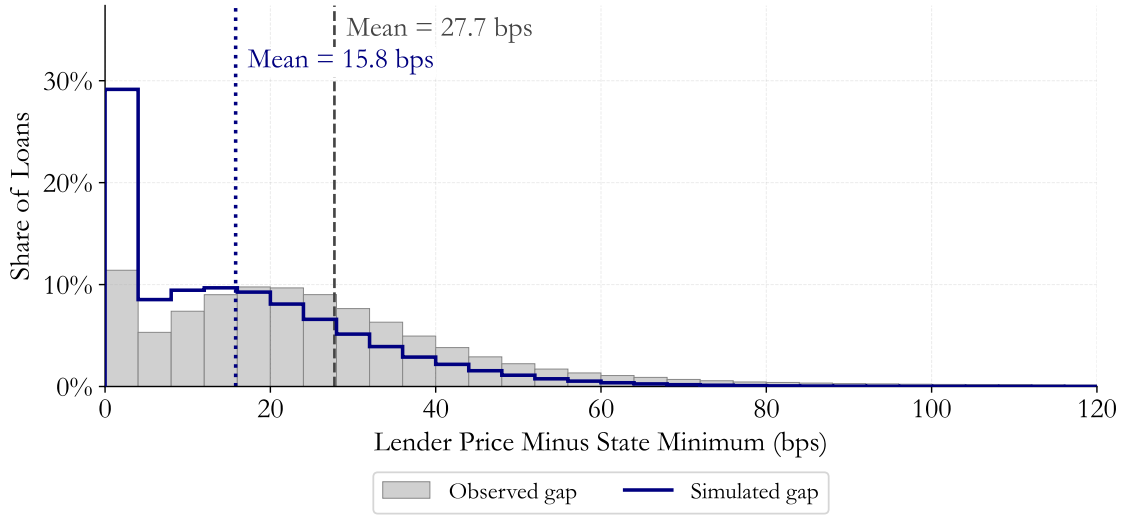


Figure 7: The Steering Gap

This figure compares actual lender choices with a simulated diligent-shopping benchmark and relates the resulting steering gaps to realized mortgage rates. The observed gap is the chosen lender's predicted spread minus the lowest predicted spread among lenders active with at least 10 brokers in the state during the previous year and the focal broker's own lenders. Prices are predicted using lender-specific coefficients on income, credit score, loan amount, property value, combined loan-to-value ratio, debt-to-income ratio, occupancy, property type, and number of units, and broker, state-year, and lock-month effects. The simulated gap is the expected gap of a diligent broker choosing the lowest quote, with quotes equal to predicted spreads plus independent normal errors with a standard deviation of 35 basis points. The steering gap is the observed gap minus the simulated gap. Panel A compares observed gaps (gray bars) and simulated gaps (dark blue line), with vertical lines marking their means. Panel B shows the distribution of steering gaps in 4-basis-point bins (bars, left axis) and realized rate spreads in 30 bins (blue dots, right axis), adjusted for borrower and loan controls in Table 4 and county × lock-month effects. The red dashed line marks a steering gap of zero. All gaps and spreads are in basis points. The sample includes 308,810 brokered conventional conforming fixed-rate loans matched across Optimal Blue, CoreLogic, and HMDA, 2018–2025, with usable lender-price predictions. Means include observations outside the displayed range.

Panel A: Observed and Simulated Gaps



Panel B: Distribution of the Steering Gap and Realized Rates

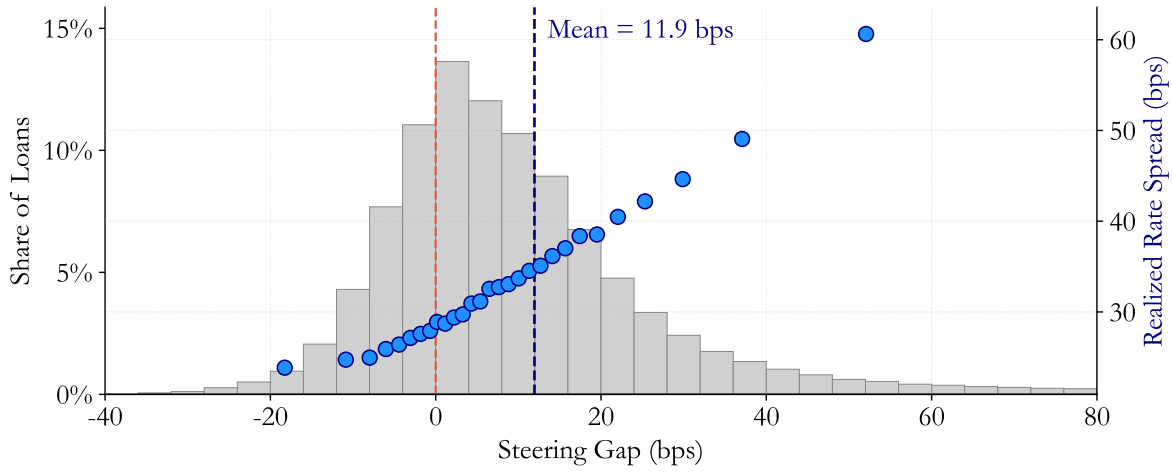
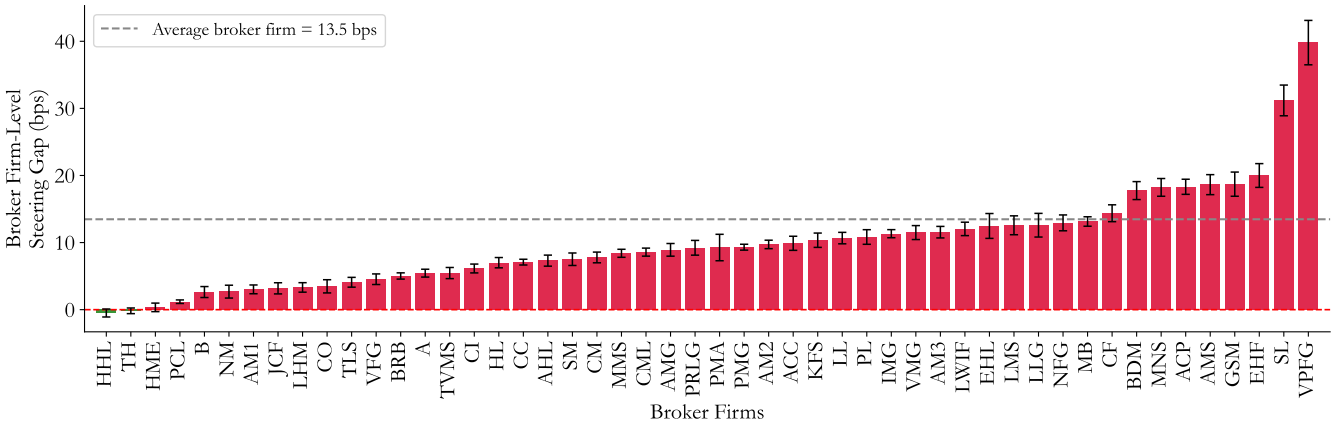


Figure 8: Broker Firm Steering Gaps and Price Levels

This figure examines differences in steering gaps across broker firms and their relationship with broker pricing. The firm-level steering gap is the average gap across a firm's loans, defined as the chosen lender's predicted price minus the expected price selected by a simulated diligent broker, whose quotes include independent normal pricing errors with a standard deviation of 35 basis points. Panel A shows steering gaps for the 50 largest broker firms, ordered from lowest to highest, with 95% confidence intervals based on within-firm variation. Firms are labeled by their initials; red bars indicate positive gaps and green bars negative gaps. The gray dashed line marks the equally weighted average among firms with at least 10 loans, and the red dashed line marks zero. Panel B plots firm price levels relative to retail from equation (2) against firm-level steering gaps, for firms with at least 10 loans in both analyses. Dots are binned means weighted by firm loan counts, with a linear fit and 95% confidence band based on robust standard errors. Gaps and price levels are in basis points. The sample includes brokered conventional conforming fixed-rate loans matched across Optimal Blue, CoreLogic, and HMDA, 2018–2025, with usable lender-price predictions.

Panel A: Broker Firm-Level Steering Gaps of the 50 Largest Firms



Panel B: Broker Firm Price Levels and Firm-Level Steering Gaps

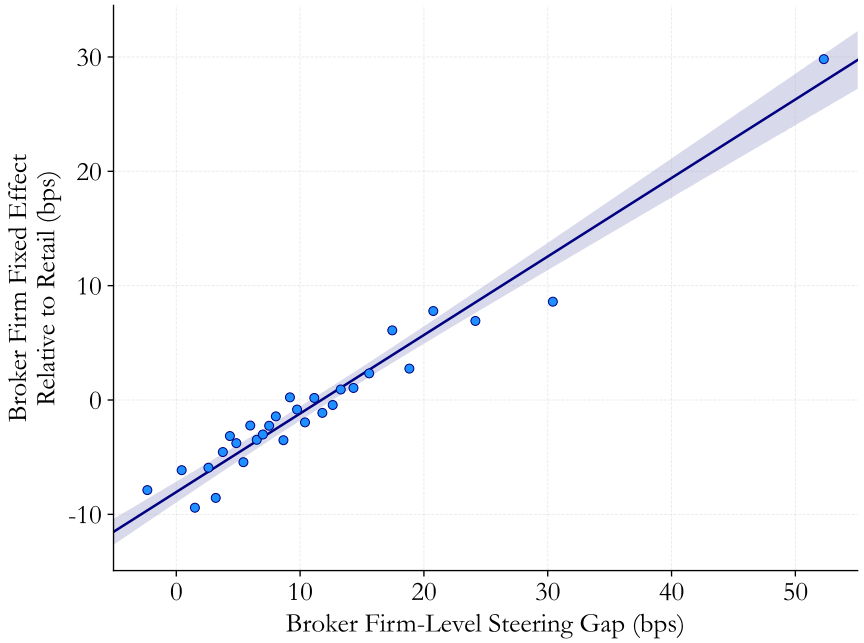
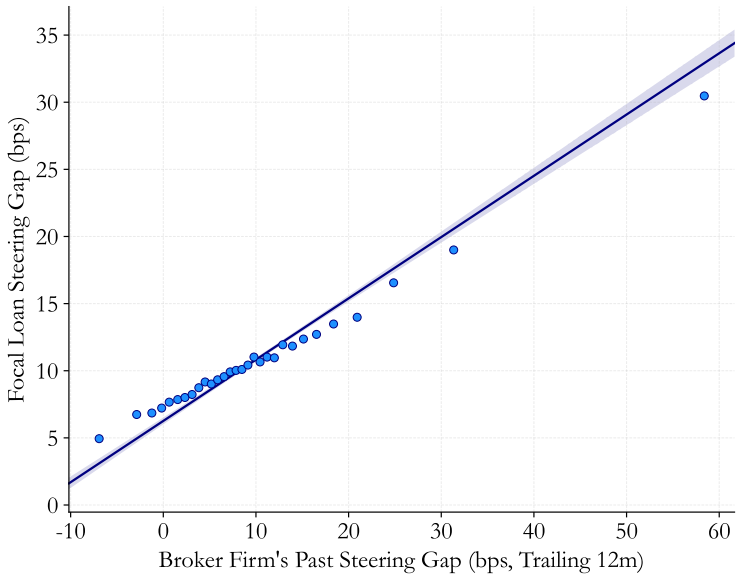


Figure 9: Persistence in Broker Firm Steering Gaps

This figure examines whether broker firms' steering gaps persist across loans and over time. Panel A plots each loan's steering gap against its firm's average gap over the previous 12 months, adjusting for the borrower and loan controls in Table 4, lender concentration, county × lock-month fixed effects, and broker firm fixed effects. Panel B compares firm-level average gaps in months $t-24$ to $t-13$ (horizontal axis) with those in months $t-12$ to $t-1$ (vertical axis), adjusting for county × month fixed effects. Each observation in Panel B is a firm-month, with at least five loans required in each window. In both panels, dots represent binned means, with linear fits and shaded 95% confidence bands. The steering gap is the chosen lender's predicted spread minus the expected predicted spread under simulated diligent shopping, with independent normal pricing errors of 35 basis points in standard deviation. Gaps are in basis points. The sample includes brokered conventional conforming fixed-rate loans matched across Optimal Blue, CoreLogic, and HMDA, 2018–2025, with usable lender-price predictions. Standard errors are clustered by county.

Panel A: Past Firm Steering Gap and Focal Loan Steering Gap



Panel B: Firm Steering Gap, 12 Months Apart

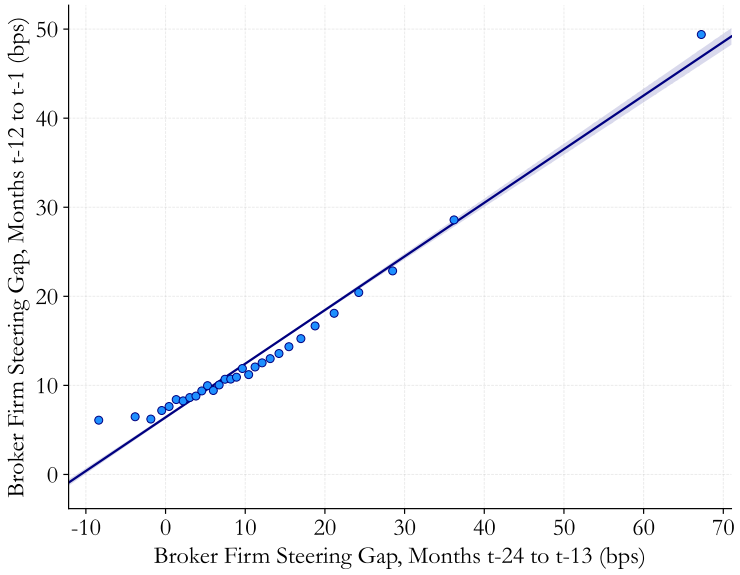


Figure 10: Past Firm Steering Gap and Mortgage Rates

This figure examines the relationship between brokers' past steering practices and the mortgage rates paid by subsequent borrowers. The horizontal axis shows the broker firm's average steering gap over the previous 12 months, requiring at least five loans, and the vertical axis shows the realized mortgage rate spread. Dots represent 30 binned averages, with a linear fit and 95% confidence band. The plot adjusts for borrower and loan controls in Table 4, lender concentration, county × lock-month fixed effects, and broker firm fixed effects. The steering gap is the chosen lender's predicted spread minus the expected predicted spread under simulated diligent shopping, with independent normal pricing errors of 35 basis points in standard deviation. Spreads and gaps are in basis points. The sample includes brokered conventional conforming fixed-rate loans matched across Optimal Blue, CoreLogic, and HMDA, 2018–2025, with usable lender-price predictions. Standard errors are clustered by county.

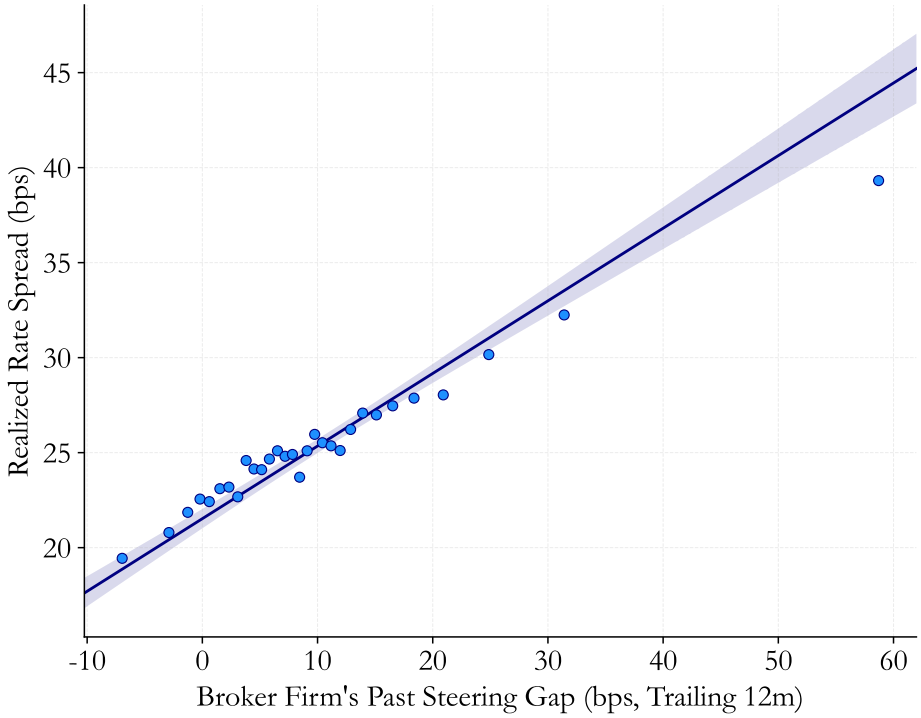
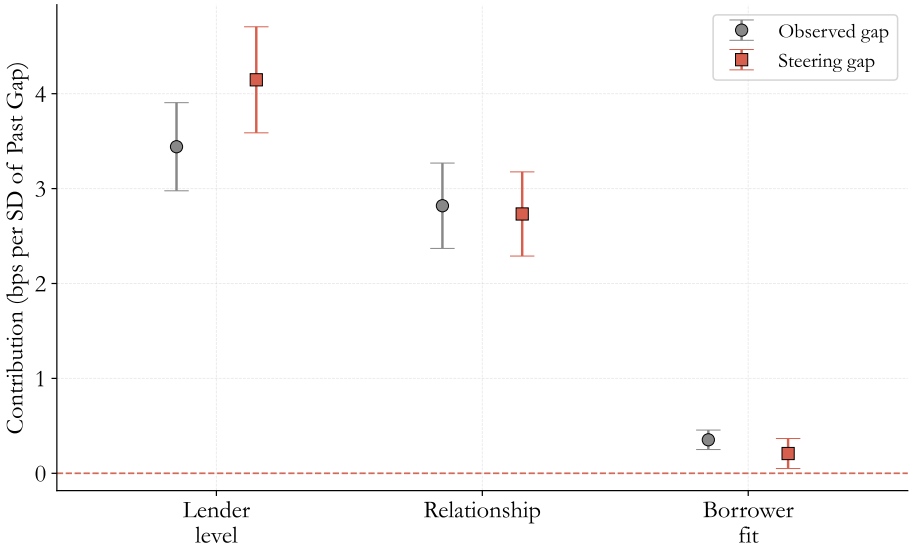


Figure 11: Decomposing Past Steering Gaps and Mortgage Rates

This figure decomposes the relationship between brokers' past lender choices and the mortgage rates paid by subsequent borrowers. Panel A separates general lender pricing, broker-lender relationship pricing, and borrower-lender fit. Panel B separates lender choice within the broker's existing network from the choice of lender network relative to the state lender set. Gray circles represent the observed gap, measured against the lowest predicted lender price, and red squares represent the steering gap, measured against the expected price under simulated diligent shopping with independent normal pricing errors of 35 basis points. Past gaps and components are firm averages over the previous 12 months, requiring at least five loans, standardized by the standard deviation of the past gap. Each contribution equals the component's coefficient in a joint regression of rate spreads on all components, multiplied by its coefficient from a regression on the total past gap using the same controls. Contributions sum to the coefficient on the past gap and are measured in basis points per standard deviation, with 95% confidence intervals. Regressions include the borrower and loan controls of Table 4, lender concentration, county × lock-month fixed effects, and broker firm fixed effects. The sample includes brokered loans with measured steering gaps. Standard errors are clustered by county.

