Price Discrimination and Mortgage Choice*

Jamie Coen[†], Anil Kashyap[‡], and May Rostom[§]

November 2023

Abstract

We characterize the large number of mortgage offers for which people qualify in Britain. The proliferation of offers comes from lenders' attempts to serve heterogeneous consumers while respecting constraints on how much price discrimination is legally permitted. While few people pick the cheapest mortgage, failing to do so is not too costly. Some, however, do make very costly choices. These expensive picks are more common when the menu offered has more expensive options in it. These expensive menus tend to be presented to highly leveraged borrowers, who are often younger or first-time home-buyers.

Keywords: Price discrimination, consumer choice, mortgages.

JEL Codes: G21, G51, G53, D12.

---

*We thank João Cocco, Patrick Coen, Kris Gerardi, Robert Hall, John Krainer, Guido Menzio, David Miles, Daniel Paravisini, Morten Ravn, Francesc Rodriguez-Tous, Maximilian Schleritzko, Sergio Vicente, Erkki Vihriälä and Tomasz Wieladek for helpful conversations. We thank David Bradnum, Chiranjit Chakraborty, Alastair Firrell, and Simon Pittaway for in-depth conversations and help regarding the data. We thank Tom Cui for outstanding research assistance and seminar participants at the Bank of England, Bank of Lithuania, Chicago Booth, the Federal Reserve Bank of Dallas, the FDIC, Imperial College London, the London School of Economics, ASSA 2023, Emerging Scholars in Banking and Finance Conference 2023, Fed Day Ahead Conference 2023, Lapland Household Finance Summit 2023, NBER SI Household Finance 2022 and New Zealand Finance Meeting 2023 for comments. Kashyap thanks the Center for Research on Securities Prices and the Initiative on Global Markets at the University of Chicago Booth School of Business as well as the National Science Foundation for a grant administered by the National Bureau of Economic Research for research support. The views expressed here are those of the authors only and not necessarily of the institutions with which they are affiliated. All errors are our own.

[†]Imperial College London and Bank of England, w.coen@imperial.ac.uk.

[‡]University of Chicago, Booth School of Business, National Bureau of Economic Research, Centre for Economic Policy Research, anil.kashyap@chicagobooth.edu.

[§]Bank of England and Centre for Macroeconomics, May.Rostom@bankofengland.co.uk.

1 Introduction

We study how households in the United Kingdom choose their mortgage. Housing is the main asset most adults purchase in their lifetime. And almost everyone who buys a house does so by borrowing. Mortgages are complex financial products where the total cost of the mortgage depends on an initial interest rate, the fees charged at origination, and the prevailing interest rate once the introductory period lapses. We show that most borrowers face a bewildering set of choices for which they qualify, with considerable dispersion in the costs of different mortgages. Using a unique dataset covering the mortgages households select and the mortgages they *could have* chosen, we study how customers make mortgage choices, and how lenders set the menus they offer to customers.

We show that customers face menus with price dispersion, both within and across banks. Importantly, the choice sets show three recurrent patterns. First, all banks offer many mortgages with slightly different options. Second, most of the mortgages wind up with nearly equivalent total costs to the borrowers. Third, mixed into some menus are a group of very expensive choices, which if selected will net the bank some windfall payments.

We explain why these patterns point towards the banks recognizing consumers are heterogeneous and attempting to price discriminate subject to binding legal and informational constraints. Whilst banks can vary mortgage terms according to the types of mortgages for which customers are shopping, they cannot legally tailor menus to individual characteristics, and nor can they tell which potential customers are shopping at other lenders. So while the lenders would like to customize the terms for each borrower in various ways (for example by offering better options to someone who is comparison shopping), they cannot do that. Instead, the lenders can only offer menus with a myriad of options that aim to cater to a wide range of borrower preferences. If a lender has a sense of which customers are not able to identify cheap mortgages, all it can do to exploit that hunch is to include some expensive ones in the menu and hope they are selected. One factor limiting the percentage of expensive options is the knowledge that most borrowers will avoid such options; filling a menu with them will only drive those borrowers elsewhere. Thus, the necessity of offering the same menu to a heterogeneous set of borrowers limits the lender's ability to price discriminate.

We show that customers facing large price dispersion are typically those borrowing large amounts relative to both their income and the value of their house. These tend to be younger customers, and people who are buying a house for the first time. Lenders thus price discriminate, offering menus with greater price dispersion to customer groups who may be less able to identify and avoid expensive options, or have fewer options to go elsewhere.

There are several reasons why these patterns are not easily explained by appealing to differences in the costs (or risks) of serving different customers. First, menus cannot be tailored to individuals– mortgage options are posted and borrowers choose among the many for which they qualify. Second, the house serves as collateral and loans are with recourse, so the risk levels for all borrowers are pretty low and similar across individuals. Third, direct measures of who falls behind on mortgage payments (going into arrears) do not match the variation in the menus we document, so ex-post measures of risk do not track the menu variation. Most importantly, the big differences that drive our findings come from variation in the *range* of terms of the mortgages that different customers pick from, rather than differences in the average quality of the mortgages in the menu across customers.

Instead, it appears banks want to offer at least some cheap mortgage options to entice sophisticated customers who might be comparison shopping at multiple banks. To attract these customers they need to have some attractively priced options available. To the extent they do not know exactly which features are most important to these customers, having many close substitutes in the menu makes sense. At the same time, the banks also want to offer expensive mortgages in case someone is careless or unable to choose well. Given they can't tell these types apart and cannot tailor the mortgage menu based on personal characteristics, they offer a menu with price dispersion. For most groups of customers, the menu does not include many expensive choices. This would be expected if the lenders worry that some customers will not fully sample all the options and might opt to borrow elsewhere if they see expensive options. This kind of competitive pressure protects most customer groups. In some cases, banks suspect that certain customer groups are less able to shop around and thus present them with a menu with many expensive options. For customers in this group that do decide to proceed, they are more likely to select an expensive mortgage.

The results highlight the importance in recognizing both the heterogeneity in people's ability to select the cheapest mortgage option and the limits banks face in taking advantage of this heterogeneity. As we describe next, a growing literature proposes structural models of mortgage shopping by borrowers and lending competition by banks. We will explain why as currently constituted those models will not match our facts. Instead, current models have focused on different types of heterogeneity that appear to be less relevant than the aspects that we highlight.

1.1 Related literature

Our paper contributes to a literature studying how customers choose complex financial products, and how this affects firms' supply of these products. One strand of this literature studies how customers choose between mortgages depending on their various price characteristics. Two of these are particularly relevant for our study. Liu (2019) documents the heterogeneity in fees that we also confirm. She emphasizes that slight differences in fees allow lenders to boost profits by offering customers who are less cost-sensitive products with higher fees. While that is true, we emphasize two different aspects of the data. First, that fees are a small cost of the overall payments borrowers make, so that expensive choices cannot be purely driven by high-fee choices. Second, initial rates and fees are negatively related, such that expensive choices are more likely to come with *low* fees. For modeling household mortgage choices (which is the focus of our analysis), these two considerations are quantitatively more relevant than the variation emphasized by Liu (2019).

Guiso et al. (2022) provide evidence that lenders steer borrowers to particular mortgage products that they would like to keep on their books. This leads some borrowers who ask the lender for mortgage advice to wind up with more expensive mortgages. In the UK, lenders (and other mortgage advisers) have a fiduciary duty to the borrowers so this is illegal. Hence, these considerations do not apply in the UK market, and we identify a different mechanism that can lead to expensive choices.

Iscenko (2020) observes that many UK customers pick mortgages that are higher in all price dimensions compared to other mortgages with the same non-price characteristics. We build on this analysis in our empirical work.[1] In a related paper, Woodward and Hall (2012)

[1] A number of other papers study these issues in various product markets, including financial products (Duca and Kumar, 2014; Foà et al., 2019; Ru and Schoar, 2016) and non-financial products (Scott Morton et al., 2003).

find that customers would make large savings if they shopped around more, in particular whether they consulted more brokers. The possibility of comparison shopping features importantly our narrative.

A second branch of the literature seeks to rationalize observed choices and mortgage offerings by structurally estimating models of the mortgage market. Allen et al. (2019) and Allen and Li (2020) study the Canadian mortgage market which they argue is characterized by extensive bargaining between borrowers and local loan officers. This kind of bargaining is absent in the UK (and many other markets), so the way that they model the market does not apply in our setting. Benetton et al. (2021) build a model of mortgage choices made by heterogeneous consumers. The key source of heterogeneity is assumed to be differences in consumers' preferences over the different components of the mortgage contract. We adopt the much more conventional economic assumption that people care about the total cost of the product– not its various parts. Also, certain patterns in the data (such as the kinds of choices identified by Iscenko (2020) and confirmed and extended by us) cannot be rationalized by differences in preferences over rates and fees.

Other papers investigate the role of brokers in the UK. Myśliwski and Rostom (2022) show that customers with high search costs usually shop with a broker, and that customers with different demographic characteristics place different values on the information brokers provide. Robles-Garcia (2019) studies the trade-off between brokers facilitating competition in the mortgage market and facing incentives to distort households' decisions to profit from commissions. Because brokers have a fiduciary duty to their customers their presence can't explain the patterns we document. So while there are many interesting questions related to brokers, they are separate from the ones that motivate us.

Overall, none of the facts that we present are inconsistent with prior findings, but many are new. Our characterization of the data is important because the extant structural models do not fully explain the patterns we document. In particular, current models do not study the impact of menu heterogeneity on price dispersion. Put differently, they do not characterize lenders as posting menus to serve customer groups of varying sophistication, subject to the constraint that lenders must post the same set of prices to all customers within each group. Below, we reference the work by Menzio and Trachter (2018) which outlines a theoretical model in this spirit. Structural models should move in this direction to fully describe the patterns we see in the data.

A purely empirical paper using US data that is similar to ours is Bhutta et al. (2021). They study shopping patterns in the market for 30-year fixed rate mortgages in 20 US cities from 2016 to 2019 using data from a lending platform. As in our paper, they find considerable dispersion in the prices of mortgages that co-exist in the market; people with low credit scores, high loan-to-value and debt-to-income ratios, or first-time buyers select more expensive mortgages than others. They use novel survey data to show customers who shopped intensively and knew relatively more about the mortgage market got better rates.

Our paper primarily focuses on a different question from theirs. Using their data it is not possible to construct the menu available to a borrower at the lender from which they borrowed. This means they cannot disentangle the separate role that the menu offered to the customer plays vis-à-vis the customer's ability to pick from the menu - a key focus of our paper. By controlling for borrower characteristics and offered price dispersion, we find that the bulk of the differences in mortgage outcomes across borrower groups are driven by the menu they were offered, rather than the choice they made. Further, our distinction between expensive choices within and across banks means we can distinguish between the customer's ability to pick well from a menu from whether or not they borrowed at a bank with mostly cheap options. We find that both matter, but avoiding borrowing from an expensive bank is more important.[2]

We also contribute to the large literature documenting and studying price dispersion. A vast literature, starting with Stigler (1961) rationalizes price dispersion through search frictions. More recently, Kaplan and Menzio (2015) document widespread price dispersion both within and across firms. Menzio and Trachter (2018) rationalize this kind of price dispersion in a model where customers vary in their ability to shop at different times and across multiple firms. Because some customers may walk away if they find a bad option, the presence of these customers disciplines the sellers.

[2] Importantly, we scale our estimated cost differences by income and focus on cases where this scaled cost difference is large. This allows us to focus on choices that can reasonably be described as expensive. Put differently, for people with very different incomes, we do not suppose the same nominal monthly cost difference of a mortgage, e.g. £50 per month, is equivalent.

As has been shown for many markets in different countries, we document price dispersion both within and across mortgage lenders. We show that lenders vary the extent of this price dispersion across product types, so customers with different characteristics wind up facing different menus. As predicted by Menzio and Trachter (2018), these menus are set so lenders can attract a range of customers to their firm while making it easy for unsophisticated customers to choose expensive products. Importantly, competition appears to be a strong enough force that for most customer types, the menus have relatively few expensive options.

In the following section we describe our data and the relevant institutional features of the UK mortgage market. In Section 3, we explain how we characterize customers' choice sets and rank the choices by cost. We then study how well customers choose, and which customers make particularly expensive choices. We show that this is driven by both the quality of the customer's choice and the quality of the menu they face. In Section 4, we study what determines the menus customers face, and in Section 5 we study which customers pick expensive mortgages. In Section 6, we discuss our results and interpret them in terms of price discrimination.

2 Data and institutional setting

2.1 Data

Our sample period is 2009 to 2014. Our main data source is the Product Sales Database (PSD), a loan-level administrative dataset capturing all newly issued mortgages in the UK. The data contain information recorded by financial institutions at the time of mortgage take-out. This includes information on the borrower characteristics, such as income and age; information on the property, such as postcode and house price; and loan details, such as the amount borrowed, initial rate, and the mortgage term.

The PSD, however, misses information on product fees and the standard variable rate (SVR), the rate the mortgage resets to when the fixation period expires. To get this information, we merge the PSD with a secondary data source, Moneyfacts, which records the set of mortgages on offer in the UK at any given time. Moneyfacts at this time had a strong commercial interest in accurately identifying which mortgages were actually available, because their revenue came from selling this information to lenders and brokers who wanted to compare offerings across the market. This enables us to construct the set of mortgages on offer to customers when they shopped. We also use it to merge in the minimum and maximum loan amount sizes for each product.[3] We match the two datasets together using a matching algorithm that uses the name of the bank, the product type, the initial rate, the length of initial period and whether the purchase date fell in the time period the mortgage product was on offer in the market.

In all of our analysis, we focus on loans granted by the six largest lenders for which reliable information is available. Appendix A1 describes how the dataset is formed and summarizes key variables. We restrict the analysis to loans of no more than £1 million, with loan-to-value ratios (LTV) between 65 and 95%.[4] Our final dataset comprises just under 900,000 mortgages between 2009 and 2014.

2.2 Institutional setting

Most mortgages in the UK amortize over a period of 25 years in our sample period. In their loan offerings, lenders advertise initial promotional rates that would apply for loans up to a given amount, and conditional on the amount borrowed relative to the value of a home (LTV). For example, a bank might be willing to lend up to £1 million to any borrower who makes a down-payment of at least 20 percent (so that the LTV would be no more than 80%) provided the borrower pays a fee of £995. Borrowers typically have the option to pay this fee up front, or add it to the mortgage. A different rate might apply for a different, maximum loan size or LTV. The combinations of different fees, loan amounts, rates and LTV limits means that most people qualify for many different mortgages, and these mortgages have different required mortgage payments.

Table 1 shows an example of a typical menu a borrower might face at two banks. Several

[3] Several papers have used the PSD data for research. These include Benetton (2021), Benetton et al. (2020), Cloyne et al. (2019), Robles-Garcia (2019), Iscenko (2020) and Bracke and Tenreyro (2021). More recent vintages of the PSD have included many of these additional variables, but as our Moneyfacts data runs until 2014, our analysis stops there.

[4] The £1 million limit rules out only a small minority of borrowers while making sure that most banks have a mortgage on offer for each loan category in most time periods. In many cases, low LTV loans are associated with only small amounts being borrowed. This doesn't yield many expensive choices, but more generally menu offerings and incentives to shop around may be quite different when the stakes are lower.

key features of the data are evident. In particular, lower fees and higher maximum loan amounts tend to be associated with higher initial rates. Nonetheless, there are some low fee mortgages that have low initial rates, and some high fee mortgages that have high initial rates. The range of initial rates and fees differs noticeably between the two banks and at each of them the borrower has at least 15 mortgages (at the indicated LTV) that are available.

After having selected a mortgage, customers pay the initial promotional rate for a set period, after which the interest rate changes. Table A5 in Appendix A1 shows the distribution of the length of the initial period across our sample. 58% percent are fixed for two years and another 17% are fixed for three. The remaining loans are either fixed for five years or are a floating rate.[5]

Once the initial period expires for fixed rate mortgages (FRMs), the borrower moves on to the firm's Standard Variable Rate, or reset rate, which fluctuates depending on the prevailing Bank Rate set by the Bank of England, macroeconomic conditions, and banks' own idiosyncrasies, such as their funding costs.[6] Banks have discretion to change the reset rate at any time but, at the time the loan is granted, all that the borrower is told is the current value of the SVR. In other words, she will not know what the SVR will be when the initial period expires. The payment schedule the borrower would receive presumes that the payments will revert to the currently posted SVR.

Some borrowers may plan to refinance the loan once the promotional period expires. Whether that will be possible depends on a number of factors, including the overall interest rate environment and whether the value of the home or the borrower's income has changed. So the ability to refinance and avoid facing the reset rate is uncertain.

This is just one example of a more general choice problem issue that can arise because the borrower will know more about her own circumstances than we do. Below we explain the various different analyses we undertake to show our finding cannot be explained by such private information.

[5]In more recent years as interest rates have dropped, the share of five year fixed rate mortgages has grown significantly.

[6]Typically the SVR is common across all a bank's loans. However, as we discuss below in some cases our banks have two subsidiaries with different SVRs.

In practice, many borrowers do wind up having their rate switch to the reset rate. Reset rates vary across lenders and over time. Figure 1 shows the reset rate during our sample period for the 6 lenders that we analyze. To facilitate a comparison, the rates are shown as deviations from the average level (across the six banks) in each month. There is considerable dispersion at each point in time across the lenders, but their relative positions are quite stable. So although a borrower would not know the rate to which her mortgage will reset when she enters the contract, if she had done some investigation she could know whether her lender's reset rate is likely to be relatively low or high.

