BANK OF ENGLAND

Staff Working Paper No. 967

Value of information, search, and competition in the UK mortgage market

Mateusz Myśliwski and May Rostom

March 2022

BANK OF ENGLAND

Staff Working Paper No. 967

Value of information, search, and competition in the UK mortgage market

Mateusz Myśliwski[1] and May Rostom[2]

Abstract

We formulate a structural model of search with lender and borrower heterogeneity to estimate the value of information provided to UK households by mortgage brokers. Using administrative data on loans originating in 2016 and 2017, we document the existence of a substantial degree of unexplained price dispersion, and observe that while mortgages obtained from brokers are cheaper, borrowers who use intermediaries pay more once commissions are factored in. Assuming that borrowers with high search costs are more likely to use brokers, we nonparametrically estimate the distributions of search, and the banks' costs of providing these loans. Our results show that search costs vary by demographic groups, and that broker presence exerts negative pressure on lenders' market power. Compared to a world where broker advice is unavailable, we estimate their presence reduces average monthly mortgage costs by 21%, and welfare losses arising from search frictions by 70% – although the results differ by borrower and loan characteristics. We also find that regulation in support of market centralization halves lenders' markups and lowers monthly costs of an average mortgage by 4.4%.

Key words: Mortgage markets, consumer search, intermediation, auction estimation.

JEL classification: C57, D83, G21, L85.

(1) Norwegian School of Economics. Email: mateusz.mysliwksi@nhh.no

(2) Bank of England. Email: may.rostom@bankofengland.co.uk

First version: September 2018. This version: March 2022. We thank Jaap Abbring, Lars Nesheim, Áureo de Paula, Fabien Postel-Vinay, Nikita Roketskiy, Fred Schroyen, and Sorawoot Srisuma for helpful comments and advice. We also thank seminar participants at UCL, Bank of England, NHH, Sciences Po, Imperial College London, NES, HSE, Aalto, City University of London, and University of Surrey for valuable feedback. Any views expressed are solely those of the author(s) and so cannot be taken to necessarily represent the views of the Bank of England or to state Bank of England Policy. All errors are our own.

The Bank's working paper series can be found at www.bankofengland.co.uk/working-paper/staff-working-papers

Bank of England, Threadneedle Street, London, EC2R 8AH
Email enquiries@bankofengland.co.uk

ISSN 1749-9135 (on-line)

1 Introduction

For most households, a mortgage contract is the largest financial investment they will undergo in their lifetime. A large and recent body of literature considers the role search frictions play when consumers pick a mortgage (Allen, Clark, and Houde (2013, 2017), Woodward and Hall (2012), Agarwal, Grigsby, Hortaçsu, Matvos, Seru, and Yao (2020), Alexandrov and Koulayev (2018), Guiso, Pozzi, Tsoy, Gambacorta, and Mistrulli (2021), Deltas and Li (2018)). This is of little surprise: searching for a mortgage is hard and costly. It involved tedious administrative tasks, product comparison-shopping, and time-consuming tasks of submitting applications and undergoing lender interviews, all weighed against the risk that with each rejected application their credit score could fall. Shopping around is further exacerbated by complicated price structures, and lenders offering large menus of near-identical products. For example, Coen, Kashyap, and Rostom (2021) report that the median British household taking out a mortgage faces a choice of 70 different options, with over half picking worse than the average option available to them. No wonder then, that many households choose to outsource this decision to intermediaries, or brokers, in the hope of finding the best deal.

Do consumers benefit from outsourcing this decision to brokers? If so, which consumers? This is what we study in our paper. Specifically, we aim to answer three questions. First, what quantifiable savings do brokers provide to consumers who use them? Second, which consumers benefit the most from using brokers? And, third, does a competitive broker market compared to, say, a scenario where all prices were known to everyone, make agents better- or worse-off?

The UK mortgage market provides an ideal setting to study this. Unlike the US or Canada, an unusually high percentage of UK mortgages originate from brokers. For example, the Intermediary Mortgage Lenders Association (IMLA) reports that in the second quarter of 2015, 67% of borrowers used broker services, corresponding to 71% of the total value of all new mortgages in that period (IMLA, 2015). Indeed, this figure is comparable to what we observe in our data.

We start by documenting a significant proportion of unexplained price dispersion in our data.[1] This variation motivates our structural model of search, the framework we rely on to answer our questions. In our model, heterogeneous consumers decide whether or not to pay a fee and use a broker; and lenders– who privately observe their marginal costs and can always sell a loan directly to the consumer instead– set one price for both direct and broker markets.[2] Each broker is then treated as a platform, enabling borrowers who use them to find the cheapest product, thereby reducing lender monopoly power.[3]

We leverage the literature on nonparametric estimation of auctions in Myśliwski, Sanches, Silva Junior, and Srisuma (2020) to uncover the unobserved distribution of borrower search

[1]Other studies also find this, e.g. see Coen, Kashyap, and Rostom (2021).

[2]Frankel (1998) calls this *price coherence*, see also Edelman and Wright (2015).

[3]Under these assumptions, the model becomes an extension of that proposed by MacMinn (1980). Namely, the pricing problem is equivalent to a first-price procurement auction with an unknown number of competitors, see also Salz (2022).

costs and lender heterogeneity.[4] Aided by 1.3 million mortgage contracts from 2016 and 2017, we can quantify the cost of searching for a mortgage. Specifically, our nonparametric approach means we remain agnostic about the shape and modality of the search cost distribution, and allow for non-linearities in the data to drive the relationship. The results are striking.

Three main findings emerge from these comparisons. First, search cost distributions differ substantially across demographic groups. Older borrowers face higher search costs in rural areas, but lower costs in urban areas. Incomes play a limited role, and only in rural areas. Low income borrowers from rural areas have higher median search costs, but in cities this difference is almost always negligible. Moreover, urban, non-first time buyers face higher search costs on average, than first-time buyers– most likely, because they have a higher opportunity cost for time. For instance, compared to first-time buyers (FTB), they will be selling their old home too, or are in a point in their life-cycle with more familial or professional responsibilities; or as urban dwellers, lead busier lives than rural borrowers.

Second, we find a large variation in search costs. Take the average monthly mortgage cost of £300, the median cost of obtaining an additional quote ranges from as little as 5% to almost one-quarter. This disparity is consistent with priors that searching is more costly for some than others. Alternatively, not everyone will benefit in the same way when seeking assistance from a broker.

Third, lenders' margins exhibit dispersion across mortgage types. More leveraged loans or those carrying longer tenures are, on average, less profitable. More generally on the supply side, despite high market concentration, it is relatively competitive with an average markup of 10.37%.

After quantifying these relationships, we then turn our attention to two questions that, as far as we are aware, are still pending in the literature. First, we ask whether brokers improve welfare, compared to a world where their advice is unavailable. Simply put, can we quantify the value of information brokers provide? In a second related question, we ask whether an alternative to the traditional intermediary is better. Online brokerages and comparison tools have become increasingly popular, but what happens to prices and margins? Who benefits and who loses?

To answer to our first question, we simulate optimal prices and search behaviour in a new equilibrium where no intermediation exists. The value of information is therefore the difference between the expected consumer surplus in our baseline and counterfactual case. Our results show that, on average, brokers are a net-positive. Mortgagors save £72.31 per month in sunk expenditures on a median-sized mortgage.[5] About one-third of these savings are down to brokers finding cheaper prices. Another sixth are due to lower search costs; searching for a mortgage is hard and costly, and brokers fill that gap. However, not everyone benefits equally from the current market structure. Markers of financial acuity and experience matter– with younger,

[4] We estimate these distributions after conditioning on a set of borrower and loan characteristics.

[5] Sunk expenditures are defined as those not related to paying off the mortgage principal.

lower-income, and first time buyers benefiting the most. Remarkably, borrowers choosing longer fixed-rate deals or shorter amortization periods experience only a slightly lower price when using brokers, but the commissions they pay exceed their counterfactual search cost.

This net positive effect can be attributed to the externality brokers impose on the direct market (Salz, 2022). The existence of intermediaries reduces lenders' market power, who are unable to price discriminate between informed and uninformed consumers. This explanation is reinforced by looking at the counterfactual distribution of price-cost margins. Without intermediation, the average Lerner index almost reaches 24% and a fourth of all mortgages have margins exceeding 33.5%.

For our second scenario, we study the effects of a hypothetical market centralization. We assume lenders post all prices, and consumers are automatically matched with the best offer. Direct sales are no longer possible, and are replaced by a free, market-wide platform. We find that average prices in a centralized market would decrease by 4.4%– saving borrowers almost £15 a month. Conversely, lenders' margins drop by almost half. These online platforms, which rely on machine learning technology, stand-in for the human knowledge dispensed from traditional brick and mortar brokers. The total welfare effects would need to weigh-in the modest reduction in prices and search expenditure against any sunk cost of physical brokers exiting the market, against the value gained from long-term relationships in a market where mortgagors refinance often. It is likely that over the longer-term, these modest benefits entirely disappear.

All in all, our paper makes two main contributions. First, we provide a quantifiable estimate for the value of information mortgage intermediaries provide, uncovering heterogeneity across demographic groups and loan types. We can show from the rich patterns of the data and our nonparametric estimation, that while the net effects of brokers' presence is positive, not every borrower is necessarily better off. Second, the novelty of our structural approach provides an attractive framework for studying welfare effects in industries with two-sided platforms and search frictions. The mortgage market is one, but our technique can be applied to other industries, such as insurance. Importantly, the estimators do not require any optimization, the structural features are identified in closed form, and the results are robust to distributional assumptions about search costs and firm heterogeneity.

■ **Related literature.** We contribute to several strands in the literature. First, there is a growing body of empirical papers using structural models of consumer search to study mortgage markets. Allen, Clark, and Houde (2017) is perhaps methodologically closest to ours. The authors consider a search and bargaining framework with bilateral heterogeneity. However, they focus on the role of loyalty advantage and do not study intermediation, choosing instead to exclude brokered loans from their analysis. On the opposite side of the spectrum, is the paper by Woodward and Hall (2012) which only studies brokered mortgages. They conclude that mortgagors in the US would benefit from shopping at multiple brokers. We abstract from the search for brokers, assuming that intermediaries operate in a competitive sector and have no incentive to provide dishonest advice. Rather, by including borrowers who do and do not use brokers, we compare whether, *on the whole*, brokers confer a benefit.

In another recent study, Agarwal, Grigsby, Hortaçsu, Matvos, Seru, and Yao (2020) uses data on actual search behaviour and rejected mortgage applications to document that, contrary to predictions stemming from standard search models, more search does not always result in lower prices. To explain this finding, the authors introduce screening and the probability of getting one's mortgage application rejected into a standard search model, finding that a standard framework is only able to recover true search costs, scaled by the probability of approval. While our data do not inform us about rejected applications, we remain agnostic whether our search cost estimates also indirectly account for the probability of being rejected. Alexandrov and Koulayev (2018) investigate the interplay of search and preference for non-price characteristics (such as brand effects) to explain sub-optimal shopping efforts in the US market. And although Thiel (forthcoming) focuses on financial advice more generally, he shows that banning financial advisors from receiving commissions leads to a reduction in consumer surplus in the long-run when advisors exit the market.

There are also a number of empirical papers that examine mortgage price shopping. Coen, Kashyap, and Rostom (2021) use the same data, from an earlier time period, to study how consumers shop for mortgages. They find that young and inexperienced consumers face a large amount of price dispersion, and that households who pick badly do so because they are presented with menus containing many expensive options. They do not examine the role of brokers in the mortgage market, however. In a similar paper, Bhutta, Fuster, and Hizmo (2020) also document a large degree of price dispersion, with the least financially sophisticated borrowers massively overpaying relative to market rates. They suggest that rising borrowing costs encourages search. Malliaris, Rettl, and Singh (2020) also show that while increased mortgage competition is financially beneficial to both sophisticated and naive borrowers, by encouraging lenders to include attractive offerings, it does not remove costly products from menu offerings that unsophisticated borrowers are more likely to pick. They conclude that lender competition is not a substitute for borrower sophistication. Similarly, Andersen, Campbell, Nielsen, and Ramadorai (2020) find that households' mortgage choices depend on their characteristics. Using Danish administrative data, they show that poorer, older, and less educated households are less likely to refinance their mortgage, missing out in potential savings as a result. These findings are consistent with earlier work by Lusardi & Mitchell documenting the strong correlation between personal characteristics and degree of financial literacy (e.g. see Lusardi and Mitchell (2014)). Guiso, Pozzi, Tsoy, Gambacorta, and Mistrulli (2021) uses Italian data to study whether in-house bank advisers distort advice, steering borrowers into taking up more risky and expensive adjustable rate mortgages compared to fixed rate mortgages. Whereas they do find welfare losses associated with sub-optimal advice, they also conclude that banning advice altogether would result in an average annual loss of €998. This number is lower than our calculations, but it is consistent. Since the advice in our model is akin to being fully impartial, we think our estimate is informative of the upper bound on the change in consumer surplus. Finally, Deltas and Li (2018) present empirical evidence on how search costs in the US mortgage market can be reduced by network externalities.

We also contribute the academic literature, which studies the role brokers play in retail financial markets (Bergstresser, Chalmers, and Tufano (2007), Inderst and Ottaviani (2012b), Egan, Matvos, and Seru (2018) Egan (2018)). The contemporaneous work of Robles-Garcia (2020) studies bargaining over commissions between lenders and brokers in the mortgage market. She finds that a ban on brokers would lead to a 24% increase in prices. This lends credence to our finding of 21%, especially as she defines prices differently, proposes a different model and estimation methodology, and has more granular data on brokers.

Our work also touches on IO literature, where search models have been used to study welfare effects of intermediation in other industries. For example, Gavazza (2016) investigates the role of dealers in the secondary market for business aircraft. In another paper, Byrne and Martin (2021) argue for the importance of consumer protection for different types of households, especially for poorer households who typically do not search. Importantly for our work, Salz (2022) looks at the role of brokers in contracting trade waste removal in New York City. In particular, the structural model in our paper resembles Salz's framework where the same firms participate in both direct and brokered markets and cannot charge different prices. The main finding in our paper corroborates Salz's conclusion that overall, intermediation reduces information frictions and can be seen as a positive externality in reducing market power. However, we are also able to show that the effects can be negligible or even negative for certain types of consumers. Our identification strategy relies on a weaker set of assumption and hence differs from Salz's approach. We discuss the econometric differences in detail in section 6.1 of the paper.

Finally, our paper is tangentially related to two strands of the theoretical literature: an array of papers studying the effects of middlemen (e.g. Rubinstein and Wolinsky (1987), Biglaiser (1993), Yavaş (1994), Spulber (1995), Hall and Rust (2003), Thiel (forthcoming)), and an active literature on multi-sided platforms (e.g. Armstrong (2006), Rochet and Tirole (2006), Galeotti and Moraga-González (2009), Edelman and Wright (2015), de Cornière and Taylor (2017)). The way we treat brokers in the model is reminiscent of a platform with endogenous buyer entry.

The paper is organized as follows: Section 2 outlines main institutional features of the industry, describes the data and provides some reduced-form evidence on price dispersion and the impact of brokers on transaction prices. Section 3 and 4 discuss the data and the reduced form findings. Section 5 introduces the theoretical model, and in Section 6, we discuss nonparametric identification of its primitives and outline the estimation method. Our main results are presented in Sections 7 and 8. Section 9 concludes and proposes directions for future research.

2 The UK mortgage market

The UK mortgage market is relatively concentrated, with a total share of the six biggest banks exceeding 70%. As in the US, mortgage terms in the UK typically amortize over 25 years, although longer durations are also common[6]. However, unlike the US, the most contracts are short-term, and refinancing is common. The most common products are 2-, 3-, and 5-year fixed rate mortgages (FRM), and 2-year adjustable rate mortgages (ARM). Since the Great Financial Crisis (GFC), FRMs make up the vast majority of mortgages, and in our sample period of 2016-17, they account for over 90% of all mortgage contracts (2-year FRMs being the most popular). Upon expiration of the initial contract period, borrowers can negotiate a new contract with the same or a different lender.

For each type of product, banks post quoted rates that vary by contract period and loan-to-value ratios, or LTV. When selecting a mortgage, borrowers can either choose to pay an upfront fee to the lender in return for slightly lower monthly payments, or pay no fee but pay more per month. Lender fees are small, however, especially relative to the size of the mortgage. Median loan fees are £999, and 40% pay no fees at all. As documented by Coen, Kashyap, and Rostom (2021) and Iscenko (2018), among others, lenders typically offer broad product portfolios with different combinations of fees and interest rates. In addition to that, they also offer loans with optional cashback (a one-off lump sum payment to new borrowers) or flexible repayment schemes (i.e. possibility of over- or underpayment) which are priced differently.