Panel A: Lender Level, Relationship, and Borrower Fit



Panel B: Lender Choice Within the Menu and Menu Choice

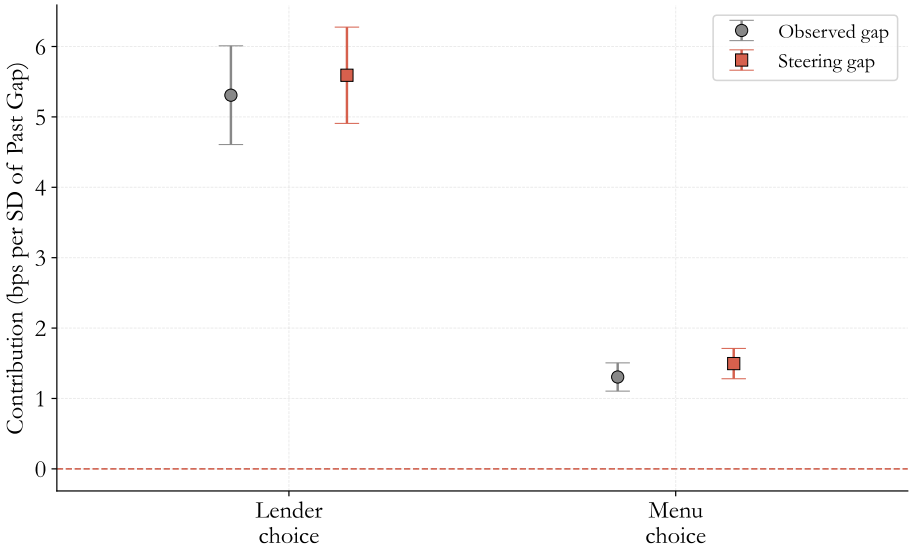


Figure 12: The Steering Gap and Broker Compensation

This figure relates lender-paid broker compensation to the rate spread across deciles of the steering gap. Loans are grouped into deciles of the steering gap, and compensation and the rate spread are each regressed on decile indicators:

$$y_i = \sum_{k=2}^{10} \beta_k \, \mathbb{1}(\mathit{Decile}_i = k) + X_i'\gamma + \delta_{c(i),t(i)} + \zeta_{f(i)} + \varepsilon_i,$$

where y_i is compensation in basis points of the loan amount or the rate spread in basis points, and X_i denotes the borrower and loan controls in Table 4. Decile 1 is the omitted category, so β_k is the difference between decile k and decile 1. Each dot plots a decile's adjusted mean rate spread against its adjusted mean compensation, the average prediction with every loan assigned to that decile. The inner axes are in basis points. The outer axes convert them to dollars for a $330,000 loan at a 3.38% note rate, the sample medians: compensation is its share of the loan amount, and borrowing costs are the present value, discounted at the note rate, of a 30-year fixed-rate mortgage priced at the rate spread above that rate and repaid after 11 years, including the larger balance repaid at the end. Bars are 95% confidence intervals. The steering gap is the chosen lender's predicted spread minus the expected predicted spread under simulated diligent shopping, with independent normal pricing errors with a standard deviation of 35 basis points. The sample includes brokered conventional conforming fixed-rate loans with recorded lender-paid compensation and usable lender-price predictions, matched across Optimal Blue, CoreLogic, and HMDA, 2018–2025. Standard errors are clustered by county.

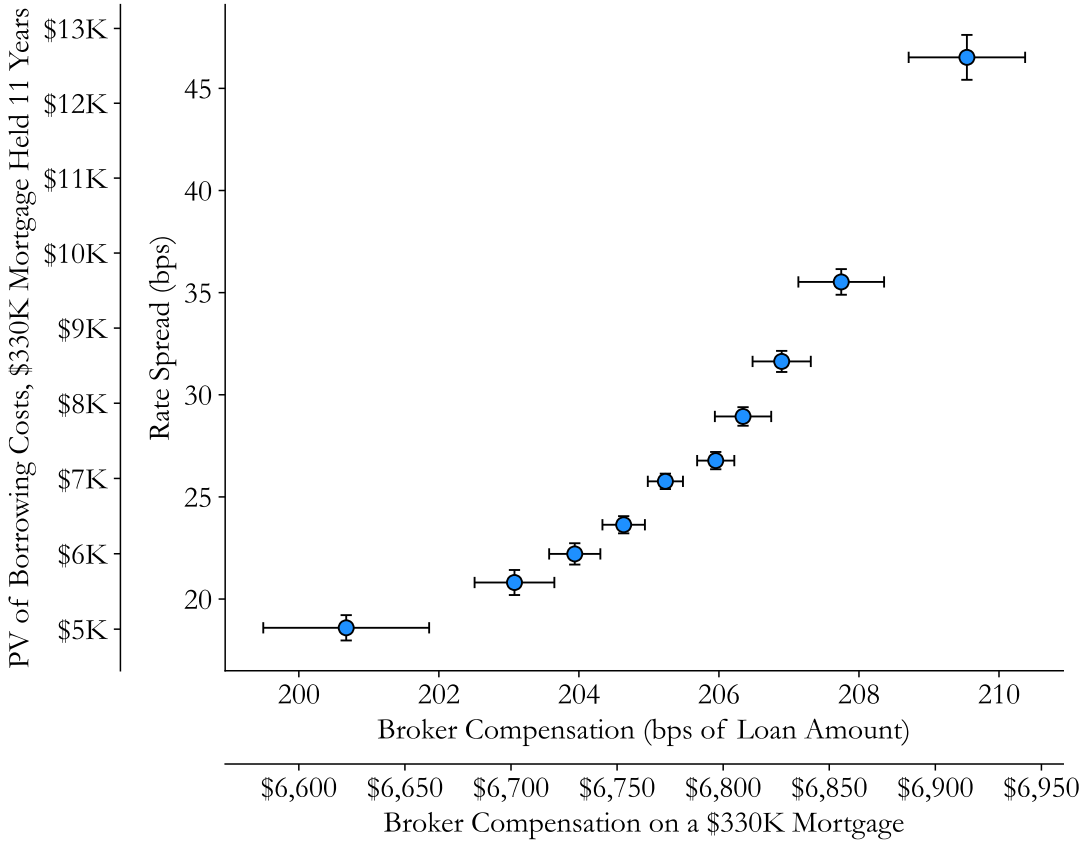


Table 1: Summary Statistics of Brokered Loans

This table summarizes mortgage prices, broker compensation, borrower and loan characteristics, and broker-lender networks. Panel A reports mortgage prices and compensation; Panel B, borrower characteristics; Panel C, loan characteristics; and Panel D, broker firm and loan officer lender networks. Rate spread is APR minus APOR. Broker compensation is recorded in Optimal Blue in basis points of the loan amount. Compensation paid by the lender is an indicator equal to one for lender-paid and zero for borrower-paid compensation, among loans with an identified payer. Variables marked (1/0) are indicators. Minority identifies non-White applicants among those reporting race, with joint applications coded as non-minority. Lender counts and HHI are calculated using all brokered CoreLogic loans in the 12 months before the lock month. HHI is the sum of squared lender loan shares, multiplied by 10,000. The sample includes conventional conforming fixed-rate loans matched across Optimal Blue, CoreLogic, and HMDA, closing in 2018–2025, with complete borrower and loan controls. Rate spread, income, combined loan-to-value ratio, and origination charges are winsorized at the 1st and 99th percentiles by closing year. Obs. reports the available number of observations for each variable.

	Mean	SD	P10	Median	P90	Obs.
Panel A: Prices						
Rate spread (bps)	31.3	48.2	−17.8	22.7	92.0	335,997
Note rate (%)	3.99	1.53	2.62	3.38	6.62	335,991
Origination charges ($)	3,381	3,344	197	1,992	8,295	335,997
Broker compensation (bps of loan amount)	207.6	51.0	137.5	200.0	275.0	186,991
Compensation paid by the lender (1/0)	0.633	0.482	0.000	1.000	1.000	320,440
Panel B: Borrower Characteristics						
Credit score	752	44	688	761	802	335,997
Income ($000s)	130.4	86.5	52.0	109.0	229.0	335,997
Debt-to-income ratio (%)	37.4	7.7	30.0	37.0	47.0	335,997
Combined loan-to-value ratio (%)	73.0	17.9	46.9	77.2	95.0	335,997
Minority applicant (1/0)	0.175	0.380	0.000	0.000	1.000	290,929
Female applicant (1/0)	0.236	0.424	0.000	0.000	1.000	318,253
Has co-applicant (1/0)	0.444	0.497	0.000	0.000	1.000	335,940
Panel C: Loan Characteristics						
Loan amount ($000s)	352.7	164.6	160.0	330.0	569.4	335,997
Property value ($000s)	523.1	316.8	225.0	455.0	875.0	335,997
Purchase loan (1/0)	0.485	0.500	0.000	0.000	1.000	335,997
Cash-out refinance (1/0)	0.193	0.395	0.000	0.000	1.000	335,997
30-year term (1/0)	0.857	0.350	0.000	1.000	1.000	335,997
Principal residence (1/0)	0.954	0.210	1.000	1.000	1.000	335,997
Panel D: Broker Lender Networks, Previous 12 Months						
Lenders used by the broker firm	6.5	6.6	2.0	5.0	13.0	324,802
Broker firm lender HHI (0–10,000)	4,895	2,480	2,099	4,357	9,232	324,802
Lenders used by the loan officer	3.6	2.4	1.0	3.0	7.0	301,859
Loan officer lender HHI (0–10,000)	5,737	2,639	2,595	5,122	10,000	301,859
Brokered loans	335,997					
Broker firms	15,409					
Broker loan officers	31,377					
Lenders	334					

Table 2: Broker Firm and Lender Fixed Effects in Brokered Rates

This table examines how much variation in brokered mortgage rates is explained by broker firm and lender fixed effects. All columns regress brokered-loan rate spreads on income, credit score, loan amount, property value, combined loan-to-value ratio, debt-to-income ratio, origination charges, race, sex, and county × lock-date fixed effects. Column (1) is the baseline; Columns (2) and (3) add lender and broker firm fixed effects, respectively; Column (4) includes both. The share of baseline residual variance explained measures the reduction in degrees-of-freedom-adjusted residual variance relative to Column (1). Rate spread is APR minus APOR, in basis points. The sample includes brokered loans matched across Optimal Blue, CoreLogic, and HMDA, 2018–2025.

Dep. Variable:	Rate Spread (bps)			
	(1)	(2)	(3)	(4)
R^2	0.557	0.599	0.631	0.658
Adjusted R^2	0.383	0.441	0.458	0.496
Residual SD (bps)	35.07	33.38	32.89	31.72
Share of baseline residual variance explained		9.4%	12.1%	18.2%
County × Lock-Date FE	✓	✓	✓	✓
Lender FE		✓		✓
Broker Firm FE			✓	✓
Borrower and loan controls	✓	✓	✓	✓
Broker firms			7,969	7,969
Observations	207,573	207,573	207,573	207,573
Dep. Var. Mean	27.0	27.0	27.0	27.0
Dep. Var. SD	44.7	44.7	44.7	44.7

Residual SDs are degrees-of-freedom corrected for the absorbed fixed effects.

Table 3: High-Cost Brokers and Mortgage Rates: Differential Distance IV

This table estimates mortgage rate differences between borrowers using high- and low-cost broker firms, using OLS and a differential-distance IV. Firms are classified by their previous-year price effects relative to retail, estimated using the controls in Figure 2 and county × lock-date fixed effects, with estimates shrunk toward the monthly mean. High- and low-cost firms are in the top and bottom terciles across branches in the lock month; middle-tercile firms are excluded. Panel A reports OLS estimates. Panel B instruments high-cost broker use with distance to the nearest low-cost branch minus distance to the nearest high-cost branch, measured in 10-mile units. The bottom rows report first-stage F-statistics. The dependent variable is rate spread (APR minus APOR), in basis points. Column (1) includes origin county × lock-month fixed effects; Column (2) adds mortgage county × lock-month fixed effects; Columns (3)–(5) additionally include lock-date effects, lender × lock-month effects, or both, respectively. Borrower controls include income, loan amount, property value, origination charges, HMDA race and sex indicators, and credit-score, combined loan-to-value, and debt-to-income deciles. Origin controls include nearest-branch distance, log median household income, and the 16 neighborhood characteristics in Figure 5. The search origin is the previous address for purchases (from Verisk and DataAxle) and the mortgaged property for refinances. The sample includes owner-occupied brokered loans matched across Optimal Blue, CoreLogic, and HMDA, 2019–2025, with traced origins, lagged firm classifications, top- or bottom-tercile broker choice, and absolute differential distance below its 95th percentile. Standard errors are clustered by mortgage county.

Dep. Variable:	Rate Spread (bps)				
	(1)	(2)	(3)	(4)	(5)
Panel A: OLS					
$\mathbb{1}$(High-Cost Broker Firm)	15.79***	15.01***	14.90***	12.88***	12.84***
	(0.62)	(0.67)	(0.63)	(0.68)	(0.64)
Panel B: 2SLS					
$\widehat{\mathbb{1}(\text{High-Cost Broker Firm})}$	23.46***	23.99***	23.65***	29.09***	26.85***
	(5.40)	(5.43)	(5.14)	(5.81)	(5.60)
Origin County × Lock-Month FE	✓	✓	✓	✓	✓
Mortgage County × Lock-Month FE		✓	✓	✓	✓
Lock-Date FE			✓		✓
Lender × Lock-Month FE				✓	✓
Borrower controls	✓	✓	✓	✓	✓
Origin controls	✓	✓	✓	✓	✓
First-stage F-statistic	77.7	76.1	73.7	72.1	66.3
Observations	90,473	85,569	85,369	82,009	81,755
Dep. Var. Mean	22.97	22.15	22.13	21.77	21.76
Dep. Var. SD	43.86	42.73	42.68	42.11	42.07

Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: The Steering Gap and Realized Rates

This table examines the relationship between lender-choice steering gaps and realized mortgage rate spreads, both in basis points. The steering gap is the chosen lender's predicted spread minus the expected predicted spread under simulated diligent shopping across the state lender set, with independent normal pricing errors with a standard deviation of 35 basis points. Column (1) includes county × lock-month fixed effects; Column (2) adds broker firm fixed effects; Column (3) adds standardized broker firm HHI measured over the previous 12 months; and Column (4) adds lock-date fixed effects. HHI is the sum of squared lender loan shares within the firm. Controls include income, credit score, loan amount, property value, combined loan-to-value ratio, debt-to-income ratio, race, ethnicity, sex, occupancy, property type, number of units, loan purpose, and loan term. The sample includes brokered conventional conforming fixed-rate loans matched across Optimal Blue, CoreLogic, and HMDA, 2018–2025, with usable lender-price predictions. Standard errors are clustered by county.

Dep. Variable:	Rate Spread (bps)			
	(1)	(2)	(3)	(4)
Steering gap (bps)	0.786***	0.729***	0.730***	0.726***
	(0.017)	(0.016)	(0.016)	(0.016)
Broker firm HHI (std.)			0.587***	0.479***
			(0.149)	(0.155)
County × Lock-Month FE	✓	✓	✓	✓
Lock-Date FE				✓
Broker Firm FE		✓	✓	✓
Borrower and loan controls	✓	✓	✓	✓
Observations	308,810	304,526	297,131	296,937
Adjusted R^2	0.497	0.547	0.547	0.573
Dep. Var. Mean	30.9	30.5	30.5	30.5
Dep. Var. SD	49.7	49.1	49.0	49.0

Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Persistence in Broker Firm Steering Gaps

This table examines whether brokers' past steering gaps predict their current lender choices. Panel A regresses each loan's steering gap on its firm's average gap over the previous 12 months. All columns include borrower and loan controls and standardized broker firm HHI. Column (1) includes county × lock-month fixed effects; Column (2) adds broker firm fixed effects; and Column (3) adds lock-date fixed effects. Panel B regresses firm average gaps in months $t-12$ to $t-1$ on those in months $t-24$ to $t-13$. Observations are firm-months assigned to the firm's modal county for that year, with at least five loans required in each window. Columns (1)–(3) include county × month, state × month, and month fixed effects, respectively, without firm fixed effects. The steering gap is the chosen lender's predicted spread minus the expected predicted spread under simulated diligent shopping, with independent normal pricing errors with a standard deviation of 35 basis points. Gaps are in basis points. Borrower and loan controls include income, credit score, loan amount, property value, combined loan-to-value ratio, debt-to-income ratio, race, ethnicity, sex, occupancy, property type, number of units, loan purpose, and loan term. The sample includes brokered conventional conforming fixed-rate loans matched across Optimal Blue, CoreLogic, and HMDA, 2018–2025, with usable lender-price predictions. Standard errors are clustered by county.