Summary statistics for our data are shown in Table 2. Fees are typically set in terms of a fixed number of pounds, rather than as a percentage of the amount borrowed. The level of fees overall are low compared to the US. The median borrower pays about £760, and 17% pay no fees at all. Less than 5% of the sample pays more than £1,100. Given the small size of these fees, in the UK it is hard to make a really expensive choice just because of fees. The median borrower in the sample is taking out a mortgage of £136,000 and making a down payment of 20%. The median income is £37,000 after tax, meaning that for the median person we will only label mortgages as expensive if they cost at least £925 ($0.025 \times \$37,000$) per year more than the reference choice we describe below. [7] First-time buyers constitute 40% of the sample. The sample is mostly made up of younger borrowers, no doubt in part because older ones would be re-financing mortgages that have lower LTVs than we consider.

Banks in the UK cannot tailor the choice set facing individual borrowers. Legal restrictions prohibit lenders from customizing loans to individual borrowers. If two borrowers qualify for a mortgage at a given lender, with a mortgage of the same size and a home of the same value, they would generally have the same set of mortgage options from which to choose.

Some borrowers may use independent mortgage brokers to secure a mortgage. Brokers have networks of lenders with whom they have a relationship (Robles-Garcia, 2019). Brokers are bound by a fiduciary responsibility to their customers to present all suitable options within their network of lenders for which they are eligible, and legally cannot steer borrowers

[7] We subtract income tax from gross income based on the UK's tiered income tax system, described at https://www.gov.uk/income-tax-rates. We follow the industry practice when calculating loan-to-income ratios and base this variable on gross income.

towards a particular lender or product. We would expect that the largest lenders - on whom we focus here - are present in most brokers' networks. Brokers typically receive their payment from lenders in the form of procuration fees, and from borrowers as broker fees. Procuration fees are similar across all the big lenders, and do not vary by product type. Broker fees are normally a fixed sum, regardless of the loan size, and are usually small.[8] We cannot identify who in our sample uses brokers, but during our sample period it would be conservative to assume half the borrowers used a broker to find a mortgage. For all these reasons, brokers are not going to lead their customers to expensive loans and we explain how their presence might influence the results in Section 6.

Alternatively, borrowers may consult loan officers at the bank, but these employees are also bound by conduct guidelines. The Financial Conduct Authority sets these guidelines which are very similar to how they also regulate brokers. They cannot be paid based on whether someone picks an expensive mortgage, nor do they get fees or commissions for the loans they initiate. Rather, many are salaried employees, and if they do receive bonuses these would be tied to other performance considerations such as effectiveness of work, quality of key judgements, service levels, and meeting compliance objectives. They may be rewarded in part for loan volume, however to preserve incentives for maintaining standards and not taking excessive risk, this component of a bonus is likely to be small. So if the two borrowers mentioned above wind up with different mortgages, this will be due to choices they made rather than because of a decision directly controlled by the lender (or broker).

3 The choice problem

To characterize the choice problem we need to define the choice set customers face and establish a metric by which to compare their options.

To define a customer's choice set we first identify all mortgages that were on offer when they were actively shopping.[9] We then restrict this set to include only mortgages with the

[8]See https://www.fca.org.uk/publication/policy/ps20-01.pdf for expectations set out by the FCA regarding broker conduct. See https://www.legalandgeneral.com/adviser/mortgage-club/lenders/procuration-fees/ for recent information on current procuration fees, although these are subject to regular updates.

[9]In the UK, it is rare to put in an offer on a house before a mortgage has been secured, and for the vast majority, a mortgage is usually agreed 4 months before the closing date. In the data, however, we observe

same initial period as the mortgage they chose, available for the amount they borrowed and with the lowest LTV band for which they qualify. For example if a customer took out a FRM worth £160k with a 2-year initial period on a house worth £200k, we would define their choice set as all 2-year FRMs on offer for loan sizes greater than or equal to £160k, at a maximum allowable LTV of 80%, at the time they were shopping.

We consider two different potential choice sets: within banks and across banks. The within-bank menu includes only the mortgages a customer qualified for at the bank that granted the mortgage they selected. The across-bank menu considers the mortgages that the customer qualified for at all banks. Each menu is informative about different questions. The within-bank menu allows us to study individual banks' price discrimination, because that bank can control the offerings that its customers see, and to explore how well borrowers pick from a set of choices that were definitely available.

Having characterized the choice set, we need a metric by which to compare mortgages. We use three different metrics to rank choices given a customer's LTV. In our 'baseline' approach, we take the four key elements of the mortgage contract - the fees, initial promotional period, initial interest rate, and the reset rate - and compute the present value of the payments for the borrower over the first seven years of the mortgage. Seven years is about the half-life of the stock of mortgages outstanding for people buying a house with a mortgage.[10] Calculating the payments only over the first seven years has two benefits. First, as a practical matter most borrowers do refinance at some point, so our baseline approach takes that into account. Second, if we did the calculation over the full length of the mortgage, the level of the reset rate would dominate the size of the payments since it would be the operative rate for the vast majority of the payment.

In the calculation, we assume the reset rate remains constant, which is the assumption embedded in the initial monthly payment the borrower will be given upon signing the contract.[11] We use the seven year LIBOR rate to discount the payments. Hence the formula

---

the reset rate on the closing day. To ensure all rates are from the time the customer made the mortgage decision, we lag the reset rate by 4 months.

[10] Gianinazzi (2019) finds that 61% of borrowers in the UK were on their bank's SVR. Given the average initial period of the mortgages in our sample is 2.6 years (Table A5) and most mortgages amortise over 25 years, this is consistent with customers refinancing once every 8 to 9 years.

[11] Because of the option to refinance and the variability of the reset rate there is risk associated with the

we use to compute the net present value is:

$$NPV = fee + \sum_{t=1}^{T_F} \frac{IP}{(1+i)^t} + \sum_{t=T_F+1}^{84} \frac{RP}{(1+i)^t} \tag{1}$$

where fee is the initial fee, T_F is the length in months of the initial promotional period over which the initial payment is fixed, IP is the initial monthly payment during the initial period, RP is the monthly payment after the initial period (implied by the reset rate), and i is the seven year LIBOR rate, scaled to its monthly equivalent. For the 2 year ARMs, we assume the initial payments are identical for the first two years because that is the way the payment would be computed for a borrower who asks to see a payment schedule.[12]

We now describe two other ranking algorithms that are included to demonstrate that the seven year cutoff is not responsible for our main findings. Our first alternative ranking assumes people automatically refinance once the initial period is over. In this 'immediate refinancing' approach, the reset rate will no longer matter for any of the calculations. This extreme assumption means we will never assume someone has made an expensive choice because they cannot refinance. Rather, we suppose they can always avoid that problem. Even though this assumption is counter to what we know about the frequency of refinancing, it is nontheless helpful in seeing whether our description of the choice sets and the choices made are particularly sensitive to the refinancing assumption.

Our last approach to evaluating choices builds on Iscenko (2020) by comparing the chosen mortgage to an alternative that dominates it in at least one dimension. For instance, if instead of the mortgage someone picked there existed a competing one with an identical reset rate, fee, and promotion period, but a lower initial rate, we would conclude the selected mortgage was dominated by at least one better option– in other words, it would be a dom-

choices households make. However, those rates are extremely persistent: the first three autocorrelations of the average reset rates are all greater than 0.95. We calibrate a Markov process that matches these correlations closely and then simulate the distribution of rates which a borrower would face at different horizons. The differences between the simple expected value of the rates and the rates that are adjusted for risk (to deliver a certainty equivalent mortgage payment) are very small. Hence, we abstract from that uncertainty.

[12]As part of the mortgage contract, banks are required to tell the borrower the mortgage payments for the duration of their mortgage term.

inated choice. We further refine this definition by saying the choice is strongly dominated if the alternative mortgage would lead to savings of at least 2.5% of the borrower's income over the first seven years of the loan. Our refinement is intended to separate loans that might have trivial cost consequences from those that would make a material difference in the mortgage payments.

This 'strongly dominated' ranking removes the possibility a borrower had private information that could justify a choice. This definition has the disadvantage of overlooking some potentially very expensive choices that are not strongly dominated. For instance, if the initial interest rate is relatively high but that mortgage has no fees, then this mortgage by construction cannot be dominated by any mortgages that include fees, even if they are much cheaper. In contrast, our baseline approach and immediate refinancing alternative suppose the borrower cares only about the total cost of the mortgage rather than the various components that contribute to the overall cost. Regardless, our main findings are present using each of the these ranking criteria.

In cushioning our analysis by (i) choosing a mortgage that isn't the cheapest in a customer's menu to compare their choice to; (ii) characterizing an expensive choice as costing at least 2.5% of net income; (iii) having three different ranking algorithms to characterize the choice set, we are confident that the patterns we observe in the data are not driven by how we compute the menu nor how expensive choices are defined. Rather, they are robust features of the data with important ramifications for future modelling efforts.

3.1 Customer choices

Table 3 summarizes the size of the menus and describes the cost of the choices people make. The median customer faces a menu with over 15 options at a single bank, and more than 70 across banks. Customers do not pick particularly well - only 5% pick the cheapest mortgage in their choice set at their bank and 51% pick worse than the median choice.

The quality of a customer's choice depends on how much money their choice causes them to save or lose - picking a poorly ranked mortgage isn't important if it's not much more expensive than the cheapest mortgage. To understand the financial consequences of customers' decisions, in our baseline ranking we define a 'reference mortgage' against which we evaluate customers' choices. We set this reference to be the 15th percentile option in their menu, where options are ordered from cheapest to most expensive. We then compute the amount of money a customer saves or leaves on the table relative to this reference as a percentage of the customer's monthly income after tax.

This reference point reflects a couple of considerations. Given the fluctuating size of the choice set across banks and over time, we would like a reference point that scales with the choice set. The 25th and 75th percentile of the within-bank choice set size is 11 and 23 respectively, so this cut off means that we are using the 2nd or 3rd cheapest mortgage for most people rather than the absolutely cheapest mortgage.

Figure 2 shows the distribution of customers' savings relative to the reference choice within banks (left panel) and across banks (right panel). Most customers' mortgage choices do not save or cost them a large amount of money. At their chosen bank, for example, 85% of customers' possible choices neither save them more than 0.5% of their monthly income nor cost them more than 1% (shown in the shaded area in the figure). Looking across banks, 66% of the choices are within that same range.

It's therefore clear that while customers don't pick well in ordinal terms, for most borrowers, the cost consequences of their decisions are relatively minor. We attribute this to the role of competition, which disciplines the banks and protects customers. The fact that customers can shop elsewhere prevents banks from offering primarily expensive products: if a customer picks an expensive option to use in comparisons with other lenders, the bank will likely lose the customer. As a result, the menu customers are given is often full of products with similar prices, meaning they won't lose a large amount of money even if they fail to pick particularly well. Offering a variety of choices increases the chance any particular borrower can find a product that suits her well and can make borrowing at that bank appealing.

If we characterize the choice sets and choices made using the other two ranking criteria, similar patterns are present. Figure A1 in the appendix replicates Figure 2 under the immediate refinancing assumption. The within-bank results barely change, whilst across banks the share of customers that neither lose nor gain a large amount of money from their choice is even larger. The same must also be true if we consider only mortgages that are strongly dominated: if few customers leave a large amount of money on the table, then even fewer must leave a large amount of money on the table *and* make choices that are dominated (by the reference mortgage) on all price dimensions.

3.2 Expensive choices

Despite the fact that most mortgages yield similar mortgage payments, some customers do make expensive choices (Figure 2). Within banks, 2.3% of borrowers make choices that cost them more than 2.5% of their income, and across banks this figure is 6.7%.[13] For the average borrower, this amounts to around £88 per month - a meaningful amount of money given the net income of the typical borrower.

In what follows, we study the subset of customers that makes these expensive choices. In our baseline rankings, we define an *expensive choice* as a choice that costs a borrower more than 2.5% of their monthly household take home pay relative to the reference mortgage. Figure 3a plots the average probability of an expensive choice through time, both within and across banks. There is significant time series variation, with expensive choices more prevalent early and in the middle of the sample.[14] Figure 3b shows how the monthly percentages of strongly dominated choices within and across banks vary over time, where again there is significant time variation, with the strongly dominated picks being least common at the end of our sample.

A customer can make an expensive choice for two reasons: choosing badly from a given menu, or when facing a menu with many poor choices making a more typical choice that is expensive. To isolate the role of the menu in driving expensive choices we define a new variable, *bad tail*, which is the percent of mortgages on the customer's menu that - if chosen - would represent an expensive choice for that customer. The variable bad tail is a function of price dispersion. If prices are identical, bad tail is, by definition, 0. If there is a large amount of price dispersion, then some of the choices are likely to be expensive, and bad tail is large and positive.

[13]Under the immediate refinancing assumption, these figures are 1.9% and 3.5%. We provide more details on the strongly dominated choices below.

[14]From Figure 1, we know that the dispersion of reset rates was larger at the start of the sample and one way to make an expensive choice (across banks) is to borrow from a bank with a high reset rate when there are other banks with noticeably lower reset rates.

Figure 4 plots the distribution of bad tail through time. Many borrowers face menus with almost no bad choices. A small subset of customers face menus that are filled with many bad choices. As with expensive choices, there is significant variation through time, with the choice sets being worse at the start of our sample.[15]

Our most remarkable finding about the time series patterns, however, is the strong influence the menu quality plays in driving expensive choices. In Figure 5, we show a strong and positive correlation between the monthly percentage of expensive choices and the average size of the bad tail in that month, both within and across banks. As the menu quality deteriorates, the percentage of expensive choices rises. Simply put, customers make expensive choices when banks make it easy for them to do so. Figure 6 shows the equivalent plot for the strongly dominated criterion. The same basic patterns documented in these figures are present if we rank choices with the immediate refinancing assumption. The analogous pictures are shown in the appendix in Figures A2 to A4. In particular, expensive choices remain rare, but they are tightly tied to the quality of the menu that people face. Specifically, the left panel shows the within-bank (monthly) scatter plot of the probability of strongly dominated choice against the size of the *strongly dominated tail*, defined as the percentage of choices in the menu that are dominated by the cheapest option, and cost 2.5% of income more than it. As with our baseline ranking there is a strong positive correlation between the two. The across-bank results, shown in the right panel, similarly mimic the patterns using the baseline ranking.

Overall, we conclude that our use of the baseline ranking algorithm is not responsible for our conclusions about the importance of menus. In particular, the percentage of strongly dominated choices is highly variable over time and when the tail is larger, more strongly dominated choices are selected.

These conclusion lead to two obvious questions. First, what leads banks to price discriminate and give some customers menus with more expensive options? In other words, what determines the menu structure and who gets the ones with more expensive options? Second, for any given menu, what explains why some customers pick an expensive mortgage?

[15]In the interest of brevity, we omit the time series pictures showing how the strongly dominated tail varies over time. They exhibit the same basic patterns as the bad tail.

4 Setting the mortgage menu

In this section we assess what determines the choice set a customer faces, and what drives the heterogeneity in choice sets across customers.

Legal constraints limit lenders' abilities to fine-tune the menu they offer to different customers. Although it's illegal to vary menu prices by certain individual characteristics (such as gender or race), they can price discriminate by altering the characteristics of the mortgage contract presented to all borrowers. We focus on two key dimensions of the mortgage: loan-to-value and loan-to-income (LTI) ratios. The LTV is directly relevant to the loan contract, and as we noted earlier, for a given a loan amount, the cost of borrowing rises with the LTV, reflecting increased risk of mortgages with higher leverage. The LTI is indirectly relevant because banks' feed this into their internal risk models to judge a borrower's credit risk and ultimately their ability to make their mortgage payments.[16]

Figure 7 plots the average size of bad tail and the strongly dominated tail according to customers' LTVs and LTIs at the banks they shopped at. Here, we define high LTV to be 85% and above, while high LTI is a ratio of 4 or more.

Customers with both low LTV and LTI ratios make up half the sample, and receive good menus with very few bad choices using either metric of menu quality (Figure 7). This is to be expected. These customers have low credit risk and can probably qualify for a mortgage at many lenders. Banks are unlikely to have any market power with respect to these customers. Where either the loan-to-value or the loan-to-income ratio becomes high, the size of the bad tail or strongly dominated tail doubles.[17] For customers who are borrowing a lot relative to both their income and their house value, over 6% of choices on the menu would represent an expensive choice and about 13% of the menu would be strongly dominated.[18] This also is somewhat expected. These borrowers would not necessarily sail through a mortgage approval

[16]It is worth noting that in 2014, the Bank of England's Financial Policy Committee instituted a set of rules for mortgage lending in the UK that prevented banks from offering more than 15% of new mortgages in a given quarter to borrowers with LTIs above 4.5. See Kashyap (2020) and Peydró et al. (2020).

[17]See Figure A5 for the corresponding figures under the immediate refinancing ranking, where the patterns are very similar to the baseline.

[18]Note the reference mortgage used for the two metrics differ. Whereas the strongly dominated set compares the menu to the cheapest mortgage, the bad tail compares the menu to the cost of the 15$^{\text{th}}$ percentile of the menu.

process at any other bank.

This pattern of menu quality across customers is consistent with banks offering menus to customers who are potentially constrained, and might not have many options elsewhere. Customers with high LTV and LTI ratios cannot borrow more against their income, and didn't save enough to make a large deposit. They may not qualify for many mortgages at other banks. Thus these customers' outside options are probably inferior compared to customers who are borrowing less against their home and/or their income.