A striking feature of the UK market is that almost 70% of mortgages are accessed via brokers. This number is significantly bigger than the share of brokered mortgages in the US (Alexandrov and Koulayev (2018) report roughly 10%) or Canada (Allen, Clark, and Houde (2017) have 28% of brokered contracts in their data). The *Intermediary Mortgage Lenders Association* (IMLA) report an upward trend in the fraction of borrowers who use intermediaries, noting that the value share of mortgages originating from brokers increased from about 50% before the GFC to 71% in the second quarter of 2015 (IMLA, 2015).

Applying for a mortgage directly with a bank typically involves interviews with loan officials and lengthy applications, weighed against the backdrop that with each rejected application credit scores could fall. Brokers, on the other hand, help zero-in on the most suitable products and assist borrowers through the application process. The market for intermediation is competitive and geographically dispersed. As noted by IMLA (2015), *the UK mortgage broking business is dominated by small firms serving local client bases. According to data from the Financial Adviser Confidence Tracking Index in September 2015, 69% of broking firms employed only 1 or 2 mortgage advisers with another 20% employing 3 to 5.*". While there is no regu-

[6]The median first time buyer amortizes over 30 years.

lation in place that obliges brokers to search through all available mortgage products[7], broker services offer affordability comparisons across banks that are different from those of lenders' in-house advisers. In return, intermediaries are compensated in one of three ways. They can receive commissions directly from borrowers, procure a charge from lenders, or do both.[8] In our data, the median fee paid by borrowers to brokers is relatively small: a lump sum of £349, or an average of £10 per month over the duration of the initial period.

3 Data

We use loan-level administrative data from the Product Sales Database (PSD) of all new mortgage originations in the UK. The data contain information from the mortgage application, including borrower characteristics such as age and income; loan details such as the issuing bank, interest rate, and loan size; and property details such as the purchase price and location. The data begin in 2005, but quality is patchy until 2008. Following the GFC, collection substantially improved, and in 2016, information on direct sales, brokers, and fees were added. For this reason, our main analysis begins in 2016.[9].

Our final sample includes over 1.3M mortgage contracts issued in 2016 and 2017. We cannot identify specific brokers, so cannot tell who searches the entire market and who does not. Our solution is to focus on the big six lenders, as their products are available to every potential borrower in the country, and easy to find. Another reason we focus on the big players is because, in our model, we abstract from lenders' budget constraints and capital requirements which are much more important for small lenders (Benetton, 2021). We also exclude ARMs, loans with non-standard FRM lengths, and loans with LTVs greater than 95%. All constitute a very small fraction of the market. Further details on the sample construction and summary statistics can be found in appendix A.1.

3.1 Mortgage costs

We define a cost metric to compare the mortgage costs faced by borrowers in a consistent and unified way. Constructing this scalar measure of cost will turn out to be vital for the structural model, since all estimates can be ordered relative to it. Mortgage costs vary along several dimensions. First, in whether or not they have upfront fees. Second, in the interest rate paid during their fixed period, and third, in the length of the fixed period.[10]

[7] The intermediaries that do are known as *whole-of-market* brokers.

[8] Woodward and Hall (2012) argue that, in the US, brokers are indifferent about their source of compensation. A different strand of (mostly theoretical) literature studies how different compensation schemes can alter brokers' incentives (see e.g. Inderst and Ottaviani (2012a)). Our study abstracts from this issue by assuming that payments from lenders to brokers constitute part of their costs which are eventually passed onto borrowers in the form of higher prices.

[9] Due to restrictions on data access, we are only able to get the data until 2017

[10] A calculation of total mortgage costs over the entire mortgage horizon would also include the variable rate the mortgage resets to after the deal period expires, and the mortgage term (usually 25 years). However, this would assume no one refinances and given how short fixed term

The monthly economic ("sunk") cost is the interest component of the monthly payment plus any upfront fees added onto the loan by the lender:

$$p = iL + \frac{Fee}{N}, \tag{1}$$

where N is the initial period of the mortgage contract (24, 36, or 60 months), L is the size of the loan, and i is the fixed interest rate. Since in the structural model we take the loan size as given, to adequately compare costs of mortgages with different initial loan amounts, we normalize the monthly cost of the loan to correspond to a median loan value in the sample, £150,000. Our approach is similar to Allen, Clark, and Houde (2017) who normalize their price variable to correspond to the monthly payment on a $100,000 loan.

4 Reduced form findings

Using PSD data, we provide descriptive evidence of several features of the UK mortgage market in support of the modelling framework. First, we show a substantial degree of price dispersion in transacted prices. Second, we show that borrowers who used brokers have, on average, lower monthly mortgage costs, but total costs are higher once we factor in broker fees. Finally, we show that observable borrower and product characteristics are poor predictors for whether or not to use a broker. Overall, the type of evidence we present is akin to that in Section 3 of Salz (2022), suggesting that high search costs may be why some households use a broker, and justifying our modelling assumptions.

4.1 Price dispersion

To see whether there are differences in price dispersion by choice of sales channel, we look at the level of unexplained variation after regressing mortgage prices on observed characteristics. We do this for borrowers who got their mortgage directly from lenders from those who used a broker. We run the following hedonic regression:

$$p_{ijt} = \mathbf{X}'_{ijt}\boldsymbol{\beta} + \psi_t + \xi_j + u_{ijt} \tag{2}$$

where p_{ijt} is the mortgage price for household i, from bank j, at time t. $\mathbf{X}_{ijt}$ is a vector of household and loan characteristics, e.g. household income, LTV, and the mortgage term.[11] ψ_t and ξ_j are time and bank fixed effects[12].

periods are in the UK, this would put undue weight on the period after which the FRM expires.

[11] More specifically, we control for household income, house price, loan size, LTV (included as a set of dummy variables corresponding to LTV thresholds), first-time buyer (FTB) status, region, mortgage type, length, and other product characteristics, as well as their interactions and allow for potential nonlinearities.

[12] We also run the same specification without bank fixed effects.

We define our dependent variable, p_{ijt}, in two ways. In one calculation we use the interest rate in basis points[13]; and in the other we use normalized interested payments in £, as defined in Section 3.1. Table 1 reports the level of unexplained variation (as captured by $1 - R^2$) and the coefficient of variation, for both measures of the dependent variable, and during the initial period of the loan.

Panel A and B in Table 1 report results for p_{ijt} as measured by interest rates and interest payments respectively. Overall, the level of unexplained variation, $1 - R^2$, is about 30% although its lower for brokers and especially when interest rates are the dependent variable.[14] This proportion is quantitatively similar to the percent of unexplained variation in the Canadian data reported by Allen, Clark, and Houde (2017) who report $1 - R^2$ of 0.39.[15] The table also compares the results with and without lender fixed effects. Fixed effects allows us to control for bank heterogeneity, but leaves within bank variation unexplained. Comparing columns (1) with (3), and (2) with (4), we can see that adding fixed effects substantially reduces the proportion of residual variation from direct sales, but has virtually no effect on the R^2 in the regression using broker data. This finding is consistent with our priors that brokers help borrowers find the most suitable product across lenders. Suppose there exists a lender that, on average, sets higher interest rates than its competitors. Adding lender fixed effects will help explain more of the unobserved variation in prices in the direct segment of the market, especially if consumers do not shop around. However, in the brokered segment, intermediaries compare products against competing offers, making it unlikely for borrowers who use a broker to go with the more expensive lender.

Table 1: Price dispersion by sales channel.

	Panel A: Interest rate			
	No FE		With FE	
	Direct	Broker	Direct	Broker
	(1)	(2)	(3)	(4)
$1 - R^2$	0.316	0.181	0.232	0.172
Coefficient of variation	0.307	0.316	0.307	0.316

	Panel B: Interest payments			
	No FE		With FE	
	Direct	Broker	Direct	Broker
$1 - R^2$	0.369	0.364	0.283	0.355
Coefficient of variation	0.294	0.305	0.294	0.305

Note: Table presents 1-R^2 from the regression defined by 2, separately by direct and broker sales channels and for two different definitions of price. The second row in each panel is the coefficient of variation defined the ratio of the standard deviation to the mean.

[13] We control for the level of upfront fees on the right-hand side.

[14] The decrease in unexplained variation in Panel A for brokers is likely down to controlling for broker fees in the regressions in Panel A.

[15] See also Allen, Clark, and Houde (2014) for a detailed study of price dispersion in the Canadian market.

4.2 Do brokers offer cheaper prices?

To establish whether brokered mortgages are cheaper, we check whether households who used a broker received a lower rate than those who did not after controlling for a flexible function of individual and product characteristics. Table 2 reports regression results on the dependent variable p_{ijt}, as measured using two ways– mortgage interest rates, or the normalized monthly mortgage payments (see Sections 3.1 and 4.1). In all cases, the coefficient is negative and significant suggesting that those who shopped with a broker received a cheaper product. However, the monetary savings appear to be modest and are about 7 basis points when measured by the interest rate and about £5 per month when measured in terms of monthly payments.[16]

Table 2: Price benefits of using a broker.

Dependent variable:	(1) Interest	(2) Interest	(3) Interest	(4) Monthly Payment	(5) Monthly Payment
Used a broker	-7.761***	-6.710***	-7.428***	-5.103***	-3.562***
	(0.0824)	(0.0822)	(0.0816)	(0.119)	(0.122)
Lender Fees	Linear	Linear	Non-linear	-	-
Controls	Yes	Yes	Yes	Yes	Yes
Regional FE	No	Yes	Yes	No	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
Observations	1,309,067	1,309,067	1,309,067	1,309,067	1,309,067
R^2	0.768	0.772	0.778	0.627	0.632

Note: *** denotes significant at 1% level. Robust standard errors in parentheses. Interest is measured in basis points. Monthly interest is the component of the initial monthly payment that goes towards payment of the interest, including lender fees, and normalized by the size of the loan. Controls are income, house price, loan size, LTV, first time buyer and mortgage term. Time fixed effects are at the monthly level. Regional fixed effects are at the Government Office Region level and include a flag for an urban region. Non-linearities in lender fees are controlled for using a fifth-order spline.

The dependent variable used in the regressions presented in Table 2 is constructed in a way to control for lender fees only and our definition of monthly cost does not include broker commissions. Once we add broker fees divided by the number of months in the deal period to reflect monthly cost, the sign of the coefficient switches to positive (see table 3).

The findings summarized in Tables 2 and 3 are in line with the descriptive evidence Salz (2022) used to justify the assumption that buyers with higher search cost select themselves into the brokered market. Brokers seem to offer broadly lower prices, but once their commissions are factored in, borrowers end up paying more than they would in the direct market. This finding is crucial to justify that borrowers with higher search costs are more likely to use brokers. Without the sign reversal, standard models of search would have difficulties explaining why brokers are not used by everyone. To provide some intuition, suppose that one always expects to pay less by going to the broker. Then borrowers with low search costs would have an incentive

[16] Since some of the brokers in our sample are compensated directly by borrowers while others only receive commissions from the lenders, we test whether different broker compensation schemes affect their incentives to provide unbiased advice. We run the same regression for two subsamples of the data – one which only includes brokers who are only paid by the lenders, and one which only includes those who are not receiving any commissions from the banks. The sign and the magnitude of the effect measured by the coefficient of interest do not change by much across the subsamples, suggesting that brokers on average offer cheaper loans, regardless of who they are paid by. The results are presented in appendix A.4.

Table 3: Impact of using a broker on price plus broker fees.

Dependent variable:	(1)	(2)	(3)	(4)
	Monthly payment + broker fee			
Used a broker	1.729***	2.985***	7.049***	11.40***
	(0.122)	(0.126)	(0.140)	(0.152)
Controls	Yes	Yes	Yes	Yes
Regional FE	No	Yes	No	Yes
Time FE	Yes	Yes	Yes	Yes
Observations	1,309,067	1,309,067	792,023	792,023
R^2	0.607	0.610	0.626	0.632

Note: *** denotes significant at 1% level. Robust standard errors in parentheses. Monthly interest is the component of the initial monthly payment that goes towards payment of the interest, including lender and broker fees, and normalized by the size of the loan. Columns (1) and (2) show estimates obtained using the entire sample whereas columns (3) and (4) restricts the sample to only brokers who charge borrowers directly (i.e. the fees are non-zero).

to pretend that their cost is high and use them as well.

4.3 Predicting broker use

The final fact we document in this section is that observable characteristics of the borrower do not really help predict who uses brokers. Table A.6 in Appendix A.3 reports results from a linear probability model where we regress an indicator variable for using a broker on a set of personal characteristics (e.g. age and income), mortgage product characteristics (e.g. product type and mortgage term), and regional fixed effects. Because of potential identification issues, we do not attempt to extrapolate our interpretation of the effects beyond conditional correlations. Irrespective of the observable characteristics that we do control for, the R^2 is never greater than 0.13, even if we allow for multiple interactions between variables.[17]

All in all, these findings are consistent with our hypothesis that observables have little predictive power in understanding who uses brokers. This opens up scope for an unobserved component, such as search cost, to be a more important driving force behind borrowers' decisions.

5 Model

We introduce a stylized model of mortgage pricing when consumers can search across different lenders directly or use a broker. As in Allen, Clark, and Houde (2017) we assume that there exists an initial period outside the model where the borrower chooses the property she wishes to purchase, associated loan size, and her preferred mortgage criteria, e.g. duration and term maturity. Therefore, the dimension of search we consider is one where the borrower can compare similar products across different banks. The assumptions on intermediation technology

[17] The R^2 is not a perfect measure of predictive power. We also plotted propensity score distributions for using a broker for borrowers who in reality did not use a broker and those who *did*. We found a large degree of overlap between them (see Figure A.1). Iscenko and Nieboer (2018)) plot a similar chart.

closely follow the ones in Salz (2022). We treat brokers as non-strategic players, and assume they act in the borrower's best interest by choosing the cheapest offer in the market at a given time.[18] This assumption allows us to treat brokers similarly to a price comparison platform, or, in the parlance of Baye, Morgan, and Scholten (2006), as an *information clearinghouse*.[19]

The framework we present here extends MacMinn (1980) and Myśliwski, Sanches, Silva Junior, and Srisuma (2020) search models with bilateral heterogeneity by adding another stage to the consumer's problem where they decide whether to use an intermediary or search. Consider an environment with a finite number of J lenders and a continuum of borrowers with unit demands. Borrowers, indexed by i, receive iid draws from a continuous search cost distribution $\kappa_i \sim \mathcal{G}(\cdot|\mathbf{x}^G)$. $\mathbf{x}^G$ is a vector of observables which can shift the distribution of search cost. These are covariates defining consumer type, and include characteristics, such as age, income, house location, and if they are a first time buyer (FTB).

Lenders are heterogenous in their marginal cost of providing the loan, $c_{ij} \sim \mathcal{H}(\cdot|\mathbf{x}^H)$, which is their private information. $\mathcal{H}$ is continuously distributed on a compact support $[\underline{c}; \bar{c}]$.[20] Given we are considering a market with posted prices, $\mathbf{x}^H$ is a vector of covariates which includes key characteristics of the mortgage[21], but could also include some elements of $\mathbf{x}^G$ if price discrimination or bargaining are an important feature of the market (Allen, Clark, and Houde, 2017). While direct price negotiation is not typical in the UK[22], its effects are somewhat mimicked by the fact that lenders typically have broad product menus[23]; it is virtually costless to introduce a new mortgage with a slightly different iteration. Therefore, the fact that the marginal cost is transaction-specific (i.e. varies across lenders and borrowers) should be seen as an approximation to residual product differentiation which is not captured by the conditioning variables.

5.1 Borrowers

Having drawn their search cost, borrowers decide whether to engage in a non-sequential search or use a broker. The search technology is such that i chooses the optimal number of price draws, k, to solve:

$$\min_{k \geq 1} (k-1)\kappa_i + \mathbb{E}\left[p_{(1:k)}|\mathbf{x}^G, \mathbf{x}^H\right] \tag{3}$$

[18]This assumption does not allow us to study the consequences of *distorted* financial advice as in Guiso, Pozzi, Tsoy, Gambacorta, and Mistrulli (2021), so we can interpret our estimates as the upper bound on the value of brokers. If in addition to charging commissions they also provided sub-optimal advice, their impact on consumer welfare would be lower. In other words, we abstract from brokers potentially facing conflicts of interest between providing the best advice and being compensated by the lender, as discussed by Inderst and Ottaviani (2012c) and Woodward and Hall (2012)

[19]Early theoretical models which consider price dispersion in markets where some consumers can access sellers directly or use such a clearinghouse include Salop and Stiglitz (1977), Rosenthal (1980), Varian (1980), and Baye and Morgan (2001) among others.