Panel A: Past Firm Steering Gap and Focal Loan Steering Gap

Dep. Variable:	Focal Loan Steering Gap (bps)		
	(1)	(2)	(3)
Past firm steering gap (bps, 12 months)	0.713***	0.456***	0.455***
	(0.010)	(0.019)	(0.020)
Broker firm HHI (std.)	0.038	0.028	0.036
	(0.072)	(0.119)	(0.118)
County × Lock-Month FE	✓	✓	✓
Lock-Date FE			✓
Broker Firm FE		✓	✓
Borrower and loan controls	✓	✓	✓
Observations	260,440	260,013	259,822
Adjusted R^2	0.459	0.496	0.499
Dep. Var. Mean	10.89	10.88	10.88
Dep. Var. SD	19.36	19.35	19.35

Panel B: Autocorrelation of the Firm Steering Gap

Dep. Variable:	Firm Steering Gap, Months $t-12$ to $t-1$ (bps)		
	(1)	(2)	(3)
Firm steering gap, months $t-24$ to $t-13$ (bps)	0.602***	0.622***	0.655***
	(0.025)	(0.019)	(0.019)
County × Month FE	✓		
State × Month FE		✓	
Month FE			✓
Observations	64,232	74,601	74,964
Adjusted R^2	0.557	0.521	0.460
Dep. Var. Mean	13.29	13.37	13.33
Dep. Var. SD	16.41	16.46	16.44

Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Past Firm Steering Gap and Mortgage Rates

This table examines whether broker firms' past steering practices predict the mortgage rates paid by subsequent borrowers. The past firm steering gap is the firm's average steering gap over the previous 12 months, requiring at least five loans, standardized by its standard deviation of 18.6 basis points. The steering gap is the chosen lender's predicted spread minus the expected predicted spread under simulated diligent shopping, with independent normal pricing errors with a standard deviation of 35 basis points. All columns include county × lock-month fixed effects. Columns (3)–(5) include broker firm fixed effects, and Column (5) adds lock-date fixed effects. Columns (2), (4), and (5) control for standardized broker firm HHI, calculated from squared lender loan shares over the previous 12 months. The dependent variable is the realized rate spread, in basis points. Controls include income, credit score, loan amount, property value, combined loan-to-value ratio, debt-to-income ratio, race, ethnicity, sex, occupancy, property type, number of units, loan purpose, and loan term. The sample includes brokered conventional conforming fixed-rate loans matched across Optimal Blue, CoreLogic, and HMDA, 2018–2025, with usable lender-price predictions. Standard errors are clustered by county.

Dep. Variable:	Rate Spread (bps)				
	(1)	(2)	(3)	(4)	(5)
Past firm steering gap (std.)	9.984***	9.997***	7.056***	7.087***	6.772***
	(0.460)	(0.462)	(0.422)	(0.425)	(0.397)
Broker firm HHI (std.)		−0.113		0.645***	0.520**
		(0.167)		(0.239)	(0.237)
County × Lock-Month FE	✓	✓	✓	✓	✓
Lock-Date FE					✓
Broker Firm FE			✓	✓	✓
Borrower and loan controls	✓	✓	✓	✓	✓
Observations	246,342	244,929	245,890	244,492	244,297
Adjusted R^2	0.435	0.436	0.486	0.486	0.518
Dep. Var. Mean	27.7	27.8	27.7	27.8	27.8
Dep. Var. SD	45.2	45.2	45.1	45.2	45.2

Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: The Steering Gap and Broker Compensation

This table examines whether costly lender choices are associated with higher lender-paid broker compensation. Compensation is measured in basis points of the loan amount, and the steering gap in basis points of rate spread. The steering gap is the chosen lender's predicted spread minus the expected predicted spread under simulated diligent shopping across the state lender set, with independent normal pricing errors with a standard deviation of 35 basis points. Column (1) has no fixed effects. Column (2) includes county × lock-month and broker firm fixed effects; Column (3) adds lock-date fixed effects. Controls include income, credit score, loan amount, property value, combined loan-to-value ratio, debt-to-income ratio, race, ethnicity, sex, occupancy, property type, number of units, loan purpose, and loan term. The sample includes lender-paid brokered conventional conforming fixed-rate loans with recorded compensation and usable lender-price predictions, matched across Optimal Blue, CoreLogic, and HMDA, 2018–2025. Standard errors are clustered by county.

Dep. Variable:	Broker Compensation (bps of Loan Amount)		
	(1)	(2)	(3)
Steering gap (bps)	0.282***	0.114***	0.114***
	(0.012)	(0.013)	(0.013)
County × Lock-Month FE		✓	✓
Lock-Date FE			✓
Broker Firm FE		✓	✓
Borrower and loan controls	✓	✓	✓
Observations	197,611	178,648	178,451
Adjusted R^2	0.259	0.888	0.888
Dep. Var. Mean	207.2	205.4	205.4
Dep. Var. SD	51.1	51.0	51.0

Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

For Online Publication

Internet Appendix For:
"Do Mortgage Brokers Help Borrowers Get Better Rates?"

A Sample Construction

The sample combines the CoreLogic, Optimal Blue, and HMDA data described in Section 1.1. Figure IA.2 summarizes the two matching steps.

Step 1: Optimal Blue and CoreLogic. Optimal Blue locks carry no borrower identity, so I match them to CoreLogic mortgages on loan characteristics that both sources record. A CoreLogic mortgage is a candidate for a lock if it has the same property ZIP code and the same loan amount and was signed between 10 days before and 240 days after the lock date. A candidate is rejected if the two records disagree on the loan purpose, the loan program, the loan term, or whether the rate is fixed or adjustable. The remaining candidates are scored on secondary features of the property and the loan, and a lock is matched only when its best candidate is clearly better than the next one. Later rounds use the county instead of the ZIP code or allow a small difference in the loan amount, for locks and mortgages that are still unmatched. When several locks match the same mortgage, as with relocks or a borrower who locked with more than one lender, I keep the best-scoring lock.

Step 2: adding HMDA. I then attach HMDA records to the matched lock-mortgage pairs. HMDA reports loan amounts in $10,000 bins and no dates, so in the first round a candidate HMDA record must have the same activity year, census tract, and loan-amount bin as the pair, and the same note rate, occupancy, loan purpose, loan program, and term as the Optimal Blue lock. Candidates are scored on the lender and on other fields that both records report. Later rounds search the county instead of the tract, the neighboring year's HMDA file for loans closing in December or January, and, last, candidates whose note rate differs; each of these rounds also requires the same lender as the CoreLogic record. The sample keeps the most reliable matches: first-round matches whose best candidate has a high score and is the only candidate or clearly better than the next one, and county-round matches, which also require the same loan-to-value ratio. These are 14.1 million of the 15.8 million matched loans (89.4%). In a placebo that matches the 2022 locks to the 2021 HMDA file, about 1% of locks find a match, providing a diagnostic for accidental matches.

Brokered and retail loans. CoreLogic records up to two companies on each mortgage but does not record which is the lender and which is the broker, and the order of the two fields changes over time. I

assign the lender as the company whose name matches the lender of record on the deed, using NMLS registry names, and the broker as the other company, with a list of firms that fund loans to break ties. Brokered loans are then classified as in Section 1.1. Each brokered loan is assigned to the branch where its loan officer worked in the lock month, from the NMLS employment and branch records.

Sample filters. The main sample includes conventional conforming fixed-rate loans closing from 2018 to 2025, with complete borrower and loan controls, and excludes singleton county × lock-date cells. Rate spread, income, combined loan-to-value ratio, and origination charges are winsorized at the 1st and 99th percentiles within closing year. The steering analysis retains 308,810 brokered loans with usable lender-price predictions. The IV analysis uses owner-occupied brokered loans locked from 2019 to 2025, with a located search origin and a previous-year firm price level. Origins are previous addresses for purchases and mortgaged properties for refinances. Additional restrictions appear in the exhibit notes.

B Lender Pricing and the Diligent-Broker Benchmark

The 100 lenders with the most brokered loans have separate pricing models, estimated on their retail and brokered loans; the remaining lenders share a pooled model. The chosen lender's predicted price leaves out the focal loan. Predictions require sufficient observations in the relevant state-year and lock-month cells. Figure IA.10 reports model fit, and Figure IA.11 compares estimated prices of borrower characteristics.

To predict prices for lenders outside a broker's observed network, I decompose the lender × broker effect into a broker component, a lender component, and a relationship-specific component. All lender predictions include the broker component. A relationship-specific component is included only for lenders on the broker's observed menu. The state set combines lenders used by at least 10 brokers in the state during the previous 12 months with the focal broker's own lenders there. It is a market benchmark, rather than a record of offers made to the borrower.

The diligent-broker calculation is a robustness exercise for pricing information observed by lenders but absent from the research data. It assumes independent normal pricing errors with a common standard deviation of 35 basis points, calibrated to the 35.1 basis point out-of-sample root mean squared

prediction error on brokered loans. This assigns the full prediction-error scale to lender-specific variation, although part may reflect risk priced alike by lenders and therefore leave their ranking unchanged. I compute the probability that each lender offers the lowest quote by numerical integration over this distribution. The distribution in Figure 7, Panel A, weights each possible lender's predicted-price gap by this probability; it is not the distribution of loan-level expected gaps. The loan-level expected gap is subtracted from the observed gap to construct the steering-gap proxy in Panel B. Table IA.15 varies the pricing-error assumption and the lender set.

Figure IA.1: A Broker Firm's Website

This figure shows the home page of one broker firm's website, an example of how mortgage brokers describe their service to borrowers. The broker advertises that it compares rates across many wholesale lenders on the borrower's behalf. Figure 6, Panel A, summarizes the phrases used across all broker websites.

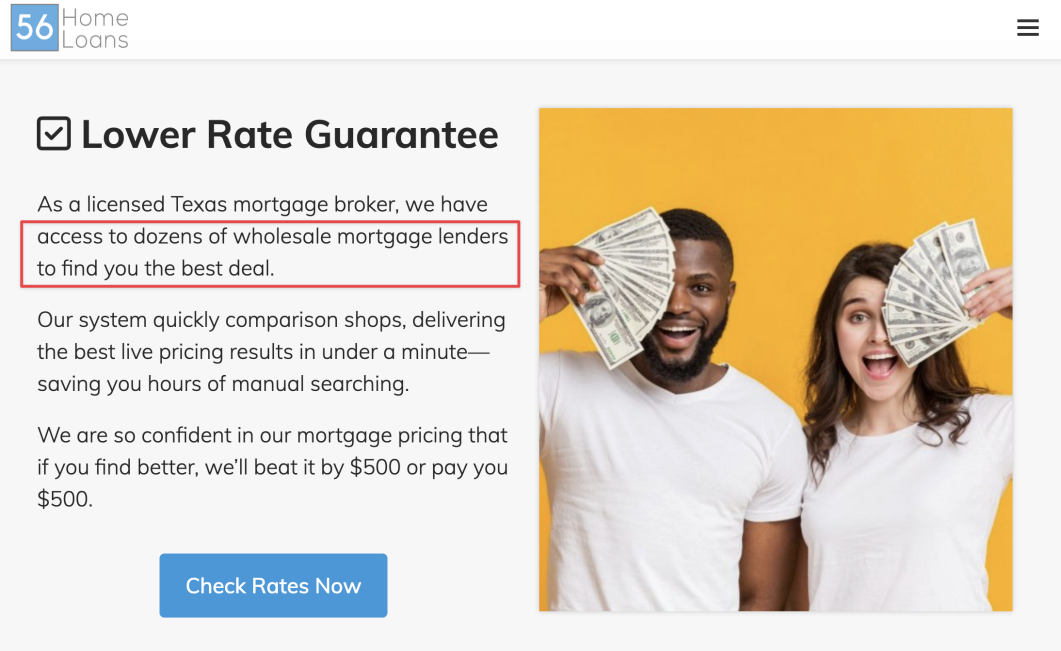
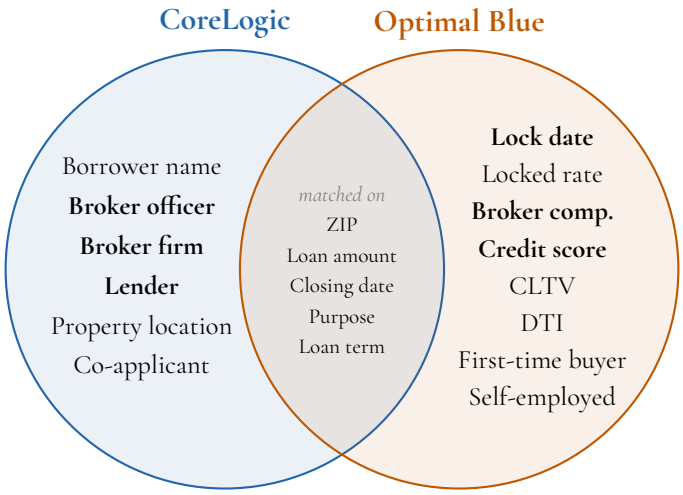


Figure IA.2: Linking Borrowers, Brokers, and Lenders

This figure shows the two steps that link the three main data sources, described in Section A. Panel A matches Optimal Blue rate locks to CoreLogic mortgages. Panel B adds HMDA records to the matched lock-mortgage pairs. In each panel, the circles list what each source records, the center lists the matching variables, and the variables in bold are the ones the analysis requires from each source.

Panel A: Matching Optimal Blue Rate Locks to CoreLogic Mortgages



Panel B: Adding HMDA Records to the Matched Lock-Mortgage Pairs

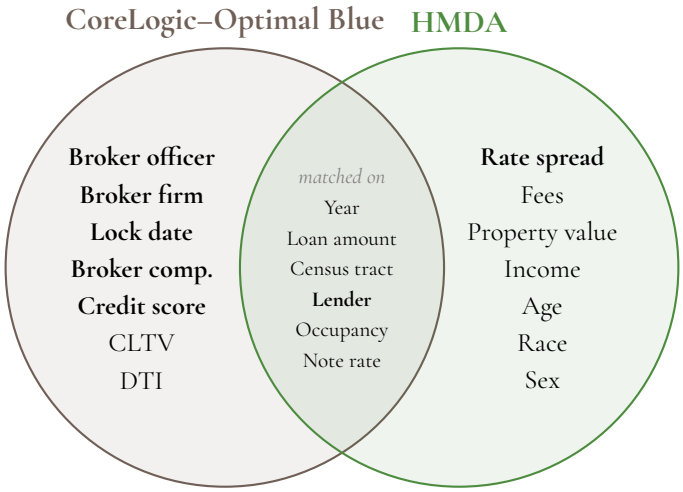


Figure IA.3: Residual Rate Dispersion with Census Tract × Lock-Date Fixed Effects

This figure compares residual rate dispersion for retail and brokered loans using census tract × lock-date fixed effects. Residuals come from a pooled regression of rate spread on a brokered indicator, income, credit score, loan amount, property value, combined loan-to-value ratio, debt-to-income ratio, origination charges, race, and sex. Gray bars represent retail loans and the red line brokered loans. Rate spread is APR minus APOR, in basis points; the legend reports residual standard deviations. The displayed range matches the county-date comparison in Figure 1. The data are matched Optimal Blue, CoreLogic, and HMDA loans, 2018–2025.

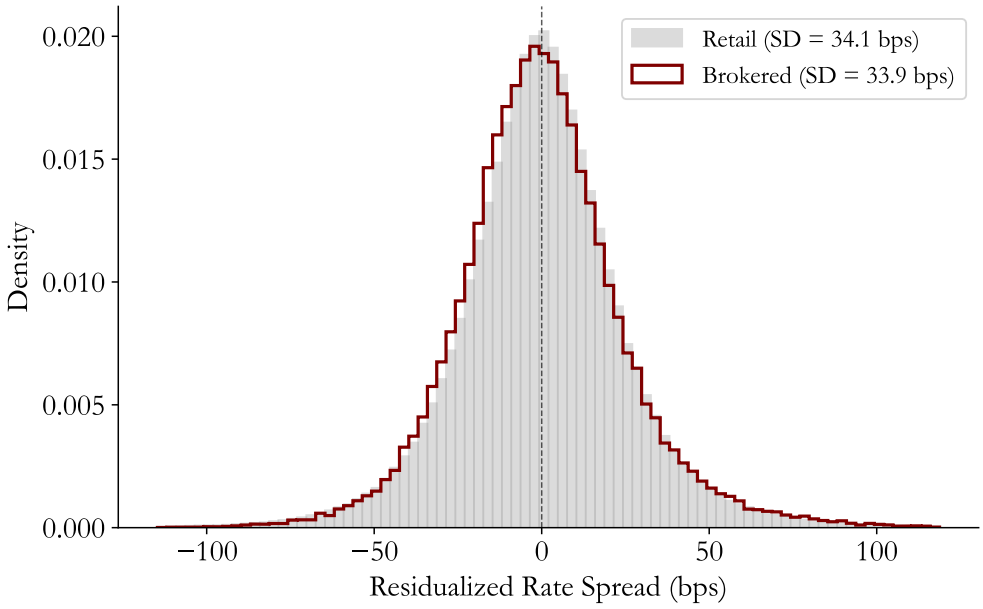


Figure IA.4: Rate Spreads of Retail and Brokered Loans by Credit Score

This figure compares the average rate spread of retail and brokered loans across borrower credit scores. I regress the rate spread on indicators for each channel and credit-score bin, with no other controls, so each point is the mean rate spread of the channel in the bin; the bars are 95% confidence intervals. Gray circles are retail loans and red squares brokered loans. The rate spread is in basis points. The sample is the matched Optimal Blue, CoreLogic, and HMDA sample of rate locks from 2018 to 2025. Standard errors are clustered by state.