We summarize customers' outside options according to their LTV and LTI in Figure 8. We define a new variable - *outside tail* - which captures the quality of a customer's outside options. We take all mortgages a customer could have chosen at all banks *except* the one from which they borrowed, and compute the percentage of these mortgages that would represent an expensive choice relative to the reference mortgage at their chosen bank. Figure 8 plots the average of outside tail by LTV and LTI bucket, along with the average of bad tail at their chosen bank. The same pattern of deteriorating choices at the customer's own bank is true for the choice sets at other banks. This is consistent with banks all making similar conjectures about which types of customers would be able to qualify for loans at competing banks and pricing accordingly.

For three reasons, this menu variation likely reflects price discrimination and not variation in risk across customers. First, all mortgages in the UK are made with recourse to the borrower, making default extremely rare. Even in 2009, when house prices fell by 20% and unemployment rose to 8%, banks suffered few losses on mortgage loans, and the default rate rose by only 1 percentage point, to 1.5%.[19] Second, the probability of going into arrears is very similar across higher and lower LTVs and LTIs.[20] This suggests default cannot be driving the kind of menu variation that we observe in Figure 7. Third, our measure of menu quality is relative, and thus relates to price *dispersion* - not *average* prices. While

[19]Default is defined as mortgage payments that have fallen into arrears of more than 6 months. See Aron and Muellbauer (2016) for the default statistics, and Bank of England (2010) for house price and unemployment statistics.

[20]We directly calculate the percentage of mortgages going into arrears by LTI and LTV for a large sample of borrowers in 2015. We find that across the four categories of borrowers in Figure 7 the percentages are all between 1.9 and 3.9 percent and whilst the levels are higher for high-LTV loans they are no higher for high-LTI loans. So the arrears pattern does not follow the monotonic pattern for the bad tail size shown in Figure 7. We use PSD data from 2015 as this is the first available year for which arrears data is collected.

variation in risk may provide some rationale for lenders to increase average prices for high-LTV customers, risk alone cannot explain why price dispersion is also higher for high-LTV customers. In contrast, a price discrimination motive does predict that price dispersion should vary in the way that it does.

Setting the menus according to loan characteristics results in the menus varying with demographic characteristics. In Table 4, we show who takes out high loan-to-value and loan-to-income mortgages. We run probit regressions of a customer taking out one of these mortgages based on the three borrower characteristics that we observe in our data: whether the customer is young or old, a first-time buyer or not, and rich or poor. Table 4 reports the marginal effects that these variables have on mortgage choice.[21]

Young people and first-time buyers are significantly more likely to take out mortgages where they're borrowing a large amount relative to their income and house value. These differences are economically large. A customer under 30 years old is around 5 percentage points more likely to take out a high-LTV and high-LTI mortgage than a customer over 45 years old. This effect is large relative to the population average probability of around 5%. The effect is similarly large for first-time buyers. While low income customers are more likely to borrow a large amount relative to their income, they're less likely to borrow a large amount relative to their house value.

Young people and first-time buyers thus opt for mortgages that come with bad menus. Given the evidence in the literature that young people tend to have lower financial knowledge (Lusardi and Mitchell, 2011) and make more financial mistakes than the middle aged (Agarwal et al., 2009), this would be consistent with banks offering customers who are less able to pick effectively mortgage menus where the consequences of a bad choice are greater.[22]

Table 5 shows how banks vary the menu across different customers. It summarizes the dispersion customers' face in all three price components of a mortgage, within and across banks. The main dimension of dispersion is the initial rate: for the median borrower the difference between the 85^{th} and 15^{th} percentiles of the initial rate distribution is 1 percentage point at the bank where they took out their mortgage, and 1.3 percentage points across banks.

[21] The table shows results for our within-bank sample. Results for the across-bank sample are equivalent.

[22] Lusardi and Mitchell (2011) and Agarwal et al. (2009) document an inverted U-shape in financial literacy and decision making, with performance increasing in age up to age 50 and declining thereafter.

Across banks customers also face significant variation in the reset rates they face.

Table 6 relates the size of the bad and strongly dominated tails that borrowers face to the dispersion in price. The key determinant of the size of the bad tail is the dispersion in initial rates across products. Combining the information in Tables 5 and 6, a customer whose initial rate dispersion is at the 75$^{\text{th}}$ percentile of the sample distribution has a bad tail at the bank from which she borrowed that is 4.8 percentage points larger than a customer at the 25$^{\text{th}}$ percentile. This difference is double the average size of bad tail in our sample. Taking Tables 5 and 6 together it is clear that banks predominantly rely on the initial rate to vary their menu offering.

In the case of the strongly dominated tail, at the selected bank an increase in fee dispersion reduces the size of the strongly dominated tail. This is essentially mechanical because if the menu includes any zero fee options, then the candidate pool of dominating mortgages is reduced to only including other zero fee mortgages. Across banks, fee dispersion is still associated with a larger strongly dominated tail of mortgages. More importantly, the rate and reset rate dispersion are again correlated (both within and across banks) with a worse menu. Quantitatively, the dispersion in initial rates is the most important factor in explaining the size of the strongly dominated tail. Moving from the 25$^{\text{th}}$ percentile to the 75$^{\text{th}}$ percentile of the (within bank) sample would raise the size of the strongly dominated tail by 6.6 percentage points, which is one and half times the size of the mean of that variable. So if we rank mortgages according to whether they are strongly dominated we reach the same conclusion about the importance of avoiding high initial rates.

5 Customer choices

Having established how banks tailor the menu they offer to different customers, we now ask how customers pick from a given menu. What leads to an expensive choice? Who makes expensive choices? Were these expensive choices driven by the menu the customer was offered, or some aspect of the choice they made?

5.1 Expensive choice mechanisms

Table 7 reports the marginal effects from probit regressions of making an expensive choice on the various loan features that a customer faces when choosing. For each customer, we calculate the distribution of the fees, initial rates and reset rates on the menu she faces. Picking a high price is defined as picking a product whose price is greater than the 85^{th} percentile on offer, and picking a low price is defined as picking better than the 15^{th} percentile. We control for the menu quality by including bad tail as a regressor.

As would be expected from the prior tables, the choice of initial interest rate is the key driver of making a poor choice: choosing a high rate increases the likelihood of an expensive choice by 5 percentage points relative to picking a low rate within a bank, and 15 percentage points across banks. These effects are more than double the probability of making an expensive choice in the population. Controlling for the other price dimensions, the choice of fee has close to zero impact on expensive choices, as fees are generally not large enough to materially impact mortgage cost. Across banks there is a quantitatively significant role for the reset rate - picking a high reset rate makes a customer 8 percentage points more likely to make an expensive choice than picking a low reset rate.[23]

Given the key to making an expensive choice is choosing a high initial rate, what do mortgages with high initial rates look like when they are selected? Table 8 answers this question for within-bank expensive choices. The first column regresses the likelihood of picking a product with a high initial rate on the other two price dimensions. Customers that pick products with low fees are significantly more likely to pick a product with a high initial rate. These effects are large: a customer who picks a low fee at their bank is over 20 percentage points more likely to pick a product with a high rate than one who picks a high fee, an effect which is roughly equal to the proportion of customers that pick high rates in our sample. The second column of Table 8 regresses the likelihood of an expensive choice on the customer's fee choice. Given low fee products tend to have high rates, picking a low fee

[23]An analogous set of results for strongly dominated mortgages is shown in the appendix in Table A9. Given how dominated choices are defined, picking a low fee, low initial rate or low reset rate, necessarily reduces the odds of making a dominated choice, yet the same patterns hold using that ranking criteria. Table A6 in the appendix shows the same regressions using the immediate refinancing criteria, where the patterns for fees and interest rates are very similar to the baseline results. The reset rate plays no role using the immediate refinancing critieria, as all borrowers by assumption refinance at the end of the initial period.

increases the chances of making an expensive choice.[24] Even though many of the expensive choices arise because of mortgages that have low fees and high initial rates, the coefficient on low fees controlling for bad tail is modest for two reasons. First, about 40% of all the mortgages (at the selected bank) that are not in the bad tail have low fees. So picking one of those mortgages does not cost the borrower too much. Second, there are also some expensive choices that are strictly dominated, and some of these have higher fees and a high initial rate. These two considerations bring down the coefficient, even though many of the within bank expensive choices combine low fees and high initial rates. Across banks some people also make an expensive choice by picking a mortgage with a very high reset rate.

Figure 9 sets out how the role of the fee choice in driving expensive choices depends on the initial period of a product. For each length of fixation period in our sample, we compute the percentage of customers that chose a product with a low fee conditional on whether the customer made an expensive choice or not.[25] Where the initial period is only 2 years, picking a low fee is not associated with making an expensive choice. For products with longer initial periods, however, picking a low fee leads to expensive choices. This is consistent with the evidence in Table 8, which shows that low-fee products come with high initial rates. The longer the customer must pay this initial rate, the more likely it is that picking a low-fee product proves costly enough that it leads to an expensive choice.

At most banks there is only one reset rate available. In some cases, however, a banking group may have several distributional channels which are branded differently. This typically happens when there is a merger and the acquiring bank allows the target bank to operate using its existing policies for some time. So in several cases, there is a transition period - usually between three and eighteen months - where a lender offers its own reset rate and a different reset rate for the legacy bank it has acquired.

Picking a low reset rate, where there are two or more available in that banking group, is generally associated with picking a high initial rate, though the magnitude of this effect is smaller than the effect of the fee choice. Cases of multiple reset rates are uncommon, so

[24] It makes no sense to replicate these tests for strongly dominated choices because picking low fees will lower the odds a choice will be dominated. The results for the immediate refinancing ranking are very similar to the baseline results and are omitted to save space.

[25] We group 2 year fixed and adjustable rate mortgages together.

cannot be the drivers of the main results in the paper. For instance, if we rerun the within-bank specification in Table 7 without controlling for multiple reset rates, the coefficients on the high and low fee are effectively identically to those we report in Table 7.

We now assess the drivers of expensive choices across banks. Table 9 shows the incidence of expensive choices within and across banks. Just about as many borrowers who pick poorly at their own bank do not make an expensive choice when the choice is defined across banks. This can only happen if some banks offer enough cheap options that picking badly at that bank is costly, but the chosen mortgage judged against the universe of offerings elsewhere is not that expensive. Likewise, there are some people (about 1 in 18) who do not pick badly relative to the choice set at their bank, but wind up with an expensive mortgage judged against what is available elsewhere. This pair of facts suggests that in any customer's choice set there are some relatively cheaper banks and others that are more expensive.

Tables 10 and 11 disaggregate the roles of picking the wrong bank and picking poorly at a given bank in driving expensive choices. To capture how well a customer chose their bank, we compute the cost of the average mortgage at the bank where they shopped minus the cost of the average mortgage at the cheapest bank they could have shopped at, scaled by their income. To capture who well they chose at the bank where they shopped, we compute the difference between the cost of the mortgage they chose and the cost of the 15$^{\text{th}}$ percentile mortgage at the bank where they shopped, scaled by income.

Table 10 summarizes these two variables. At most of the banks, for most customers even picking at the 90$^{\text{th}}$ percentile would not lead to an expensive choice. In contrast, when comparing different banks the dispersion in mortgage costs is greater.

Table 11 shows how expensive choices vary according to the quality of a customer's choice of bank and of their choice at their bank. Both play a quantitatively significant role in driving expensive choices, though the choice of bank plays the larger role. Combining the information in Table 10 and the 5$^{\text{th}}$ column of Table 11, a customer whose quality of choice of bank is at the 75$^{\text{th}}$ percentile of the distribution is 4 percentage points more likely to make an expensive choice than one at the 25$^{\text{th}}$ percentile. A customer whose within-bank choice is at the 75$^{\text{th}}$ percentile is 1.4 percentage points more likely to make an expensive choice than one at the 25$^{\text{th}}$ percentile. Both of these effects are significant relative to the average probability of making an expensive choice, 6.7%.

These results thus establish the ways in which customers choose expensive mortgages. Customers that fail to shop across banks, or even upon doing so pick an expensive bank, are significantly more likely to make an expensive choice. Conditional on the bank a customer chooses, the driver of expensive choices is the initial interest rate. Customers that focus on paying a low fee, rather than a low initial rate, are thus liable to make expensive choices.

5.2 Expensive choices by customer type

Having established how customers make expensive choices, we now ask who makes expensive choices? Table 12 reports the marginal effects from regressions of making an expensive and strongly dominated choice on customers' loan-to-value and loan-to income ratios, with and without controlling for the quality of the menu, within and across banks. The first four columns show the results for expensive choices. Customers borrowing large amounts relative to their house and/or income are significantly more likely to make expensive choices, both within and across banks. The effects are economically large: customers with high LTV and LTI are 9 percentage points more likely to make an expensive choice within-bank than a customer with low LTV and LTI, which is over four times the average probability of expensive choices in our sample. Within banks, this effect is almost wholly due to the quality of the menu - the differences disappear once we control for the menu quality.

Across banks, while most of the variation in expensive choices can be explained by the menu, customers with high LTIs are more likely to make an expensive choice even after controlling for the quality of the menu. The effects once we control for the menu remain significant, though relatively small, with a customer with high LTV and LTI two percentage points more likely to make an expensive choice than one with low LTV and LTI, relative to a sample average of 7%. This suggests that, given the same choice set as other customers, these customers are marginally worse at shopping across banks.

The final four columns of Table 12 shows the analogous regressions for strongly dominated choices. The combination of high LTI and LTV increases the probability of a strongly dominated choice by 20 percentage points at their chosen bank. Even controlling for menu quality, there is almost a 5 percentage point increase in chances of a strongly dominated choice; recall the mean rate of strongly dominated choices within-bank is 8 percentage points, so this is still a meaningful effect. The across-bank results are similarly large.

Table 13 reports the results of probit regressions of expensive and strongly dominated choices on customer demographics. The first four columns show the results for expensive choices. Young people and first-time buyers are significantly more likely to make expensive choices both within and across banks, though the economic magnitudes are modest. This is driven almost wholly by the menus they are given, with little variation in the likelihood of expensive choices across demographics once we control for the tail of the menu. The final four columns repeat these regressions with a dependent variable that identifies strongly dominated choices. Once again the menu quality is the dominant predictor of who makes strongly dominated choices.

6 Discussion

We summarize our findings as follows. Despite most people facing a bewildering number of choices, the cost implications of their decision are fairly small. There are, however, some people who do face menus that includes some very expensive options. The details of how we identify these expensive mortgages are not important, as all three of our measurement strategies point to four robust facts about the mortgage market.

First, menus with lots of pricey options are more common for borrowers with higher LTV and LTI ratios. Second, borrowers with higher LTVs and LTIs are typically younger, more likely to be first time buyers, and have lower incomes. Third, the expensive choices primarily arise from having a high interest rate during the initial promotional period. Moreover, those who pick mortgages with higher rates tend to pick products with low fees. Fourth, there are two ways by which people wind up with an expensive mortgage. A minority of the cases come from people who select a mortgage that is expensive despite have a menu that did not have an exceptional number of expensive choices. More often, these people simply faced a menu with many expensive products to start of with.

The existing literature fails to fully explain these facts. One can appeal to cognitive challenges in explaining why some people choose expensive options when cheaper ones are present. Yet, that approach can neither explain why the menus combine so many similar cost products with the expensive ones nor why the size of the bad tail varies in the way that we establish.

Here, instead, is a more promising way to model the market. Suppose the lender assumes there are two types of customers: sophisticated customers and randomizers. Randomizers walk into a bank, and pick a random choice from the menu. They don't shop at other banks, perhaps because they find it too costly to shop around, are unaware of alternatives, or don't qualify for mortgages with other lenders. Since they prefer having a mortgage to not having one, they take the mortgage that they randomly select. Sophisticated customers go to all banks, consider all options and pick the cheapest available.

How should a lender design their menu in this environment? It must balance two considerations: providing cheap options in order to entice sophisticated customers to shop at the bank, and to offer expensive options to profit from the randomizers. The menu on offer will feature price dispersion, with good options for sophisticated customers and bad options to profit from the randomizer. The higher the percentage of randomizers, the more they want to fill the mortgage menu with bad options.

This characterization is inspired by Menzio and Trachter (2018), who present a model in which customers vary in their ability to shop within and across firms. They find that the equilibrium price distribution will have price dispersion both within and across firms.

Our evidence is consistent with this framing of the problem. We find that customers who are borrowing a lot relative to their income and home value are given worse choice sets. These customers are more likely to resemble the randomizers than those who are not borrowing much relative to their home or income. They didn't save for a large deposit, which would have given them a cheaper mortgage. They can't afford a bigger house, and may not qualify for mortgages at many banks. They're disproportionately likely to be taking out a mortgage for the first time, and tend to be younger. As a result, these customers are the type that lenders would like to exploit, in order to profit from them either because of their lack of sophistication or because of their possible lack of choices. The market equilibrium will mean that these customers are more likely to wind up with expensive mortgages.

6.1 Alternative explanations

There are a number of other potential explanations for our evidence. We briefly explain why the most obvious ones are inconsistent with our findings.

The first concern is that the variation in menu quality across borrowers could reflect variation in risk or costs across borrowers, and not price discrimination. In Section 4 we explain in detail why such concerns cannot explain our findings.

A second concern is that the use of brokers might be driving the results. This could be true for two reasons. One possibility is that the menu could depend on who shops with a broker. Alternatively, making use of brokers could be correlated with some other factor that changes the probability of making an expensive choice. We dispense with each in turn.