[20]We allow the support to differ for different $\mathbf{x}^H$.

[21]Specifically: Mortgage term; LTV band; FTB status; initial deal length; and indicators for a flexible or cashback mortgage.

[22]See the discussion in Benetton (2021), and this Guardian newspaper article for anecdotal evidence that few rates are negotiated (accessed 15/09/2018).

[23]For example, Coen, Kashyap, and Rostom (2021) report that the median borrower can choose from 16 different loans within the same lender.

Just like in Hong and Shum (2006), we assume that the first draw is costless[24], and the valuation of all consumers is equal to the upper bound of the support of the marginal costs, i.e. highest observed price. $\mathbb{E}\left[p_{(1:k)}|\mathbf{x}^G,\mathbf{x}^H\right]$ is the expected lowest among k prices drawn from the equilibrium distribution $\mathcal{F}(p|\mathbf{x}^G,\mathbf{x}^H)$, which arises as a result of lender's profit-maximizing pricing decisions given borrowers optimal search behaviour. Unlike Burdett and Judd (1983) where firms and consumers are *ex ante* identical, the equilibrium price dispersion arises as a result of search *and* lender heterogeneity.

The cost of using an intermediary is the expected rate paid for the mortgage suggested by the broker plus any commission charged for using the service:

$$\mathbb{E}\left[p^B|\mathbf{x}^G,\mathbf{x}^H\right] + \varrho(\mathbf{x}^G,\mathbf{x}^H) \tag{4}$$

Under the assumption that brokers inform the borrower of the best possible deal, we can treat them as auctioneers holding reverse first-price auctions. Therefore, $\mathbb{E}\left[p^B|\mathbf{x}^G,\mathbf{x}^H\right] = \mathbb{E}\left[p_{(1:J)}|\mathbf{x}^G,\mathbf{x}^H\right]$, so the price is the expected price obtained by searching all J lenders.

Let $k^*(\kappa_i)$ be the optimal number of searches for an individual with unit search cost equal to κ_i. Then the choice of direct search versus using an intermediary is the solution to the following cost minimization problem:

$$\min_{\{\text{Broker, Direct}\}} \left\{\mathbb{E}\left[p^B|\mathbf{x}^G,\mathbf{x}^H\right] + \varrho(\mathbf{x}^G,\mathbf{x}^H), (k^*(\kappa_i)-1)\kappa_i + \mathbb{E}\left[p_{(1:k^*)}|\mathbf{x}^G,\mathbf{x}^H\right]\right\}, \tag{5}$$

Following the insight of Hong and Shum (2006) and lemma 1 in Salz (2022), the linearity of search costs in the number of searches and because $\mathbb{E}\left[p_{(1:k)}|\mathbf{x}^G,\mathbf{x}^H\right]$ is non-increasing in k, borrowers in equilibrium will endogenously sort themselves into types defined by the number of searches, by forming cut-off points along the search cost distribution:

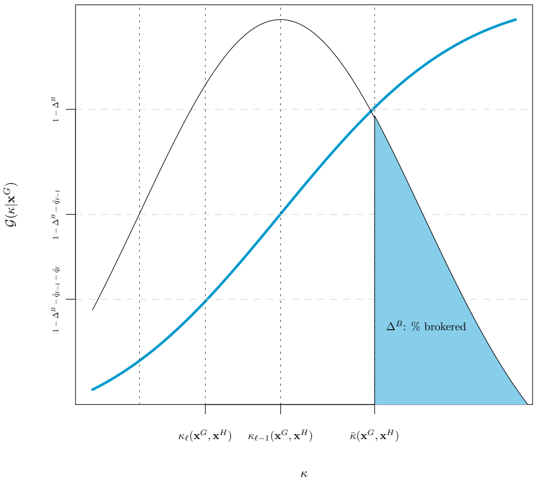
$$0 \leq \kappa_J(\mathbf{x}^G,\mathbf{x}^H) < \kappa_{J-1}(\mathbf{x}^G,\mathbf{x}^H) < \cdots < \kappa_k(\mathbf{x}^G,\mathbf{x}^H) < \kappa_{k-1}(\mathbf{x}^G,\mathbf{x}^H) < \bar{\kappa}(\mathbf{x}^G,\mathbf{x}^H) \leq \infty$$

In the above, κ_k should be understood as the highest search cost so that everyone with $\kappa \in [\kappa_k, \kappa_{k-1}]$ searches exactly k firms. Because we assume brokers are used by individuals with high search costs, $\bar{\kappa}$ is the search cost of the consumer who is indifferent between searching $k-1$ firms and delegating search efforts to a broker.

Define $\Delta^B(\mathbf{x}^G,\mathbf{x}^H) = 1 - \mathcal{G}(\bar{\kappa}(\mathbf{x}^G,\mathbf{x}^H)|\mathbf{x}^G)$ as the proportion of borrowers who access the mortgage using a broker. Let $\mathbf{q} = \left[q_1(\mathbf{x}^G,\mathbf{x}^H),\ldots,q_J(\mathbf{x}^G,\mathbf{x}^H)\right]$ be the proportion of borrowers who search $1,\ldots,J$ times. Figure 1 below illustrates the equilibrium sorting of borrowers according to the number of searches.

[24] In the context of our application, this could be interpreted as the offer from the consumer's primary bank. This consequently brings our model closer to Allen et al. (2017), where borrowers first receive a free offer from their home bank and then decide whether or not to search. One important difference is we do not assume the home bank has a cost advantage over its competitors.

Figure 1: Equilibrium sorting.



Note: Figure shows equilibrium sorting of buyers into types defined by the number of searches and use of brokers. The proportion of borrowers who use brokers is the blue shaded area under the search cost density. The areas between the cutoffs on the x-axis determine the proportions of buyers who search different number of firms. PDF not drawn up to scale.

5.2 Lenders and equilibrium

Assume that all $j = 1, \ldots, J$ lenders have an equal probability of being sampled in the search market. In the price-setting model, we depart from two key assumptions used by Salz (2022). Firstly, firms are not able to set different prices in the search and brokered markets, which is reminiscent of the notion of price coherence of Edelman and Wright (2015). Secondly, the same firms participate in the direct and brokered markets. Since we assumed that non-price characteristics were chosen by the borrowers outside of the model, the total profit function is additively separable in profits from every single transaction. Therefore, the same pricing game is played in each of the product-markets defined by the conditioning variables, so for clarity of exposition, we suppress the conditioning sets. Firm j with marginal cost c_{ij} solves:

$$\max_p \; \Delta^B \cdot \Pi^B(p, c_{ij}) + (1-\Delta^B) \cdot \Pi^D(p, c_{ij}; \mathbf{q}), \tag{6}$$

where $\Pi^B(\cdot)$ and $\Pi^D(\cdot)$ are the profits in the brokered and direct search market, respectively. Suppose all lenders have equal probabilities of being found by borrowers. Then the probability of being searched by a borrower who samples ℓ lenders is $\frac{\ell}{J}$ and we can restrict our attention to symmetric equilibria. Following Burdett and Judd (1983) and assuming that brokers hold first-price auctions, we can rephrase the problem as:

$$\max_p \; \Delta^B \cdot (p - c_{ij})(1-\mathcal{F}(p))^{J-1} + (1-\Delta^B) \cdot (p - c_{ij}) \sum_{\ell=1}^{J} q_\ell \frac{\ell}{J} (1-\mathcal{F}(p))^{\ell-1} \tag{7}$$

Due to costs being lenders' private information drawn from the same distribution, the probability that a lender with $\ell - 1$ competitors wins the contract (is the cheapest amongst ℓ firms) is $(1 - \mathcal{F}(p))^{\ell-1}$.

Let $\tilde{q}_\ell = (1-\Delta^B)q_\ell$ for $\ell = 1, \ldots, J-1$, and $\tilde{q}_J = \Delta^B + (1-\Delta^B)q_J$. Then the maximization problem simplifies to:

$$\max_p (p - c_{ij}) \sum_{\ell=1}^{J} \tilde{q}_\ell \frac{\ell}{J} (1 - \mathcal{F}(p))^{\ell-1}, \tag{8}$$

which is the same as the pure search problem considered in Myśliwski et al. (2020) with distorted search probabilities $\tilde{\mathbf{q}} = [\tilde{q}_1, \ldots, \tilde{q}_J]$. Let $\beta(c_{ij}; \tilde{\mathbf{q}})$ denote the optimal strategy given beliefs about borrowers' search decisions. Using the envelope theorem with the boundary condition that $\beta(\bar{c}; \tilde{\mathbf{q}}) = \bar{c}$ yields the optimal pricing strategy:

$$\beta(c_{ij}; \tilde{\mathbf{q}}) = c_{ij} + \frac{\sum_{\ell=1}^{J} \tilde{q}_\ell \ell \int_{s=c_{ij}}^{\bar{c}} (1 - \mathcal{H}(s))^{\ell-1} ds}{\sum_{\ell=1}^{J} \tilde{q}_\ell \ell (1 - \mathcal{H}(c_{ij}))^{\ell-1}}. \tag{9}$$

The equilibrium price distribution, $\mathcal{F}$, emerges as a result of lenders pricing according to (). Defining a symmetric Bayesian-Nash equilibrium of the game:

DEFINITION (Myśliwski, Sanches, Silva Junior, and Srisuma, 2020). *The pair $(\tilde{\mathbf{q}}, \beta(\cdot; \tilde{\mathbf{q}}))$ is a symmetric Bayesian Nash equilibrium if:*

(i) for every $\tilde{\mathbf{q}}$ when all firms apart from j use pricing strategy $\beta(\cdot; \tilde{\mathbf{q}})$, $\beta(\cdot; \tilde{\mathbf{q}})$ is the best response for firm j;

(ii) given the price distribution induced by $\beta(\cdot; \tilde{\mathbf{q}})$, $\tilde{\mathbf{q}}$ is a vector of proportions of consumers' optimal search.

We restrict our attention to monotone pure-strategy equilibria. The action space is compact and the pricing functions are strictly increasing in cost so the existence results from Reny (2011) apply. In general, lenders' payoff function can be seen as a mixture of two auctions – one where all firms participate (broker), and one where the number of competitors is unknown (direct search). Mixing probabilities are then determined in equilibrium by optimal search decisions made by borrowers.

6 Identification and estimation

This section discusses the identification of the model's primitives, that is the set of conditional search cost distributions $\mathcal{G}(\cdot|\mathbf{x}^G)$ and distribution of cost of providing the loan, $\mathcal{H}(\cdot|\mathbf{x}^H)$. We argue that the model imposes enough structure on the data for the aforementioned distributions to be nonparametrically identified.

Throughout the section, we assume that we observe the price, p_{ij}, for each mortgage (whether or not it was accessed via an intermediary)[25], and the values of the borrower and loan characteristics, $\mathbf{x}^G$ and $\mathbf{x}^H$, respectively. In addition to that, we assume that we observe (or construct from the data) lender market shares for each of the combinations of conditioning variables and broker commissions. The goal of nonparametric identification is then to establish a mapping from these data to the unobserved primitives using the theoretical restrictions imposed by the model. The main identification theorems for a pure search model are presented in (Myśliwski et al., 2020), and draw on findings from the literature on nonparametric auction estimation, in particular Guerre, Perrigne, and Vuong (2000).[26] We therefore only devote space in this section to emphasise certain aspects of the identification strategy which have not been discussed in previous literature.

6.1 Key assumptions

Our identification strategy and estimation techniques relies on three key assumptions on the supply side of the model.[27]

First, we assume that lenders offer the same loans directly or through a broker, and we only focus on the biggest six lenders. We therefore restrict our attention to symmetric equilibria, where all firms have the same underlying cost distribution. This assumption allows us to use the data from both brokered and direct mortgages when estimating $\mathcal{H}$, and is a reasonable one to make considering that, as mentioned earlier, borrowers can access identical mortgages in either market.[28]

Second, we nonparameterically estimate costs separately for each combination of $\mathbf{x}^H$.[29] We therefore capture potential nonlinear patterns of pricing and underlying cost distributions. For example, with the linear index restriction, the estimated cost distributions for 70% LTV and 90% LTV loans could differ only in their means while the fully nonparametric approach we adopt here allows also for different higher moments.

Third, we use data on prices and market shares together with a technique that minimizes the distance between market shares predicted by the model and the data to identify the search cost distributions. We then recover the marginal cost distributions using a closed-form expression involving the proportions of borrowers searching different numbers of lenders.[30]

[25] In practice, the information on the proportion of brokered loans should suffice.

[26] See also Athey and Haile (2007) for a comprehensive overview.

[27] These assumptions is where our model departs from Salz (2022).

[28] Unlike Salz (2022), we do not need to *ex ante* classify firms into high and low types. His identifying assumption is that the econometrician needs to observe at least one firm of each type in the brokered and direct markets. Then, he uses the structure imposed by broker auctions to recover the distribution of costs for low and high type firms. As such, he only uses brokered data to recover firms' cost distributions.

[29] Alternatively, we do not work with residualized prices, i.e. residuals from a hedonic regression of prices on a vector of loan characteristics. Therefore, we do not need to assume costs are additively separable in a linear index of characteristics.

[30] Conversely, Salz (2022) first estimates the distribution of costs from prices of brokered contracts, and then uses those estimates to recover proportions of businesses (consumers in his model) searching for different numbers of waste disposal firms. Our approaches are therefore

6.2 Role of exclusion restrictions

In the model exposition and the ensuing empirical analysis, we refer to two types of conditioning variables: $\mathbf{x}^G$, which shift the consumer search cost distribution and are variables defining consumer type; and $\mathbf{x}^H$, which are primarily mortgage characteristics that directly affect lender's cost of providing the loan. The two sets of variables can contain common elements, or, in the most extreme case, fully overlap.

A natural question is whether a lack of exclusion restrictions precludes identification of the unobserved cost distribution. In general, the answer is no, but it might lead to a situation where the distribution of search cost is only identified at very few points on its support. This discussion is related to the finding in Hong and Shum (2006) which was further elaborated by Moraga-González, Sándor, and Wildenbeest (2013) – namely, that absent any other dimension of variation in the data (e.g. local markets or time), one is unable to identify the search cost distribution beyond a set of $J-1$ points where J is the number of firms. A solution to this problem is to pool estimates from multiple markets, as shown e.g. by Sanches, Silva Junior, and Srisuma (2016).

Exclusion restrictions help generate such markets. In principle, for each consumer type ($\mathbf{x}^G$) we can pool estimates from different mortgage types ($\mathbf{x}^H$). To provide a specific example, assume all first-time buyers, aged 30 and above, who live in cities, and their income is above the median, have the same distribution of search cost. One can first obtain different sets of estimates for each type of mortgage, and then combine them to obtain a smoothed version of the search cost CDF. Absent any other variation in the data, one might have to resort to a parametric specification to be able to conduct meaningful counterfactual inference.[31]

6.3 Estimation steps

Our estimation algorithm can be boiled down to three steps.[32] We first back out the proportion of borrowers searching different number of lenders, by obtaining the empirical distributions of prices, in conjunction with data on market shares.[33] Second, as discussed in the preceding subsection, we pool the data across markets to obtain the estimate of the full search cost CDF. Finally, we use an equilibrium bidding function to construct pseudo-costs, and employ kernel techniques to obtain their density. For details on the estimation algorithm, see Appendix C.1.

non-nested and valid under different sets of assumptions.

[31] Moraga-González, Sándor, and Wildenbeest (2013) suggest that pooling data is possible across much more heterogeneous markets than the ones in our application: *"(...) to estimate the costs of search in the market for carpentry, one could pool data from the various professional services needed to refurbish a house: a carpenter, an electrician, a painter, a plumber, a bricklayer, a tiler, etc."*.

[32] See Myśliwski, Sanches, Silva Junior, and Srisuma (2020) for more detail.

[33] In this step, we need to impose the constraint that the proportion of borrowers who know offers from all lenders is no smaller than the proportion of brokered loans in the data.

7 Results

We apply our model and estimation strategy to the data. We are interested in recovering search and marginal cost distributions conditional on observed heterogeneity.[34] But as discussed in Section 6.2, although exclusion restrictions are not necessary for theoretical identification of the model's primitives,[35] they are useful in practice. With some variables excluded from $\mathbf{x}^G$, but included in $\mathbf{x}^H$, we are able to pool search cost cutoff estimates and the corresponding values of the survival function originating from different markets defined by different $\mathbf{x}^H$. This allows us to identify the distribution on a wider support, instead of on a few discrete points.[36]

Table C.1 in Appendix summarizes borrower and loan characteristics, and the associated bins used in discretizing the space in the structural model. Our choice is largely driven by the trade-off between computational feasibility and willingness to accommodate rich borrower and product-level heterogeneity.