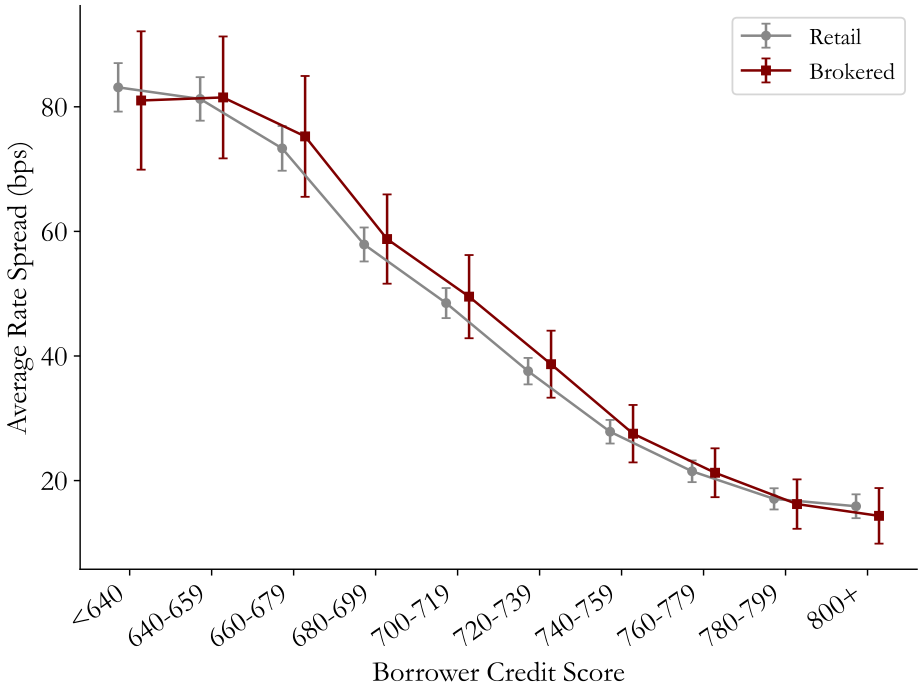
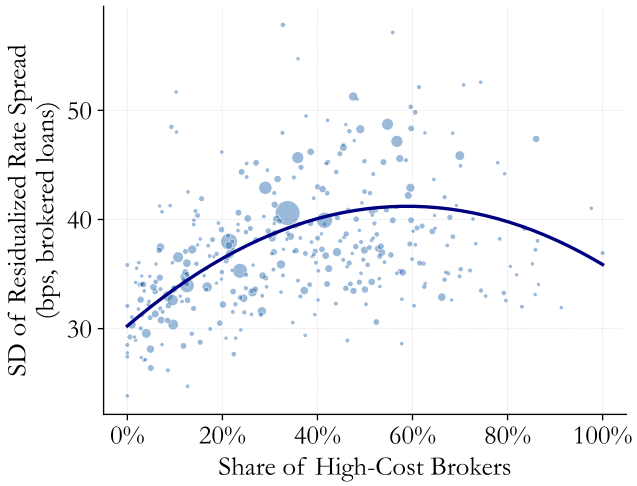


Figure IA.5: Broker Firms with High Price Levels and Brokered Rates Across Counties

This figure relates the share of a county's brokered loans made by broker firms with high price levels to the dispersion and the level of brokered rates in the county. Each dot is a county with at least 100 brokered loans, and its size is proportional to the county's brokered loans. The horizontal axis is the share of the county's brokered loans made by broker firms in the top third of price levels across brokered loans. A firm's price level is its mean brokered-minus-retail rate gap, defined below, over its brokered loans in other counties, shrunk toward the overall mean by empirical Bayes, so none of the county's own loans enters the share. Panel A plots the standard deviation of the residual rate spread of the county's brokered loans, from a regression of the rate spread on income, credit score, loan amount, property value, combined loan-to-value ratio, debt-to-income ratio, occupancy, property type, and number of units with county × lock-month fixed effects, corrected for leverage; the line is a loan-weighted quadratic fit. Panel B plots the brokered-minus-retail rate gap, the mean over the county's brokered loans of the rate spread adjusted for the same controls minus the mean adjusted rate spread of retail loans in the same county and lock month; the line is a loan-weighted linear fit, and the vertical axis is cut at −20 and 30 basis points, and the fit uses all counties. The sample is the matched Optimal Blue, CoreLogic, and HMDA sample of rate locks from 2018 to 2025, pooled over years.

Panel A: Dispersion of Brokered Rates



Panel B: Brokered Minus Retail Rate

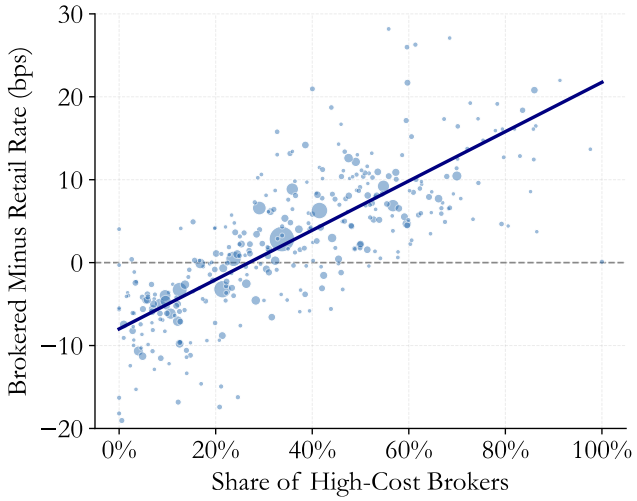


Figure IA.6: Borrower and Neighborhood Characteristics by the Sign of the Instrument

This figure compares borrower and neighborhood distributions by relative proximity to high-cost and low-cost branches. Differential distance dd is distance to the nearest low-cost branch minus distance to the nearest high-cost branch; high-cost (low-cost) firms are in the top (bottom) tercile of previous-year price effects. Blue represents $dd < 0$ and red $dd \geq 0$. Each panel displays one characteristic without controls or fixed effects, and dashed lines mark group means. Upper panels report borrower characteristics; lower panels report origin block-group and tract characteristics from the American Community Survey. The origin is the previous address for purchases and the mortgaged property for refinances. Displayed ranges cover the pooled 2nd–98th percentiles. The sample is owner-occupied brokered loans with traced origins and a chosen firm in the top or bottom tercile in the matched Optimal Blue, CoreLogic, and HMDA IV data, 2019–2025.

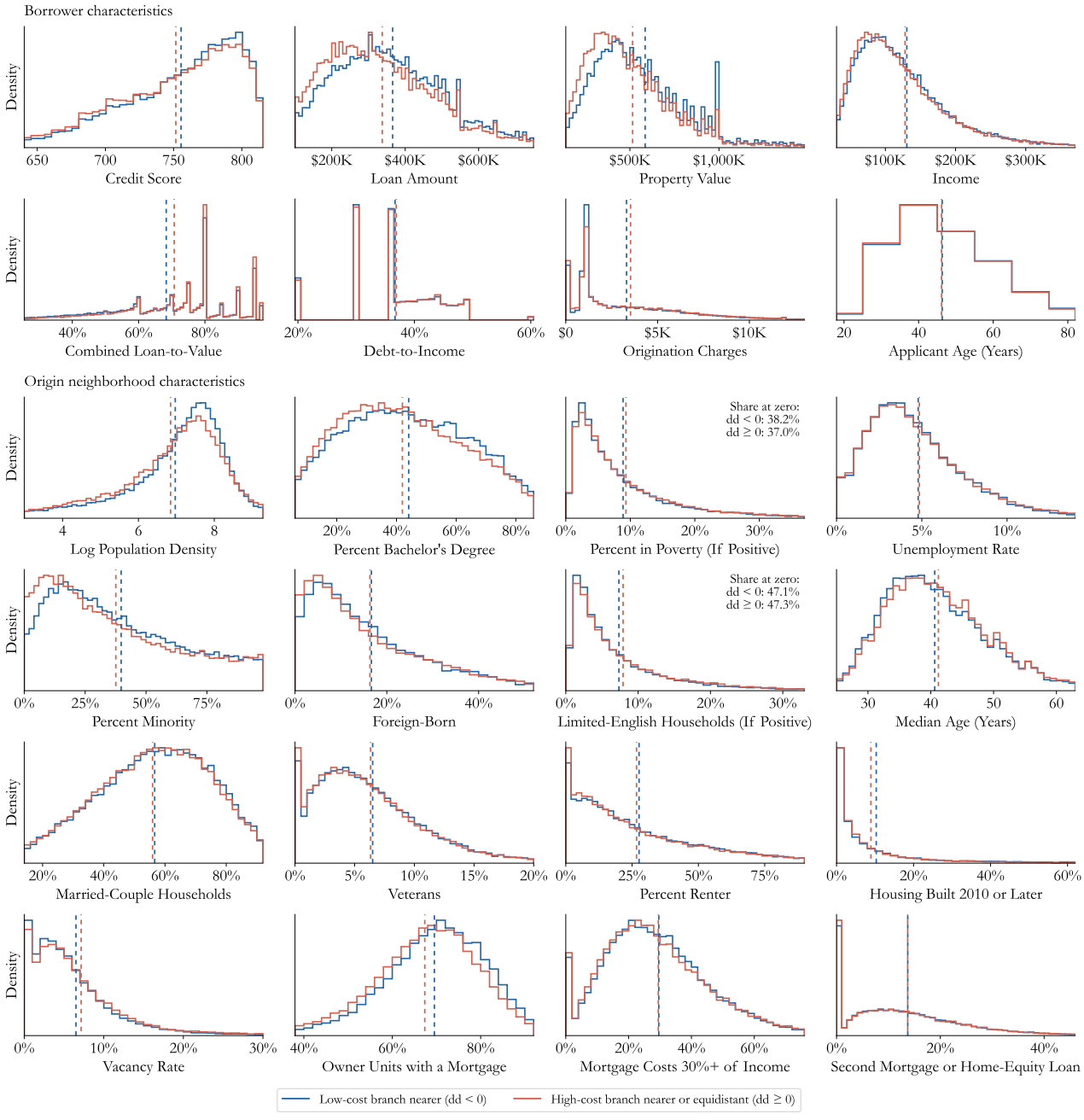


Figure IA.7: The Differential Distance Instrument: An Example

This figure illustrates differential distance for a borrower in Travis County, Texas, who locked a mortgage rate in June 2022. Triangles mark potential broker branches. Red branches belong to high-cost firms, classified by whether their previous-year price effect is at or above the monthly median across branches; blue branches belong to low-cost firms. The nearest low-cost branch is 1 mile from the origin and the nearest high-cost branch is 4 miles away, giving a differential distance of $1 - 4 = -3$ miles. The origin is the borrower's previous address, traced using Verisk and DataAxle. The map also marks the mortgaged property, and tract shading shows population density.

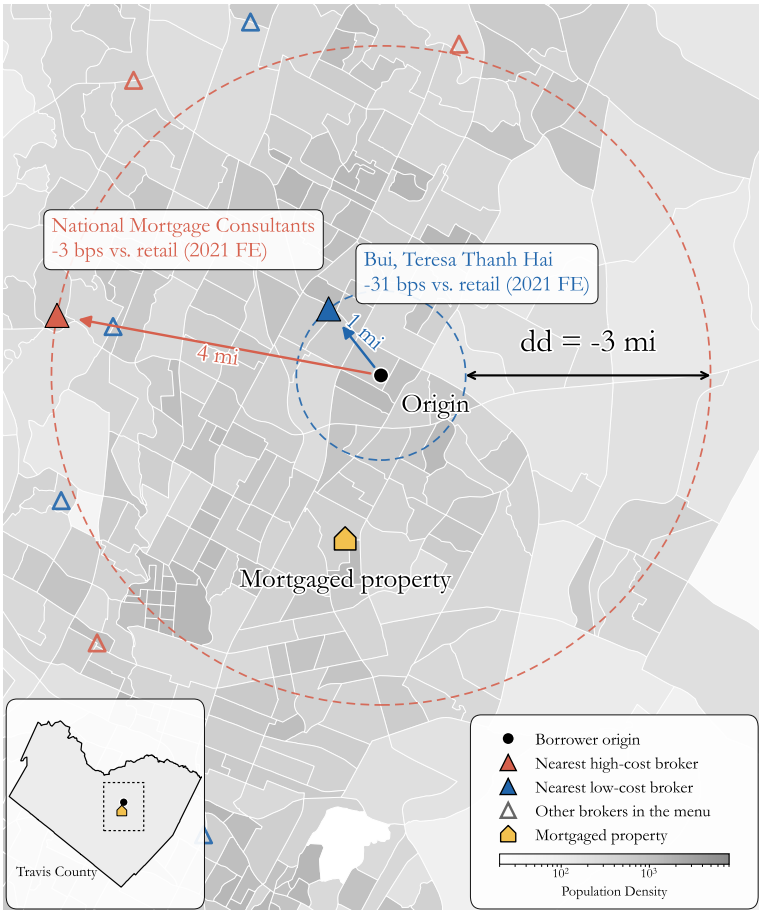
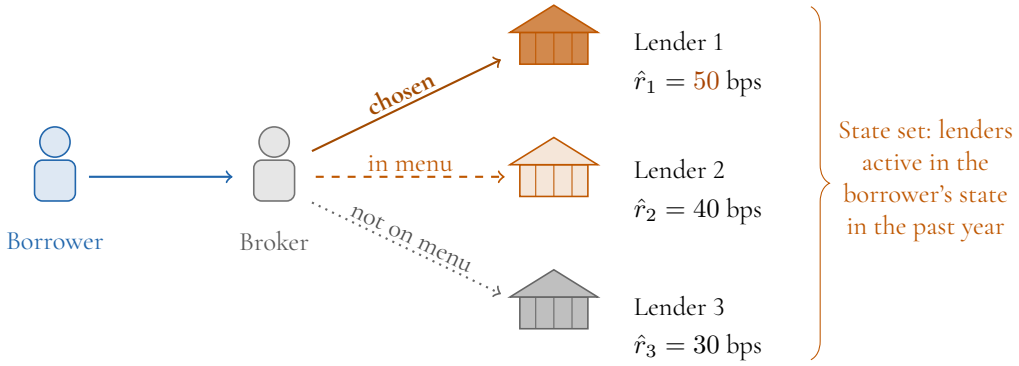


Figure IA.8: Measuring the Steering Gap

This figure illustrates the observed, simulated, and steering gaps for a hypothetical loan. The chosen lender is Lender 1. Lenders 1 and 2 are on the broker's menu; Lender 3 is in the state set but not on the menu and receives no relationship-specific pricing adjustment. All three prices are predicted rate spreads, including the chosen lender's. The observed gap is the chosen price minus the lowest predicted price, $50 - 30 = 20$ basis points. The diligent broker obtains a quote from each lender, equal to its predicted price plus independent normal noise with a standard deviation of 35 basis points, and selects the lowest. The table shows three hypothetical draws of the noise; in Draw 2, Lender 1 quotes lowest although its predicted price is the highest. Over all draws, the three lenders quote lowest with probabilities of approximately 0.22, 0.32, and 0.46, so the expected predicted price selected by the diligent broker is 37.6 basis points and the simulated gap is $37.6 - 30 = 7.6$. The steering gap is $20 - 7.6 = 12.4$, equivalently $50 - 37.6$. Using the mean predicted price of 40 as the reference instead gives an observed gap of 10 and a simulated gap of -2.4, leaving the steering gap unchanged. All values are illustrative.



	Lender 1	Lender 2	Lender 3	Lowest quote
Predicted spread $\hat{r}_\ell$	50	40	30	Lender 3
Quotes $\hat{r}_\ell + \sigma v_\ell$, three draws of the noise ($\sigma = 35$ bps)				
Draw 1	$50 - 7.0 = 43.0$	$40 + 14.0 = 54.0$	$30 + 10.5 = \mathbf{40.5}$	Lender 3
Draw 2	$50 - 31.5 = \mathbf{18.5}$	$40 + 3.5 = 43.5$	$30 + 17.5 = 47.5$	Lender 1
Draw 3	$50 + 21.0 = 71.0$	$40 - 24.5 = \mathbf{15.5}$	$30 + 7.0 = 37.0$	Lender 2
Probability of the lowest quote p_ℓ	0.22	0.32	0.46	

Diligent broker's expected predicted price $= 0.22 \times 50 + 0.32 \times 40 + 0.46 \times 30 = 37.6$ bps

Observed gap $= 50 - 30 = 20$ bps Simulated gap $= 37.6 - 30 = 7.6$ bps

Steering gap $= 20 - 7.6 = 50 - 37.6 = 12.4$ bps

Figure IA.9: Lender Choice Within the Menu and Menu Choice

This figure splits the observed and steering gaps of a hypothetical loan into lender choice within the broker's menu and menu choice. Predicted rate spreads are on the horizontal axis, in basis points. The chosen lender, Lender 1, and Lender 2 are on the broker's menu, with predicted spreads of 50 and 40 basis points; Lender 3, at 30 basis points, is in the state set but not on the menu. The loan is that of Figure IA.8. The observed gap of 20 is lender choice of $50 - 40 = 10$, the chosen price minus the lowest menu price, plus menu choice of $40 - 30 = 10$, the lowest menu price minus the lowest state-set price. The diligent brokers obtain quotes equal to predicted prices plus independent normal noise with a standard deviation of 35 basis points and select the lowest. A diligent broker that shops the state set selects the three lenders with probabilities of approximately 0.22, 0.32, and 0.46, an expected predicted price of 37.6; one that shops only the menu selects Lender 1 or Lender 2 with probabilities of approximately 0.42 and 0.58, an expected predicted price of 44.2. The steering gap of $50 - 37.6 = 12.4$ is lender choice of $50 - 44.2 = 5.8$ plus menu choice of $44.2 - 37.6 = 6.6$. All values are illustrative.

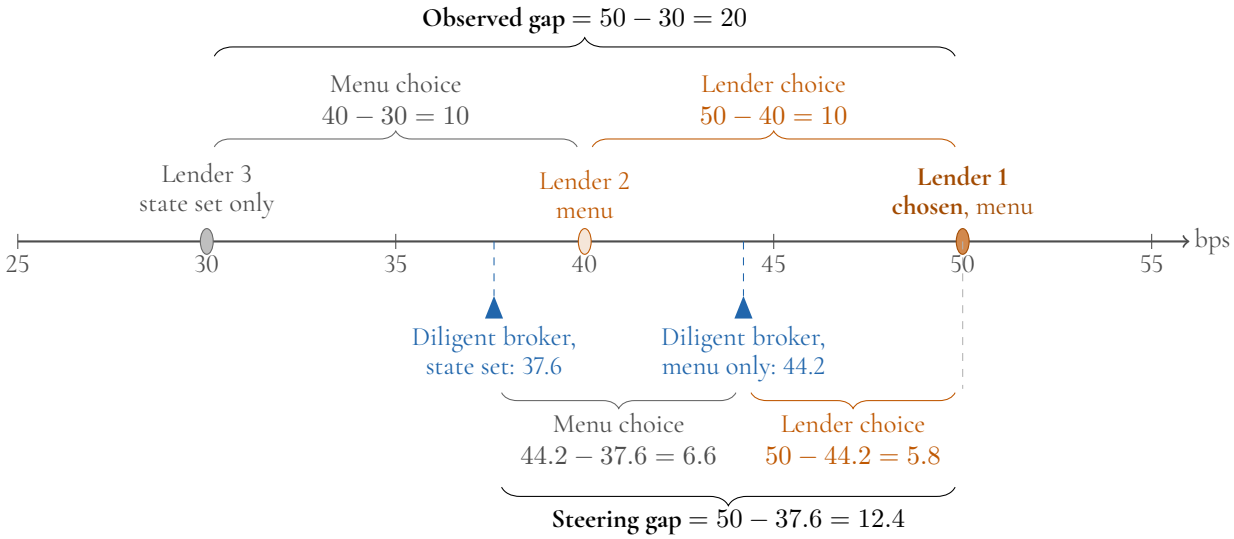


Figure IA.10: Fit of the Lender Pricing Models

This figure relates the fit of the lender pricing models of equation (6) to the number of loans each model is estimated on. Each of the 100 lenders with the most brokered loans has its own pricing model, estimated on its retail and brokered loans (dots); the remaining lenders share one pooled model (diamond). The vertical axis is the R^2 of the model over all of its loans, and the line is the median R^2 within quintiles of the number of loans across the 100 lender models. The horizontal axis is on a log scale. The sample is retail and brokered loans in the matched Optimal Blue, CoreLogic, and HMDA sample of rate locks from 2018 to 2025.