As explained already, borrowers can use brokers' help in selecting a mortgage. But this does not affect the set of mortgages for which a borrower qualifies. Brokers might not have all 6 lenders in their network of lenders (Robles-Garcia, 2019), but all our key results hold within lenders, where brokers have a fiduciary duty to help the borrower find a good mortgage. As a consequence, our finding that high-LTV and high-LTI borrowers (and hence young borrowers and first-time buyers) are offered menus with higher price dispersion cannot be explained by the presence of brokers.

Brokers can, however, affect how well borrowers choose from an offered set. Brokers' fiduciary duty to borrowers, and reputation in a market with frequent re-mortgaging, mean that it is likely that brokers help them pick *well*, not poorly. This implies brokers reduce the likelihood of making an expensive choice.

In fact, as mentioned earlier, it might be reasonable to assume that anyone who uses a broker will never make a dominated or expensive choice. In that case, the number of people picking expensive products should be compared only to the set of borrowers who did not use a broker. So the percentage of these expensive choices would be roughly double the numbers we have reported. Whether or not this rough adjustment is correct, it seems unlikely that the expensive choices arise *because* some borrowers use brokers.

Third, one could worry that differences in loan approval standards across lenders drive our results. Maybe a borrower did not pick a cheaper mortgage not because they chose poorly, but instead because they could not get a mortgage at the other bank. We do not observe lender approval standards, so cannot rule this out of our across-bank analysis. However, the key patterns that we observe in price dispersion and borrower choices are all observed within-lender. Conditional on size of the loan and the house value, loan approvals do not typically vary for products within a lender. This, coupled with the fact that lenders do not typically vary the menu they offer across different borrowers, gives us confidence that variation in loan approval standards cannot explain our main results.

Another concern about our interpretation of the findings is that the menu variation that we emphasize could be vestige of the way we construct the choice set. In particular, we construct a borrower's choice set as the set of menus with the lowest LTV cap possible conditional on the loan they took out. In principle, a customer could choose a mortgage product with a higher LTV cap, but because mortgage rates rise with leverage, they rarely do this. For example, a borrower needing to borrow only 80% of the cost of the home who picks a mortgage available to borrowers with LTV of 85% would almost always face a higher initial interest rate with the higher LTV. So in our main analysis, we only consider the 80% LTVs. In the robustness exercises shown in Appendix A3, we demonstrate that our main results carry through when we consider the even larger set of available choices a qualified borrower could theoretically pick from. In our example, this would expand the choice offerings to include 85%, 90%, and 95% loan-to-value offerings.[26]

Finally, one might worry that the choice set is mismeasured in way that explains the expensive choices. For instance, if borrowers are constrained or do not qualify for a mortgage, then that might seem like a plausible alternative interpretation. Borrowing constraints, however, seem unlikely to explain the findings because we have shown that expensive choices arise primarily when the menu quality, based on the number of very costly mortgages available, is poor. The failure to pick a very cheap mortgage is not driving the results. Loosely speaking, even picking the median mortgage in a choice set would not result in an expensive mortgage. So this explanation is unlikely to explain our findings.

All in all, the robustness of the results across and within bank, implies that a particular

[26]91% of borrowers pick a mortgage with the lowest possible LTV cap for which they qualify at their chosen bank. See Tables A11 and A12 for the results with the larger choice set which can be compared to Tables 12 and 13.

type of price discrimination best explains the patterns we observe in the data. Going forward, this suggests modeling mortgage choices ought to evolve in two ways. First, introducing heterogeneity in people's borrowing experience and allowing for choices to differ according to income levels seems appropriate. Second, it is important to model the constraints on how menus can be tailored, that limit lenders' ability to customize offerings. Absent either of these two features it will not be possible to fit the facts we have presented.

7 Conclusion

Despite the importance of mortgage choices for personal financial well-being, relatively little is known about how people choose their mortgages. A key reason for the dearth of evidence on this question is the difficulty in observing the choice sets that borrowers face in selecting among loans. Usually all that can be analyzed is the mortgage that is chosen and little is known about the other available options. In this paper, we assemble a unique data set that allows us to see the other mortgages that were on offer both at the bank where the loan was taken out and other banks offering similar mortgages. We establish a number of facts about mortgage selection.

The number of possible options that most people face is large, even at a single lender. But lenders constantly vary their product mix and even for a given loan type the number of choices fluctuates over time. Few people pick the absolutely cheapest loan that is available, but the vast majority pick a loan that is not much more expensive than the best option.

About seven percent of people pick a mortgage that is much more expensive than others that are available. The best predictor of when these expensive choices occur is whether the borrower has a relatively large number of expensive options in the menu they are facing. The variation in menus seems designed with price discrimination in mind. Banks try to make it easy for customers who might be prone to select badly to do so, without scaring away other borrowers that they expect have the ability to shop at other lenders. This competitive pressure seems to explain why most borrowers can find a reasonable mortgage even if they do not pick particularly well.

The borrowers that are presented with these unfavorable menus are seeking loans that are large multiples of their incomes and involve high loan amounts relative to the value of the house. They tend to be first-time buyers and to be younger. There is some evidence that the expensive choices come from focusing more on fees associated with a loan instead of the promotional interest rate and not paying sufficient attention to the interest rate that prevails once an introductory, promotional interest rate expires.

It may be surprising to some people to discover that some borrowers select such expensive (or even dominated) choices. Yet, we know from other domains this kind of behavior also occurs. Perhaps the best example is the way slot machines are priced in casinos. The typical casino not only has many different types of slots, but also has multiple versions of each of those types. Within a given type of slot, the expected percentage that the casino retains (the "hold percentage") varies. Indeed, in some jurisdictions (e.g the state of Nevada), that hold percentage has to be displayed on the machine. Yet, most people do not choose which particular machine to play by comparing hold percentages and playing only the low ones. If you walk through a large casino in Las Vegas you will find a mix of high and low hold machines within each type. You will also find that hold percentages vary across types of slots in a given casino and across casinos. So the type of results we find are not confined to the particular market that we have studied. Instead the noteworthy aspect of our results is that this kind of behavior persists even in a high stakes situation.

We know that there are many other circumstances where people must select between multiple outlets that are offering fairly similar products: cars, wedding venues and charter vacations are just a few examples. It would be interesting to assess the extent to which these same patterns in menu offerings and consumer choice are present in those markets. With further analysis of these markets, some lessons for financial literacy education could be drawn.

References

Agarwal, S., Driscoll, J. C., Gabaix, X., and Laibson, D. (2009). The age of reason: Financial decisions over the life cycle and implications for regulation. Brookings Papers on Economic Activity, 2009(2):51–117.

Allen, J., Clark, R., and Houde, J.-F. (2019). Search frictions and market power in negotiated-price markets. Journal of Political Economy, 127(4):1550–1598.

Allen, J. and Li, S. (2020). Dynamic competition in negotiated price markets. Bank of canada staff working paper.

Andersen, S., Badarinza, C., Liu, L., Marx, J., and Ramadorai, T. (2022). Reference dependence in the housing market. American Economic Review, forthcoming.

Aron, J. and Muellbauer, J. (2016). Modelling and forecasting mortgage delinquency and foreclosure in the UK. Journal of Urban Economics, 94:32–53.

Bank of England (2010). Inflation Report - November 2010.

Benetton, M. (2021). Leverage regulation and market structure: An empirical model of the UK mortgage market. Journal of Finance, forthcoming.

Benetton, M., Eckley, P., Garbarino, N., Kirwin, L., and Latsi, G. (2020). Capital requirements and mortgage pricing: Evidence from Basel II. Journal of Financial Intermediation.

Benetton, M., Gavazza, A., and Surico, P. (2021). Mortgage pricing and monetary policy. Bank of England Staff Working Paper No. 936.

Bhutta, N., Fuster, A., and Hizmo, A. (2021). Paying too much? Borrower sophistication and overpayment in the US mortgage market. CEPR Discussion Paper No. 14924.

Bracke, P. and Tenreyro, S. (2021). History dependence in the housing market. American Economic Journal: Macroeconomics, 13(2):420–443.

Cloyne, J., Huber, K., Ilzetzki, E., and Kleven, H. (2019). The effect of house prices on household borrowing: a new approach. American Economic Review, 109(6):2104–36.

Duca, J. V. and Kumar, A. (2014). Financial literacy and mortgage equity withdrawals. Journal of Urban Economics, 80:62–75.

Foà, G., Gambacorta, L., Guiso, L., and Mistrulli, P. E. (2019). The supply side of household finance. The Review of Financial Studies, 32(10):3762–3798.

Gianinazzi, V. (2019). Framing bias in mortgage refinancing decisions and monetary policy pass-through. Working paper, Nova School of Business and Economics.

Guiso, L., Pozzi, A., Tsoy, A., Gambacorta, L., and Mistrulli, P. E. (2022). The cost of steering in financial markets: Evidence from the mortgage market. Journal of Financial Economics, 143(3):1209–1226.

Iscenko, Z. (2020). Strictly dominated mortgage choices: Evidence from the UK. Financial Conduct Authority Occasional Paper No. 33.

Kaplan, G. and Menzio, G. (2015). The morphology of price dispersion. International Economic Review, 56(4):1165–1206.

Kashyap, A. K. (2020). My reflections on the FPC's strategy. Journal of Money, Credit and Banking, 52(S1):63–75.

Liu, L. (2019). Non-salient fees in the mortgage market. Bank of England Staff Working Paper No. 819.

Lusardi, A. and Mitchell, O. S. (2011). Financial literacy and retirement planning in the United States. NBER Working Paper No. 17078.

Menzio, G. and Trachter, N. (2018). Equilibrium price dispersion across and within stores. Review of Economic Dynamics, 28:205–220.

Myśliwski, M. and Rostom, M. (2022). Value of information, search, and competition in the uk mortgage market. Bank of England Staff Working Paper No. 967.

Peydró, J.-L., Rodriguez Tous, F., Tripathy, J., and Uluc, A. (2020). Macroprudential policy, mortgage cycles and distributional effects: Evidence from the UK. CEPR Discussion Paper No. 15275.

Robles-Garcia, C. (2019). Competition and incentives in mortgage markets: The role of brokers. Working paper, Stanford Graduate School of Business.

Ru, H. and Schoar, A. (2016). Do credit card companies screen for behavioral biases? NBER Working Paper No. 22360.

Scott Morton, F., Zettelmeyer, F., and Silva-Risso, J. (2003). Consumer information and discrimination: Does the internet affect the pricing of new cars to women and minorities? Quantitative marketing and Economics, 1(1):65–92.

Stigler, G. J. (1961). The economics of information. Journal of Political Economy, 69(3):213–225.

Woodward, S. E. and Hall, R. E. (2012). Diagnosing consumer confusion and sub-optimal shopping effort: Theory and mortgage-market evidence. American Economic Review, 102(7):3249–76.

Tables & Figures

Table 1: Example of products on offer

	Initial rate (%)	Fee (£)	Reset rate (%)	Max LTV (%)	Max loan (£000)
Bank A	3.39	999	3.94	90	400
Bank A	3.59	1,499	3.94	90	400
Bank A	3.64	999	3.94	90	400
Bank A	4.15	0	3.94	90	400
Bank A	4.15	599	3.94	90	400
Bank A	4.19	599	3.94	90	400
Bank A	4.19	399	3.94	90	400
Bank A	4.19	299	3.94	90	400
Bank A	4.19	0	3.94	90	400
Bank A	4.59	0	3.94	90	400
Bank A	4.99	399	3.94	90	400
Bank A	3.64	999	3.94	90	3,000
Bank A	3.69	999	3.94	90	3,000
Bank A	4.19	399	3.94	90	3,000
Bank A	4.19	599	3.94	90	3,000
Bank A	4.19	299	3.94	90	3,000
Bank A	4.59	0	3.94	90	3,000
Bank B	3.29	0	3.99	90	500
Bank B	4.99	999	3.99	90	750
Bank B	4.99	1,264	3.99	90	750
Bank B	3.94	1,260	3.99	90	1,000
Bank B	3.99	1,260	3.99	90	1,000
Bank B	4.54	265	3.99	90	1,000
Bank B	4.59	0	3.99	90	1,000
Bank B	4.79	1,260	3.99	90	1,000
Bank B	4.89	0	3.99	90	1,000
Bank B	5.19	1,260	3.99	90	1,000
Bank B	5.19	0	3.99	90	1,000
Bank B	5.39	0	3.99	90	1,000
Bank B	5.39	1,260	3.99	90	1,000
Bank B	5.69	0	3.99	90	1,000
Bank B	5.69	1,260	3.99	90	1,000

Note: This table reports example choices for a given customer across banks. For simplicity only a subset of the menu customers typically face is shown.

Table 2: Summary statistics

	Mean	Std. dev.	25^{th} pctile	Median	75^{th} pctile
Demographics					
Young (%)	36	48	0	0	100
Old (%)	11	31	0	0	0
First-time buyer (%)	40	49	0	0	100
Net income (£000s)	42	26	28	37	50
Loan characteristics					
Loan value (£000s)	157	90	100	136	190
House price (£000s)	201	119	125	172	242
Loan-to-value (%)	79	8	74	80	85
Loan-to-income ratio	3.2	0.9	2.6	3.2	3.8
Prices					
Fee (£000s)	0.66	0.57	0.10	0.76	1.00
Initial rate (%)	4.0	1.0	3.2	3.9	4.7
Reset rate (%)	4.1	0.4	4.0	4.0	4.2

Note: This table summarizes the key variables used in our analysis. Young customers are under 30. Old customers are over 45. A customer's net income is measured as reported (gross) income minus tax. First-time buyers are buying a house for the first time. Loan value and house price are reported in the mortgage contract, and loan-to-value (LTV) is the ratio of the loan value to the house price in percent. The loan-to-income ratio follows the industry convention and is calculated by dividing the loan amount by reported gross income. The fee, initial rate and reset rate are also taken from the mortgage contract.

Table 3: Characteristics of choice sets and choices made

	Within		Across	
	Choice set size	Pctile chosen	Choice set size	Pctile chosen
25^{th} pctile	11	33	46	27
Median	16	53	73	47
75^{th} pctile	23	75	101	70

Note: This table summarizes customers' choice sets and choices made. Within-bank figures restrict a customer's choice set to mortgages on offer at the bank that granted their mortgage. Across-bank figures include mortgages on offer at all banks in a customer's choice set. For each customer we rank the mortgages in their comparison set from cheapest to most expensive. The variable *Pctile chosen* is equal to the rank of the mortgage they choose as a percentage of the number of mortgages in their comparison set. The table summarizes the distribution of this variable, along with the size of the choice set, across the sample.

Table 4: Probit regressions of high loan-to-value and high loan-to-income on borrower types

	High LTV	High LTI	High LTV & LTI
	(1)	(2)	(3)
Young	0.071***	0.023***	0.016***
	(0.001)	(0.001)	(0.001)
Old	−0.095***	−0.079***	−0.035***
	(0.002)	(0.001)	(0.001)
First-time buyer	0.234***	0.037***	0.042***
	(0.001)	(0.001)	(0.001)
Poor	−0.076***	0.065***	−0.003***
	(0.001)	(0.001)	(0.001)
Rich	0.032***	−0.067***	−0.014***
	(0.001)	(0.001)	(0.001)
Bank dummies	Yes	Yes	Yes
Product dummies	Yes	Yes	Yes
Pseudo R-squared	0.12	0.05	0.05
Mean dependent variable	0.32	0.2	0.05
Observations	894,901	894,901	894,901

Note: This table reports average marginal effects from probit regressions with the dependent variables shown at the top of each column. High LTI customers have loan-to-income above 4. High LTV customers have loan-to-value above 85%. Young customers are under 30. Old customers are over 45. Poor (rich) customers have income in the lower (upper) tertile of the income distribution. Product dummies are indicator variables for 2, 3 and 5 year fixed rate mortgages and 2 year adjustable rate mortgages. Bank dummies are indicator variables for each of the 6 banks. *, ** and *** indicate that the coefficient is different from zero at the 10, 5 and 1 percent level of significance respectively.

Table 5: Dispersion in the price components of mortgage contracts

	Within			Across		
	25^{th} pctile	Median	75^{th} pctile	25^{th} pctile	Median	75^{th} pctile
Initial fee (£000s)	0.80	0.99	1.00	0.99	1.14	1.25
Initial rate (pp)	0.70	1.00	1.35	1.07	1.30	1.65
Reset rate (pp)	0.00	0.00	0.00	0.75	0.85	1.05

Note: This table summarizes the dispersion customers face in fees, initial rates and reset rates. We take each unique value of these three elements of a mortgage in a customer's choice set and compute the difference between the 85^{th} and the 15^{th} percentiles of the distribution of each of these variables. The table summarizes the distributions of these dispersions across customers. Within-bank figures restrict a customer's choice set to mortgages on offer at the bank that granted their mortgage. Across-bank figures include mortgages on offer at all banks in a customer's choice set.