Our price variable, taken from section 3.1, is the monthly mortgage cost. To remove dispersion from macroeconomic shocks (e.g. changes to Bank rate), we detrend prices. We also deflate them to January 2016 prices (£).[37] Ultimately, the monetary magnitudes of all our results should be referenced relative to a monthly interest and fee payment on a median-sized mortgage denominated in prices from January 2016.

7.1 Borrowers search costs

We present our nonparametric estimation for borrowers' search cost distributions in Table 4. The results show the median monthly cost (in £) of contacting an additional bank and obtaining a price quote.

Median costs for all borrowers range from £13.63 (young, high income, non-first time buyers in rural areas) to £72.05 (older, rural, low-income first time buyers). In relative terms, they represent between 5 and and 21.5% of the median interest-only payment.

These results are higher, but still quantitatively similar to other estimates provided in the literature. For instance, Allen, Clark, and Houde (2017) estimate a mean search cost of \$29/month, while Agarwal, Grigsby, Hortaçsu, Matvos, Seru, and Yao (2020) estimate an equivalent of \$27/month on a representative loan. Our estimates are higher because more than 70% of mortgages in our data are brokered, and because we assume consumers with high search costs use brokers[38].

[34] Note that the number of different search cost distributions will equal the number of bins defined by the chosen partition of $\mathbf{x}^G$. Correspondingly, the number of distinct marginal cost distributions will equal the cardinality of $\mathbf{x}^H$.

[35] $\mathcal{G}$ and $\mathcal{H}$ are identified even if $\mathbf{x}^G = \mathbf{x}^H$.

[36] See Moraga-González, Sándor, and Wildenbeest (2013) for a discussion on identifying search cost distributions using data from multiple markets and Sanches et al. (2016) for the theoretical properties of the method employed here.

[37] We do this obtaining the residuals from separate regressions of prices in each $(\mathbf{x}^G, \mathbf{x}^H)$ cell on a full set of monthly dummies.

[38] This also causes higher percentiles of the search cost distribution to be poorly identified. The third quartile for two out of sixteen distributions cannot be identified, hence why IQR estimates are missing for states 3 and 7 in the table. While the data are informative about the fraction of borrowers with $\kappa > \bar{\kappa}$, without parametric assumptions, we cannot identify the shape of the distribution above $\bar{\kappa}$. Nevertheless, variation in $\bar{\kappa}$ induced by different combinations of $\mathbf{x}^H$ is helpful although this does not allow for identification of $\mathcal{G}$ on its entire support.

Table 4: Nonparametric estimates of search cost distributions

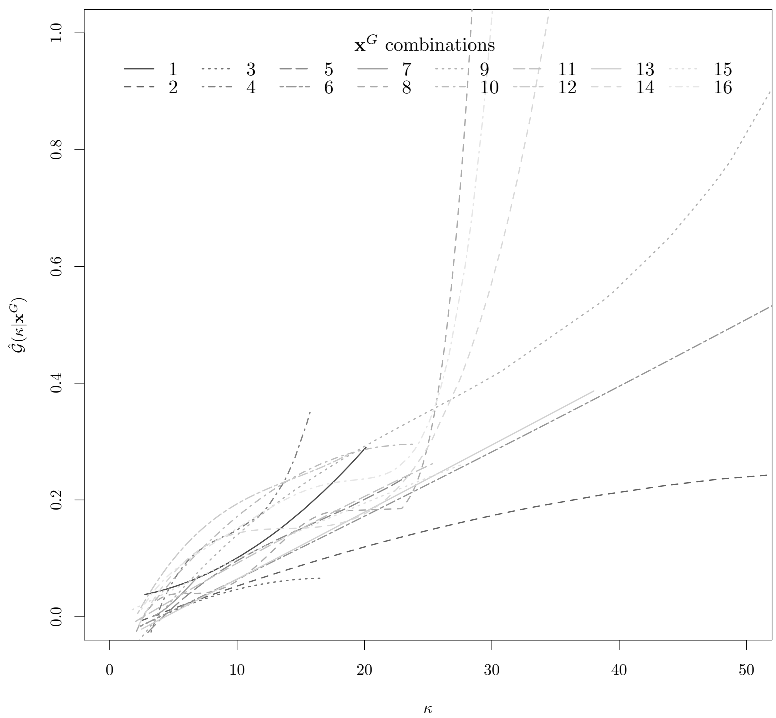
#	$\mathbf{x}^G$ bin Age	Inc	FTB	Urb	Median	Median searchers	IQR	% median price	% mean price
1	L	L	Y	R	27.40 (6.08)	12.23 (1.42)	16.45 (2.25)	7.89%	7.56%
2	H	L	Y	R	72.05 (18.62)	21.03 (4.44)	16.12 (19.91)	21.52%	20.55%
3	L	H	Y	R	22.08 (27.47)	22.08 (7.07)	0.00 (20.24)	6.91%	6.60%
4	H	H	Y	R	17.31 (4.68)	7.16 (3.03)	5.01 (7.44)	5.68%	5.39%
5	L	L	N	R	34.62 (6.66)	15.31 (6.02)	16.17 (6.93)	11.41%	10.82%
6	H	L	N	R	49.19 (13.27)	19.68 (2.81)	43.28 (22.34)	16.99%	16.29%
7	L	H	N	R	13.63 (14.50)	13.63 (3.79)	0.00 (13.18)	4.81%	4.53%
8	H	H	N	R	26.50 (7.76)	22.27 (10.73)	3.14 (1.18)	10.02%	9.46%
9	L	L	Y	U	36.31 (7.58)	11.16 (0.94)	30.90 (9.20)	10.13%	9.78%
10	H	L	Y	U	27.56 (16.41)	8.65 (0.62)	12.14 (25.35)	7.94%	7.63%
11	L	H	Y	U	49.44 (11.82)	13.72 (4.19)	50.64 (26.40)	14.89%	14.35%
12	H	H	Y	U	32.81 (13.73)	6.17 (0.69)	32.44 (18.26)	10.37%	9.94%
13	L	L	N	U	47.79 (14.87)	18.48 (1.31)	43.06 (19.29)	15.61%	14.91%
14	H	L	N	U	29.01 (7.72)	20.09 (3.88)	8.26 (6.02)	9.89%	9.53%
15	L	H	N	U	57.54 (22.34)	16.93 (2.94)	62.40 (23.13)	20.71%	19.49%
16	H	H	N	U	27.13 (4.60)	11.73 (5.43)	6.59 (4.00)	10.42%	9.84%

Note: Table presents selected features of nonparametrically estimated search cost distributions for 16 different borrower types (referred to as $\mathbf{x}^G$ bins). Age: L (below 30)/H (over 30). Inc(ome): L (below median)/H (above median). FTB (first time buyer status): Yes/No. Urb(an): U (urban area)/R (rural area). Column 6 contains the median search cost in £/month in the initial period. Column 7 reports the median search cost among borrowers who do not use brokers. These expressed in relative terms (divided by median and average monthly payment, respectively) are in columns 9 and 10. Interquartile ratio (column 8) measures the dispersion of the distribution. IQR=0.00 means that the third quartile is not identifiable in the data for that subsample. Bootstrap standard errors in parentheses based on 500 replications.

The median search costs among the borrowers who search across banks are substantially lower, ranging from £6 to about £22. The parts of the distributions corresponding to searchers are plotted in figure 2. The results point to a high degree of heterogeneity across different demographics. We also find that some of the distributions are clearly bimodal, with the first peak below £10, showing that the consumers who do not use brokers can efficiently search on their own. To compare the distributions across different demographics, we constructed an additional array of graphs in which we compare distributions across one trait keeping the other ones fixed (see figure B.2 in appendix B). The three main findings that emerge from these comparisons are that: (1) everything else equal, older borrowers face higher search costs in rural areas, but lower in urban areas; (2) low income borrowers from rural areas have higher median search costs, and in cities the difference is almost always insignificant; (3) in cities, non-first time buyers face on average higher search costs but they are less dispersed than those of FTBs.

Our first finding supports two hypotheses. First, the presence of physical bank branches is lower in rural areas, and would be more commonly used by older borrowers. Younger borrowers prefer digital channels, making mortgage comparisons easier. In urban areas, however, age isn't a distinguishing factor between borrowers who prefer brick-and-mortar over online services. This may be because physical bank branches are readily available, older borrowers who live in cities may be more computer-literate, or because young urban dwellers tend to be richer

Figure 2: Estimated search cost CDFs.



Note: Search cost denominated in January 2016 GBP (£) per month. Bernstein sieves were used to impose shape restrictions (non-decreasingness). The respective distributions are identified on $[\kappa_{min}(\mathbf{x}^G, \mathbf{x}^H), \kappa_{max}(\mathbf{x}^G, \mathbf{x}^H)]$, that is the lowest and highest cutoff estimated in the data.

than their rural cousins. But we can use it as a proxy for experience using financial services–especially because, as the last four rows of Table 4 indicate, median search costs are higher for non-FTBs. Second, there will likely be differences in financial literacy, which is typically either directly and indirectly correlated with income (e.g. see Hastings et al. (2013) and Lusardi and Mitchell (2014)). While we have not expressed search costs relative to actual incomes, no material difference in the result in urban areas can be discerned. Third, our findings reject the possible role of learning in the reduction of future search costs. The gap between first and subsequent refinancing is large, typically several years. Product offerings, saving accumulation, and mortgage qualifications will have changed, and experience from the first purchase may not be relevant at all. Our dataset does not only include borrowers refinancing with their current loan provider, so our search cost estimates will partly absorb some of the unobserved switching costs.[39] Lastly, to insure our findings are robust to different age and income bin definitions, we estimated search costs with more finely discretized grids. Namely, we considered 4 different age buckets and 4 income quartiles. This provided us no additional insight into the effects of age and income on search cost, while keeping the conclusions on the effect of FTB status and location virtually unchanged.

[39] Several authors studied the role of switching costs in the banking industry – see e.g. Kim et al. (2003), Deuflhard (2016), Honka et al. (2017).

7.2 Lenders' costs and margins

We now present the estimates of the supply-side primitives, that is lenders' marginal costs and associated markups. As a sense-check, we examine some of the aggregate statistics of marginal distributions that matter most.[40] For example, we expect riskier mortgages (e.g. higher LTV) to be associated with higher risk premia, which we expect to be captured by these cost estimates. Table 5 summarizes our findings.

Table 5: Summary statistics for estimated marginal costs and margins.

$\mathbf{x}^H$ category	Marginal cost			Price-cost margin		
	Mean	Median	IQR	Mean	Median	IQR
LTV						
≤ 70	266.4	268.2	66.2	11.15	7.79	8.91
71-75	296.8	286.9	82.3	6.58	5.44	6.50
76-80	294.7	286.2	69.8	8.59	6.11	6.72
81-85	322.9	311.9	72.6	6.06	5.06	6.37
86-90	410.9	407.6	67.7	6.17	5.13	5.71
91-95	517.8	523.6	50.6	5.04	3.11	3.36
Deal						
2 years	311.7	283.3	119.8	7.10	5.80	6.88
3 years	282.9	277.9	65.5	14.63	10.89	10.17
5 years	310.1	302.6	62.3	12.64	7.66	9.84
Term						
≤ 10 years	273.5	273.4	69.0	12.67	9.32	11.50
(10;15]	265.9	265.0	66.0	13.41	9.13	10.46
(15;20]	276.6	273.2	69.3	12.23	7.94	9.06
(20;25]	303.5	289.3	92.8	9.56	6.56	6.58
(25;30]	326.9	307.0	116.1	7.49	5.68	6.14
(30;35]	366.9	344.1	153.6	2.69	3.82	5.47
Value						
Q1	316.8	300.4	95.2	10.81	8.26	9.31
Q2	325.2	304.0	112.4	8.97	6.55	7.36
Q3	311.2	292.8	98.4	8.29	5.69	6.81
Q4	291.0	272.7	94.2	7.32	5.52	6.26
Flexible						
Regular	315.8	294.8	109.4	8.88	6.54	7.79
Flexible	277.9	276.4	66.3	8.54	5.52	6.78
Cashback						
No cashback	312.0	293.3	98.7	8.23	6.07	7.20
Cashback	300.8	281.1	115.9	12.92	9.10	10.06

Note: Means, medians and interquartile ranges of estimated marginal cost and price-cost margins defined as $PCM_{ij} = \frac{p_{ij} - c_{ij}}{p_{ij}}$. Costs expressed in £, PCMs in %.

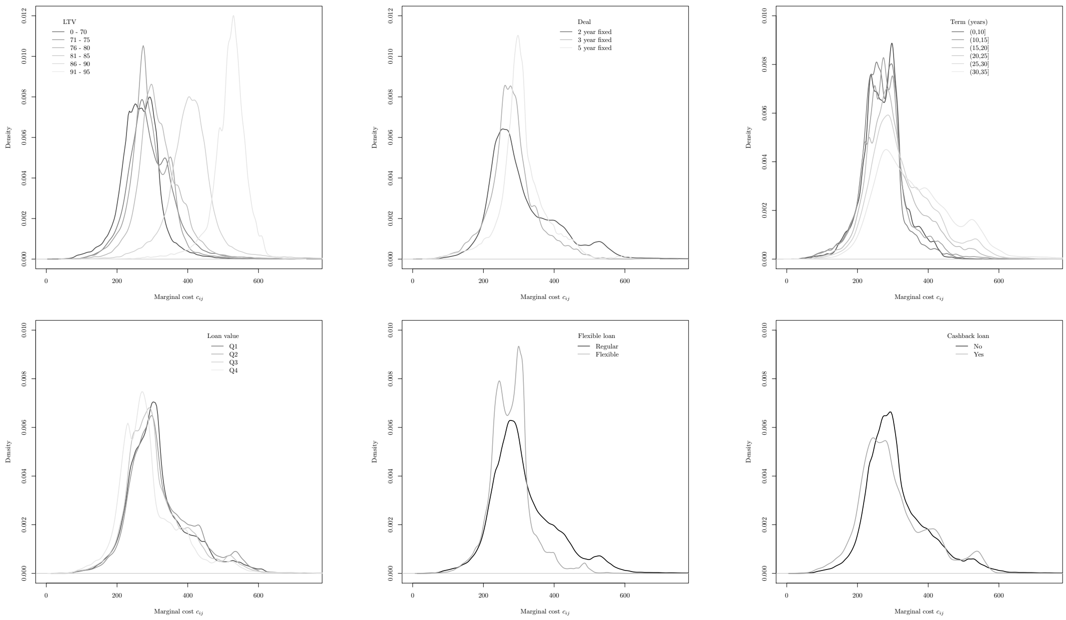
Intuitively, we find that, on average, estimated costs increase in LTV and length of amortization, and are higher for 5-year fixed rate loans than for loans of shorter duration. In our definition of price, we normalized the loan quantity to correspond to the median value, so it is of no surprise find no big cost differences across the four quartiles of the loan value distribution. Neither do we find major differences in cost distributions for mortgages offering flexible repayment schemes or cashback. Densities of marginal[41] distributions of costs are shown in Figure 3. LTVs are the main indicator of loan riskiness and the main driver of higher costs. One can see this from the top right panel of Figure 3, which shows an increase in cost variance

[40] We do this instead of rather than showcasing the full set of conditioning variables, $\mathbf{x}^H$.

[41] Here we use the word *marginal* as in the statistical definition of a marginal distribution.

as the amortization term increases. As most mortgages are refinanced, first-time buyers are predominantly those who hold a duration of 25 years or more. We can thus also interpret this finding in terms of the cost of servicing loans, being more idiosyncratic for first-time borrowers who amortize over longer periods of time.

Figure 3: Marginal distributions of marginal costs.



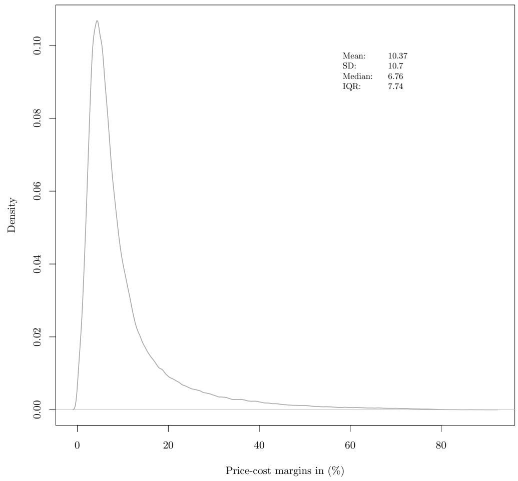
Note: Kernel estimates of marginal costs for each conditioning variable in $\mathbf{x}^H$ integrated over the remaining covariates.