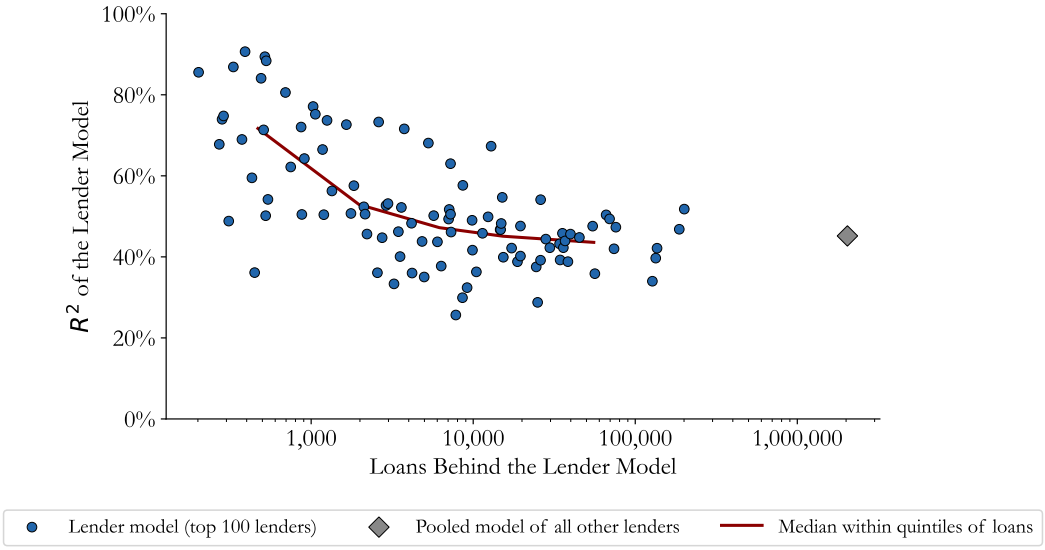


Figure IA.11: Borrower Characteristics in the Lender Pricing Models

This figure reports the coefficients on borrower and loan characteristics in the lender pricing models of equation (6), for the 30 lenders with the most brokered loans. Each panel shows one characteristic: each point is one lender's coefficient with its 95% confidence interval, and lenders are sorted by the coefficient. Coefficients on credit score, combined loan-to-value ratio, debt-to-income ratio, income, and loan amount are the change in the rate spread, in basis points, for a one-standard-deviation increase in the characteristic, with standard deviations taken over brokered loans with a measured steering gap; the loan amount coefficient holds the combined loan-to-value ratio fixed. Coefficients on investment property, second home, and two units are relative to a principal residence with one unit. The gray dashed line is the coefficient in the pooled model of the lenders outside the top 100, and the red dashed line marks zero. The sample is retail and brokered loans in the matched Optimal Blue, CoreLogic, and HMDA sample of rate locks from 2018 to 2025. Standard errors are clustered by county.

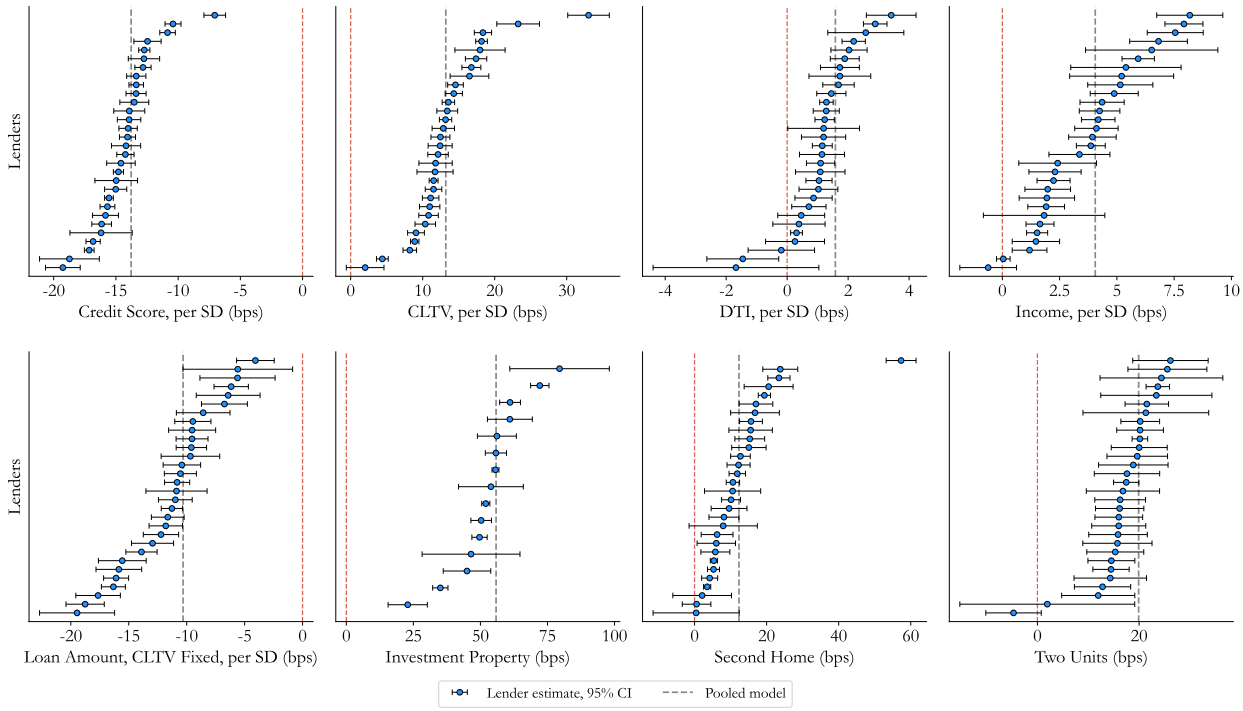


Figure IA.12: Observed and Simulated Gaps Against the State Mean

This figure repeats Figure 7, Panel A, with the mean predicted spread in the state set as the benchmark in place of the lowest. The observed gap is the chosen lender's predicted spread minus the mean predicted spread in the state set, which includes lenders active with at least 10 brokers in the state during the previous year and the focal broker's own lenders. Gray bars show observed gaps, and the dark line shows gaps under diligent-broker choices, weighting each possible lender by its probability of quoting lowest when quotes equal predicted spreads plus independent normal noise with a standard deviation of 35 basis points. Vertical lines mark the respective means. The benchmark is the same for both gaps, so the difference between the two means equals the mean steering gap in Figure 7. Gaps are in basis points, in 4-basis-point bins. The sample consists of brokered conventional conforming fixed-rate loans in the matched Optimal Blue, CoreLogic, and HMDA data, 2018–2025, with usable lender-price predictions. There are 308,810 loans; means include observations outside the displayed range.

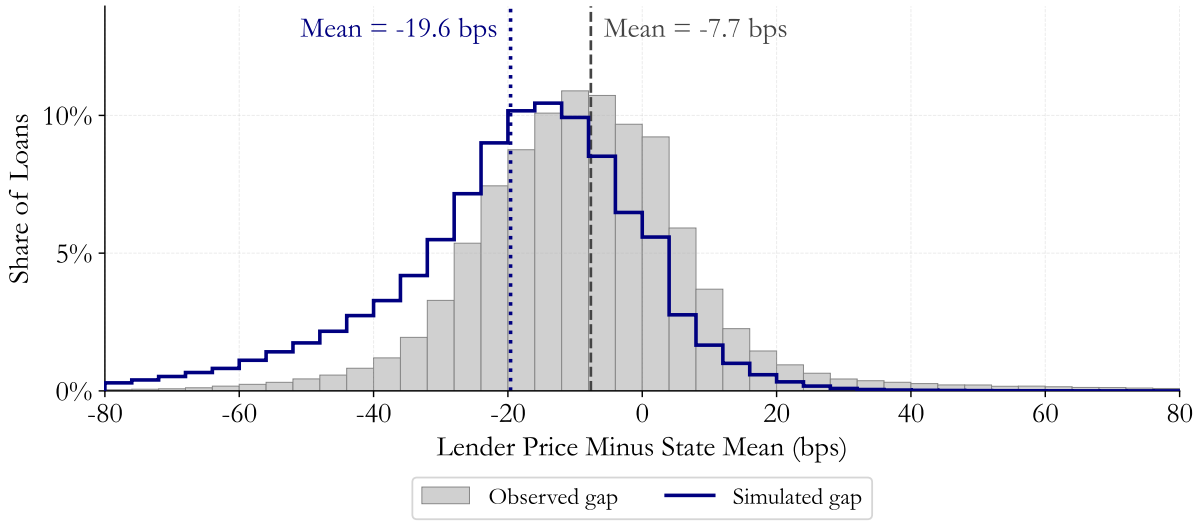


Figure IA.13: Observed, Simulated, and Steering Gaps of the Largest Broker Firms

This figure reports the firm-level observed gaps, each firm's mean across its loans, of the 50 largest broker firms in the steering analysis, in the order and with the labels of Figure 8, Panel A. Each bar's height is the firm's observed gap, shown with its 95% confidence interval. The blue part is the firm's simulated gap and the red part its steering gap. Where a firm's steering gap is negative, its bar ends below its simulated gap, and a hatched green outline marks the difference. The observed gap compares the chosen lender's predicted spread with the lowest predicted spread in the state set. The simulated gap is the expected observed gap when the broker selects the lowest of the state lenders' noisy quotes, with independent normal noise of 35 basis points. The steering gap in Figure 8 is their difference. The gray dashed and black dotted lines report the observed and simulated gaps of the average broker firm, weighting equally the broker firms with at least 10 loans in the steering analysis. Gaps are in basis points, and confidence intervals use within-firm variation. The sample consists of brokered conventional conforming fixed-rate loans in the matched Optimal Blue, CoreLogic, and HMDA data, 2018–2025, with usable lender-price predictions.

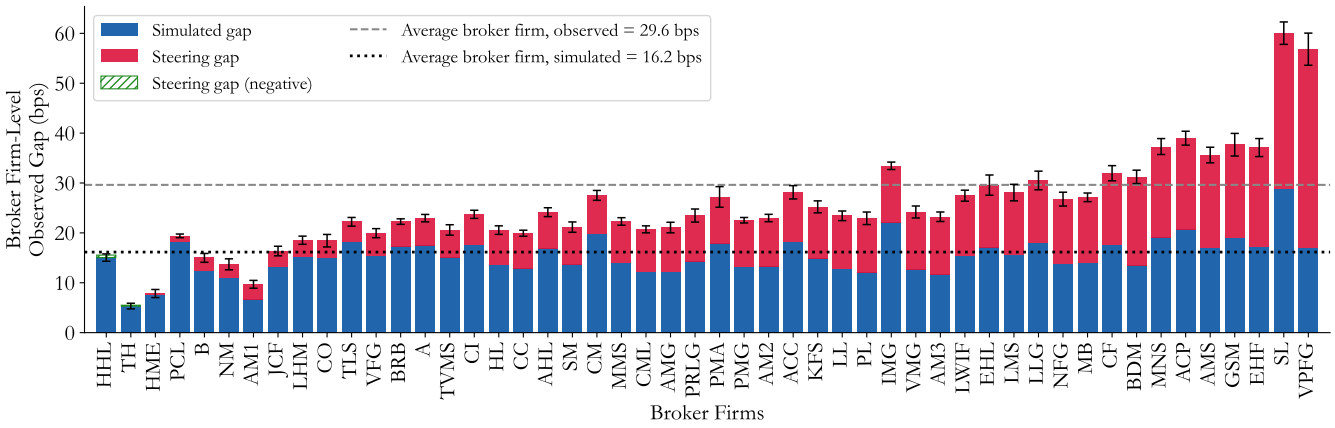


Table IA.1: Summary Statistics of Retail Loans

This table summarizes retail loans and compares their means with brokered loans. Retail loans are matched loans not classified as brokered. Panel A reports prices, Panel B borrower characteristics, and Panel C loan characteristics. Rate spread is APR minus APOR, in basis points. Minority identifies non-White applicants among those reporting race, with joint applications coded as non-minority. The final columns report the brokered mean and the brokered-minus-retail difference. The sample consists of conventional conforming fixed-rate loans in the matched Optimal Blue, CoreLogic, and HMDA data closing in 2018–2025, with complete borrower and loan controls. Rate spread, income, combined loan-to-value ratio, and origination charges are winsorized at the 1st and 99th percentiles by closing year. Statistics use the available observations for each variable.

	Retail Loans					Brokered	Brokered
	Mean	SD	P10	Median	P90	Mean	− Retail
Panel A: Prices							
Rate spread (bps)	31.2	47.9	−18.8	26.8	90.3	31.3	0.0
Note rate (%)	4.19	1.51	2.75	3.62	6.62	3.99	−0.20
Origination charges ($)	2,449	2,386	684	1,579	5,449	3,381	932
Panel B: Borrower Characteristics							
Credit score	753	44	689	762	803	752	−1
Income ($000s)	124.5	81.5	50.0	104.0	221.0	130.4	5.9
Debt-to-income ratio (%)	36.9	7.7	30.0	36.0	47.0	37.4	0.4
Combined loan-to-value ratio (%)	74.9	17.8	48.8	80.0	95.0	73.0	−1.9
Minority applicant (1/0)	0.130	0.337	0.000	0.000	1.000	0.175	0.045
Female applicant (1/0)	0.242	0.429	0.000	0.000	1.000	0.236	−0.007
Has co-applicant (1/0)	0.468	0.499	0.000	0.000	1.000	0.444	−0.025
Panel C: Loan Characteristics							
Loan amount ($000s)	307.7	152.8	135.8	280.8	510.4	352.7	45.0
Property value ($000s)	436.3	810.0	195.0	385.0	735.0	523.1	86.8
Purchase loan (1/0)	0.586	0.493	0.000	1.000	1.000	0.485	−0.100
Cash-out refinance (1/0)	0.143	0.350	0.000	0.000	1.000	0.193	0.050
30-year term (1/0)	0.849	0.358	0.000	1.000	1.000	0.857	0.008
Principal residence (1/0)	0.945	0.227	1.000	1.000	1.000	0.954	0.009
Retail loans	7,814,055						
Lenders	2,692						
Retail loan officers	197,250						
Counties	2,215						

Table IA.2: Matched Brokered Loans Compared with All Brokered Loans in HMDA

This table compares matched brokered loans with comparable brokered loans in HMDA. Panels A–C report prices, loan characteristics, and borrower characteristics; Panel D reports closing-year shares. The reweighted mean applies the matched sample's year weights to HMDA means. The normalized difference divides the difference in means by the square root of the average of the two variances. Both groups use HMDA variable definitions. The HMDA benchmark consists of originated first-lien conventional conforming fixed-rate loans whose applications were not submitted directly to the reporting lender, closing in 2018–2025. Rate spread, note rate, property value, income, combined loan-to-value ratio, and origination charges are winsorized using the HMDA benchmark's annual 1st and 99th percentiles. Debt-to-income is the upper endpoint of the reported range. Each variable uses its available observations. The matched data link Optimal Blue, CoreLogic, and HMDA.

	Matched Brokered		HMDA Brokered				
	Mean	SD	Mean	SD	Reweighted Mean	Diff. in Means	Normalized Diff.
Panel A: Prices							
Rate spread (bps)	32.4	53.7	32.2	65.2	29.3	0.2	0.004
Note rate (%)	3.99	1.53	4.52	1.84	4.07	−0.53	−0.311
Origination charges ($)	3,446	3,526	3,228	3,404	3,088	218	0.063
Panel B: Loan Characteristics							
Loan amount ($000s)	353.4	164.6	331.8	172.2	330.5	21.6	0.128
Property value ($000s)	517.1	278.3	493.8	285.0	499.9	23.4	0.083
Income ($000s)	131.1	91.0	134.4	99.0	132.5	−3.3	−0.035
Debt-to-income ratio (%)	37.4	7.7	37.3	8.0	37.0	0.1	0.008
Combined loan-to-value ratio (%)	73.0	17.9	71.9	17.8	71.0	1.0	0.059
Purchase loan (1/0)	0.485	0.500	0.501	0.500	0.448	−0.016	−0.031
Cash-out refinance (1/0)	0.193	0.395	0.185	0.389	0.198	0.008	0.021
30-year term (1/0)	0.857	0.350	0.824	0.381	0.807	0.033	0.090
Principal residence (1/0)	0.954	0.210	0.841	0.365	0.851	0.113	0.378
Panel C: Borrower Characteristics							
Minority applicant (1/0)	0.175	0.380	0.218	0.413	0.218	−0.042	−0.106
Female applicant (1/0)	0.236	0.424	0.232	0.422	0.230	0.004	0.009
Has co-applicant (1/0)	0.444	0.497	0.428	0.495	0.432	0.016	0.032
Panel D: Share of Loans by Closing Year (%)							
2018	4.3		6.8			−2.5	
2019	8.2		10.5			−2.3	
2020	26.3		22.0			4.3	
2021	33.3		24.5			8.8	
2022	12.3		10.3			2.0	
2023	4.0		7.0			−3.0	
2024	5.4		8.9			−3.6	
2025	6.3		10.0			−3.8	
Loans	335,997		7,049,923				
Matched / HMDA brokered loans (%)	4.8						

Table IA.3: Rate Levels of Brokered and Retail Loans

This table compares the rates of brokered and retail loans made to observably similar borrowers. Each column regresses the rate spread on an indicator for brokered loans and borrower and loan controls, with location-time fixed effects that become finer across columns: state × lock month in Column (1), county × lock month in Column (2), county × lock month plus lock date in Column (3), county × lock date in Column (4), and county × lock date plus census tract in Column (5). Borrower and loan controls are income, credit score, loan amount, property value, combined loan-to-value ratio, debt-to-income ratio, origination charges, race, and sex. The rate spread is in basis points. The sample is the matched Optimal Blue, CoreLogic, and HMDA sample of rate locks from 2018 to 2025. Standard errors are clustered by state.