Table 6: Determinants of the size of the tail for each borrower

	Bad tail		Strongly dominated tail	
	Within	Across	Within	Across
	(1)	(2)	(3)	(4)
Fee dispersion (£000s)	0.103***	1.960***	−0.564***	0.902***
	(0.011)	(0.035)	(0.018)	(0.056)
Rate dispersion (pp)	7.440***	13.900***	10.200***	7.930***
	(0.011)	(0.014)	(0.018)	(0.022)
Reset rate dispersion (pp)	0.657***	3.250***	3.200***	0.599***
	(0.023)	(0.022)	(0.038)	(0.035)
Bank dummies	Yes	No	Yes	No
Product dummies	Yes	Yes	Yes	Yes
Mean dependent variable	2.2	7.83	4.53	13.06
R-squared	0.42	0.6	0.32	0.16
Observations	894901	883459	894901	883459

Note: This table reports coefficients from OLS regressions with the size of borrowers' bad and strongly dominated tails as the dependent variables. We take each unique value of the three elements of a mortgage - the initial rate, the fee and the reset rate - in a customer's choice set and order them from low to high. The dispersion is the difference between the 85^{th} and the 15^{th} percentiles of the distribution of each of these variables. The *bad tail* measures the percentage of a customer's choice set that would represent an expensive choice whilst the *strongly dominated tail* computes the fraction of mortgages in a customer's choice set that are strongly dominated by the cheapest mortgage on offer. Expensive and strongly dominated choices are defined in Section 3. The first and third columns take the choice set to consist only of mortgages on offer at the bank that granted the customer their mortgage, whilst the second and fourth include all banks. Dummy variables are defined as in Table 4. *, ** and *** indicate that the coefficient is different from zero at the 10, 5 and 1 percent level of significance respectively.

Table 7: Probit regressions of expensive choice on choices of price components

	Expensive choice within	Expensive choice across
	(1)	(2)
Low Fee	-0.003^{***}	-0.009^{***}
	(0.0002)	(0.0003)
High Fee	0.0003^{*}	0.027^{***}
	(0.0002)	(0.001)
Low Inital Rate	-0.017^{***}	-0.037^{***}
	(0.0003)	(0.001)
High Initial Rate	0.032^{***}	0.122^{***}
	(0.0004)	(0.001)
Low Reset Rate	-0.006^{***}	-0.023^{***}
	(0.0003)	(0.0004)
High Reset Rate	0.006^{***}	0.057^{***}
	(0.0002)	(0.0004)
Bad tail	0.124^{***}	0.314^{***}
	(0.001)	(0.001)
Bank dummies	Yes	No
Product dummies	Yes	Yes
Pseudo R-squared	0.82	0.69
Mean dependent variable	0.023	0.067
Observations	894,901	883,459

Note: This table reports average marginal effects from probit regressions with the dependent variables shown at the top of each column. We take each unique value of the three elements of a mortgage - the initial rate, the fee and the reset rate - in a customer's choice set and order them from low to high. Choices above the 85^{th} percentile are high, whilst choices below the 15^{th} percentile are low. The *bad tail* measures the percentage of a customer's choice set that would represent an expensive choice, where expensive choices are defined in Section 3. The first column takes the choice set to consist only of mortgages on offer at the bank that granted the customer their mortgage, whilst the second includes all banks. Dummy variables are defined as in Table 4. *, ** and *** indicate that the coefficient is different from zero at the 10, 5 and 1 percent level of significance respectively.

Table 8: Determinants of within-bank expensive choices

	High rate within	Expensive choice within
	(1)	(2)
Low Fee	0.206***	0.002***
	(0.001)	(0.0002)
High Fee	−0.032***	0.001***
	(0.001)	(0.0002)
Low Reset Rate	0.008***	
	(0.001)	
High Reset Rate	−0.061***	
	(0.001)	
Bad tail		0.116***
		(0.001)
Bank dummies	Yes	Yes
Product dummies	Yes	Yes
Pseudo R-squared	0.09	0.69
Mean dependent variable	0.216	0.023
Observations	894,901	894,901

Note: This table reports average marginal effects from probit regressions with the dependent variables shown at the top of each column. We take each unique value of the initial rate, the fee and the reset rate in a customer's choice set and order them from low to high. Choices above the 85^{th} percentile are high, whilst choices below the 15^{th} percentile are low. The *bad tail* measures the percentage of a customer's choice set that would represent an expensive choice, where expensive choices are defined in Section 3. The dependent variable in column (1) is a dummy variable equal to 1 if the customer picks a product with a high initial rate, and in column (2) is a dummy variable equal to 1 if the customer makes an expensive choice. The choice set consists only of mortgages on offer at the bank that granted the customer their mortgage. Dummy variables are defined as in Table 4. *, ** and *** indicate that the coefficient is different from zero at the 10, 5 and 1 percent level of significance respectively.

Table 9: Expensive choices within and across banks

	Expensive choice across	Not
Expensive choice within	1.2	1.1
Not	5.5	92.2

Note: This table shows the distribution of expensive choices within and across banks. A customer's within-bank choice set consists only of mortgages on offer at the bank that granted the customer their mortgage, whilst their choice set across banks includes all banks. An expensive choice costs a customer at least 2.5% of their income more than the 15^{th} percentile in their choice set. The figures show the percentages of the sample in each of the four combinations of expensive and inexpensive choices.

Table 10: Choice quality within and across banks

	25th pctile	Median	75th pctile	90th pctile
Cost difference within bank	0.09	0.32	0.69	1.20
Cost difference vs best bank	0.06	0.63	1.45	2.23

Note: This table summarizes the quality of customers' choices within and across banks. The *cost difference vs best bank* measures the difference in cost between the mean product at the bank where the customer took out their mortgage and the cheapest bank they could have shopped at – where banks are ranked by the mean cost of the products they offer that customer – as a percentage of the customer's income. The *cost difference within bank* measures the difference in cost between the mortgage the customer chose and the 15th percentile mortgage on offer at their bank. The entries in the table show the values of these variables at different points in the customer distribution.

Table 11: Determinants of across-bank expensive choices

	Expensive choice across				
	(1)	(2)	(3)	(4)	(5)
Cost difference within bank			0.016***		0.023***
			(0.0002)		(0.0002)
Cost difference vs. best bank				0.023***	0.029***
				(0.0002)	(0.0002)
Bad tail		0.304***	0.279***	0.224***	0.150***
		(0.001)	(0.001)	(0.001)	(0.001)
Bank dummies	No	No	No	No	No
Product dummies	Yes	Yes	Yes	Yes	Yes
Pseudo R-squared	0.09	0.56	0.63	0.59	0.71
Mean dependent variable	0.067	0.067	0.067	0.067	0.067
Observations	883,459	883,459	883,459	883,459	883,459

Note: This table reports average partial effects from probit regressions with the dependent variable being an expensive choice across banks. The *bad tail* measures the percentage of a customer's choice set that would represent an expensive choice, where expensive choices are defined in Section 3. The *cost difference vs best bank* measures the difference in cost between the mean product at the bank where the customer took out their mortgage and the cheapest bank they could have shopped at, as a percentage of the customer's income. The *cost difference within bank* measures the difference in cost between the mortgage the customer chose and the 15th percentile mortgage on offer at their bank. Dummy variables are defined as in Table 4. Bank dummies are indicator variables for each of the 6 banks. *, ** and *** indicate that the coefficient is different from zero at the 10, 5 and 1 percent level of significance respectively.

Table 12: Expensive & strongly dominated choices and loan characteristics

	Expensive choice				Strongly dominated choice			
	Within	Within	Across	Across	Within	Within	Across	Across
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
High LTV & High LTI	0.093***	−0.002***	0.228***	0.023***	0.208***	0.049***	0.361***	0.179***
	(0.002)	(0.0003)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.002)
High LTV & Low LTI	0.044***	0.0001	0.076***	0.001	0.091***	0.017***	0.125***	0.034***
	(0.001)	(0.0002)	(0.001)	(0.0005)	(0.001)	(0.001)	(0.001)	(0.001)
Low LTV & High LTI	0.025***	0.002***	0.087***	0.008***	0.068***	0.031***	0.249***	0.138***
	(0.001)	(0.0003)	(0.001)	(0.0005)	(0.001)	(0.001)	(0.001)	(0.001)
Bad tail		0.118***		0.299***				
		(0.001)		(0.001)				
Strongly dominated tail						0.363***		0.781***
						(0.001)		(0.002)
Bank dummies	Yes	Yes	No	No	Yes	Yes	No	No
Product dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R-squared	0.35	0.69	0.15	0.57	0.2	0.38	0.08	0.17
Mean dependent variable	0.023	0.023	0.067	0.067	0.081	0.081	0.277	0.277
Observations	894,901	894,901	883,459	883,459	894,901	894,901	883,459	883,459

Note: This table reports average partial effects from probit regressions with the dependent variables shown at the top of each column. High (low) LTI customers have loan-to-income above (below) 4. High (low) LTV customers have loan-to-value above (below) 85%. The *bad tail* measures the percentage of a customer's choice set that would represent an expensive choice whilst the *strongly dominated tail* computes the fraction of mortgages in a customer's choice set that are strongly dominated by the cheapest mortgage on offer. Expensive and strongly dominated choices are defined in Section 3. Columns headed 'within' take the choice set to consist only of mortgages on offer at the bank that granted the customer their mortgage, whilst columns headed 'across' include all banks. Dummy variables are defined as in Table 4. *, ** and *** indicate that the coefficient is different from zero at the 10, 5 and 1 percent level of significance respectively.

Table 13: Expensive & strongly dominated choices and borrower characteristics

	Expensive choice				Strongly dominated choice			
	Within	Within	Across	Across	Within	Within	Across	Across
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Young	0.005***	0.001***	0.018***	0.005***	0.022***	0.010***	0.043***	0.015***
	(0.0004)	(0.0002)	(0.001)	(0.0004)	(0.001)	(0.001)	(0.001)	(0.001)
Old	−0.008***	−0.0003	−0.031***	−0.006***	−0.022***	−0.006***	−0.075***	−0.036***
	(0.0004)	(0.0003)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
First-time buyer	0.006***	−0.0003	0.005***	−0.005***	0.010***	−0.002***	−0.003***	−0.030***
	(0.0004)	(0.0002)	(0.001)	(0.0004)	(0.001)	(0.001)	(0.001)	(0.001)
Poor	0.0005	0.001***	0.003***	0.001**	0.007***	0.008***	0.034***	0.010***
	(0.0004)	(0.0002)	(0.001)	(0.0004)	(0.001)	(0.001)	(0.001)	(0.001)
Rich	−0.0001	−0.001***	−0.006***	−0.006***	−0.010***	−0.010***	−0.043***	−0.029***
	(0.0003)	(0.0002)	(0.001)	(0.0004)	(0.001)	(0.001)	(0.001)	(0.001)
Bad tail		0.117***		0.303***				
		(0.001)		(0.001)				
Strongly dominated tail						0.382***		0.870***
						(0.001)		(0.002)
Bank dummies	Yes	Yes	No	No	Yes	Yes	No	No
Product dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R-squared	0.3	0.69	0.09	0.56	0.16	0.37	0.04	0.16
Mean dependent variable	0.023	0.023	0.067	0.067	0.081	0.081	0.277	0.277
Observations	894,901	894,901	883,459	883,459	894,901	894,901	883,459	883,459

Note: This table reports average partial effects from probit regressions with the dependent variables shown at the top of each column. Young customers are under 30. Old customers are over 45. Poor (rich) customers have net income in the lower (upper) tertile. The *bad tail* measures the percentage of a customer's choice set that would represent an expensive choice whilst the *strongly dominated tail* computes the fraction of mortgages in a customer's choice set that are strongly dominated by the cheapest mortgage on offer. Expensive and strongly dominated choices are defined in Section 3. Columns headed 'within' take the choice set to consist only of mortgages on offer at the bank that granted the customer their mortgage, whilst columns headed 'across' include all banks. Dummy variables are defined as in Table 4. *, ** and *** indicate that the coefficient is different from zero at the 10, 5 and 1 percent level of significance respectively.

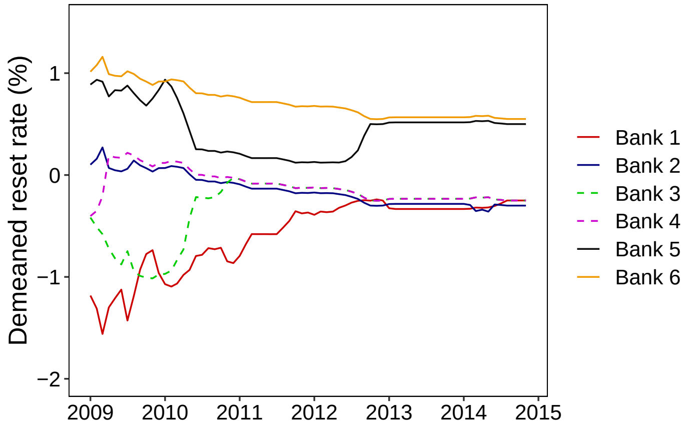
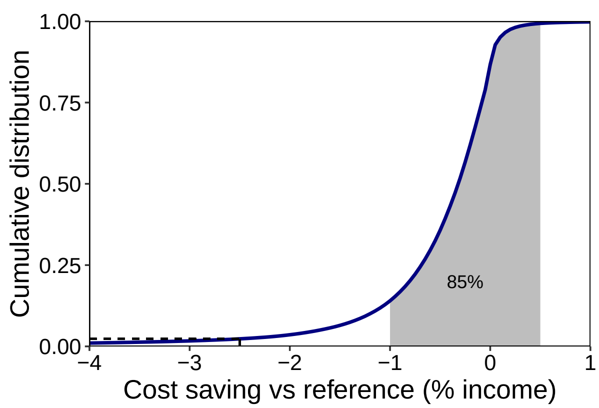
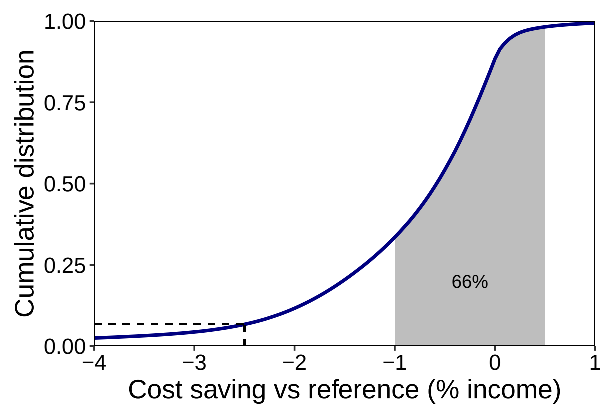


Figure 1: Across-bank variation in reset rates

Note: This figure shows the deviation of bank-specific reset rates relative to the average reset rate for each month. For each bank we compute the average reset rate across all products they offer in a given month. We then demean these by the simple average of the reset rate across banks, and plot these series.



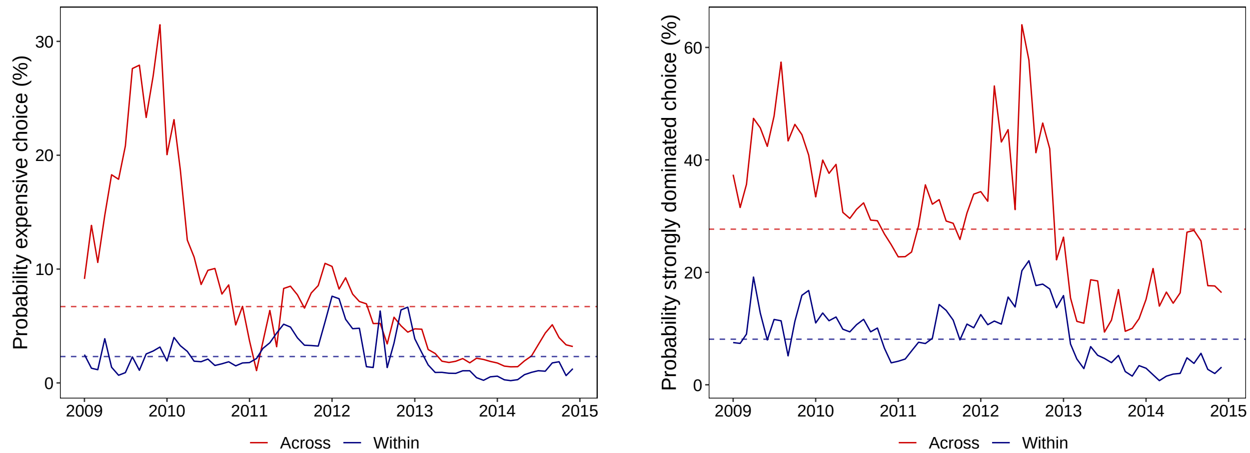
(a) Within each bank's choice set



(b) Across all banks' choice sets

Figure 2: Cost savings of chosen mortgage vs reference (% of net income)

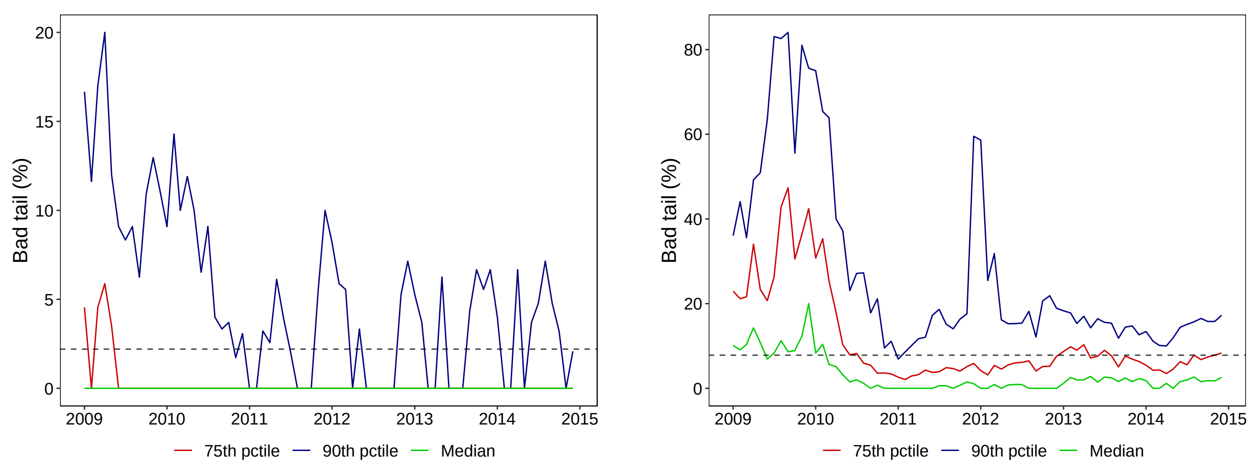
Note: These figures plot the distribution of the amount a customer saves relative to a reference mortgage as a percentage of their income, at the bank where the customer shopped and across banks. We first compute the present value of the mortgage that a customer chooses using equation (1), subtract it from the cost of the 15$^{\text{th}}$ percentile mortgage in a customer's choice set (where mortgages are ordered from cheapest to most expensive), and divide by the customer's net income. The figures plot the cumulative distribution of this figure across all customers, where the choice set consists only of mortgages on offer at the bank that gave them their mortgage (left) and of mortgages on offer across banks (right). The shaded areas show the fraction of each sample that fall between savings of 0.5% of net income and a cost of 1% of net income. The dotted lines show where the choice costs more than 2.5% of net income relative to the reference mortgage.