The distribution of markups is right-skewed with an average of 10.37% and median of 6.76% (see Figure 4 and more detailed results in Figure B.1 in Appendix B). Despite high market concentration, the six biggest banks do not seem to be able to exert substantial market power. While the estimates might appear to be low, it is worth emphasising the model does not define the PCM in the canonical way as the markup over the interbank swap rate (i.e. LIBOR) or the Bank of England base rate. Instead, since our definition of price includes upfront fees, and we do not model lenders' fixed costs, we believe the estimate is closer to the cost of servicing the loan over the fixed-rate period (including mortgage application processing costs, or hedging against default risk). Yet, our estimates are higher than those obtained by Allen et al. (2017) in a recent study of the Canadian mortgage market, where the average Lerner index is estimated to be 3.2%. This is still more than 3 times lower than the average and 2 times lower than the median of our estimates.

Overall, these results are consistent with other recent evidence[42] showing the market becoming increasingly competitive, and lenders less able to enjoy high margins. From the perspective of the structural model's mechanics, low markups emerge as an artefact of a high proportion

[42] See, for example, The Guardian (27/04/2017): *Low rates, tight margins: the mortgage market looks worryingly familiar*: https://www.theguardian.com/business/2017/apr/23/mortgage-market-lower-rates-tight-margins-worryingly-familiar (accessed 10/09/2018).

Figure 4: Distribution of price-cost margins.



Note: Kernel estimate of the density of price-cost margins defined as $PCM_{ij} = 100 \cdot \frac{p_{ij} - c_{ij}}{p_{ij}}$.

of borrowers using brokers– whose presence, by construction, stimulates competition between lenders. We will return to this discussion in the following section.

8 Counterfactuals

In the current market setting, borrowers possess different information sets due to heterogeneous search costs. Those with low search costs are able to efficiently search through products offered by competing lenders, while those with higher search costs either have only one mortgage to choose from or need to use an intermediary. We run two counterfactual experiments.

In the first experiment, we quantify the value of information provided by brokers to consumers with high search costs. We simulate market outcomes in a scenario where brokers are not present in the marketplace. As intermediaries reduce the monopoly power of lenders, prices and markups should be higher in the new equilibrium. Borrowers, on the other hand, will no longer have to pay broker commissions, but their total search cost will change.

The second experiment concerns market centralization. Suppose the regulator establishes a platform, where all lenders must post their prices, and borrowers are matched to the best offer. This setting is equivalent to a pure first-price procurement auction with no search involved. We study how such a regulation would affect consumer surplus and the prices set by the lenders.

8.1 Value of information provided by brokers

In the absence of intermediation, the borrower chooses the optimal number of searches given its cost. Finding a new equilibrium involves solving the fixed-point problem defined in the space of (new) optimal search proportions denoted as $\dot{\mathbf{q}}$:

$$\dot{q}_\ell = \begin{cases} 1 - \mathcal{G}\left[\int \mathcal{H}\left(\xi\left(p, \dot{\mathbf{q}}\right) \mid \mathbf{x}^H\right)\left(1 - \mathcal{H}\left(\xi\left(p, \dot{\mathbf{q}}\right) \mid \mathbf{x}^H\right)\right) dp \mid \mathbf{x}^G\right] & \text{for } \ell = 1 \\ \mathcal{G}\left[\int \mathcal{H}\left(\xi\left(p, \dot{\mathbf{q}}\right) \mid \mathbf{x}^H\right)\left(1 - \mathcal{H}\left(\xi\left(p, \dot{\mathbf{q}}\right)\right) \mid \mathbf{x}^H\right)^{\ell} dp \mid \mathbf{x}^G\right] & \text{for } \ell > 1 \\ -\mathcal{G}\left[\int \mathcal{H}\left(\xi\left(p, \dot{\mathbf{q}}\right) \mid \mathbf{x}^H\right)\left(1 - \mathcal{H}\left(\xi\left(p, \dot{\mathbf{q}}\right)\right) \mid \mathbf{x}^H\right)^{\ell+1} dp \mid \mathbf{x}^G\right] & \end{cases} \tag{10}$$

As discussed in Myśliwski et al. (2020), Brouwer's theorem guarantees existence of a fixed point. While uniqueness cannot be proved, we experimented with different starting points finding that the algorithm converges to the same solution in the interior of the simplex.[43] The new search proportions then feed into the firms' pricing functions to generate a set of counterfactual conditional price distributions $\dot{\mathcal{F}}(\cdot|\mathbf{x}^G, \mathbf{x}^H)$ which can then be sampled from.

The auction model assumes that for all consumers, the valuation of the mortgage is at least as high as the upper limit of the support of the cost distribution. Therefore, realized consumer surplus for a borrower paying p is:

$$CS = \bar{v} - p - SE$$

SE is the search expenditure and is equal to $\kappa(k-1)$ if the borrower with search cost κ accessed the loan directly by contacting k banks, or ϱ if she used a broker and paid commission equal to ϱ.

Without intermediation,

$$\dot{CS} = \bar{v} - \dot{p} - \dot{SE}$$

where $\dot{p}$ is the new price drawn from $\dot{\mathcal{F}}$ and $\dot{SE}$ is the new realized search expenditure which now does not include the possibility of using a broker. We now define the value of information (VoI)[44] as the difference between the expected CS and $\dot{CS}$:

$$\text{VoI} = \mathbb{E}(CS - \dot{CS}) = \mathbb{E}(\dot{p} - p) + \mathbb{E}(\dot{SE} - SE) \tag{11}$$

Tables 6 and 7 report results from our counterfactual experiments. We estimate the average value of information provided by brokers in this market to be £72.31. Given our definition of price, this means that the existence of brokers helps the average mortgagor save over £72.31 a month (or £868 a year) in *sunk* expenditures (i.e. those not related to paying off the principal). If brokers were not present in the market, borrowers would be paying 21.16% more, on average, in monthly instalments and forgoing an additional 70.66% in search cost. These calculations suggest that the role borrowers themselves play in limiting lender's monopoly power, which

[43] In several cases we observed the FP iteration to converge to a degenerate (monopoly) solution with $\dot{q}_1 = 1$. In such cases we tried different starting points to find a solution in the interior. If this method failed, we took the monopoly outcome to be the only solution to the problem, otherwise we discarded it.

[44] The definition adopted here is slightly different from e.g. Baye, Morgan, and Scholten's (2006) discussion of the Varian (1980) model who define value of information as the difference between the expected price of consumers who access the clearinghouse and those who do not.

arises when consumers do not search enough. Importantly, intermediation generates a positive externality for borrowers who search directly.[45]

So far, these results are for the average borrower. But we also show that borrower heterogeneity matters for whether they are winners or losers. Tables 6 and 7 reproduce our estimates, disaggregated by borrower characteristics and mortgage type, respectively.

Table 6: Value of information: breakdown by borrower types.

	VOI	$\%\Delta p$	$\%\Delta SE$
Overall	**72.31**	**+21.16%**	**+70.66%**
Age			
<30	114.08	+30.39%	+163.44%
30+	63.19	+19.14%	+50.39%
Income			
Low	119.81	+31.64%	+120.18%
High	54.89	+17.31%	+52.50%
FTB			
FTB	88.66	+25.56%	+119.16%
Non-FTB	56.75	+16.96%	+119.67%
Location			
Urban	69.14	+19.94%	+79.44%
Rural	84.24	+25.74%	+37.59%

Note: Second column of the table reports the estimated average value of information as defined in equation (11) in GBP per month. The third and fourth columns report the (weighted) average percentage change in prices and search expenditures, respectively. Calculations made by simulating new prices and search behaviour from the new equilibrium, assuming that lenders had the same marginal costs as in the baseline scenario.

One can see from Table 6 that young, low-income, first-time buyers benefit most from having brokers in the market. The counterfactual price they would pay increases markedly in a world with no intermediaries, reaching up to almost 32%. These are also paired off by significant changes in their total cost of search.

More interestingly, in Table 7, we find that not everyone benefits from intermediation. Borrowers with 3- and 5-year fixed rate deals with short amortization periods of under 20 years gain little benefit, and sometimes become even worse off. These results are driven by modest changes in equilibrium prices which come with massive reductions in total search expenditure. They imply that the level of consumer search for those products is low, and consequently– even with brokers present– commissions, market power, and prices are high.[46]

Moreover, the presence of intermediaries substantially affects pricing of mortgages with longer amortization periods. A world without brokers doubles prices of mortgages with amortization terms of 30 years or more. Similarly, brokers help buyers with less popular mortgage products, e.g. flexible repayment schemes, or cashback. They do so by exerting negative pressures on lenders' prices, and by reducing overall search expenditures.

Overall, as our model does not deliver predictions for some general equilibrium effects, our results should be interpreted with three caveats. First, we treat broker fees as exogenous.

[45] This is a natural consequence of the price coherence assumption and is somewhat different from the same finding in Salz (2022) who allowed separate price setting in the two market segments.

[46] Relatively higher mean markup estimates in Table 5 confirms this hypothesis.

Table 7: Value of information: breakdown by loan characteristics.

	VOI	$\%\Delta p$	$\%\Delta SE$
Overall	**72.31**	**+21.16%**	**+70.66%**
LTV			
≤70	56.90	+19.33%	-10.36%
71-75	94.21	+28.36%	+192.95%
76-80	56.72	+17.43%	+57.27%
81-85	97.23	+26.71%	+220.56%
86-90	107.33	+22.84%	+177.51%
91-95	38.80	+8.02%	-58.06%
Deal			
2 years	101.64	+29.27%	+133.04%
3 years	42.18	+14.66%	-35.63%
5 years	4.48	+2.25%	-70.26%
Term			
≤10 years	50.04	+19.11%	-70.55%
(10;15]	26.39	+11.29%	-77.65%
(15;20]	26.82	+9.92%	-76.19%
(20;25]	52.51	+15.98%	-0.98%
(25;30]	87.97	+25.16%	+95.67%
(30;35]	156.39	+40.15%	+409.79%
Value			
Q1	87.01	+26.36%	-30.94%
Q2	63.12	+18.15%	+14.86%
Q3	72.63	+20.59%	+70.43%
Q4	66.15	+19.36%	+218.25%
Flexible			
Flexible	25.67	+11.52%	-68.58%
Regular	79.72	+22.68%	+92.78%
Cashback			
No cashback	76.07	+21.88%	+89.83%
Cashback	47.45	+16.39%	-56.51%

Note: Second column of the table reports the estimated average value of information as defined in equation (11) in GBP per month. The third and fourth columns report the (weighted) average percentage change in prices and search expenditures, respectively. Calculations made by simulating new prices and search behaviour from the new equilibrium, assuming that lenders had the same marginal costs as in the baseline scenario.

And while our results make it tempting to conclude that increasing them by about £100/month would make borrowers better off than in a hypothetical scenario without intermediation, doing that would drastically reduce the demand for broker services and force many brokers to exit the market. Our analysis remains agnostic about what happens then.

Second, we do not allow switching to different mortgage types in our counterfactual. It would be reasonable to assume that some consumers could switch mortgages if they knew brokers almost exclusively provide value when shopping for 2-year fixed rate deals, for example.

Third, our model does not provide an estimate for total broker payoffs[47], so we do not attempt a full welfare analysis. With all that in mind, our result can still be interpreted in terms of value of information provided by brokers to borrowers under the current market structure.

[47] Woodward and Hall (2012) argue that brokers are indifferent between the main source of compensation (i.e. contributions from the lender versus borrower), caring only about total compensation. In our model, procuration fees– provided they are passed onto the borrowers– can be seen as part of estimated lenders' costs. In section A.4 we provide a robustness check where we adjust the estimated cost distributions by potential savings faced by the lenders assuming full pass-through of procuration fees. Overall, the average VOI drops from £112 to £97, and all the results are quantitatively similar to the ones presented here.

8.2 Market centralization

In the second experiment, we consider a hypothetical market centralization. Recently, startups such as Habito[48] have facilitated mortgage search by creating a free online platform propelled by machine learning algorithms, matching borrowers' needs with best prices on offer.[49] Unlike traditional price comparison websites like Moneyfacts, which only list prices, Habito mimics broker services, even helping borrowers through the mortgage application process.

We simulate the effects of extending such a technology to the entire market. We stop lenders from offering products directly, but only through the public platform. In a centralized market, lenders price according to the standard first-price procurement bid formula:

$$\beta(c|\mathbf{x}^G, \mathbf{x}^H) = \beta(c|\mathbf{x}^H) = c + \frac{\int_{s=c}^{\bar{c}} (1 - \mathcal{H}(s|\mathbf{x}^H))^{J-1} ds}{(1 - \mathcal{H}(c|\mathbf{x}^H))^{J-1}} \tag{12}$$

Canonical results from auction theory Milgrom and Weber (1982) assure that the symmetric equilibrium of the bidding game is unique. Therefore, solving for the counterfactual is straightforward, only involving finding the best responses defined by 12 without having to determine optimal consumer search behaviour.

We look at projected benefits from such market regulation, assuming platform access is costless and the environment completely frictionless. The results are summarized in Table 8.

Table 8: Price changes in a centralized market.

	Δp	$\%\Delta p$	ΔSE		Δp	$\%\Delta p$	ΔSE
Overall	**-14.75**	**-4.39%**	**-7.47**	**LTV**			
Age				$\leq$70	-22.45	-7.16%	-8.87
<30	-6.92	-1.51%	-6.79	71-75	-2.64	-0.49%	-5.08
30+	-16.46	-5.01%	-7.62	76-80	-13.46	-3.92%	-5.88
Income				81-85	-5.22	-1.11%	-5.83
Low	-8.97	-1.98%	-6.78	86-90	-8.31	-2.61%	-6.61
High	-16.87	-5.27%	-7.23	91-95	-8.14	-1.44%	-6.13
FTB				**Deal**			
FTB	-15.83	-4.31%	-9.54	2 years	-5.65	-1.95%	-8.30
Non-FTB	-13.73	-4.45%	-5.51	3 years	-25.39	-7.58%	-11.36
Location				5 years	-35.69	-9.96%	-5.17
Urban	-14.05	-4.15%	-7.60	**Value**			
Rural	-17.38	-5.27%	-6.99	Q1	-21.55	-5.99%	-13.72
Term				Q2	-16.63	-4.73%	-7.01
$\leq$10 years	-30.74	-9.63%	-15.48	Q3	-12.79	-3.99%	-5.17
(10;15]	-33.53	-10.55%	-10.72	Q4	-5.22	-1.11%	-5.83
(15;20]	-28.82	-9.01%	-8.11	**Flexible**			
(20;25]	-18.98	-5.54%	-6.66	Flexible	-7.72	-2.49%	-8.23
(25;30]	-9.98	-2.61%	-6.07	Regular	-15.87	-4.68%	-7.35
(30;35]	-14.21	-4.51%	-5.27	**Cashback**			
				No cashback	-13.75	-4.09%	-7.35
				Cashback	-21.37	-6.33%	-8.28

Note: The second column of each panel shows the average absolute difference between prices charged by lenders in a centralized market and prices observed in the data. The third column is the same difference but in relative terms. The fourth column shows the average search expenditure savings a fully frictionless market would generate (per loan in GBP/month).

[48]For the of description of Habito's business model see e.g. The Financial Times: https://www.ftadviser.com/mortgages/2017/01/24/habito-secures-5-5m-to-create-mortgage-platform/

[49]The so called *robo-advice*.

In a market without search frictions, consumers would pay £14.75 less per month on average (or 4.39% less than currently). The benefits are further compounded by search expenditure savings of roughly £7.47 per loan. The sum of these two numbers corresponds to increase in consumer surplus quantitatively very close to the $27.92 Allen et al. (2017) find when eliminating search frictions and limiting banks' market power in the Canadian market. The average reduction of price is double the welfare gains from eliminating search frictions,[50] suggesting that centralization would have a greater impact on competition between lenders than reducing information asymmetries between borrowers.

As in the first counterfactual, the magnitude of the change varies across borrower and product types. Richer and older borrowers would benefit more from market centralization than younger and low income borrowers, which is expected given that this exercise is the flipside of the first one where we found that currently mostly the latter group benefits from brokerage. Increased competition between lenders would render 3- and 5-year fixed mortgages significantly cheaper (by 7.5 and 10%, respectively). Finally, high LTV borrowers and those with higher loan value and longer term would not see a major difference if the entire market was centralized.

Establishing a market-wide platform would certainly stimulate competition between lenders and make borrowers better off. However, in our framework, mortgagors are the only market participants who benefit from this regulation. Banks' markups do get affected, and in the following section we examine what market centralization means for them. However, to comprehensively assess the cost of the regulation, we would need to take a stance on the profits of brokers and the potential sunk costs they would be facing if they had to exit the market.