Dep. Variable:	Rate Spread (bps)				
	(1)	(2)	(3)	(4)	(5)
$\mathbb{1}$(Brokered)	1.24	0.85	0.95	0.81	0.41
	(0.91)	(0.88)	(0.88)	(0.91)	(0.88)
State × Lock-Month FE	✓				
County × Lock-Month FE		✓	✓		
Lock-Date FE			✓		
County × Lock-Date FE				✓	✓
Census Tract FE					✓
Borrower and loan controls	✓	✓	✓	✓	✓
Observations	8,150,052	8,150,052	8,150,052	8,150,052	8,145,021
Adjusted R^2	0.297	0.317	0.344	0.363	0.389
Dep. Var. Mean	31.2	31.2	31.2	31.2	31.2
Dep. Var. SD	47.9	47.9	47.9	47.9	47.9

Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table IA.4: Correlation of Broker Firm Price Levels Across Periods and Loan Types

This table relates a broker firm's price level on one side of a sample split to its price level on the other side. For each split, the broker firm effects of Figure 2 are estimated separately on side A and side B, each relative to retail loans on the same side in the same county on the same day. Each column is one split: lock years 2018–2021 against 2022–2025, purchase against refinance loans, rate-and-term against cash-out refinances, principal residences against second homes and investment properties, 30-year against 10- to 25-year terms, credit scores of 740 or above against below 740, combined loan-to-value ratios of 80% or below against above 80%, and conventional against government-backed loans. The raw correlation is between the two sets of estimated effects. The correlation and slope net of noise use only covariances between effects estimated on loans in different lock months, to adjust for estimation noise. The persistent standard deviation is the standard deviation of firm effects net of estimation error. Panel A uses one effect per broker firm. Panel B uses one effect per broker firm and lock year, demeaned by lock year, and compares the two sides within the same firm and year; the lock-year split has no Panel B version. Firms need at least 10 brokered loans on each side. The sample is the matched Optimal Blue, CoreLogic, and HMDA sample of rate locks from 2018 to 2025. Standard errors are from a bootstrap over broker firms.

Split	Lock year	Purpose	Refi type	Occupancy	Term	Score	CLTV	Loan type
Side A	2018–21	Purchase	Rate-term	Primary	30-year	740+	≤80%	Conv.
	vs	vs	vs	vs	vs	vs	vs	vs
Side B	2022–25	Refinance	Cash-out	Other	Shorter	<740	>80%	Govt.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Broker Firm Effects								
Correlation, raw	0.56	0.62	0.60	0.48	0.65	0.77	0.74	0.58
Correlation, net of noise	0.65	0.75	0.77	0.54	0.85	0.95	0.90	0.69
	(0.03)	(0.02)	(0.08)	(0.15)	(0.02)	(0.01)	(0.02)	(0.02)
Slope of B on A, net of noise	1.08	0.55	0.89	0.88	0.86	1.06	0.99	1.10
	(0.06)	(0.02)	(0.04)	(0.16)	(0.04)	(0.03)	(0.03)	(0.05)
Mean effect, B − A (bps)	4.73	−1.55	0.26	−2.47	0.11	−0.70	0.88	−7.28
	(0.55)	(0.33)	(0.36)	(1.59)	(0.45)	(0.21)	(0.27)	(0.61)
Persistent SD, side A (bps)	11.11	13.76	10.02	12.89	11.85	11.89	11.92	13.88
Persistent SD, side B (bps)	18.46	10.13	11.69	21.07	12.06	13.22	13.23	22.16
Broker firms	1,370	2,033	1,254	179	987	2,277	1,833	2,039
Brokered loans, side A	150,066	114,636	66,093	60,963	147,290	169,517	156,370	208,287
Brokered loans, side B	54,000	120,081	37,098	3,569	29,348	79,899	69,608	84,941
Panel B: Broker Firm-Year Effects, Demeaned by Lock Year								
Correlation, raw		0.63	0.64	0.35	0.67	0.74	0.70	0.54
Correlation, net of noise		0.81	0.84	0.40	0.85	0.97	0.90	0.69
		(0.03)	(0.03)	(0.17)	(0.03)	(0.02)	(0.02)	(0.04)
Slope of B on A, net of noise		0.66	0.89	1.03	0.90	1.09	0.96	1.00
		(0.03)	(0.05)	(0.20)	(0.04)	(0.03)	(0.04)	(0.07)
Mean effect, B − A (bps)		−0.17	2.02	1.38	0.28	0.17	0.34	−9.09
		(0.36)	(0.40)	(2.89)	(0.42)	(0.31)	(0.34)	(0.82)
Persistent SD, side A (bps)		12.77	9.87	9.29	11.12	11.98	12.74	15.22
Persistent SD, side B (bps)		10.32	10.34	24.11	11.75	13.42	13.51	22.08
Broker firm-years		1,665	942	94	877	2,127	1,808	1,832
Broker firms		962	643	67	608	1,121	892	839
Brokered loans, side A		50,230	40,289	17,550	66,915	100,445	88,211	96,017
Brokered loans, side B		72,222	21,406	1,560	20,563	46,524	40,999	45,004

Standard errors in parentheses.

Table IA.5: Distance and the Choice of Broker Branch

This table relates the probability that a borrower uses a broker branch to the distance between the borrower and the branch. Each observation is a loan paired with one broker branch in the loan's choice set, and the dependent variable equals one if the borrower used that branch. Distance is measured from the borrower's origin to the branch, and loan fixed effects absorb each loan's choice set. Columns (1)–(4) are linear probability models weighted by inverse sampling probabilities, with the dependent variable multiplied by 100 so that coefficients are in percentage points; Columns (1) and (2) measure distance in units of 10 miles, capped at 50 miles, and Columns (3) and (4) use the log of one plus the distance in miles. Because of the weights, the dependent-variable mean in Columns (1)–(4) is the probability that a branch in the full choice set is chosen, one over the number of branches per choice set. Columns (5) and (6) are conditional logit models with a sampling correction; Column (6) adds broker branch constants on a smaller sample of alternatives, and their dependent-variable means are the shares of chosen branches among the sampled alternatives. The choice set, the stratified sample of alternatives, and the origin are as in Figure 4. The sample is loans in the matched Optimal Blue, CoreLogic, and HMDA sample with lock dates from 2018 to 2025. Standard errors are clustered by origin county.

Dep. Variable:	$\mathbb{1}$(Branch Is Chosen)					
Model:	Linear Probability (ppt)				Conditional Logit	
	(1)	(2)	(3)	(4)	(5)	(6)
Distance (10 miles)	−0.102***	−0.109***				
	(0.024)	(0.021)				
Log(1 + distance in miles)			−0.066***	−0.089***	−1.06***	−1.08***
			(0.011)	(0.011)	(0.021)	(0.019)
Loan FE	✓	✓	✓	✓	✓	✓
Branch × Lock-Month FE		✓		✓		
Branch FE						✓
Loans (choice sets)	218,192	218,192	218,192	218,192	218,192	218,180
Branches per choice set	1,681	1,681	1,681	1,681	1,681	1,681
Observations	20,452,623	20,373,529	20,452,623	20,373,529	20,452,623	5,507,731
Dep. Var. Mean (ppt)	0.06	0.06	0.06	0.06	1.07	3.96
Dep. Var. SD (ppt)	2.44	2.44	2.44	2.44	10.3	19.5

Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table IA.6: High-Cost Brokers and Mortgage Rates: First Stage

This table reports first-stage regressions of high-cost broker use on differential distance. A firm is high-cost if its shrunk previous-year price effect is in the top tercile across branches in the lock month, and low-cost if it is in the bottom tercile; loans at firms in the middle tercile are excluded. Differential distance is distance to the nearest low-cost branch minus distance to the nearest high-cost branch, in units of 10 miles. The dependent variable is a zero-one indicator. Column (1) includes origin county × lock-month effects; Column (2) adds mortgage county × lock-month effects; Columns (3)–(5) add lock-date effects, lender × lock-month effects, or both. All columns include income, loan amount, property value, origination charges, indicators for HMDA race and sex categories, credit-score, combined loan-to-value, and debt-to-income deciles, nearest-branch distance, and origin-neighborhood controls. The origin is the borrower's previous address for purchases, traced using Verisk and DataAxle, and the mortgaged property for refinances. The sample is owner-occupied brokered loans with traced origins, lagged firm classifications, and a chosen firm in the top or bottom tercile in the matched Optimal Blue, CoreLogic, and HMDA data, 2019–2025, excluding the top 5% of absolute differential distance. Standard errors are clustered by mortgage county.

Dep. Variable:	$\mathbb{1}$(High-Cost Broker Firm)				
	(1)	(2)	(3)	(4)	(5)
Differential distance (10 miles)	0.0467***	0.0481***	0.0483***	0.0419***	0.0419***
	(0.0053)	(0.0055)	(0.0056)	(0.0049)	(0.0051)
Origin County × Lock-Month FE	✓	✓	✓	✓	✓
Mortgage County × Lock-Month FE		✓	✓	✓	✓
Lock-Date FE			✓		✓
Lender × Lock-Month FE				✓	✓
Borrower controls	✓	✓	✓	✓	✓
Origin controls	✓	✓	✓	✓	✓
First-stage F-statistic	77.7	76.1	73.7	72.1	66.3
Observations	90,473	85,569	85,369	82,009	81,755
Dep. Var. Mean	0.402	0.398	0.398	0.396	0.396
Dep. Var. SD	0.490	0.490	0.490	0.489	0.489

Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table IA.7: High-Cost Brokers and Mortgage Rates: Other Definitions of High-Cost Firms

This table reports 2SLS estimates of the relationship between high-cost broker use and rate spreads under other definitions of high-cost broker firms. Firms' previous-year price effects are those of Table 3. Panel A compares firms in the top quintile of the previous-year price effect across branches in the lock month with firms in the bottom quintile; the top-quintile indicator, marked with a hat, is instrumented with distance to the nearest bottom-quintile branch minus distance to the nearest top-quintile branch, in units of 10 miles, and loans at firms between the two groups are excluded. Panel B compares firms at or above the monthly median with firms below it, using every branch and loan; the above-median indicator is instrumented with distance to the nearest below-median branch minus distance to the nearest above-median branch. Panel C instruments the previous-year price effect, in basis points, with the differential distance of Panel B on the same loans. The dependent variable is APR minus APOR, in basis points. The fixed effects and borrower and origin controls follow Table 3. Standard errors are clustered by mortgage county.

Dep. Variable:	Rate Spread (bps)				
	(1)	(2)	(3)	(4)	(5)
Panel A: Top vs. Bottom Quintile					
$\widehat{\mathbb{1}}$(Top-Quintile Broker Firm)	21.47***	21.43***	21.78***	23.23***	23.27***
	(7.09)	(7.32)	(6.64)	(8.38)	(7.02)
First-stage F-statistic	44.0	39.8	38.5	19.1	18.3
Observations	51,142	47,882	47,584	45,617	45,231
Dep. Var. Mean	22.60	21.66	21.59	21.15	21.06
Dep. Var. SD	45.08	43.84	43.67	43.05	42.74
Panel B: Above vs. Below Median					
$\widehat{\mathbb{1}}$(Above-Median Broker Firm)	11.00**	10.35**	9.95**	11.79**	10.99**
	(4.73)	(4.72)	(4.76)	(5.07)	(5.20)
First-stage F-statistic	94.2	86.8	85.9	100.0	100.6
Observations	143,584	137,328	137,157	131,894	131,712
Dep. Var. Mean	23.28	22.66	22.65	22.39	22.38
Dep. Var. SD	43.14	42.29	42.28	41.90	41.88
Panel C: Previous-Year Price Effect					
$\widehat{\text{Previous-Year Price Effect (bps)}}$	0.725**	0.687**	0.663**	0.786**	0.733**
	(0.298)	(0.300)	(0.301)	(0.310)	(0.321)
First-stage F-statistic	49.2	45.0	44.0	41.7	40.9
Observations	143,584	137,328	137,157	131,894	131,712
Dep. Var. Mean	23.28	22.66	22.65	22.39	22.38
Dep. Var. SD	43.14	42.29	42.28	41.90	41.88
Origin County × Lock-Month FE	✓	✓	✓	✓	✓
Mortgage County × Lock-Month FE		✓	✓	✓	✓
Lock-Date FE			✓		✓
Lender × Lock-Month FE				✓	✓
Borrower controls	✓	✓	✓	✓	✓
Origin controls	✓	✓	✓	✓	✓

Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table IA.8: High-Cost Brokers and Mortgage Rates: Leave-Year-Out Firm Effects

This table repeats the median-split comparison of Table IA.7, Panel B, with OLS and 2SLS estimates and high-cost firms defined by leave-year-out price effects. A firm's price effect for lock year y is estimated relative to retail with the borrower and loan controls of Figure 2 and county × lock-date effects on the firm's loans locked in every year other than y, and shrunk toward the mean for year y. A firm is high-cost if this effect is at or above the monthly median across branches. Panel A reports OLS estimates. Panel B instruments the high-cost indicator, marked with a hat, with distance to the nearest low-cost branch minus distance to the nearest high-cost branch, in units of 10 miles, with branches classified by the same effects. The bottom rows report first-stage F-statistics. The dependent variable is APR minus APOR, in basis points. Column (1) includes origin county × lock-month effects; Column (2) adds mortgage county × lock-month effects; Columns (3)–(5) additionally include lock-date effects, lender × lock-month effects, or both. Borrower controls are income, loan amount, property value, origination charges, indicators for HMDA race and sex categories, and credit-score, combined loan-to-value, and debt-to-income deciles. Origin controls include nearest-branch distance, log median household income, and the 16 neighborhood characteristics in Figure 5. The origin is the borrower's previous address for purchases, traced using Verisk and DataAxle, and the mortgaged property for refinances. The sample is owner-occupied brokered loans in the matched Optimal Blue, CoreLogic, and HMDA data, 2018–2025, with traced origins, a leave-year-out firm effect, and absolute differential distance below its 95th percentile. Standard errors are clustered by mortgage county.

Dep. Variable:	Rate Spread (bps)				
	(1)	(2)	(3)	(4)	(5)
Panel A: OLS					
$\mathbb{1}$(High-Cost Broker Firm)	12.45***	11.82***	11.73***	9.97***	9.97***
	(0.41)	(0.43)	(0.42)	(0.45)	(0.44)
Panel B: 2SLS					
$\widehat{\mathbb{1}(\text{High-Cost Broker Firm})}$	10.53	12.65*	13.13**	15.85**	15.77**
	(6.83)	(6.81)	(6.34)	(7.35)	(6.85)
Origin County × Lock-Month FE	✓	✓	✓	✓	✓
Mortgage County × Lock-Month FE		✓	✓	✓	✓
Lock-Date FE			✓		✓
Lender × Lock-Month FE				✓	✓
Borrower controls	✓	✓	✓	✓	✓
Origin controls	✓	✓	✓	✓	✓
First-stage F-statistic	49.4	43.8	46.2	39.3	41.3
Observations	179,934	172,300	172,107	165,215	165,008
Dep. Var. Mean	25.38	24.71	24.71	24.44	24.44
Dep. Var. SD	45.05	44.18	44.15	43.93	43.90

Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table IA.9: High-Cost Brokers and Mortgage Rates: Other Measures of Relative Proximity

This table repeats the 2SLS estimates of Table 3 with other measures of a borrower's relative proximity to high-cost and low-cost branches as the instrument. High-cost (low-cost) firms have a previous-year price effect in the top (bottom) tercile across branches in the lock month; loans at middle-tercile firms are excluded. Panel A uses $\ln(1+d_L)-\ln(1+d_H)$, where d_L and d_H are the distances in miles to the nearest low-cost and high-cost branches. Panel B uses the mean distance to the three nearest low-cost branches minus the mean distance to the three nearest high-cost branches, in 10-mile units, and controls for the mean distance to the three nearest branches of either type. Panels C and D use the high-cost share of the high-cost and low-cost branches within 10 and 5 miles of the origin, control for the log number of these branches, and exclude borrowers with none. Panel E winsorizes the instrument of Table 3 at its 1st and 99th percentiles instead of excluding the top 5% of its absolute values; Panels A and B exclude the top 5% of their own absolute values. Column (1) includes origin county × lock-month effects; Column (2) adds mortgage county × lock-month effects; Columns (3)–(5) additionally include lock-date effects, lender × lock-month effects, or both. All columns include the borrower and origin controls of Table 3. The origin is the borrower's previous address for purchases and the mortgaged property for refinances. The sample is owner-occupied brokered loans in the matched Optimal Blue, CoreLogic, and HMDA data, 2019–2025, with traced origins, a lagged firm classification, and a chosen firm in the top or bottom tercile. Standard errors are clustered by mortgage county.

Dep. Variable:	Rate Spread (bps)				
	(1)	(2)	(3)	(4)	(5)
Panel A: Log differential distance, $\ln(1+d_L)-\ln(1+d_H)$					
$\widehat{\mathbb{1}}$(High-Cost Broker Firm)	18.47***	20.55***	19.18***	22.37***	19.16***
	(5.97)	(5.99)	(5.91)	(6.19)	(6.20)
First-stage F-statistic	70.1	68.1	63.6	51.5	48.1
Observations	88,491	83,678	83,488	80,124	79,876
Dep. Var. Mean / SD	22.9 / 43.9	22.1 / 42.8	22.1 / 42.7	21.7 / 42.2	21.7 / 42.1
Panel B: Mean distance to the three nearest branches of each type					
$\widehat{\mathbb{1}}$(High-Cost Broker Firm)	18.51***	19.90***	18.96***	23.68***	21.28***
	(5.90)	(5.91)	(5.82)	(6.66)	(6.75)
First-stage F-statistic	61.9	59.9	59.3	59.5	58.2
Observations	89,724	84,852	84,651	81,240	80,978
Dep. Var. Mean / SD	23.0 / 43.9	22.2 / 42.7	22.1 / 42.7	21.8 / 42.1	21.8 / 42.0
Panel C: High-cost share of branches within 10 miles					
$\widehat{\mathbb{1}}$(High-Cost Broker Firm)	19.26***	18.82***	14.15**	21.51***	17.07**
	(6.65)	(6.86)	(6.67)	(8.05)	(8.19)
First-stage F-statistic	42.4	41.1	37.8	31.5	28.3
Observations	77,676	73,817	73,616	70,737	70,469
Dep. Var. Mean / SD	21.7 / 42.9	20.9 / 41.9	20.9 / 41.8	20.5 / 41.3	20.5 / 41.2
Panel D: High-cost share of branches within 5 miles					
$\widehat{\mathbb{1}}$(High-Cost Broker Firm)	16.94**	18.86***	17.11**	17.89**	14.01**
	(6.98)	(6.95)	(6.75)	(7.01)	(6.78)
First-stage F-statistic	52.2	54.9	61.6	53.4	56.1
Observations	54,144	51,321	50,970	49,055	48,567
Dep. Var. Mean / SD	20.3 / 42.1	19.5 / 41.1	19.4 / 40.9	19.0 / 40.2	18.8 / 39.9
Panel E: Differential distance winsorized at 1% and 99%					
$\widehat{\mathbb{1}}$(High-Cost Broker Firm)	22.13***	22.42***	22.99***	26.98***	25.74***
	(5.34)	(5.45)	(5.10)	(5.69)	(5.62)
First-stage F-statistic	82.5	78.2	73.6	78.0	69.8
Observations	94,440	89,416	89,220	85,754	85,506
Dep. Var. Mean / SD	23.1 / 43.9	22.3 / 42.7	22.2 / 42.7	21.9 / 42.1	21.9 / 42.1
Origin County × Lock-Month FE	✓	✓	✓	✓	✓
Mortgage County × Lock-Month FE		✓	✓	✓	✓
Lock-Date FE			✓		✓
Lender × Lock-Month FE				✓	✓
Borrower and origin controls	✓	✓	✓	✓	✓

Standard errors in parentheses.
* $p<0.10$, ** $p<0.05$, *** $p<0.01$

Table IA.10: High-Cost Brokers and Other Loan Outcomes: Differential Distance IV

This table reports 2SLS estimates of the relationship between high-cost broker use and other loan outcomes, with the instrument, controls, and sample of Table 3. High-cost (low-cost) firms have a previous-year price effect in the top (bottom) tercile across branches in the lock month; loans at middle-tercile firms are excluded. Each row is a separate regression, and the FE column gives the column of Table 3 whose fixed effects are used. Panel A reports rate and broker-compensation outcomes; its first row repeats Table 3, Column (2). Note rates and lender-paid compensation (comp.) are in basis points, the latter as a share of the loan amount on loans for which the lender pays the broker. Panel B examines whether lower upfront fees offset the higher rates paid by borrowers using high-cost brokers. Origination charges, discount points, lender credits, and total loan costs are HMDA dollar amounts, winsorized at the 1st and 99th percentiles by year, with missing discount points and lender credits coded as zero; these regressions exclude origination charges from the controls. Standard errors are clustered by mortgage county.