(a) Expensive choices (b) Strongly dominated choices

Figure 3: Expensive & strongly dominated choices through time

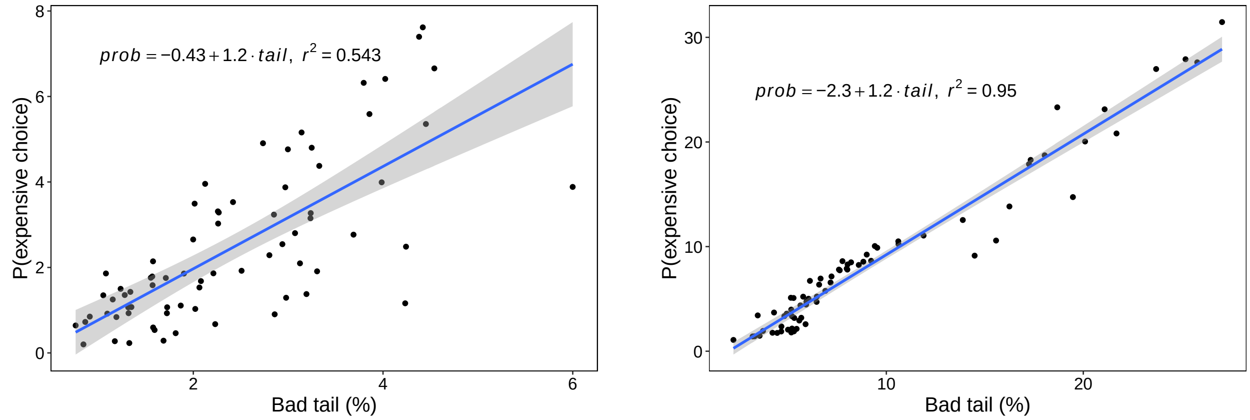
Note: These figures plot the percentage of customers that make expensive choices (left) and strongly dominated choices (right) each month, at the bank where they shopped and across banks. Expensive and strongly dominated choices are defined in Section 3. The blue line takes the choice set to consist only of mortgages on offer at the bank that gave the customer their mortgage, and the red line includes mortgages on offer across all banks. The horizontal lines plot the means over the sample period for each comparison.



(a) Within each bank's choice set (b) Across all banks' choice sets

Figure 4: Bad tail through time

Note: These figures summarize the distribution of the fraction of a customer's choice set that would represent an expensive choice if chosen, both within and across banks, through time. The *bad tail* computes the fraction of the mortgages in a customer's choice set that would represent an expensive choice, where an expensive choice is defined in Section 3. The figures plot percentiles of the distribution of *bad tail* through time for the median (green) 75$^{\text{th}}$ percentile (red) and 90$^{\text{th}}$ percentile (blue), where the choice set consists only of mortgages on offer at the bank that gave the customer their mortgage (left panel) and of mortgages on offer across banks (right panel). The horizontal lines plot the means over the sample period.

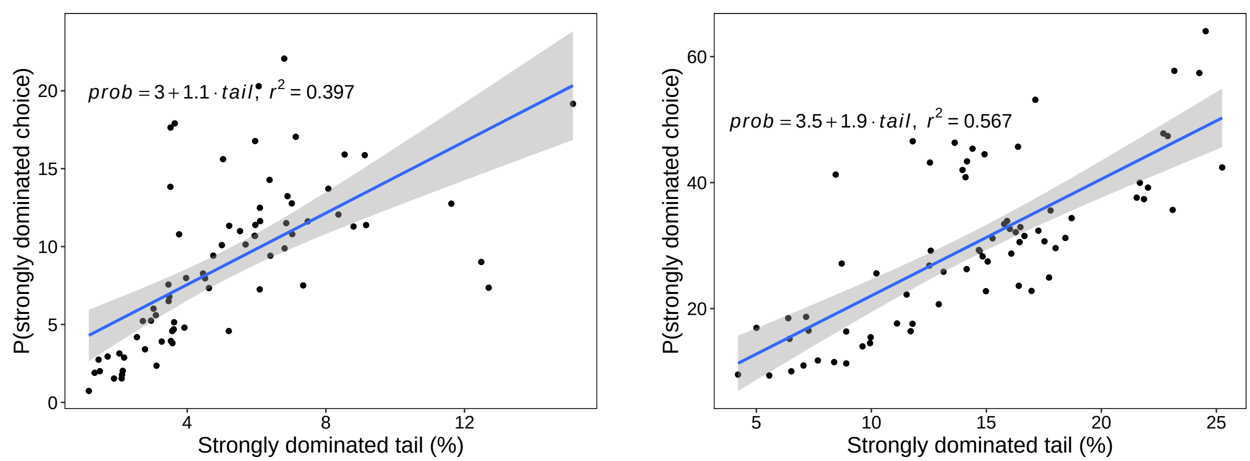


(a) Within each bank's choice set

(b) Across all banks' choice sets

Figure 5: Expensive choices and bad tails

Note: These figures summarize the relationship between the frequency of expensive choices and the average quality of customers' choice sets in a given month. The *bad tail* computes the fraction of the mortgages in a customer's choice set that would represent an expensive choice, where an expensive choice is defined in Section 3. The figures plot the percentage of customers that make expensive choices in a month against the average size of *bad tail* in that month, where the choice set consists only of mortgages on offer at the bank that gave the customer their mortgage (left panel) and of mortgages on offer across banks (right panel). The blue line shows a linear regression of the probability of an expensive choice on the size of *bad tail*, with equation displayed in each panel. The shaded area represents the 95% confidence interval.



(a) Within each bank's choice set

(b) Across all banks' choice sets

Figure 6: Strongly dominated choices and tails

Note: These figures summarize the relationship between the frequency of strongly dominated choices and the average quality of customers' choice sets in a given month. The *strongly dominated tail* computes the percentage of a customer's choice set that is strongly dominated by the cheapest option in their choice set, where strong domination is defined in Section 3. The figures plot the percentage of customers that make strongly dominated choices in a month against the average size of the strongly dominated tail in that month, where the choice set consists only of mortgages on offer at the bank that gave the customer their mortgage (left panel) and of mortgages on offer across banks (right panel). The blue line shows a linear regression of the probability of a strongly dominated choice choice on the size of the strongly dominated tail, with equation displayed in each panel. The shaded area represents the 95% confidence interval.

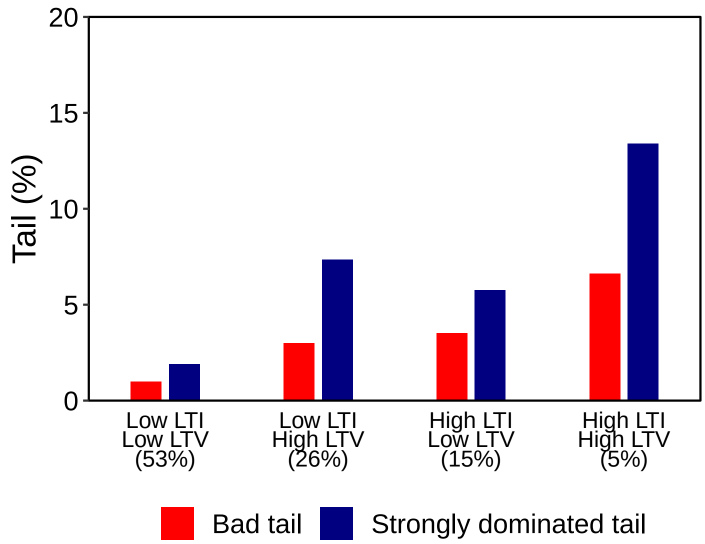


Figure 7: Bad and strongly dominated tail by loan-to-value and loan-to-income ratio

Note: This figure summarizes the average quality of customers' choice sets by combinations of their LTV and LTI ratios. The *bad tail* computes the fraction of the mortgages in a customer's choice set that would represent an expensive choice, whilst the *strongly dominated tail* computes the percentage of a customer's choice set that is strongly dominated by the cheapest option in their choice set. Expensive and strongly dominated choices are defined in Section 3. This figure plots the average of *bad tail* and *strongly dominated tail* according to a customer's LTV and LTI. High LTV is defined as LTV$>$ 85%, and low LTV as LTV$<$ 85%. High LTI is defined as LTI$>$ 4, and low LTI as LTI$<$ 4. The numbers in parentheses are the percentages of the sample in each bin.

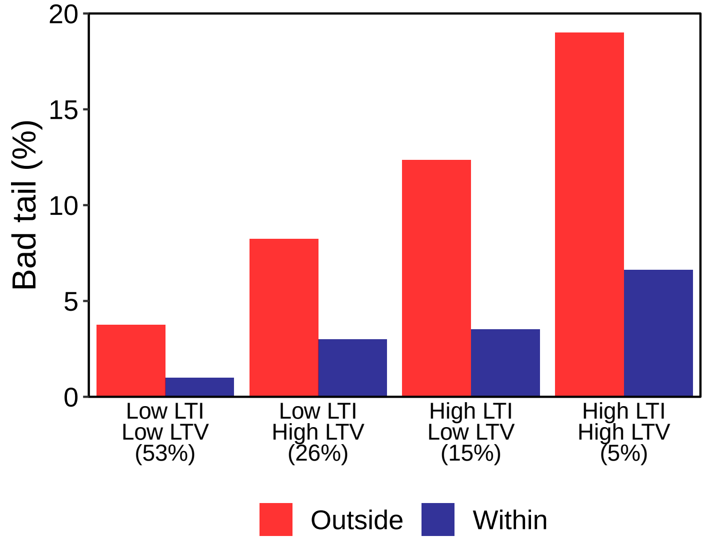


Figure 8: Tails within and outside banks by loan-to-value and loan-to-income ratio

Note: This figure summarizes the average quality of customers' choice sets by their LTV and LTI, both at the bank that granted them their mortgage and at the banks where they did not borrow. The *bad tail* computes the fraction of the mortgages in a customer's choice set that would represent an expensive choice, where an expensive choice is defined in Section 3. The within-bank bad tail (blue) is the average of *bad tail* at the bank where the customer borrowed. The variable *outside tail* (red) computes the fraction of the mortgages the customer could have chosen at the other five banks that would represent an expensive choice, relative to the same reference of the 15^{th} percentile at the customer's chosen bank. This figure plots the average of the within-bank *bad tail* and *outside tail* according to a customer's LTV and LTI. High LTV is defined as LTV$>85\%$, and low LTV as LTV$<85\%$. High LTI is defined as LTI>4, and low LTI as LTI<4. The numbers in parentheses are the percentages of the sample in each bin.

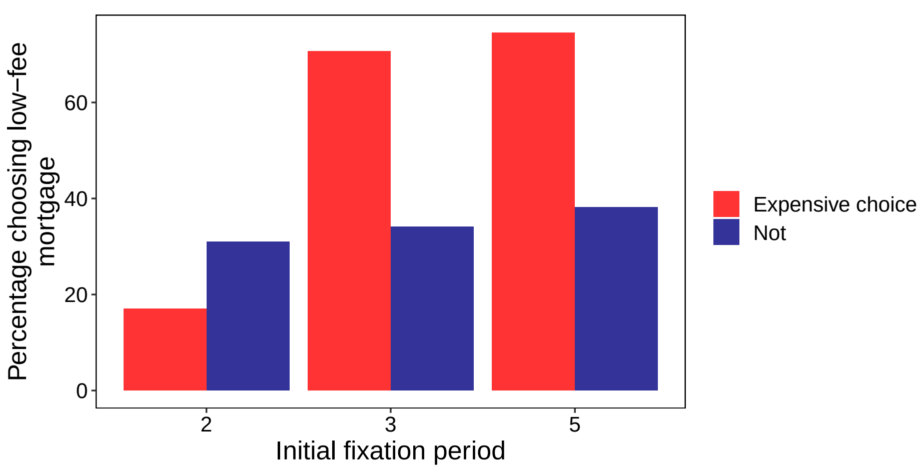


Figure 9: Expensive choices by choice of fee

Note: This figure shows the percentage of borrowers that choose low-fee mortgages for mortgages with different initial fixation periods. A mortgage is low-fee if its fee is below the 15^{th} percentile of the distribution of unique fees in the customer's choice set. A mortgage is high-fee if its fee is above the 85^{th} percentile. Expensive choices are defined in Section 3.

Internet Appendix

A1 Sample formation

Our main data source is the Product Sales Database (PSD), a loan-level administrative dataset capturing all newly issued mortgages in the UK. A typical set of choices an individual would be presented with is shown in Table 1.

The PSD contains information recorded by financial institutions at the time of mortgage take-out. In principle, this includes borrower characteristics, such as income and age; information on the property, such as its postcode and value, and loan details such as the initial interest rate, length of fixation period, and the issuing institution. Table A1 shows the raw data between 2009 and 2014, after discarding observations for which no initial rate information is available. Crucially, PSD contains no information on the reset or standard variable rate.

We merge the PSD with a secondary data source, Moneyfacts, which records the reset rates on all products at a given point in time. We also merge in product fees, and the minimum and maximum loan size available for each product. Most loans have no minimum loan requirement, and a maximum loan requirement of £1 million. We match the two data sources together using a matching algorithm that uses the institutional name, product type, initial rate, fixation period, and whether the purchase date fell during the period the mortgage product was on the market. Not all this information is populated in the data, but we are able to successfully match 73% or 2.6 million observations out of the 3.5 million. Table A2 shows the resulting dataset broken down by year and product type.

The most popular products are the 2, 3 and 5 year fixed-rate mortgages (FRMs) and the 2 year adjustable-rate mortgage (ARM). So we focus on only these 4 types of mortgages, and drop the roughly 475,000 other bespoke mortgages that would be difficult to compare and may not be available at many lenders.[27] We further drop observations that have missing data or are outliers. These screens exclude cases with very small loan values, very high loan-to-income values or if income or loan values are missing. The sample characteristics

[27] The 'OTHER' category contains less popular products such as ARMs with different initial period durations and ARMs with an upper cap on the initial rate.

after these filters are applied are shown in Table A3.

Although there are many small mortgage lenders, most of the mortgage market is dominated by a handful of top players. Given our interest in contrasting choices within the set of loans made by an individual lender, we want enough loans every month to make meaningful comparisons. We eliminate peripheral lenders, and focus on six large mortgage providers that provide the most loans in our sample. Table A4 shows the loan characteristics of those six lenders. Focusing on these lenders only shifts the sample from just over 2 million loans to just under 1.6 million loans.

We further restrict our sample to households with loan amounts less than £1 million, who have a loan-to-value ratio between 65% - 95% LTV, and who have at least 5 products to choose from at their bank. This reduces the sample to one shown in Table A5. Around a third of our sample are taking out a mortgage for the first time. The 2yr FRM is by far the most popular product in our sample period, representing over half the sample.[28] For the across-bank analysis we restrict our sample to customers with at least 25 products to choose from if they choose a 2yr FRM, and 15 if they chose the other products, reflecting the larger menus for the 2yr FRM. Our final samples are 894,901 observations for analysis within banks, and 883,459 observations for analysis across banks.

A2 Calculations

To compare mortgage options, we first specify the set of all possible mortgages a household could have chosen from. We do this by creating a dataset with unique product-level observations, arranged by month, six LTV buckets (65%, 70%, 75%, 80%, 85%, 90% and 95%), the minimum and maximum loan sizes allowable, and the product type. Our final sample of mortgage choices is then compared against with eligible product.

The observed date in the PSD is the completion date on the house purchase. For the majority of homes, a mortgage is usually agreed 3-5 months in advance of the sale. The initial rates and fees associated with these completion dates will therefore be from offers available 3 to 5 months ago. But the observed reset rate will be at the time of completion,

[28]After the sample ended, as mortgage rates fell further and the yield curve became extremely flat, many borrowers shifted to 5 year fixed rate mortgages. So currently that mortgage is the most common.

not at the time of the mortgage offer. We therefore lag the reset rate by 4 months to ensure it coincides with the quoted SVR at the time of the mortgage offer, instead of the time of closing.