8.3 Summary of findings

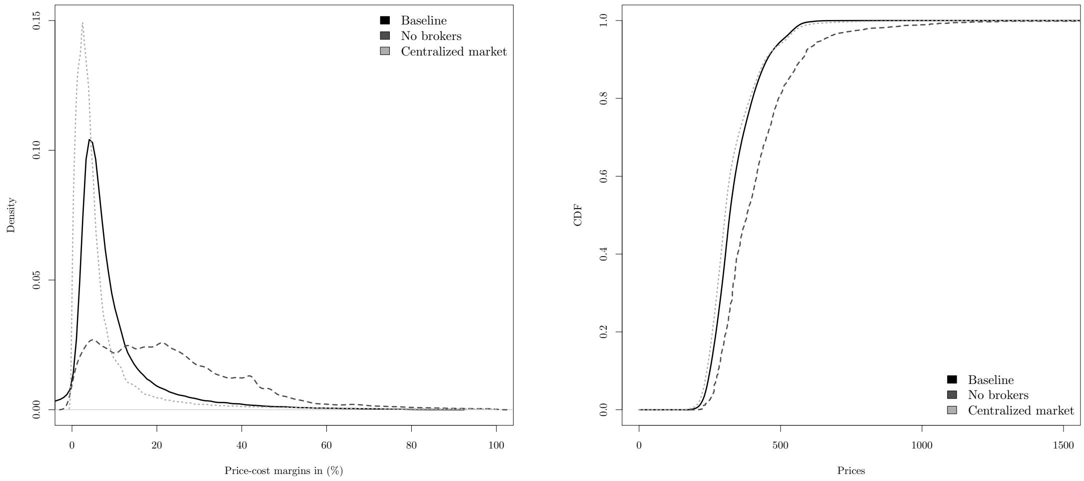
Clearly, the current market structure lies somewhere in between the two extreme cases we consider in our counterfactual experiments. But the real-world set-up is likely much more closely aligned with a centralized market given that 70% of all mortgages are currently brokered. Figure 5 displays the distributions of markups and prices in the data, and the two counterfactual scenarios.

Without brokers, lenders would enjoy much more market power. The average Lerner index would increase to 24.03% and dispersion would also be larger, with 25% of borrowers facing margins of 33.5% or more. In a centralized market, the median PCM is only 3.95%, nearly half of the baseline estimate of 6.4%.

We conclude that the market is currently much more competitive than it would be if brokers were not present. While complete centralization would reduce mortgage prices and lenders' margins, the overall change would be modest and may not sufficiently compensate for the (po-

[50] In fact, it might appear that ΔSE is small compared to our estimates of median search costs from section 7. It should be however noted that in our framework, the first offer is free, and a substantial fraction of borrowers does not engage in search at all.

Figure 5: Distributions of prices and price-cost margins in the counterfactual scenarios.



Note: Kernel estimates of the density of price-cost margins defined as $PCM_{ij} = \frac{p_{ij}-c_{ij}}{p_{ij}}$ (left panel) and estimated CDFs of prices (right panel).

tentially high) costs of establishing such a platform. Overtime, the emergence (and success) of online brokerages may be all we need at present to ensure a fully competitive market. Our policy conclusion is therefore one where the regulator should focus on facilitating broker competition, easing barriers to entry, but without necessarily banning lenders from forming direct sales.

9 Conclusion

This paper estimates the value of information provided by brokers using a structural model of borrower search. Using administrative data on all mortgage originations in 2016 and 2017, we document the existence of price dispersion and the modest pecuniary benefits that arise from shopping with a broker. We show that a large part of deciding whether to use a broker cannot be explained by observable borrower characteristics. This leads us to conclude that shopping with a broker is driven by unobserved heterogeneity in the costs of shopping for a mortgage. Our main identifying assumption is that borrowers with high search costs use brokers to find them the best deals.

Our structural model nonparametrically identifies the distribution of search and lender's costs of providing the loan. We estimate those primitives using techniques recently developed in the consumer search literature, leveraging methods used for nonparametric auction estimation. We find a large variation in search costs across different consumers, and that these costs are sometimes substantial. On the supply side, the market appears to be relatively competitive, with average markups around 10%.

We use the estimates to simulate the effects of removing intermediaries from the market. The difference in consumer surplus is what we label as value of information. On average, we find that broker advice is worth around £72.31 pounds per month, though not every borrower benefits from their presence. In the absence of brokers, firms would enjoy significantly higher monopoly power, and consumers would have to spend more on search. In the second counterfactual, we simulate the effects a hypothetical market centralization, finding that it would lead only to a modest reduction of prices and lenders' market power.

This paper makes two main contributions: first, the empirical results contribute to the policy discussion on the regulation of banks, brokers, and the mortgage market itself. Second, the structural model presented here is straightforward to estimate and simulate, and the results are robust to distributional assumptions, and can be used to study any industry where some consumers can access a platform while others purchase the good directly. This is an attractive framework for empirical studies of welfare effects of multi-sided platforms.

In future work, we hope to use recent results on the estimation of auctions with unobserved heterogeneity (Haile and Kitamura, 2019) to introduce broker heterogeneity into our model. This would allow us to relax the assumption that all intermediaries act as unbiased auctioneers and introduce potentially distorted advice. On the lender side, we could model the decision of offering the product via an intermediary using the results on auctions with endogenous entry.

References

AGARWAL, S., J. GRIGSBY, A. HORTAÇSU, G. MATVOS, A. SERU, AND V. YAO (2020): "Searching for Approval," Working Paper.

ALEXANDROV, A. AND S. KOULAYEV (2018): "No Shopping in the U.S. Mortgage Market: Direct and Strategic Effects of Providing Information," CFPB Office of Research Working Paper No. 2017-01.

ALLEN, J., R. CLARK, AND J.-F. HOUDE (2013): "The Effect of Mergers in Search Markets: Evidence from the Canadian Mortgage Industry," *American Economic Review*, 104, 3365–3396.

——— (2014): "Price Dispersion in Mortgage Markets," *Journal of Industrial Economics*, 62, 377–416.

——— (2017): "Search Frictions and Market Power in Negotiated Price Markets," *Journal of Political Economy*, forthcoming.

ANDERSEN, S., J. Y. CAMPBELL, K. M. NIELSEN, AND T. RAMADORAI (2020): "Sources of Inaction in Household Finance: Evidence from the Danish Mortgage Market," *American Economic Review*, 110, 3184–3230.

ARMSTRONG, M. (2006): "Competition in Two-Sided Markets," *RAND Journal of Economics*, 37, 668–691.

ATHEY, S. AND P. A. HAILE (2007): "Nonparametric Approaches to Auctions," in *Handbook of Econometrics*, ed. by E. E. L. James J. Heckman, Elsevier, vol. 6A, 3847–3965.

BAYE, M. R. AND J. MORGAN (2001): "Information Gatekeepers on the Internet and the Competitiveness of Homoegneous Product Markets," *American Economic Review*, 91, 454–474.

BAYE, M. R., J. MORGAN, AND P. SCHOLTEN (2006): "Information, Search, and Price Dispersion," in *Handbook of Economics and Information Systems*, ed. by T. Hendershott, Elsevier Press, Amsterdam.

BENETTON, M. (2021): "Leverage Regulation and Market Structure: A Structural Model of the U.K. Mortgage Market," *Journal of Finance*, 76, 2997–3053.

BERGSTRESSER, D., J. M. CHALMERS, AND P. TUFANO (2007): "Assessing the Costs and Benefits of Brokers in the Mutual Fund Industry," *Review of Financial Studies*, 22, 4129–4156.

BHUTTA, N., A. FUSTER, AND A. HIZMO (2020): "Paying Too Much? Price Dispersion in the US Mortgage Market," CEPR Discussion Papers 14924.

BIGLAISER, G. (1993): "Middlemen as experts," *RAND Journal of Economics*, 24, 212–223.

BURDETT, K. AND K. JUDD (1983): "Equilibrium Price Dispersion," *Econometrica*, 51, 955–969.

BYRNE, D. P. AND L. A. MARTIN (2021): "Consumer search and income inequality," *International Journal of Industrial Organization*, 79, 102716.

CHEN, S. X. (1999): "Beta Kernel Estimators for Density Functions," *Computational Statistics & Data Analysis*, 31, 131–145.

COEN, J., A. KASHYAP, AND M. ROSTOM (2021): "Price discrimination and mortgage choice," Staff Working Paper 926, Bank of England.

DE CORNIÈRE, A. AND G. TAYLOR (2017): "A Model of Biased Intermediation," Working Paper, Toulouse School of Economics.

DELTAS, G. AND Z. LI (2018): "Free Riding on the Search of Others: Information Externalities in the Mortgage Industry," Working Paper, University of Illinois.

DEUFLHARD, F. (2016): "Quantifying Inertia in Retail Deposit Markets," Working Paper.

EDELMAN, B. AND J. WRIGHT (2015): "Price Coherence and Excessive Intermediation," *Quarterly Journal of Economics*, 130, 1283–1328.

EGAN, M. (2018): "Brokers vs. Retail Investors: Conflicting Interests and Dominated Products," *Journal of Finance*, forthcoming.

EGAN, M., G. MATVOS, AND A. SERU (2018): "The Market for Financial Adviser Misconduct," *Journal of Political Economy*, forthcoming.

FRANKEL, A. S. (1998): "Monopoly and Competition in the Supply and Exchange of Money," *Antitrust Law Journal*, 66, 313–361.

GALEOTTI, A. AND J. L. MORAGA-GONZÁLEZ (2009): "Platform Intermediation in a Market for Differentiated Products," *European Economic Review*, 53, 417–428.

GAVAZZA, A. (2016): "An Empirical Equilibrium Model of a Decentralized Asset Market," *Econometrica*, 84, 1755–1798.

GUERRE, E., I. PERRIGNE, AND Q. VUONG (2000): "Optimal Nonparametric Estimation of First-Price Auctions," *Econometrica*, 68, 525–574.

GUISO, L., A. POZZI, A. TSOY, L. GAMBACORTA, AND P. E. MISTRULLI (2021): "The cost of steering in financial markets: Evidence from the mortgage market," *Journal of Financial Economics*.

HAILE, P. A. AND Y. KITAMURA (2019): "Unobserved Heterogeneity in Auctions," *The Econometrics Journal*, 22, C1–C19.

HALL, G. AND J. RUST (2003): "Middle Men versus Market Makers: A Theory of Competitive Exchange," *Journal of Political Economy*, 111, 353–403.

HASTINGS, J. S., B. C. MADRIAN, AND W. L. SKIMMYHORN (2013): "Financial Literacy, Financial Education and Economic Outcomes," *Annual Review of Economics*, 1, 347–373.

HAYFIELD, T. AND J. S. RACINE (2008): "Nonparametric Econometrics: The np Package," *Journal of Statistical Software*, 27.

HONG, H. AND M. SHUM (2006): "Using Price Distribution to Estimate Search Costs," *RAND Journal of Economics*, 37, 257–275.

HONKA, E., A. HORTAÇSU, AND M. A. VITORINO (2017): "Advertising, Consumer Awareness and Choice: Evidence from the U.S. Banking Industry," *RAND Journal of Economics*, 48, 611–646.

IMLA (2015): "The Changing Face of Mortgage Distribution," Intermediary Mortgage Lenders Association report.

INDERST, R. AND M. OTTAVIANI (2012a): "Competition through Commissions and Kickbacks," *American Economic Review*, 102, 780–809.

——— (2012b): "Financial Advice," *Journal of Economic Literature*, 50, 494–512.

——— (2012c): "How (Not) to Pay for Advice: A Framework for Consumer Financial Protection," *Journal of Financial Economics*, 105, 393–411.

ISCENKO, Z. (2018): "Choices of Dominated Mortgage Products by UK Consumers," FCA Occasional Paper No. 33.

ISCENKO, Z. AND J. NIEBOER (2018): "Effects of the advice requirement and intermediation in the UK mortgage market," FCA Occasional Paper No. 34.

KIM, M., D. KLIGER, AND B. VALE (2003): "Estimating Switching Costs: The Case of Banking," *Journal of Financial Intermediation*, 12, 25–56.

LUSARDI, A. AND O. S. MITCHELL (2014): "The Economic Importance of Financial Literacy: Theory and Evidence," *Journal of Economic Literature*, 52, 5–44.

MACMINN, R. D. (1980): "Search and Market Equilibrium," *Journal of Political Economy*, 88, 308–327.

MALLIARIS, S., D. A. RETTL, AND R. SINGH (2020): "Is competition a cure for confusion? Evidence from the residential mortgage market," *Real Estate Economics*, 50, 206–246.

MARRON, J. S. AND D. RUPPERT (1994): "Transformations to Reduce Boundary Bias in Kernel Density Estimation," *Journal of the Royal Statistical Society Series B*, 56, 653–671.

MILGROM, P. AND R. J. WEBER (1982): "A Theory of Auctions and Competitive Bidding," *Econometrica*, 50, 1081–1122.

MORAGA-GONZÁLEZ, J. L., Z. SÁNDOR, AND M. R. WILDENBEEST (2013): "Semi-Nonparametric Estimation of Consumer Search Costs," *Journal of Applied Econometrics*, 28, 1205–1223.

MYŚLIWSKI, M., F. SANCHES, D. SILVA JUNIOR, AND S. SRISUMA (2020): "Identification and Estimation of a Search Model: A Procurement Auction Approach," Working Paper.

RENY, P. J. (2011): "On the Existence of Monotone Pure-Strategy Equilibria in Bayesian games," *Econometrica*, 79, 499–553.

ROBLES-GARCIA, C. (2020): "Competition and Incentives in Mortgage Markets: The Role of Brokers," Working Paper.

ROCHET, J.-C. AND J. TIROLE (2006): "Two-Sided Markets: A Progress Report," *RAND Journal of Economics*, 37, 645–667.

ROSENTHAL, R. W. (1980): "A Model in Which an Increase in the Number of Sellers Leads to a Higher Price," *Econometrica*, 48, 1575–1580.

RUBINSTEIN, A. AND A. WOLINSKY (1987): "Middlemen," *Quarterly Journal of Economics*, 102, 581–594.

SALOP, S. C. AND J. E. STIGLITZ (1977): "Bargains and Ripoffs: A Model of Monopolistically Competitive Price Dispersion," *Review of Economic Studies*, 44, 493–510.

SALZ, T. (2022): "Intermediation and Competition in Search Markets: An Empirical Case Study," *Journal of Political Economy*, 130.

SANCHES, F., D. SILVA JUNIOR, AND S. SRISUMA (2016): "Minimum Distance Estimation of Search Costs using Price Distribution," *Journal of Business and Economic Statistics*, forthcoming.

SPULBER, D. F. (1995): "Bertrand Competition when Rivals' Costs are Unknown," *Journal of Industrial Economics*, 43, 1–11.

THIEL, J. H. (forthcoming): "Adviser Compensation, Endogenous Entry, and the Advice Gap," *American Economic Journal: Microeconomics*.

VARIAN, H. R. (1980): "A Model of Sales," *American Economic Review*, 70, 651–659.

WOODWARD, S. E. AND R. E. HALL (2012): "Diagnosing Consumer Confusion and Sub-optimal Shopping Effort: Theory and Mortgage Market Evidence," *American Economic Review*, 102, 3249–3276.

YAVAŞ, A. (1994): "Middlemen in Bilateral Search Markets," *Journal of Labor Economics*, 12, 406–429.

A Appendix: data and reduced form results

A.1 Data and summary statistics

Table A.1 summarizes our main variables of interest by broker usage for different types of mortgage products– the two-, three-, and five-year FRMs. There is variation in loan size, fees, and offered interest rate across product type, but between borrowers who go direct or use brokers. The last row in the table shows monthly interest payments normalized by the size of the loan, which is our preferred measure of calculating mortgage cost. Section 3.1 outlines the calculation in detail.

Table A.1: Summary statistics

	2 yr FRM		**3 yr FRM**		**5 yr FRM**	
	Direct	**Broker**	**Direct**	**Broker**	**Direct**	**Broker**
Interest (bps)	232	213	245	235	259	247
Loan (£)	161,070	185,578	166,804	176,525	145,680	167,747
Loan Fees (£)	373.97	575.18	434.58	578.43	495.12	620.56
Monthly payment (£)	788.47	781.00	831.52	796.33	782.82	791.66
Monthly interest (£)	317.85	358.93	345.73	363.63	316.03	355.56
Normalized payment (£)	305.84	298.95	314.16	312.54	331.70	320.59

Note: Interest is the interest rate in basis points. Loan is the size of the mortgage issued by the bank. Monthly payment is the payment of capital and interest during the initial contract period of the loan, excluding lender fees. Monthly interest is the component of the monthly payment that goes towards payment of the interest, and includes the fees. Normalized interest payment is the monthly interest payment normalized to take into account the size of the loan.