Dependent variable	FE column	2SLS	SE	First-stage F	Observations	Mean	SD
Panel A: Rates and broker compensation							
Rate spread (bps)	(2)	23.99***	(5.43)	76.1	85,569	22.15	42.73
Note rate (bps)	(2)	22.61***	(7.24)	76.0	85,840	344.41	118.81
Lender-paid comp. (bps)	(2)	57.45***	(15.48)	28.6	45,785	200.25	50.65
Lender-paid comp. (bps)	(5)	49.25**	(19.93)	22.0	43,593	199.80	50.63
Panel B: Upfront fees							
Origination charges ($)	(2)	916	(565)	75.9	85,840	3,292	3,230
Discount points ($)	(2)	497*	(266)	75.2	85,728	627	1,433
Lender credits ($)	(2)	77	(508)	74.6	84,772	2,103	2,714
Total loan costs ($)	(2)	892	(698)	76.5	85,727	5,573	3,736

Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table IA.11: High-Cost Brokers and Loan Timing: Differential Distance IV

This table estimates the relationship between high-cost broker use and loan timing by 2SLS. High-cost (low-cost) firms have a previous-year price effect in the top (bottom) tercile across branches in the lock month; loans at middle-tercile firms are excluded. The high-cost indicator, marked with a hat, is instrumented by distance to the nearest low-cost branch minus distance to the nearest high-cost branch. Origins are previous addresses for purchases and mortgaged properties for refinances. Panel A uses days from the Optimal Blue lock date to the CoreLogic mortgage date, winsorized at the 1st and 99th percentiles (6 and 102 days). Columns add origin county × month, mortgage county × month, lock-date, and lender × month effects as indicated. Panel B uses the two county-month effects and four outcomes: lock-to-close days, requested lock-period days, an indicator for a requested period above 30 days, and slack (requested period minus lock-to-close days). All columns include the borrower and origin controls of Table 3. The sample is owner-occupied brokered loans in the matched Optimal Blue, CoreLogic, and HMDA IV data, 2019–2025, with a chosen firm in the top or bottom tercile and both timing measures. Standard errors are clustered by mortgage county.

Panel A: Lock-to-Close Days

Dep. Variable:	Lock-to-Close Days				
	(1)	(2)	(3)	(4)	(5)
$\widehat{\mathbb{1}(\text{High-Cost Broker Firm})}$	3.04	2.88	2.76	3.83	3.66
	(2.93)	(3.07)	(3.11)	(3.35)	(3.33)
Origin County × Lock-Month FE	✓	✓	✓	✓	✓
Mortgage County × Lock-Month FE		✓	✓	✓	✓
Lock-Date FE			✓		✓
Lender × Lock-Month FE				✓	✓
Borrower controls	✓	✓	✓	✓	✓
Origin controls	✓	✓	✓	✓	✓
First-stage F-statistic	77.8	76.3	73.7	72.1	66.3
Observations	90,466	85,561	85,360	82,005	81,750
Dep. Var. Mean	32.25	32.39	32.39	32.39	32.40
Dep. Var. SD	17.05	17.11	17.12	16.98	16.98

Panel B: Measures of Loan Timing

Dep. Variable:	Lock-to-Close Days	Requested Lock Period (Days)	$\mathbb{1}$(Lock Period $>$ 30 Days)	Slack (Days)
	(1)	(2)	(3)	(4)
$\widehat{\mathbb{1}(\text{High-Cost Broker Firm})}$	2.88	1.05	0.012	−1.83
	(3.07)	(2.27)	(0.089)	(2.46)
Origin County × Lock-Month FE	✓	✓	✓	✓
Mortgage County × Lock-Month FE	✓	✓	✓	✓
Borrower controls	✓	✓	✓	✓
Origin controls	✓	✓	✓	✓
First-stage F-statistic	76.3	76.3	76.3	76.3
Observations	85,561	85,561	85,561	85,561
Dep. Var. Mean	32.39	36.11	0.397	3.72
Dep. Var. SD	17.11	11.69	0.489	14.84

Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table IA.12: Borrower and Neighborhood Characteristics by the Sign of the Instrument

This table compares borrower and neighborhood characteristics by relative proximity to high-cost and low-cost branches. Differential distance dd is distance to the nearest low-cost branch minus distance to the nearest high-cost branch. High-cost (low-cost) firms are in the top (bottom) tercile of previous-year price effects. The groups are $dd < 0$ and $dd \geq 0$. The table reports unconditional means, medians, mean differences, and differences divided by the characteristic's standard deviation. Origins are previous addresses for purchases and mortgaged properties for refinances; neighborhood measures come from the origin block group and tract in the American Community Survey. Figure 5 defines the linked HMDA denial. Cost-burdened owners spend 30% or more of income on housing costs, limited-English households have no member aged 14 or older who speaks only English or speaks English very well, and veterans are civilian adults who served in the armed forces. Dollar values are in thousands, neighborhood shares are in percent, and variables marked (1/0) are indicators. The sample is owner-occupied brokered loans with traced origins and a chosen firm in the top or bottom tercile in the matched Optimal Blue, CoreLogic, and HMDA IV data, 2019–2025. Standard errors are clustered by mortgage county.

	$dd < 0$		$dd \geq 0$		Difference		
	Mean	Median	Mean	Median	Diff.	SE	Diff./SD
Credit score	755	765	752	761	−3.84***	(0.679)	−0.086
Loan amount ($000s)	366	345	338	315	−28.2***	(5.35)	−0.175
Property value ($000s)	586	515	515	455	−70.4***	(13.6)	−0.207
Income ($000s)	131	112	128	108	−2.5	(1.55)	−0.030
Combined loan-to-value	68.3	71.7	70.7	74.6	2.41***	(0.448)	0.134
Debt-to-income	36.7	36	36.9	36	0.201**	(0.0927)	0.026
Origination charges ($000s)	3.31	1.99	3.53	2.3	0.221**	(0.0893)	0.066
Non-white applicant (1/0)	0.182	0	0.155	0	−0.026**	(0.013)	−0.071
Female applicant (1/0)	0.397	0	0.400	0	0.0029	(0.0066)	0.006
Applicant age (bin midpoint)	46.3	40	46.2	40	−0.176	(0.186)	−0.014
Has co-applicant (1/0)	0.470	0	0.451	0	−0.020***	(0.0054)	−0.039
First-time homebuyer (1/0)	0.141	0	0.154	0	0.013**	(0.0054)	0.036
Self-employed (1/0)	0.139	0	0.147	0	0.0076**	(0.0030)	0.022
Linked HMDA denial, same or previous year (1/0)	0.0071	0	0.0067	0	−0.0004	(0.0006)	−0.005
Log population density	6.97	7.3	6.85	7.14	−0.123**	(0.0594)	−0.083
Percent bachelor's degree	44.1	43	42	40.3	−2.11***	(0.591)	−0.100
Percent in poverty	5.49	2.51	5.87	2.75	0.38***	(0.123)	0.045
Tract: unemployment rate (percent)	4.79	4.16	4.86	4.2	0.0679	(0.0556)	0.021
Percent minority	39.8	34.2	37.6	30.9	−2.19**	(1.09)	−0.080
Tract: percent foreign-born	16.6	13	16.3	12.3	−0.308	(0.654)	−0.023
Percent limited-English households	3.89	0.992	4.17	0.924	0.285	(0.242)	0.041
Median age (years)	40.6	39.5	41.2	40.2	0.59***	(0.18)	0.067
Percent married-couple households	56.6	57.9	55.9	57.1	−0.687	(0.504)	−0.036
Percent veterans	6.51	5.56	6.33	5.4	−0.173	(0.164)	−0.036
Percent renter	27.6	20.8	26.7	19.5	−0.955*	(0.527)	−0.040
Tract: percent of housing built 2010 or later	10.3	3.73	8.92	3.59	−1.39**	(0.597)	−0.092
Tract: vacancy rate (percent)	6.5	4.8	7.15	5.18	0.655***	(0.203)	0.088
Tract: percent of owner units with a mortgage	69.5	70.5	67.4	68.4	−2.17***	(0.296)	−0.176
Percent cost-burdened mortgaged owners	29.6	27.5	29.5	26.7	−0.155	(0.409)	−0.009
Percent second mortgage or home-equity loan	13.7	11.6	13.7	11.6	−0.0119	(0.232)	−0.001
Observations	59,831		54,071		113,902		

Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table IA.13: Borrower and Neighborhood Characteristics by the Sign of the Instrument, Within County and Month

This table compares borrower and neighborhood characteristics by relative proximity to high-cost and low-cost branches, adjusting for county and month. Differential distance dd is distance to the nearest low-cost branch minus distance to the nearest high-cost branch; high-cost (low-cost) firms are in the top (bottom) tercile of previous-year price effects. Means and medians are unconditional within the $dd < 0$ and $dd \geq 0$ groups. Adjusted differences come from regressions of each characteristic on an indicator for $dd \geq 0$, with mortgage county × lock-month effects, and are also reported in standard-deviation units. Dollar values are in thousands, neighborhood shares are in percent, and variables marked (1/0) are indicators. Neighborhood characteristics are from the origin block group and tract in the American Community Survey; Figure 5 defines the linked HMDA denial. Cost-burdened owners spend 30% or more of income on housing costs, limited-English households have no member aged 14 or older who speaks only English or speaks English very well, and veterans are civilian adults who served in the armed forces. The sample is the matched Optimal Blue, CoreLogic, and HMDA IV sample with a chosen firm in the top or bottom tercile, 2019–2025. Standard errors are clustered by mortgage county.

	$dd < 0$		$dd \geq 0$		Difference Within County-Month		
	Mean	Median	Mean	Median	Diff.	SE	Diff./SD
Credit score	755	765	752	761	−1.10***	(0.415)	−0.025
Loan amount ($000s)	366	345	338	315	−1.43	(1.82)	−0.009
Property value ($000s)	586	515	515	455	−3.65	(4.95)	−0.011
Income ($000s)	131	112	128	108	−1.46	(1.12)	−0.017
Combined loan-to-value	68.3	71.7	70.7	74.6	0.239	(0.175)	0.013
Debt-to-income	36.7	36.0	36.9	36.0	0.165**	(0.070)	0.021
Origination charges ($000s)	3.31	1.99	3.53	2.30	0.015	(0.032)	0.005
Non-white applicant (1/0)	0.182	0	0.155	0	−0.0025	(0.0068)	−0.007
Female applicant (1/0)	0.397	0	0.400	0	0.0042	(0.0061)	0.009
Applicant age (bin midpoint)	46.3	40.0	46.2	40.0	0.013	(0.129)	0.001
Has co-applicant (1/0)	0.470	0	0.451	0	−0.0024	(0.0059)	−0.005
First-time homebuyer (1/0)	0.141	0	0.154	0	0.0034	(0.0035)	0.010
Self-employed (1/0)	0.139	0	0.147	0	−0.0040	(0.0040)	−0.011
Linked HMDA denial, same or previous year (1/0)	0.0071	0	0.0067	0	−0.0008	(0.0007)	−0.010
Log population density	6.97	7.30	6.85	7.14	−0.026	(0.040)	−0.018
Percent bachelor's degree	44.1	43.0	42.0	40.3	−1.74***	(0.463)	−0.082
Percent in poverty	5.49	2.51	5.87	2.75	0.336***	(0.121)	0.040
Tract: unemployment rate (percent)	4.79	4.16	4.86	4.20	0.085**	(0.043)	0.026
Percent minority	39.8	34.2	37.6	30.9	0.951	(0.747)	0.035
Tract: percent foreign-born	16.6	13.0	16.3	12.3	0.241	(0.370)	0.018
Percent limited-English households	3.89	0.992	4.17	0.924	0.351*	(0.188)	0.050
Median age (years)	40.6	39.5	41.2	40.2	−0.046	(0.141)	−0.005
Percent married-couple households	56.6	57.9	55.9	57.1	−0.417	(0.448)	−0.022
Percent veterans	6.51	5.56	6.33	5.40	−0.090	(0.075)	−0.019
Percent renter	27.6	20.8	26.7	19.5	0.401	(0.491)	0.017
Tract: percent of housing built 2010 or later	10.3	3.73	8.92	3.59	−0.506	(0.325)	−0.033
Tract: vacancy rate (percent)	6.50	4.80	7.15	5.18	0.179	(0.132)	0.024
Tract: percent of owner units with a mortgage	69.5	70.5	67.4	68.4	−0.618**	(0.248)	−0.050
Percent cost-burdened mortgaged owners	29.6	27.5	29.5	26.7	0.551***	(0.205)	0.031
Percent second mortgage or home-equity loan	13.7	11.6	13.7	11.6	−0.118	(0.145)	−0.010
Observations	59,831		54,071		113,902		
Mortgage County × Lock-Month FE					✓		

Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table IA.14: Differential Distance and Retail Borrowers' Rates: Placebo Test

This table reports a placebo test of the differential-distance instrument on retail refinance borrowers, who obtain their loans directly from a lender without a broker. Differential distance is the distance from the mortgaged property to the nearest low-cost broker branch minus the distance to the nearest high-cost broker branch, in 10-mile units, with the lock month's tercile classification of Table 3. Panel A reports, for brokered refinances in the sample of Table 3, the OLS regression of the rate spread on differential distance (the reduced form), the first-stage regression of the high-cost broker indicator on differential distance, and the 2SLS estimate. Panel B repeats the reduced-form regression for retail refinances. Because retail borrowers do not use a broker, there is no first stage, and differential distance should not affect their rates unless high-cost brokers locate near borrowers who pay higher rates regardless of the broker. In Panel B, Column (1) uses all retail refinances and excludes absolute differential distances above the 95th percentile of the brokered IV sample; Column (2) uses the retail sample's own 95th percentile. Columns (3)–(5) restrict the sample to the counties and months, the county-months, or the refinance county-months of the brokered IV sample. Column (6) adds lender × lock-month fixed effects. All regressions include mortgage county × lock-month fixed effects, the borrower controls of Table 3, and origin controls measured at the property. The sample is owner-occupied conventional conforming fixed-rate retail refinances in the matched Optimal Blue, CoreLogic, and HMDA data, 2019–2025. Standard errors are clustered by mortgage county.

Panel A: Brokered refinances in the IV sample

Dep. Variable:	Reduced Form Rate Spread (bps)	First Stage $\mathbb{1}$(High-Cost)	2SLS Rate Spread (bps)
Differential distance (10 miles)	1.113***	0.0483***	
	(0.283)	(0.0058)	
$\widehat{\mathbb{1}(\text{High-Cost Broker Firm})}$			23.06***
			(5.22)
Observations	60,871	60,871	60,871
Dep. Var. Mean	17.90	0.385	17.90
Dep. Var. SD	37.81	0.487	37.81

Panel B: Retail refinances (placebo)

Dep. Variable:	Rate Spread (bps)					
	(1)	(2)	(3)	(4)	(5)	(6)
Differential distance (10 miles)	0.105*	0.126**	0.104	0.120	0.151**	0.117**
	(0.063)	(0.056)	(0.064)	(0.073)	(0.076)	(0.052)
Own trim of the instrument		✓				
IV-sample counties and months			✓			
IV-sample county-months				✓		
Refinance IV-sample county-months					✓	
Mortgage County × Lock-Month FE	✓	✓	✓	✓	✓	✓
Lender × Lock-Month FE						✓
Borrower and origin controls	✓	✓	✓	✓	✓	✓
Observations	2,566,633	2,722,676	2,518,096	1,948,281	1,824,171	2,550,703
Dep. Var. Mean	25.55	25.29	25.43	23.23	22.39	25.53
Dep. Var. SD	39.83	39.82	39.71	38.08	37.27	39.77

Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table IA.15: The Steering Gap Under Other Diligent Brokers

This table re-estimates the main column of each table with the steering gap measured against other diligent brokers. Each panel reports the coefficient of interest from one column of a main table, with the same controls and fixed effects: Panel A from Table 4, Column (2); Panels B and C from Table 5, Panel A, Column (2), and Panel B, Column (1); Panel D from Table 6, Column (3); and Panel E from Table 7, Column (2). In Panels B to D, the past firm gap is the trailing mean of the same measure. Column (1) uses the observed gap against the lowest predicted rate spread, before the diligent-broker adjustment. Column (2) repeats the steering gap of the main tables, in which the diligent broker shops the state set and every lender quotes with a noise standard deviation of 35 basis points. Column (3) gives each lender its own noise standard deviation, the leave-one-out root mean squared error of its pricing model, and Column (4) a common standard deviation of 111 basis points. Column (5) measures the steering gap against a diligent broker that shops only the priced lenders on the broker's menu, with the same noise of 35 basis points, so it leaves out what a diligent broker gains by shopping the whole state set. The value of 111 basis points is the noise level at which simulated diligent brokers in Texas and Ohio, with quote noise of 30 basis points and rates priced by pricing models refit on the simulated loans, have a mean steering gap of zero. This level was calibrated for a steering gap measured against the broker's own menu rather than the state set, so here it is a sensitivity check only. The sample is brokered loans with a measured steering gap. Standard errors are clustered by county.