Second, we compare each chosen mortgage against mortgage products of the same product type and available in the same month. We also only compare products where the actual loan size does not exceed the maximum loan amount, nor fall short of the minimum; and the actual LTV is in the same bucket as the maximum allowable LTV on that product. For example, a customer with a LTV of 82% will have a choice set consisting of mortgages where the max LTV is 85%.[29]

Finally, we use each product's initial rate, reset rate, fee and initial promotional period, together with the borrower's loan size, to compute the net present value (NPV) of each mortgage in a borrower's choice set. To compute the NPV we first compute the borrower's monthly payments during the initial promotional period:

$$IP = \frac{(1+i)Q}{1 + \sum_{j=1}^{T}(1+\frac{r_i}{12})^{-j}}$$

where Q is the loan amount, T is the mortgage term and r_i is the initial interest rate.

The borrower's monthly payments after the initial period are given by:

$$RP = \frac{\widetilde{Q}}{1 + \sum_{j=1}^{\widetilde{T}}(1+\frac{r_r}{12})^{-j}}$$

where $\widetilde{T} = T - T_F$ is the mortgage term T minus the initial period T_F, $\widetilde{Q}$ is the loan balance remaining at the end of the initial period, r_r is the reset rate and

[29] We use the lowest LTV that is actually available for which the borrower qualifies. So in the example, if the customer's bank does not offer a product with an LTV of 85% but does have one at 90%, we will use the 90% loan terms in forming the menu. For the across-bank analysis, if in this case other banks do offer products with an LTV of 85% then the menu will be based on the set of loans with an LTV of 85%, supplemented with the 90% LTV loan the customer chose.

$$\widetilde{Q} = Q - (IP - \frac{r_i}{12}Q)T_F$$

We then compute the net present value as the discounted sum of repayments over the first 7 years of the contract:

$$NPV = fee + \sum_{t=1}^{T_F} \frac{IP}{(1+i)^t} + \sum_{t=T_F+1}^{84} \frac{RP}{(1+i)^t}$$

where i is the seven year LIBOR rate.

A3 Robustness

A3.1 Refinancing decision

In this section we demonstrate that our key results are robust to alternative assumptions regarding when customers choose to refinance their mortgage. In our main analysis we assume customers refinance their mortgage after 7 years. In this section we instead assume that customers refinance immediately after the initial period of their mortgage ends, after 2, 3 or 5 years depending on their choice of mortgage product. Under this alternative assumption, customers can only make expensive choices due to their choice of initial rate or fee, as the reset rate no longer affects the computed cost of a mortgage. This robustness exercise checks that our findings are not the result of assuming customers switch onto expensive reset rates when in reality they might refinance.

Figures A1 to A5 and Tables A6 to A8 replicate the key charts and tables of the paper under this alternative assumption. The paper's key results are unchanged. In particular, customers' expensive choices are driven by the price dispersion in the menu they are offered. Customers borrowing large amounts relative to the value of their house and/or their income are more likely to make expensive choices, and this is driven by the menus they are given by banks.

Across banks, restricting the cost calculation to the initial period reduces both the prevalence of expensive choices and the price dispersion in mortgage menus as captured by the size of the bad tail (Figures A2 and A3). Expensive choices and price dispersion within banks are largely unchanged, as there is typically only one reset rate available at a given bank. The correlation between the quality of the choice set and the frequency of expensive choices in a given month (Figure A4) remains strong.

Figure A5 shows the pattern of tail size by LTV and LTI for the alternative assumption about refinancing. The relationship is the same as in the main analysis: customers borrowing large amounts relative to the value of their home and/or their income receive menus with significantly more bad choices. Table A6 shows that expensive choices continue to be driven primarily by the choice of initial rate under this alternative assumption about refinancing. Table A7 shows that these customers continue to be more prone to make expensive choices, and that this is largely driven by the quality of the menu that they were offered. As in the main analysis, this means young people and first-time buyers are more likely to make expensive choices, and this is largely driven by the quality of the choice set they receive (Table A8).

A3.2 Strongly dominated mortgages

Throughout the main text we have emphasized that our results are robust to an alternative ranking of mortgages according to the criteria of strict dominance. Table A9 sets out how strongly dominated choices depend on a borrower's choice of initial rate, fee and reset rate. The choice of initial rate is the key driver of strongly dominated choices within banks, whilst across banks all price dimensions play a role.

A3.3 Choice set construction

We demonstrate that our key results are robust to the way we construct customers' choice sets. To define a customer's choice set we first identify all mortgages that were on offer when they were shopping around, with the same initial period as the mortgage they chose, and that were available for the amount they borrowed. In our main results we then further restrict the choice set to those mortgages with the lowest LTV band for which the customer qualifies. We do so on the basis that this is the relevant menu for most customers, on the basis that over 90% of customers choose a mortgage with the lowest LTV band for which they qualify. In this section we replicate our results without this further restriction on choice sets, so that the menu of a customer with an 85% LTV includes mortgages with a maximum LTV of 85%, but also mortgages with maximum LTVs greater than 85%.[30]

Figures A6 to A9 and Tables A10 to A12 replicate the key charts and tables of the paper with these alternative choice sets. The size of the choice sets and the tails increases, and the fact that low-LTV mortgages cost less mechanically introduces a negative correlation between LTV and the size of the tail. However, the key results - the positive correlation between expensive choices and price dispersion and the fact that high-LTV and high-LTI customers are more likely to make expensive choices - are unchanged. This is consistent with the main messages of our paper: customers borrowing large amounts relative to their house value and/or their income are more likely to make expensive choices, and this is driven by the menus that they are given by banks. The relevant menu to consider is the set of mortgages with the lowest LTV for which a customer qualifies, as this is what drives expensive choices.

The alternative choice set significantly increases the number of options in a customer's menu (Table A10). The percentile chosen decreases, which is as one would expect given most customers choose the lowest LTV option available, and the high LTV options we've added to the choice set will generally be more expensive. The likelihoods of expensive choices within and across banks are 3% and 5% respectively, versus 2% and 7% in the main results. The pattern of expensive choices through time is similar to that in the main results (Figure A6).

Expanding the choice set significantly increases the size of the bad tail (Figure A7). The average bad tail is now 8% within bank and 13% across banks, versus 2% and 8% in the main results. This is exactly as one would expect - the reason customers tend to pick low LTV products is that they come at a lower price, so expanding the choice set to include products with higher LTVs will increase price dispersion. Nonetheless, there remains a positive correlation between the quality of the choice set in a given month and frequency of expensive choices in a given month (Figure A8). The relationship within banks is somewhat weaker than in the main results. This is to be expected, as expanding the choice set to include a set of choices people rarely pick is likely to reduce the predictive power of price dispersion for expensive choices.

[30] Note that as we're changing the size of the choice set, the baseline mortgage (the 15^{th} percentile) relative to which we measure mortgage cost will also change.

Figure A9 shows the pattern of tail size by LTV and LTI for the alternative choice sets. The relationship between LTI and the size of the tail is the same as in the main results - higher-LTI customers receive menus with greater price dispersion. High-LTV customers now receive menus with lower price dispersion. This is to be expected and is largely mechanical - with the alternative choice sets low-LTV customers qualify for all mortgages with high LTVs. Given these are typically highly priced, low-LTV customers' menus have greater price dispersion - though most of them would have ignored these more expensive options.

Table A11 shows how results vary by LTV and LTI. As in the main results, customers borrowing a large amount relative to their house value and/or income are significantly more likely to make expensive choices, both within and across banks (first and third columns of Table 12). This means that the fact low-LTV borrowers have a large tail of expensive high-LTV products does not translate into an increased likelihood of making expensive choices. This would suggest that our decision to focus only on the menu of contracts at the lowest possible LTV band was a good one, as the shape of the choice set for higher LTVs is not a big driver of expensive choices. When we control for the size of the tail in columns two and four, the marginal effects of LTV and LTI are diminished, but not as dramatically as in our main results in Table 12. This again suggests that the menu variation that drives expensive choices is generally variation for mortgages with the lowest LTV for which a customer qualifies.

Table A12 shows the likelihood of expensive choices by demographic. The results are similar to the main results in Table 13 - young people and first-time buyers are slightly more likely to make expensive choices, and this is largely driven by the quality of the choice set they receive.

Table A1: Raw data

Year	Number of observations	Percentage of sample
2009	580,431	16
2010	531,273	15
2011	592,492	17
2012	587,846	17
2013	602,417	17
2014	639,672	18
Total	3,534,131	100

Note: This table summarizes the raw Product Sales Data.

Table A2: After merging with Moneyfacts

Year	Observations	%	Product	Observations	%
2009	345,746	13	2yr FRM	1,054,041	41
2010	335,026	13	3yr FRM	348,091	13
2011	417,292	16	5yr FRM	420,955	16
2012	455,004	18	2yr ARM	218,991	8
2013	493,116	19	SVR	76,360	3
2014	548,003	21	OTHER	475,749	18
Total	2,594,187	100	Total	2,594,187	100

Note: This table summarizes the dataset after merging the Product Sales Data with the Moneyfacts data.

Table A3: Top 4 products after dropping outliers and missing data

	FTB	%	NFTB	%	Total	%
By Year						
2009	64,862	11	177,858	12	242,720	12
2010	67,919	12	170,258	12	238,177	12
2011	85,259	15	219,120	15	304,379	15
2012	99,669	17	244,459	17	344,128	17
2013	121,579	21	294,940	20	416,519	21
2014	145,884	25	339,921	23	485,805	24
By Product						
2yr ARM	45,622	8	171,901	12	217,523	11
2yr FRM	305,803	52	742,986	51	1,048,789	52
3yr FRM	125,472	21	221,110	15	346,582	17
5yr FRM	108,275	19	310,559	21	418,834	21
Total	585,172	100	1,446,556	100	2,031,728	100

Note: This table summarizes the dataset after removing outliers and missing data, and retaining only 2 year, 3 year and 5 year fixed rate as well as 2 year adjustable rate mortgages from the merged PSD-Moneyfacts dataset.

Table A4: After restricting to 6 lenders

	FTB	%	NFTB	%	Total	%
By Year						
2009	54,622	12	136,351	12	190,973	12
2010	54,021	12	136,572	12	190,593	12
2011	68,686	15	181,616	16	250,302	16
2012	79,541	17	194,573	17	274,114	17
2013	99,229	21	224,330	20	323,559	20
2014	110,424	24	245,080	22	355,504	22
By Product						
2yr ARM	38,356	8	131,993	12	170,349	11
2yr FRM	257,836	55	598,572	54	856,408	54
3yr FRM	88,986	19	161,636	14	250,622	16
5yr FRM	81,345	17	226,321	20	307,666	19
Total	466,523	100	1,118,522	100	1,585,045	100

Note: This table summarizes the dataset after removing outliers and missing data, and retaining only 2 year, 3 year and 5 year fixed rate as well as 2 year adjustable rate mortgages from the merged PSD-Moneyfacts dataset, but restricting to the six lenders in our final sample.

Table A5: Final dataset

	FTB	%	NFTB	%	Total	%
By Year						
2009	39,150	11	52,901	10	92,051	10
2010	38,320	11	56,431	11	94,751	11
2011	54,158	15	88,335	17	142,493	16
2012	62,790	17	95,770	18	158,560	18
2013	79,946	22	115,750	22	195,696	22
2014	86,576	24	124,774	23	211,350	24
By Product						
2yr ARM	26,259	7	60,729	11	86,988	10
2yr FRM	209,171	58	310,905	58	520,076	58
3yr FRM	70,032	19	80,995	15	151,027	17
5yr FRM	55,478	15	81,332	15	136,810	15
Total (within)	360,940	100	533,961	100	894,901	100
Total (across)	357,044	100	526,415	100	883,459	100

Note: This table summarizes our final dataset after restricting the sample to mortgages with LTV between 65% and 95% of value less than £1mn, and removing any borrowers who had fewer than 5 options to choose from at their bank. For across-bank analysis we further remove any customers who had fewer than 15 options across banks or who selected a 2yr FRM and had fewer than 25 options.

Table A6: Probit regressions of expensive choice on choices of price components: robustness to refinancing

	Expensive choice within	Expensive choice across
	(1)	(2)
Low Fee	-0.002^{***}	-0.006^{***}
	(0.0001)	(0.0002)
High Fee	-0.001^{***}	0.002^{***}
	(0.0001)	(0.0002)
Low Inital Rate	-0.018^{***}	-0.036^{***}
	(0.0004)	(0.0001)
High Initial Rate	0.020^{***}	0.046^{***}
	(0.0002)	(0.001)
Bad tail	0.094^{***}	0.088^{***}
	(0.001)	(0.001)
Bank dummies	Yes	No
Product dummies	Yes	Yes
Pseudo R-squared	0.85	0.86
Mean dependent variable	0.019	0.036
Observations	894,901	883,459

Note: This table reports average marginal effects from probit regressions with the dependent variables shown at the top of each column. We take each unique value of the three elements of a mortgage - the initial rate, the fee and the reset rate - in a customer's choice set and order them from low to high. Choices above the 85^{th} percentile are high, whilst choices below the 15^{th} percentile are low. An expensive choice is defined as a choice that costs a customer at least 2.5% of their income more than the 15^{th} percentile in their choice set. The variable *bad tail* measures the percentage of a customer's choice set that would represent an expensive choice. The first column takes the choice set to consist only of mortgages on offer at the bank that granted the customer their mortgage, whilst the second includes all banks. Dummy variables are defined as in Table 4. *, ** and *** indicate that the coefficient is different from zero at the 10, 5 and 1 percent level of significance respectively. In contrast to the main analysis, the cost of a mortgage is computed assuming all customers refinance their mortgage after its initial period ends.

Table A7: Expensive choices and loan characteristics: robustness to refinancing

	Expensive choice within		Expensive choice across	
	(1)	(2)	(3)	(4)
High LTV & High LTI	0.094***	0.001**	0.112***	0.001***
	(0.002)	(0.0003)	(0.002)	(0.0003)
High LTV & Low LTI	0.043***	0.002***	0.060***	−0.0001
	(0.001)	(0.0002)	(0.001)	(0.0002)
Low LTV & High LTI	0.020***	0.003***	0.024***	0.004***
	(0.001)	(0.0003)	(0.001)	(0.0003)
Bad tail		0.086***		0.096***
		(0.001)		(0.001)
Bank dummies	Yes	Yes	No	No
Product dummies	Yes	Yes	Yes	Yes
Pseudo R-squared	0.43	0.74	0.3	0.78
Mean dependent variable	0.019	0.019	0.036	0.036
Observations	894,901	894,901	883,459	883,459

Note: This table reports average partial effects from probit regressions with the dependent variables shown at the top of each column. High (low) LTI customers have loan-to-income above (below) 4. High (low) LTV customers have loan-to-value above (below) 85%. The *bad tail* measures the percentage of a customer's choice set that would represent an expensive choice, where expensive choices are defined in Section 3. The first 2 columns take the choice set to consist only of mortgages on offer at the bank that granted the customer their mortgage, whilst the final 2 columns include all banks. Dummy variables are defined as in Table 4. *, ** and *** indicate that the coefficient is different from zero at the 10, 5 and 1 percent level of significance respectively. In contrast to the main analysis, the cost of a mortgage is computed assuming all customers refinance their mortgage after its initial period ends.

Table A8: Expensive choices and borrower characteristics: robustness to refinancing

	Expensive choice within		Expensive choice across	
	(1)	(2)	(3)	(4)
Young	0.006***	0.001***	0.012***	0.003***
	(0.0003)	(0.0002)	(0.0005)	(0.0002)
Old	−0.008***	−0.0005*	−0.015***	0.0002
	(0.0004)	(0.0003)	(0.001)	(0.0004)
First-time buyer	0.006***	0.001***	0.007***	−0.001***
	(0.0003)	(0.0002)	(0.0004)	(0.0002)
Poor	0.001	0.001***	0.001**	0.002***
	(0.0003)	(0.0002)	(0.0005)	(0.0002)
Rich	0.0002	−0.001***	−0.002***	−0.002***
	(0.0003)	(0.0002)	(0.0004)	(0.0002)
Bad tail		0.087***		0.095***
		(0.001)		(0.001)
Bank dummies	Yes	Yes	No	No
Product dummies	Yes	Yes	Yes	Yes
Pseudo R-squared	0.36	0.74	0.24	0.78
Mean dependent variable	0.019	0.019	0.036	0.036
Observations	894,901	894,901	883,459	883,459

Note: This table reports average partial effects from probit regressions with the dependent variables shown at the top of each column. Young customers are under 30. Old customers are over 45. Poor customers have net income in the lower tertile whilst rich customers have net income in the upper tertile. The *bad tail* measures the percentage of a customer's choice set that would represent an expensive choice, where expensive choices are defined in Section 3. The first 2 columns take the choice set to consist only of mortgages on offer at the bank that granted the customer their mortgage, whilst the final 2 columns include all banks. Dummy variables are defined as in Table 4. *, ** and *** indicate that the coefficient is different from zero at the 10, 5 and 1 percent level of significance respectively. In contrast to the main analysis, the cost of a mortgage is computed assuming all customers refinance their mortgage after its initial period ends.