Mortgage contracts in the UK are short-term, with an initial duration of 2-, 3-, or 5-years. Following the expiration of the initial period, and if the household does not refinance, the mortgage contract reverts to the bank's posted rate, or Standard Variable Rate (SVR). There are two types of contracts in the UK: fixed and variable. Fixed rate mortgages (FRM) have a fixed interest rate during the initial period, while adjustable rate mortgages (ARM) have a fluctuating rate that is a discount off of the SVR. Mortgage rates are arranged according to the length of the initial period and by LTV band. The longer the period and the higher the LTV, the more expensive the product. Table A.2 shows that, on average in our sample, households pay 230 basis points on their mortgage product, but that there is a spread of 280 basis points between the 2-year FRM at 70% LTV (cheapest) and the 5-year FRM at 95% LTV. Given that yield curves were roughly flat during this period, spreads across products have remained more or less constant.

Just over one-third of our sample are FTB, with the remainder either moving home or remortgaging their current home. But there is variation in the distribution of mortgagors at different LTV bands. Table A.3 shows that 80% of mortgagors on 95% LTV products are FTB, whereas 80% of mortgagors who took out an LTV of 70% or less are non-FTB.

Different banks also specialize in different products, with the share of longer term products more likely to be offered by some banks over others. This can be seen in table A.4

Table A.2: Interest Rates by LTV and Rate Duration

	2 yr FRM	**3 yr FRM**	**5 yr FRM**	**Total**
≤ 70	1.8	2.2	2.3	2.0
71 - 75	1.8	2.2	2.5	2.0
76 - 80	1.9	2.4	2.6	2.1
81 - 85	2.1	2.5	2.8	2.2
86 - 90	2.8	3.0	3.3	2.9
91 - 95	3.0	4.0	4.6	4.0
Total	2.2	2.4	2.5	2.3

Table A.3: Share by Household type and LTV

	Non-FTB	FTB	Total
≤ 70	82	18	100
71 - 75	60	40	100
76 - 80	68	32	100
81 - 85	56	44	100
86 - 90	36	64	100
91 - 95	19	81	100
Total	64	36	100

Table A.4: Share by Bank and Product Type

	2 yr FRM	**3 yr FRM**	**5 yr FRM**	**Total**
Bank 1	76.61	0.89	22.50	100
Bank 2	67.48	2.26	30.26	100
Bank 3	58.80	9.75	31.45	100
Bank 4	66.22	4.71	29.07	100
Bank 5	44.19	10.74	45.07	100
Bank 6	72.30	1.42	26.28	100
Total	66.27	4.33	29.40	100

A.2 Estimation sample

We restrict our sample to standard[51] fixed rate mortgage products with two-, three-, and five-year durations; and to loan sizes less than £1M. This leaves us with about 82% of the sample (1.7M loans) for analysis. We further restrict our sample to the six largest mortgage providers which made up about 75% (or 1.3M loans) in 2016 and 2017. The differences between the raw and final sample are tabulated in table A.5.

[51] These are products that include repayment of the capital.

Table A.5: Raw and Final Sample

	Big Six	%	Raw Sample	%
Total	1,539,009	100.00	2,138,754	100.00
Interest-only mortgages	43,276	2.81	81,482	3.81
Non-FRM	114,099	7.41	152,856	7.15
Not 2, 3, 5yrs	61,765	4.01	141,054	6.60
£1M+ loan	4,186	0.27	5,886	0.28
Outliers	6,606	0.43	13,892	0.65
Final Sample	**1,309,077**	85.06	1,743,584	81.52

A.3 Probability of using a broker

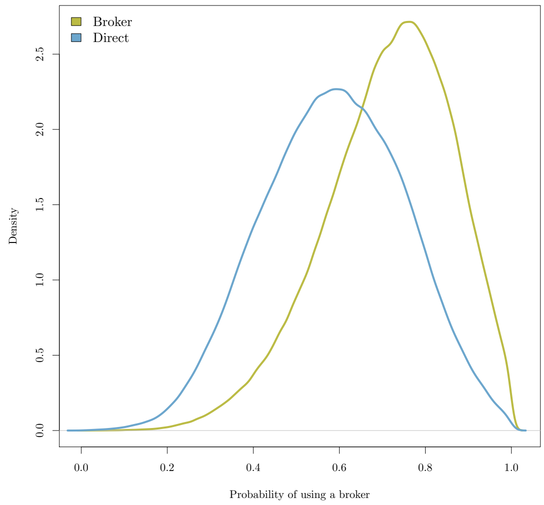
Table A.6 reports the estimates from a linear probability model where we regressed the indicator whether the contract was brokered on a number of personal and product characteristics. The first observation is that the signs are in line with intuition. For example, lower income, first time buyers, the employed, and older mortgagors are more likely to use a broker. Moving to column 2, adding product characteristics shows that mortgagors who took longer term contracts were less likely to visit brokers (the causality may also be in the other direction, so we interpret the results in terms of conditional correlations, rather than causal relationships). In fact, a recent FCA investigation (Iscenko and Nieboer, 2018) hypothesises that brokers might be more likely to suggest 2-year contracts knowing that this makes borrowers use their services more frequently in the future. A longer mortgage term is also associated with increased probability of using brokers. However, column 2 also shows that when product characteristics are added, the sign on LTV indicators is reversed from positive (column 1) to negative. In fact, the higher the LTV the less likely a household uses a broker. This may be for a number of reasons, for example, households on low LTV products typically have smaller absolute loans, therefore the costs of visiting a broker and paying a lump-sum is relatively higher. Finally, column 3 shows that even after controlling for regional fixed effects, the coefficients remain unchanged and the R^2 remains low, so the observables are rather poor predictors for broker use.

Table A.6: Probability of using a broker

	(1)	(2)	(3)
Dependent var: Used a broker	Personal Characteristics	Product Characteristics	Regional Characteristics
Income	-0.002***	-0.016***	-0.037***
First Time Buyer	0.048***	0.016***	0.009***
Aged 25 - 29	0.024***	0.028***	0.024***
Aged 30-34	0.042***	0.068***	0.063***
Aged 35- 39	0.049***	0.121***	0.115***
Aged 40 - 45	0.036***	0.179***	0.172***
Aged 45+	-0.022***	0.251***	0.241***
71 - 75 LTV	0.131***	0.063***	0.078***
76 - 80 LTV	0.042***	-0.029***	-0.012***
81 - 85 LTV	0.075***	-0.019***	-0.000
86 - 90 LTV	0.035***	-0.066***	-0.041***
91 - 95 LTV	0.028***	-0.111***	-0.084***
Employed	-0.054***	-0.046***	-0.045***
Mortgage Term		0.021***	0.020***
3 Year FRM		-0.229***	-0.228***
5 Year FRM		-0.187***	-0.184***
Flexible Mortgage		0.086***	0.083***
Urban area			-0.011***
Regional FE	No	No	Yes
Observations	1,309,067	1,309,067	1,307,538
R^2	0.020	0.124	0.130

Note: *** denotes 1% significance level. Robust standard errors used.

Figure A.1: Distributions of predicted probabilities of using a broker.



Note: Density estimates of the distributions of $\widehat{\Pr}(d_i = \text{broker}|\mathbf{X})$ based on the LPM in the third column of table A.6 for the brokered and direct subsamples.

A.4 Robustness checks

This section presents robustness checks, which examine potential effects of procuration fees paid by the lenders to the brokers. The first two tables display the results of the regression of prices on brokered dummy (table 2 in the main text) for two subsamples of the data – A.7 only uses data on brokers who are not paid by the borrowers directly and are only compensated by the lenders, while A.8 only uses data on brokers who are not paid by the lenders and are only paid directly by the borrowers. The signs on the variables of interest are negative for all specifications and subsamples. This suggests that there is no evidence that different sources of compensation can alter brokers incentives to provide advice about cheaper products.

Table A.7: Price benefits of using a broker: brokers who do not charge the borrowers.

Dependent variable:	(1) Interest	(2) Interest	(3) Interest	(4) Monthly Payment	(5) Monthly Payment
Used a broker	-6.509***	-6.370***	-7.720***	-2.091***	-2.210***
	(0.0902)	(0.0917)	(0.0927)	(0.143)	(0.153)
Lender Fees	Linear	Linear	Non-linear	-	-
Controls	Yes	Yes	Yes	Yes	Yes
Regional FE	No	Yes	Yes	No	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
Observations	940,921	940,921	940,921	940,921	940,921
R^2	0.741	0.747	0.754	0.600	0.605

Note: *** denotes significant at 1% level. Robust standard errors in parentheses. Interest is measured in basis points. Monthly interest is the component of the initial monthly payment that goes towards payment of the interest, including lender fees, and normalized by the size of the loan. Controls are income, house price, loan size, LTV, first time buyer and mortgage term. Time fixed effects are at the monthly level. Regional fixed effects are at the Government Office Region level and include a flag for an urban region. Non-linearities in lender fees are controlled for using a fifth-order spline.

Table A.8: Price benefits of using a broker: brokers who are not paid by lenders.

Dependent variable:	(1) Interest	(2) Interest	(3) Interest	(4) Monthly Payment	(5) Monthly Payment
Used a broker	-6.181***	-2.443***	-4.390***	-6.151***	-1.046***
	(0.180)	(0.177)	(0.182)	(0.244)	(0.245)
Lender Fees	Linear	Linear	Non-linear	-	-
Controls	Yes	Yes	Yes	Yes	Yes
Regional FE	No	Yes	Yes	No	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
Observations	464,012	464,012	464,012	464,012	464,012
R^2	0.689	0.698	0.704	0.628	0.636

Note: *** denotes significant at 1% level. Robust standard errors in parentheses. Interest is measured in basis points. Monthly interest is the component of the initial monthly payment that goes towards payment of the interest, including lender fees, and normalized by the size of the loan. Controls are income, house price, loan size, LTV, first time buyer and mortgage term. Time fixed effects are at the monthly level. Regional fixed effects are at the Government Office Region level and include a flag for an urban region. Non-linearities in lender fees are controlled for using a fifth-order spline.

Tables A.9 and A.10 display alternative calculations of the value of information under the assumption that removing brokers would reduce lenders' costs by the expected amount of procuration fees. To implement this, we adjust each estimated $\mathcal{H}(\cdot|\mathbf{x}^H)$ by $\overline{\Delta^B} \cdot \overline{\phi}(\mathbf{x}^H)$, where the first term is the average proportion of brokered loans with characteristics $\mathbf{x}^H$ and the second term is the average observed procuration fee for a mortgage characterized by $\mathbf{x}^H$ taken from the data. Since the result is a leftward shift of the entire distribution, equilibrium search behaviour does not change because proportions of borrowers searching different number of lenders depend on the differences in the expected prices and not the level of prices itself. The results of this exercise are valid under the assumption that the mortgage sold through a broker and directly is indeed the same product so any additional cost, such as the procuration fee if it is sold through a broker, is also indirectly passed onto consumers who obtain it directly from the lender. The numbers in the tables below should be compared to tables 6 and 7 in the main text. Overall, adjusting for procuration fees reduces the value of information by about $\pounds 15$ through a smaller increase in prices (29% vs. 33%).

Table A.9: Value of information with adjusted costs: breakdown by borrower types.

	VOI	$\%\Delta p$	$\%\Delta SE$
Overall	**97.38**	**+29.15%**	**+16.33%**
Age			
<30	183.46	+51.18%	+65.49%
30+	78.76	+24.39%	+5.70%
Income			
Low	135.64	+36.53%	-18.62%
High	83.44	+26.46%	+50.43%
FTB			
FTB	111.57	+32.88%	+62.98%
Non-FTB	83.42	+25.51%	-0.67%
Location			
Urban	99.57	+29.42%	+26.05%
Rural	89.33	+28.18%	-19.46%

Note: Second column of the table reports the estimated average value of information as defined in equation (11) in GBP per month. The third and fourth columns report the average percentage change in prices and search expenditures, respectively. Calculations made by simulating new prices and search behaviour from the new equilibrium, assuming that lenders drew had the same cost draws as in the baseline scenario. Marginal cost distributions are adjusted to account for the fact that in a world without brokers, lenders do not pay procuration fees.

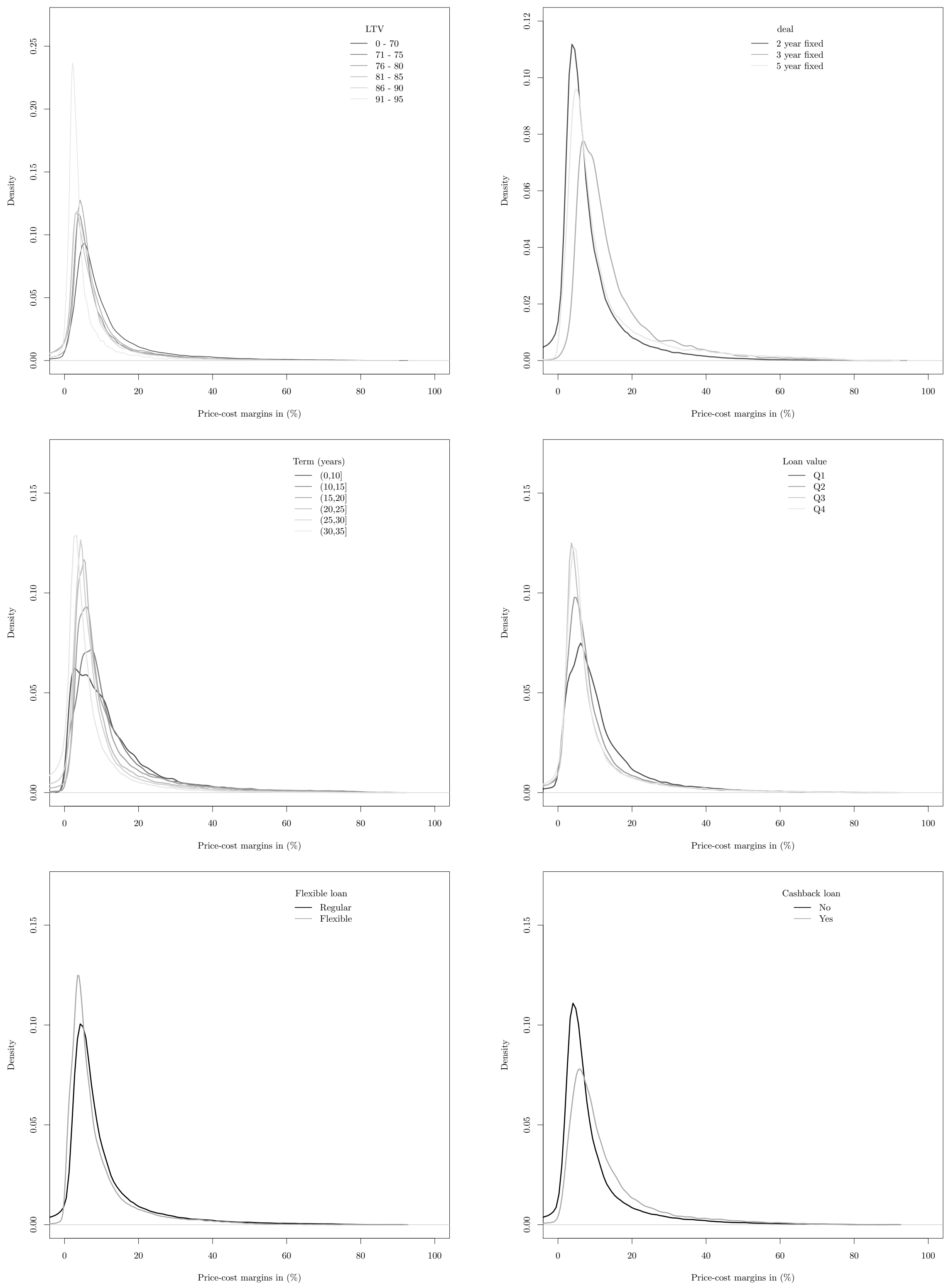
Table A.10: Value of information with adjusted costs: breakdown by loan characteristics.