Gap:	Observed (Minimum)	Steering Gap (Observed minus Simulated)			
Diligent broker shops:		State Set	State Set	State Set	Menu
Quote noise SD:		35 bps	Lender RMSE	111 bps	35 bps
	(1)	(2)	(3)	(4)	(5)
Panel A: Rate Spread on the Gap (Table 4, Column 2)					
Coefficient	0.677***	0.729***	0.740***	0.767***	0.613***
	(0.016)	(0.016)	(0.016)	(0.017)	(0.015)
Dep. Var. Mean	30.53	30.53	30.53	30.53	30.53
Dep. Var. SD	49.09	49.09	49.09	49.09	49.09
Observations	304,526	304,526	304,526	304,526	304,526
Panel B: Gap on the Past Firm Gap (Table 5, Panel A, Column 2)					
Coefficient	0.450***	0.456***	0.450***	0.449***	0.483***
	(0.019)	(0.019)	(0.019)	(0.018)	(0.019)
Dep. Var. Mean	26.71	10.88	6.90	1.39	8.25
Dep. Var. SD	21.62	19.35	19.21	16.88	19.16
Observations	260,013	260,013	260,013	260,013	260,013
Panel C: Firm Gap on Its Value 12 Months Earlier (Table 5, Panel B, Column 1)					
Coefficient	0.612***	0.602***	0.594***	0.568***	0.664***
	(0.023)	(0.025)	(0.025)	(0.029)	(0.018)
Dep. Var. Mean	29.93	13.29	9.16	2.74	9.75
Dep. Var. SD	18.06	16.41	16.37	13.90	15.98
Observations	64,232	64,232	64,232	64,232	64,232
Panel D: Rate Spread on the Past Firm Gap (Table 6, Column 3)					
Coefficient	6.57***	7.06***	7.26***	7.89***	4.15***
	(0.43)	(0.42)	(0.43)	(0.41)	(0.33)
Dep. Var. Mean	27.69	27.69	27.69	27.69	27.69
Dep. Var. SD	45.15	45.15	45.15	45.15	45.15
Observations	245,890	245,890	245,890	245,890	245,890
Panel E: Broker Compensation on the Gap (Table 7, Column 2)					
Coefficient	0.099***	0.114***	0.117***	0.130***	0.104***
	(0.011)	(0.013)	(0.013)	(0.014)	(0.012)
Dep. Var. Mean	205.39	205.39	205.39	205.39	205.39
Dep. Var. SD	50.96	50.96	50.96	50.96	50.96
Observations	178,648	178,648	178,648	178,648	178,648

Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table IA.16: The Observed Gap and Realized Rates

This table repeats Table 4 with the observed gap in place of the steering gap. It regresses realized rate spreads on the loan's observed gap, both in basis points. The observed gap is the chosen lender's predicted spread minus the lowest predicted spread in the state lender set, before the diligent-broker adjustment. Column (1) includes county × lock-month effects. Column (2) adds firm effects, Column (3) adds standardized firm HHI over the previous 12 months, and Column (4) adds lock-date effects. HHI is the sum of squared lender shares in the firm's loans. Controls are income, credit score, loan amount, property value, combined loan-to-value ratio, debt-to-income ratio, race, ethnicity, sex, occupancy, property type, number of units, loan purpose, and loan term. The sample is brokered loans with a measured steering gap. Standard errors are clustered by county.

Dep. Variable:	Rate Spread (bps)			
	(1)	(2)	(3)	(4)
Observed gap (bps)	0.722***	0.677***	0.678***	0.673***
	(0.017)	(0.016)	(0.016)	(0.015)
Broker firm HHI (std.)			0.647***	0.539***
			(0.149)	(0.155)
County × Lock-Month FE	✓	✓	✓	✓
Lock-Date FE				✓
Broker Firm FE		✓	✓	✓
Borrower and loan controls	✓	✓	✓	✓
Observations	308,810	304,526	297,131	296,937
Adjusted R^2	0.491	0.544	0.543	0.569
Dep. Var. Mean	30.9	30.5	30.5	30.5
Dep. Var. SD	49.7	49.1	49.0	49.0

Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table IA.17: Past and Current Observed Gaps

This table repeats Table 5 with the observed gap in place of the steering gap. Panel A regresses a loan's observed gap on its firm's mean observed gap during the previous 12 months. All columns include borrower and loan controls and standardized firm HHI. Column (1) includes county × lock-month effects, Column (2) adds firm effects, and Column (3) adds lock-date effects. Panel B regresses the firm's mean observed gap in months $t-12$ to $t-1$ on its mean in months $t-24$ to $t-13$. Observations are firm-months assigned to the firm's modal county of the year. Columns (1)–(3) include county × month, state × month, or month effects, without firm effects. Each window requires at least five loans. The observed gap is the chosen lender's predicted spread minus the lowest predicted spread in the state lender set, before the diligent-broker adjustment. Gaps are in basis points. Controls are income, credit score, loan amount, property value, combined loan-to-value ratio, debt-to-income ratio, race, ethnicity, sex, occupancy, property type, number of units, loan purpose, and loan term. The sample is brokered loans with a measured steering gap. Standard errors are clustered by county.

Panel A: Past Firm Observed Gap and Focal Loan Observed Gap

Dep. Variable:	Focal Loan Observed Gap (bps)		
	(1)	(2)	(3)
Past firm observed gap (bps, 12 months)	0.723***	0.450***	0.448***
	(0.009)	(0.019)	(0.019)
Broker firm HHI (std.)	−0.091	−0.001	0.002
	(0.079)	(0.141)	(0.140)
County × Lock-Month FE	✓	✓	✓
Lock-Date FE			✓
Broker Firm FE		✓	✓
Borrower and loan controls	✓	✓	✓
Observations	260,440	260,013	259,822
Adjusted R^2	0.502	0.538	0.541
Dep. Var. Mean	26.72	26.71	26.71
Dep. Var. SD	21.63	21.62	21.62

Panel B: Autocorrelation of the Firm Observed Gap

Dep. Variable:	Firm Observed Gap, Months $t-12$ to $t-1$ (bps)		
	(1)	(2)	(3)
Firm observed gap, months $t-24$ to $t-13$ (bps)	0.612***	0.628***	0.689***
	(0.023)	(0.018)	(0.018)
County × Month FE	✓		
State × Month FE		✓	
Month FE			✓
Observations	64,232	74,601	74,964
Adjusted R^2	0.596	0.567	0.485
Dep. Var. Mean	29.93	29.67	29.58
Dep. Var. SD	18.06	18.19	18.21

Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table IA.18: Past Firm Observed Gap and Mortgage Rates

This table repeats Table 6 with the observed gap in place of the steering gap. It regresses realized rate spreads on past firm observed gaps. The past gap averages the firm's loans in the previous 12 months, with at least five loans, and is divided by its standard deviation. The observed gap is the chosen lender's predicted spread minus the lowest predicted spread in the state lender set, before the diligent-broker adjustment. All columns include county × lock-month effects. Columns (3)–(5) add firm effects, and Column (5) adds lock-date effects. Columns (2), (4), and (5) control for standardized firm HHI, the sum of squared lender shares over the previous year. The dependent variable is in basis points. Controls are income, credit score, loan amount, property value, combined loan-to-value ratio, debt-to-income ratio, race, ethnicity, sex, occupancy, property type, number of units, loan purpose, and loan term. The sample is brokered loans with a measured steering gap. Standard errors are clustered by county.

Dep. Variable:	Rate Spread (bps)				
	(1)	(2)	(3)	(4)	(5)
Past firm observed gap (std.)	9.127***	9.154***	6.573***	6.613***	6.310***
	(0.443)	(0.446)	(0.428)	(0.430)	(0.399)
Broker firm HHI (std.)		0.053		0.682***	0.555**
		(0.166)		(0.242)	(0.239)
County × Lock-Month FE	✓	✓	✓	✓	✓
Lock-Date FE					✓
Broker Firm FE			✓	✓	✓
Borrower and loan controls	✓	✓	✓	✓	✓
Observations	246,342	244,929	245,890	244,492	244,297
Adjusted R^2	0.434	0.434	0.486	0.485	0.518
Dep. Var. Mean	27.7	27.8	27.7	27.8	27.8
Dep. Var. SD	45.2	45.2	45.1	45.2	45.2

Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table IA.19: The Observed Gap and Broker Compensation

This table repeats Table 7 with the observed gap in place of the steering gap. It regresses lender-paid broker compensation on the loan's observed gap. Compensation is measured in basis points of the loan amount, and the gap in basis points of rate spread. The observed gap is the chosen lender's predicted spread minus the lowest predicted spread in the state lender set, before the diligent-broker adjustment. Column (1) has no fixed effects. Column (2) includes county × lock-month and firm effects; Column (3) adds lock-date effects. Controls are income, credit score, loan amount, property value, combined loan-to-value ratio, debt-to-income ratio, race, ethnicity, sex, occupancy, property type, number of units, loan purpose, and loan term. The sample is brokered loans with a measured steering gap. Only loans with recorded compensation are included. Standard errors are clustered by county.

Dep. Variable:	Broker Compensation (bps of Loan Amount)		
	(1)	(2)	(3)
Observed gap (bps)	0.211***	0.099***	0.099***
	(0.014)	(0.011)	(0.011)
County × Lock-Month FE		✓	✓
Lock-Date FE			✓
Broker Firm FE		✓	✓
Borrower and loan controls	✓	✓	✓
Observations	197,611	178,648	178,451
Adjusted R^2	0.256	0.887	0.888
Dep. Var. Mean	207.2	205.4	205.4
Dep. Var. SD	51.1	51.0	51.0

Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table IA.20: The Steering Gap: Other Fixed Effects and Clustering

This table repeats the main within-firm specifications of the steering-gap analysis with other fixed effects and clustering. Column (1) regresses the rate spread on the loan's steering gap (Table 4, Column (2)); Column (2) regresses the rate spread on the firm's standardized past steering gap (Table 6, Column (3)); and Column (3) regresses lender-paid compensation on the steering gap (Table 7, Column (2)). These specifications include county × lock-month and broker firm fixed effects and cluster standard errors by county; each row changes one element. Panel A uses loan-officer fixed effects in place of broker firm fixed effects, or county × lock-date fixed effects in place of county × lock-month fixed effects. Panel B keeps the fixed effects and clusters standard errors by county and broker firm or by broker firm alone. The steering gap is the chosen lender's predicted spread minus the expected predicted spread under simulated diligent shopping across the state lender set, with independent normal pricing errors with a standard deviation of 35 basis points. Controls include income, credit score, loan amount, property value, combined loan-to-value ratio, debt-to-income ratio, race, ethnicity, sex, occupancy, property type, number of units, loan purpose, and loan term. The sample includes brokered conventional conforming fixed-rate loans matched across Optimal Blue, CoreLogic, and HMDA, 2018–2025, with usable lender-price predictions; Column (3) uses lender-paid loans with recorded compensation.

Dep. Variable: Regressor: Main specification:	Rate Spread Steering Gap Table 4, (2)	Rate Spread Past Firm Gap Table 6, (3)	Compensation Steering Gap Table 7, (2)
	(1)	(2)	(3)
Main specification	0.729***	7.06***	0.114***
	(0.016)	(0.42)	(0.013)
Observations	304,526	245,890	178,648
Dep. Var. Mean	30.53	27.69	205.39
Dep. Var. SD	49.09	45.15	50.96
Panel A: Fixed effects			
Loan-officer FE instead of firm FE	0.692***	6.04***	0.097***
	(0.016)	(0.42)	(0.011)
Observations	298,686	241,024	174,118
Dep. Var. Mean	30.18	27.42	205.03
Dep. Var. SD	48.53	44.68	50.81
County × lock-date FE	0.675***	4.70***	0.113***
	(0.019)	(0.60)	(0.019)
Observations	193,917	153,677	106,340
Dep. Var. Mean	27.04	24.45	197.60
Dep. Var. SD	46.10	42.55	50.71
Panel B: Clustering			
County and broker firm	0.729***	7.06***	0.114***
	(0.018)	(0.52)	(0.013)
Observations	304,526	245,890	178,648
Dep. Var. Mean	30.53	27.69	205.39
Dep. Var. SD	49.09	45.15	50.96
Broker firm	0.729***	7.06***	0.114***
	(0.013)	(0.52)	(0.009)
Observations	304,526	245,890	178,648
Dep. Var. Mean	30.53	27.69	205.39
Dep. Var. SD	49.09	45.15	50.96

Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table IA.21: The Steering Gap Excluding the Largest Wholesale Lenders

This table repeats the main within-firm specifications of the steering-gap analysis after excluding loans funded by the largest lenders. The columns are the three specifications of Table IA.20, with county × lock-month and broker firm fixed effects. Lenders are ranked by their number of loans in the sample; the largest funds 41.5% of them and each of the next four funds 4.7% to 5.7%. Excluded lenders remain in the comparison sets, and the past firm gap in Column (2) is computed from all of the firm's loans. The steering gap, controls, and sample are defined as in Table IA.20. Standard errors are clustered by county.

Dep. Variable: Regressor: Main specification:	Rate Spread Steering Gap Table 4, (2)	Rate Spread Past Firm Gap Table 6, (3)	Compensation Steering Gap Table 7, (2)
	(1)	(2)	(3)
Main specification	0.729***	7.06***	0.114***
	(0.016)	(0.42)	(0.013)
Observations	304,526	245,890	178,648
Dep. Var. Mean	30.53	27.69	205.39
Dep. Var. SD	49.09	45.15	50.96
Excluding the largest lender	0.719***	7.25***	0.123***
	(0.019)	(0.51)	(0.016)
Observations	175,102	132,027	88,481
Dep. Var. Mean	36.04	32.43	213.26
Dep. Var. SD	53.79	48.77	48.84
Excluding the 2nd-largest lender	0.742***	7.18***	0.113***
	(0.016)	(0.43)	(0.012)
Observations	284,861	230,562	166,093
Dep. Var. Mean	29.94	27.04	204.00
Dep. Var. SD	49.17	45.06	50.88
Excluding the 3rd-largest lender	0.735***	6.86***	0.112***
	(0.016)	(0.43)	(0.013)
Observations	286,942	232,140	167,346
Dep. Var. Mean	30.63	27.74	205.66
Dep. Var. SD	49.40	45.29	51.24
Excluding the 4th-largest lender	0.715***	7.38***	0.066***
	(0.017)	(0.43)	(0.009)
Observations	288,256	238,047	174,588
Dep. Var. Mean	30.99	27.99	206.10
Dep. Var. SD	49.49	45.39	50.32
Excluding the 5th-largest lender	0.741***	7.12***	0.113***
	(0.015)	(0.42)	(0.013)
Observations	290,199	234,454	168,485
Dep. Var. Mean	30.44	27.75	204.85
Dep. Var. SD	48.87	45.04	51.39
Excluding the two largest lenders	0.727***	7.08***	0.126***
	(0.018)	(0.53)	(0.014)
Observations	155,554	116,836	76,075
Dep. Var. Mean	35.66	31.78	211.62
Dep. Var. SD	54.54	49.12	48.64

Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table IA.22: The Steering Gap in Subsamples

This table repeats the main within-firm specifications of the steering-gap analysis in subsamples. The columns are the three specifications of Table IA.20, with county × lock-month and broker firm fixed effects. Panels A to C split loans by purpose, lock year, and credit score. Panel D splits broker firms at the loan-weighted median number of sample loans, 126. The steering gap, controls, and sample are defined as in Table IA.20. Standard errors are clustered by county.

Dep. Variable:	Rate Spread	Rate Spread	Compensation
Regressor:	Steering Gap	Past Firm Gap	Steering Gap
Main specification:	Table 4, (2)	Table 6, (3)	Table 7, (2)
	(1)	(2)	(3)
Main specification	0.729***	7.06***	0.114***
	(0.016)	(0.42)	(0.013)
Observations	304,526	245,890	178,648
Dep. Var. Mean	30.53	27.69	205.39
Dep. Var. SD	49.09	45.15	50.96
Panel A: Loan purpose			
Purchases	0.795***	6.73***	0.101***
	(0.016)	(0.56)	(0.015)
Observations	138,412	109,078	79,126
Dep. Var. Mean	39.74	35.91	218.24
Dep. Var. SD	55.19	50.69	48.69
Refinances	0.505***	3.98***	0.124***
	(0.012)	(0.53)	(0.021)
Observations	150,590	125,292	88,487
Dep. Var. Mean	20.82	19.70	192.46
Dep. Var. SD	39.14	37.41	49.74
Panel B: Lock year			
2018–2021	0.617***	3.50***	0.129***
	(0.015)	(0.62)	(0.014)
Observations	229,423	189,820	143,756
Dep. Var. Mean	26.50	24.53	203.72
Dep. Var. SD	42.19	40.15	51.53
2022–2025	0.745***	6.68***	0.099***
	(0.023)	(0.81)	(0.016)
Observations	73,217	55,608	33,330
Dep. Var. Mean	42.67	38.40	212.49
Dep. Var. SD	64.06	57.77	47.49
Panel C: Credit score			
Below 740	0.809***	7.75***	0.127***
	(0.028)	(0.70)	(0.021)
Observations	92,002	71,763	48,290
Dep. Var. Mean	52.30	48.42	215.80
Dep. Var. SD	56.63	51.71	48.55
740 or above	0.611***	5.70***	0.098***
	(0.014)	(0.46)	(0.011)
Observations	196,930	162,614	118,846
Dep. Var. Mean	18.80	17.23	199.49
Dep. Var. SD	39.18	36.80	51.14
Panel D: Broker firm size			
Larger firms	0.721***	7.76***	0.093***
	(0.022)	(0.74)	(0.019)
Observations	150,409	145,206	89,123
Dep. Var. Mean	26.30	25.88	203.61
Dep. Var. SD	43.76	43.45	51.40
Smaller firms	0.714***	5.53***	0.122***
	(0.018)	(0.53)	(0.015)
Observations	143,530	92,153	81,616
Dep. Var. Mean	34.22	29.66	205.80
Dep. Var. SD	53.16	46.65	50.44

Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$