	VOI	$\%\Delta p$	$\%\Delta SE$
Overall	**97.37**	**+29.15%**	**+16.33%**
LTV			
≤70	69.65	+23.09%	-26.37%
71-75	132.21	+42.63%	+90.52%
76-80	51.31	+16.38%	+26.06%
81-85	163.04	+49.09%	+98.04%
86-90	151.47	+34.97%	+48.36%
91-95	45.49	+9.13%	-50.29%
Deal			
2 years	146.15	+43.38%	+48.25%
3 years	-9.45	-0.96%	-67.62%
5 years	-9.27	-2.05%	-52.44%
Term			
≤10 years	-2.69	+2.26%	-86.19%
(10;15]	-24.38	-5.34%	-77.46%
(15;20]	-9.36	-1.33%	-69.37%
(20;25]	60.89	+19.53%	-43.84%
(25;30]	108.85	+31.75%	+7.88%
(30;35]	342.77	+98.11%	+284.55%
Value			
Q1	101.83	+30.98%	-50.97%
Q2	66.17	+19.43%	-28.31%
Q3	103.31	+29.87%	+30.91%
Q4	116.71	+35.83%	+112.17%
Flexible			
Flexible	4.47	+4.13%	-82.27%
Regular	111.71	+33.01%	+31.69%
Cashback			
No cashback	106.57	+31.59%	-70.04%
Cashback	35.87	+12.84%	+29.33%

Note: Second column of the table reports the estimated average value of information as defined in equation (11) in GBP per month. The third and fourth columns report the average percentage change in prices and search expenditures, respectively. Calculations made by simulating new prices and search behaviour from the new equilibrium, assuming that lenders drew had the same cost draws as in the baseline scenario. Marginal cost distributions are adjusted to account for the fact that in a world without brokers, lenders do not pay procurement fees.

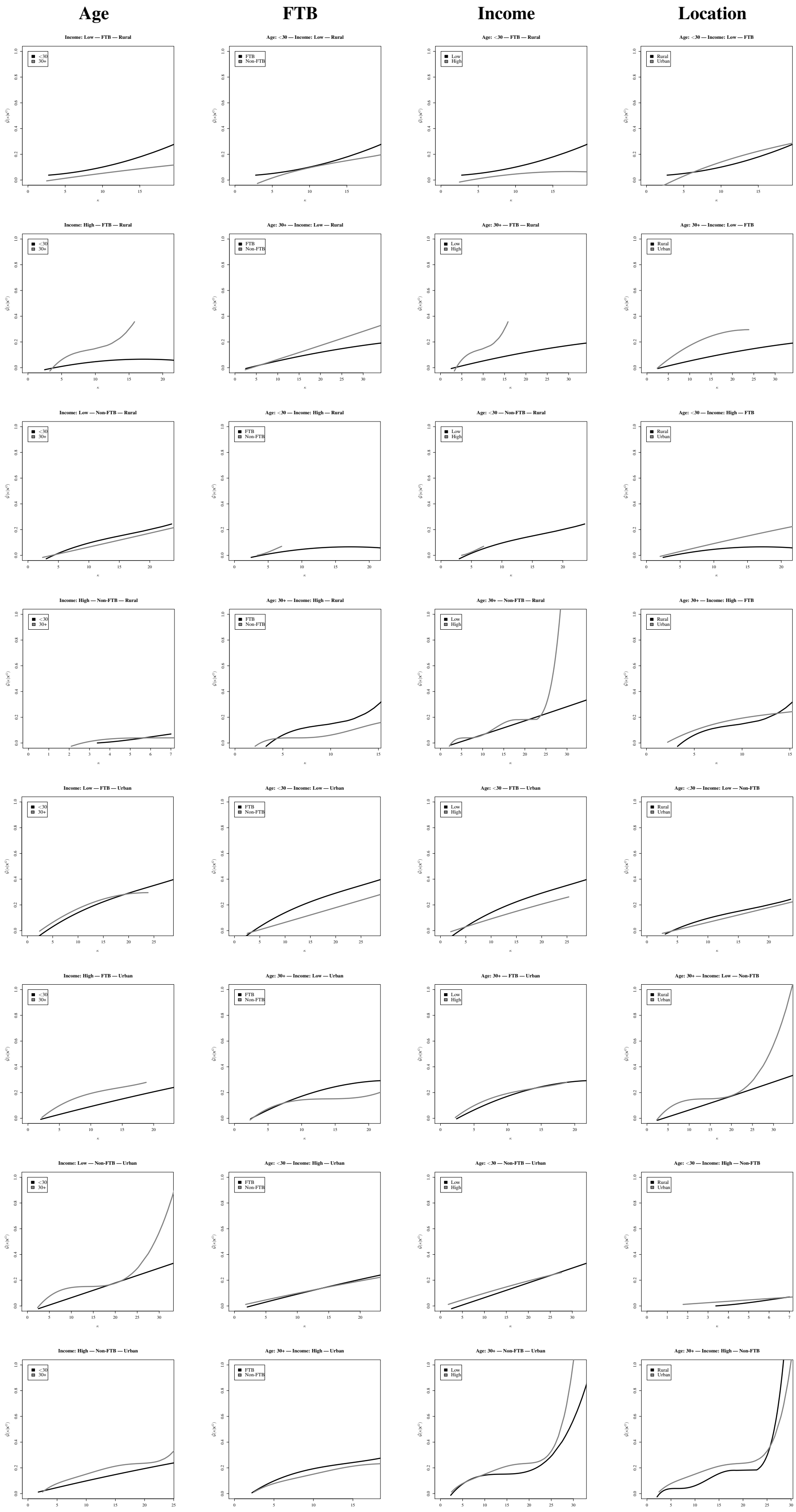
B Appendix: supplementary figures

Figure B.1: Distributions of price-cost margins.



Note: Kernel estimate of the density of price-cost margins defined as $PCM_{ij} = \frac{p_{ij} - c_{ij}}{p_{ij}}$.

Figure B.2: Pairwise comparisons of estimated search cost CDFs.



Note: Figure shows estimated search cost distributions in a way that allows to compare them across one dimension of observed heterogeneity (see top of each column), keeping all the other ones fixed at all their possible values (see graph headings).

C Appendix: Model estimation and results

C.1 Estimation algorithm

To ease the notation, let s index distinct discrete combinations of $(\mathbf{x}^G, \mathbf{x}^H)$ and, for the sake of brevity, let $\mathcal{F}_s(p) \equiv \mathcal{F}(p|s)$, $f_s(p) \equiv f(p|s)$, and $\Delta_s^B \equiv \Delta^B(s)$. The estimation algorithm consists of the following 5 steps:

1. We estimate $\mathcal{F}_s(p)$ and $f_s(p)$ separately for all $s \in \{\mathbf{x}^G \times \mathbf{x}^H\}$, where the cardinality of the set of covariates depends on how the variables are discretized (can potentially be very large). The CDF is estimated simply as:
$$\hat{\mathcal{F}}_s(p) = \frac{1}{n_s} \sum_{i=1}^{n_s} \mathbf{1}\{p_i \leq p\}.$$
To estimate the density, we need to to address the problem of bias near the lower boundary and the possibility that the density near the upper boundary may be unbounded. To tackle this, we use an asymmetric Beta kernel suggested by Chen (1999) that performs well on densities defined over compact supports[52] together with the transformation method of Marron and Ruppert (1994) near the upper boundary.

2. Using our transaction data, we construct a vector of observed market shares[53], $Y_s = (Y_{1s}, \ldots, Y_{Js})^\top$ for every s. Let $\mathbf{X}_s$ be a $J_s \times J_s$ matrix such that $(\mathbf{X}_s)_{j\ell} = \frac{\ell}{J_s}(1 - \mathcal{F}_s(p_j))^{\ell-1}$. Then under Assumption I of Myśliwski et al. (2020) we have:
$$\tilde{\mathbf{q}}(s) = \frac{\mathbb{E}\left[\mathbf{X_s}^\top \mathbf{X_s}\right]^{-1} \mathbb{E}\left[\mathbf{X_s}^\top Y_s\right]}{\iota^\top \mathbb{E}\left[\mathbf{X_s}^\top \mathbf{X_s}\right]^{-1} \mathbb{E}\left[\mathbf{X_s}^\top Y_s\right]}, \tag{C.1}$$
where ι denotes a vector of ones. To obtain an estimate of $\tilde{\mathbf{q}}_\mathbf{s}$, we use the estimated CDFs from step 1 and evaluate them at the average price charged by firm j conditional on s, obtaining $\left(\widehat{\mathbf{X}}_s\right)_{j\ell} = \frac{\ell}{J_s}\left(1 - \widehat{\mathcal{F}}_s(\bar{p}_j)\right)^{\ell-1}$ as the sample analogue of the $\mathbf{X}_s$ matrix.[54] The intuition behind this step is that, conditional on price, the observed difference in market shares can only be explained by some consumers having different number of offers to compare than others. Therefore the variation in market shares conditional on price identifies the search proportions.
To accommodate the restriction that $\tilde{q}_{J_s}(s) \geq \Delta_s^B$, we use constrained quadratic programming to solve the least squares problem (C.1). The right-hand side of the constraint, Δ_s^B is the proportion of brokered mortgages and can be directly obtained from the data.

[52] The implementation comes from the `npuniden.boundary` function from the `np` package in `R` (Hayfield and Racine, 2008).

[53] We experimented with both aggregate market shares over the entire period of the sample as well as quarterly shares. With a fine grid for $(\mathbf{x}^G, \mathbf{x}^H)$, obtaining precise estimates of quarterly shares requires a lot of data to prevent

[54] The identifying assumption suggested by MSSS is that the observed market shares are systematically related (proportional) to the ex-ante probabilities of winning the procurement auction. A slight difficulty in the empirical application using transaction data is that constructing market shares from transaction data typically requires summing over multiple transactions by the same firm, which tend to be associated with different prices. Therefore one needs to choose at which price should the CDF be evaluated. Using the average is consistent with the proportionality assumption – since lenders are assumed to have the same underlying cost distribution, we can only explain differences in aggregate market shares in the data by lower/higher draws from $\mathcal{H}$ and consequently lower/higher average prices quoted.

3. Estimate the vectors of cutoff types $\boldsymbol{\kappa}_s \equiv \boldsymbol{\kappa}(s)$ for each $s \in \{\mathbf{x}^G \times \mathbf{x}^H\}$, where for $\ell \in \{1, \ldots, J_s - 1\}$:
$$\kappa_\ell(s) = \mathbb{E}_{\mathcal{F}_s}\left[p_{(1:\ell)}\right] - \mathbb{E}_{\mathcal{F}_s}\left[p_{(1:\ell+1)}\right]$$
and the marginal type who is indifferent between using a broker and searching directly is estimated as:
$$\bar{\kappa}(s) = \frac{\varrho(s) - \left(\mathbb{E}_{\mathcal{F}_s}\left[p_{(1:k^*)}\right] - \mathbb{E}_{\mathcal{F}_s}\left[p_{(1:J)}\right]\right)}{k^* - 1}.$$
$\varrho(s)$ is the average broker commission and k^* is the equilibrium number of searches of the marginal type. To determine k^*, we find the lowest ℓ, such that $\kappa_\ell(s) < \bar{\kappa}(s)$.[55] To estimate the expectations of the order statistics, we draw repeatedly from the price distributions and calculate the sample averages of the minimum prices.
Finally, $\hat{\hat{\mathbf{q}}}(s)$ estimated in the previous step can be used to recover $\mathcal{G}$ evaluated at the cutoff points as follows:
$$\begin{aligned} \mathcal{G}(\bar{\kappa}(s)|\mathbf{x}^G) &= 1 - \Delta_s^B \\ \mathcal{G}(\kappa_{k^*}(s)|\mathbf{x}^G) &= 1 - \Delta_s^B - \hat{\hat{q}}_{k^*}(s) \\ &\vdots \\ \mathcal{G}(\kappa_{J_s-1}(s)|\mathbf{x}^G) &= 1 - \Delta_s^B - \hat{\hat{q}}_{k^*}(s) - \cdots - \hat{\hat{q}}_{J_s-1}(s) \end{aligned}$$

4. Let $\{s^H(x)\}_{x=1}^{|\mathbf{x}^G|}$ denote the collection of partitions of $\{\mathbf{x}^G \times \mathbf{x}^H\}$ such that $\mathbf{x}^G = x$. The cardinality of each of those sets, $|s^H(x)|$, is equal to the cardinality of $\mathbf{x}^H$, say n_H. For each of the sets, we now have n_H estimates of the cutoffs $\{\boldsymbol{\kappa}^t\}_{t \in s^H(x)}$ and $\{\mathcal{G}(\boldsymbol{\kappa}^t|\mathbf{x}^G)\}_{t \in s^H(x)}$. We then pool the estimates using the method suggested in Section 4 of (Sanches et al., 2016). Specifically, separately for each x, we seek to minimize the following least squares criterion function:
$$\Psi_x(g) = \frac{1}{n_H}\sum_{t=1}^{n_H}\sum_{\ell=1}^{J_t-1}\left[\mathcal{G}(\kappa_\ell^t|\mathbf{x}^G = x) - g(\kappa_\ell^t)\right]^2,$$
where g is a flexible function of the cutoffs. To impose appropriate shape restrictions on the estimated CDF, we choose Bernstein polynomials[56] to construct the sieve.
This step results in a sieve-least squares estimator for $\mathcal{G}(\cdot|\mathbf{x}^G)$, whose theoretical properties and assumptions needed for consistency are discussed in Sanches et al. (2016).

5. In the final step, we recover the distributions of lenders' marginal costs. This step is reminiscent of recovering the distribution of valuations from observed bids in a first-price auction (Guerre et al., 2000). First, for each observed price, we construct pseudo-marginal costs using the inverse

[55] Clearly, $\bar{\kappa}(s)$ is not identified if $k^* = 1$, so if a borrower is now indifferent between using a broker or not, she would not search beyond the first offer she receives for free if intermediation was not available. In this case, we replace $\bar{\kappa}(s) = \kappa_1(s) + \epsilon$ where $\epsilon \sim \text{Unif}[0, \kappa_1(s)]$.

[56] A Bernstein polynomial of order P is a set of $p = 0, \ldots, P+1$ functions where $g_{pP}(\kappa) = \frac{P!}{p!(P-p)!}\kappa^p(1-\kappa)^{P-p}$

of the bidding function:

$$\hat{c}_{ij}(s) = p_{ij} - \frac{\sum_{\ell=1}^{J_s} \hat{\tilde{q}}_{\ell}\ell \left(1 - \hat{\mathcal{F}}_s(p_{ij})\right)^{\ell-1}}{\hat{f}_s(p_{ij}) \sum_{\ell=1}^{J_s} \hat{\tilde{q}}_{\ell}\ell(\ell-1) \left(1 - \hat{\mathcal{F}}_s(p_{ij})\right)^{\ell-2}}, \tag{C.2}$$

As before, let $\{s^G(z)\}_{z=1}^{|\mathbf{x}^H|}$ be defined as the collection of partitions of $\{\mathbf{x}^G \times \mathbf{x}^H\}$ such that $\mathbf{x}^H = z$. We can now pool the generated pseudo-costs corresponding to each value of $\mathbf{x}^H$: $\{\hat{c}_{ij}(t)\}_{t \in s^G(z)}$ and proceed to estimate $\mathcal{H}(\cdot|\mathbf{x}^H)$ and $h(\cdot|\mathbf{x}^H)$. As with the price density, we estimate the density using boundary kernels to reduce the bias.

C.2 Covariate selection for the structural model

Borrower and loan characteristics used in our structural model are displayed in Table C.1. Since the model needs to be solved for each combination of $(\mathbf{x}^G, \mathbf{x}^H)$, we rely on discretizing continuous variables. A related issue is that for kernel methods to provide a reliable estimate of the pdf of observed prices we need possibly many data points in each of the bins. Therefore, out of the initial 27,648 bins we used only those with 50 or more observations. This leaves us with 3,697 combinations (86.68% of the total number of mortgages in our main sample) representing the most popular products and borrower types. While by doing this we are no longer working with the entire universe of mortgages, the scope of loans we look at is still much broader than in previous literature using search models to study mortgage markets.[57]

Table C.1: Borrower and loan characteristics.

Variable	Discretization	# bins
	$\mathbf{x}^G$ (16 combinations)	
Age	<30, 30+	2
Income	Below median, Above median	2
FTB status	FTB, Non-FTB	2
Location	Urban, Rural	2
	$\mathbf{x}^H$ (1,728 combinations)	
LTV	≤70, 71-75, 76-80, 81-85, 86-90, 91-95	6
Deal length	2-, 3-, 5-year	3
Duration	<10, (10;15], (15;20], (20;25], (25;30], (30;35]	6
Loan value	4 quantiles	4
Flexible	Yes, No	2
Cashback	Yes, No	2
	Total: 27,648 bins	

Note: Table presents the selection of conditioning variables and associated bins used in the estimation of the structural model. The total number of bins is the cardinality of the Cartesian product of the elements of $\mathbf{x}^G$ and $\mathbf{x}^H$.

[57] For example, Allen, Clark, and Houde (2013) look exclusively at FTBs taking out loans with 25 year amortization and 5-year initial deal period.