Table A9: Probit regressions of strongly dominated choice on choices of price components

	Strongly dominated choice within	Strongly dominated choice across
	(1)	(2)
Low Fee	-0.028^{***}	-0.140^{***}
	(0.001)	(0.001)
High Fee	0.017^{***}	0.155^{***}
	(0.001)	(0.002)
Low Inital Rate	-0.030^{***}	-0.101^{***}
	(0.001)	(0.001)
High Initial Rate	0.064^{***}	0.130^{***}
	(0.001)	(0.002)
Low Reset Rate	-0.069^{***}	-0.091^{***}
	(0.001)	(0.001)
High Reset Rate	-0.004^{***}	0.185^{***}
	(0.001)	(0.001)
Strongly dominated tail	0.384^{***}	0.868^{***}
	(0.001)	(0.002)
Bank dummies	Yes	No
Product dummies	Yes	Yes
Pseudo R-squared	0.42	0.25
Mean dependent variable	0.081	0.277
Observations	894,901	883,459

Note: This table reports average marginal effects from probit regressions with the dependent variables shown at the top of each column. We take each unique value of the three elements of a mortgage - the initial rate, the fee and the reset rate - in a customer's choice set and order them from low to high. Choices above the 85^{th} percentile are high, whilst choices below the 15^{th} percentile are low. The *strongly dominated tail* computes the percentage of a customer's choice set that is strongly dominated by the cheapest option in their choice set, where strong domination is defined in Section 3. The first column takes the choice set to consist only of mortgages on offer at the bank that granted the customer their mortgage, whilst the second includes all banks. Dummy variables are defined as in Table 4. *, ** and *** indicate that the coefficient is different from zero at the 10, 5 and 1 percent level of significance respectively.

Table A10: Characteristics of choice sets and choices made: alternative choice set

	Within		Across	
	Choice set size	Pctile chosen	Choice set size	Pctile chosen
25^{th} pctile	22	16	107	15
Median	37	30	176	30
75^{th} pctile	56	50	272	53

Note: This table summarizes customers' choice sets and choices made. Within-bank figures restrict a customer's choice set to mortgages on offer at the bank that granted their mortgage. Across-bank figures include mortgages on offer at all banks in a customer's choice set. For each customer we rank the mortgages in their comparison set from cheapest to most expensive. The *Pctile chosen* equals the rank of their choice as a percentage of the number of mortgages in their comparison set. The table summarizes the distribution of this variable and the size of the choice set across the sample. In contrast to the main analysis, a customer's choice set is taken to include mortgages with maximum LTV *greater than or equal to* the customer's LTV.

Table A11: Expensive choices by loan characteristic: alternative choice set

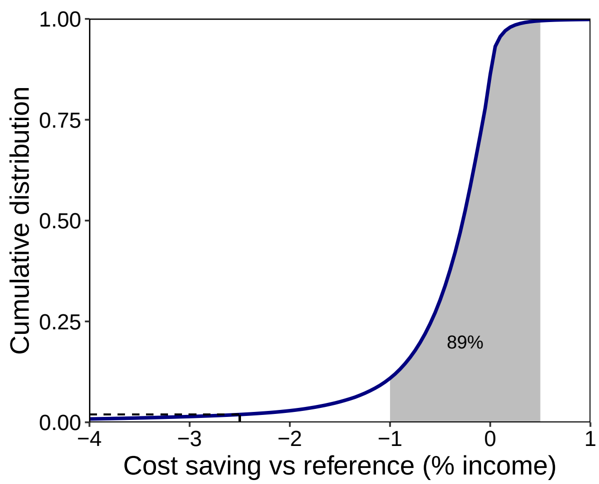
	Expensive choice within		Expensive choice across	
	(1)	(2)	(3)	(4)
High LTV & High LTI	0.065***	0.014***	0.207***	0.040***
	(0.001)	(0.001)	(0.002)	(0.001)
High LTV & Low LTI	0.031***	0.020***	0.063***	0.031***
	(0.001)	(0.0005)	(0.001)	(0.001)
Low LTV & High LTI	0.020***	−0.004***	0.053***	−0.016***
	(0.001)	(0.0002)	(0.001)	(0.0004)
Bad tail		0.140***		0.295***
		(0.001)		(0.001)
Bank dummies	Yes	Yes	No	No
Product dummies	Yes	Yes	Yes	Yes
Pseudo R-squared	0.43	0.74	0.14	0.48
Mean dependent variable	0.032	0.032	0.046	0.046
Observations	930,849	930,849	927,860	927,860

Note: This table reports average partial effects from probit regressions with the dependent variables shown at the top of each column. High (low) LTI customers have loan-to-income above (below) 4. High (low) LTV customers have loan-to-value above (below) 85%. The *bad tail* measures the percentage of a customer's choice set that would represent an expensive choice, where expensive choices are defined in Section 3. The first 2 columns take the choice set to consist only of mortgages on offer at the bank that granted the customer their mortgage, whilst the final 2 columns include all banks. Dummy variables are defined as in Table 4. *, ** and *** indicate that the coefficient is different from zero at the 10, 5 and 1 percent level of significance respectively. In contrast to the main analysis, a customer's choice set is taken to include mortgages with maximum LTV *greater than or equal to* the customer's LTV.

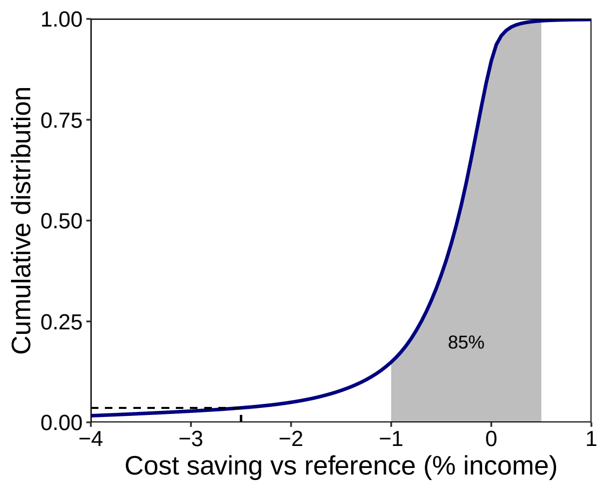
Table A12: Expensive choices by demographic: alternative choice set

	Expensive choice within		Expensive choice across	
	(1)	(2)	(3)	(4)
Young	0.006^{***}	0.003^{***}	0.016^{***}	0.010^{***}
	(0.0004)	(0.0003)	(0.001)	(0.0004)
Old	-0.005^{***}	0.00002	-0.023^{***}	-0.005^{***}
	(0.0005)	(0.0004)	(0.001)	(0.001)
First-time buyer	0.003^{***}	0.004^{***}	0.005^{***}	0.002^{***}
	(0.0004)	(0.0003)	(0.0005)	(0.0004)
Poor	0.004^{***}	-0.001^{***}	0.001^{*}	-0.004^{***}
	(0.0004)	(0.0003)	(0.001)	(0.0004)
Rich	-0.002^{***}	0.0003	-0.004^{***}	-0.003^{***}
	(0.0004)	(0.0003)	(0.0005)	(0.0004)
Bad tail		0.150^{***}		0.313^{***}
		(0.001)		(0.001)
Bank dummies	Yes	Yes	No	No
Product dummies	Yes	Yes	Yes	Yes
Pseudo R-squared	0.41	0.72	0.09	0.45
Mean dependent variable	0.032	0.032	0.046	0.046
Observations	930,849	930,849	927,860	927,860

Note: This table reports average partial effects from probit regressions with the dependent variables shown at the top of each column. Young customers are under 30. Old customers are over 45. Poor customers have net income in the lower tertile whilst rich customers have net income in the upper tertile. The *bad tail* measures the percentage of a customer's choice set that would represent an expensive choice, where expensive choices are defined in Section 3. The first 2 columns take the choice set to consist only of mortgages on offer at the bank that granted the customer their mortgage, whilst the final 2 columns include all banks. Dummy variables are defined as in Table 4. *, ** and *** indicate that the coefficient is different from zero at the 10, 5 and 1 percent level of significance respectively. In contrast to the main analysis, a customer's choice set is taken to include mortgages with maximum LTV *greater than or equal to* the customer's LTV.



(a) Within each bank's choice set



(b) Across all banks' choice sets

Figure A1: Cost savings of chosen mortgage vs reference (% of net income): robustness to refinancing

Note: These figures plot the distribution of the amount a customer saves relative to a reference mortgage as a percentage of their income, at the bank where the customer shopped and across banks. We first compute the present value of the mortgage that a customer chooses using equation (1), subtract it from the cost of the 15$^{\text{th}}$ percentile mortgage in a customer's choice set (where mortgages are ordered from cheapest to most expensive), and divide by the customer's net income. The figures plot the cumulative distribution of this figure across all customers, where the choice set consists only of mortgages on offer at the bank that gave them their mortgage (left) and of mortgages on offer across banks (right). The shaded areas show the fraction of each sample that fall between savings of 0.5% of net income and a cost of 1% of net income. The dotted lines show where the choice costs more than 2.5% of net income relative to the reference mortgage. In contrast to the main analysis, the cost of a mortgage is computed assuming all customers refinance their mortgage after its initial period ends.

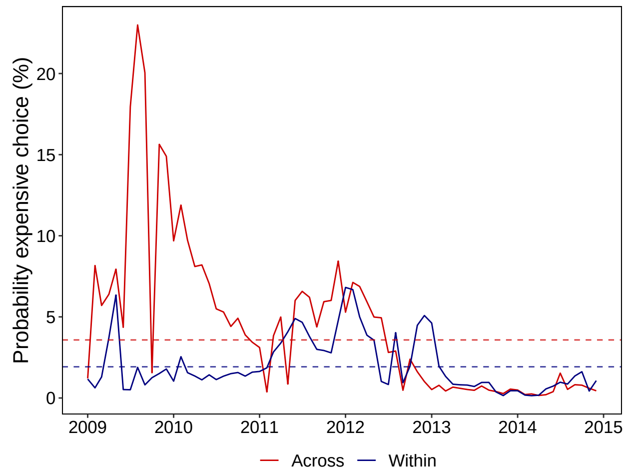
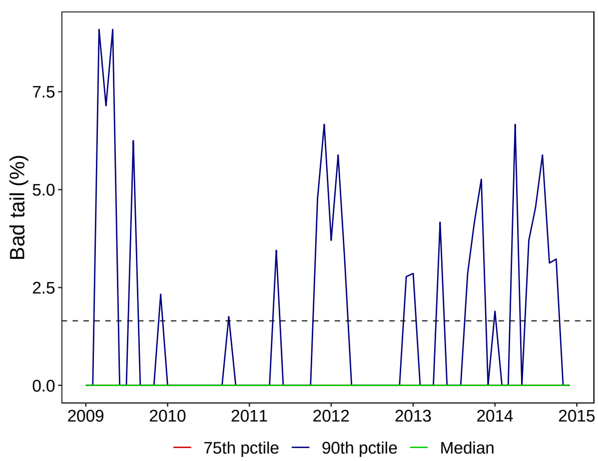
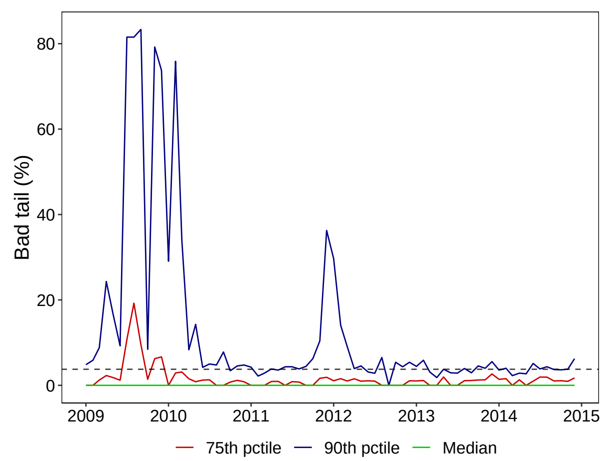


Figure A2: Expensive choices through time: robustness to refinancing

Note: This figure plots the percentage of customers that make expensive choices each month, at the bank where they shopped and across banks. Expensive choices are defined in Section 3. The blue line takes the choice set to consist only of mortgages on offer at the bank that gave the customer their mortgage, and the red line includes mortgages on offer across all banks. The horizontal lines plot the means over the sample period for each comparison. In contrast to the main analysis, the cost of a mortgage is computed assuming all customers refinance their mortgage after its initial period ends.



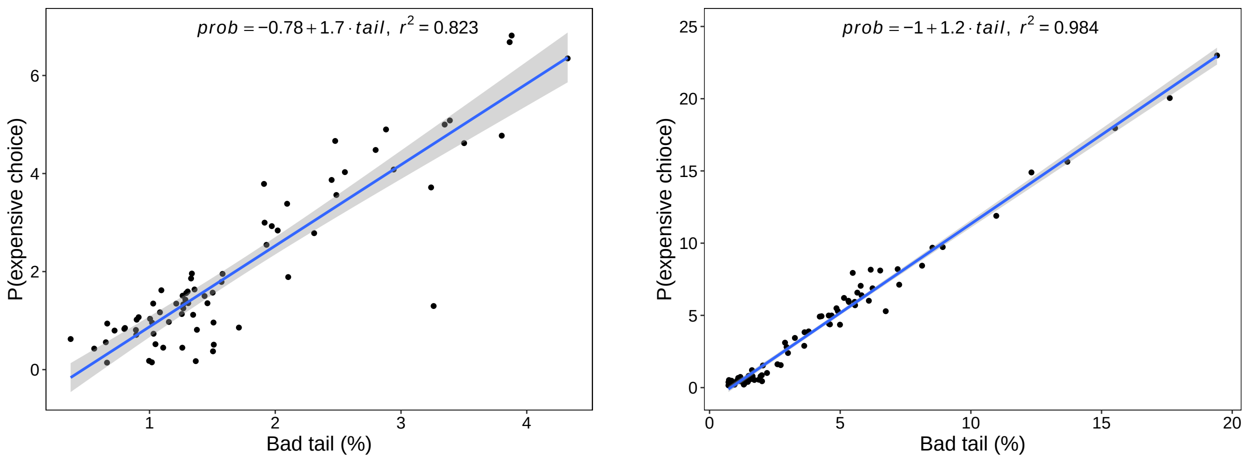
(a) Within each bank's choice set



(b) Across all banks' choice sets

Figure A3: Bad tail through time: robustness to refinancing

Note: These figures summarize the distribution of the fraction of a customer's choice set that would represent an expensive choice if chosen, both within and across banks, through time. The *bad tail* computes the fraction of the mortgages in a customer's choice set that would represent an expensive choice, where an expensive choice is defined in Section 3. The figures plot percentiles of the distribution of *bad tail* through time for the median (green) 75$^{\text{th}}$ percentile (red) and 90$^{\text{th}}$ percentile (blue), where the choice set consists only of mortgages on offer at the bank that gave the customer their mortgage (left panel) and of mortgages on offer across banks (right panel). The horizontal lines plot the means over the sample period. In contrast to the main analysis, the cost of a mortgage is computed assuming all customers refinance their mortgage after its initial period ends.



(a) Within each bank's choice set

(b) Across all banks' choice sets

Figure A4: Expensive choices and bad tails: robustness to refinancing

Note: These figures summarize the relationship between the frequency of expensive choices and the average quality of customers' choice sets in a given month. The *bad tail* computes the fraction of the mortgages in a customer's choice set that would represent an expensive choice, where an expensive choice is defined in Section 3. The figures plot the percentage of customers that make expensive choices in a month against the average size of *bad tail* in that month, where the choice set consists only of mortgages on offer at the bank that gave the customer their mortgage (left panel) and of mortgages on offer across banks (right panel). The blue line shows a linear regression of the probability of an expensive choice on the size of *bad tail*, with equation displayed in each panel. The shaded area represents the 95% confidence interval. In contrast to the main analysis, the cost of a mortgage is computed assuming all customers refinance their mortgage after its initial period ends.

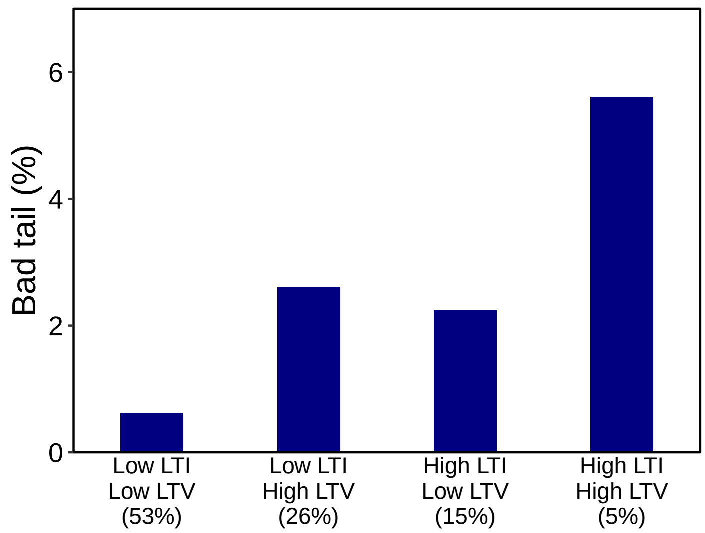


Figure A5: Bad tail by loan-to-value and loan-to-income ratio: robustness to refinancing

Note: This figure summarizes the average quality of customers' choice sets by combinations of their LTV and LTI ratios. The *bad tail* computes the fraction of the mortgages in a customer's choice set that would represent an expensive choice, where expensive choices are defined in Section 3. This figure plots the average of *bad tail* according to a customer's LTV and LTI. High LTV is defined as LTV$>$ 85%, and low LTV as LTV$<$ 85%. High LTI is defined as LTI$>$ 4, and low LTI as LTI$<$ 4. The numbers in parentheses are the percentages of the sample in each bin. In contrast to the main analysis, the cost of a mortgage is computed assuming all customers refinance their mortgage after its initial period ends.

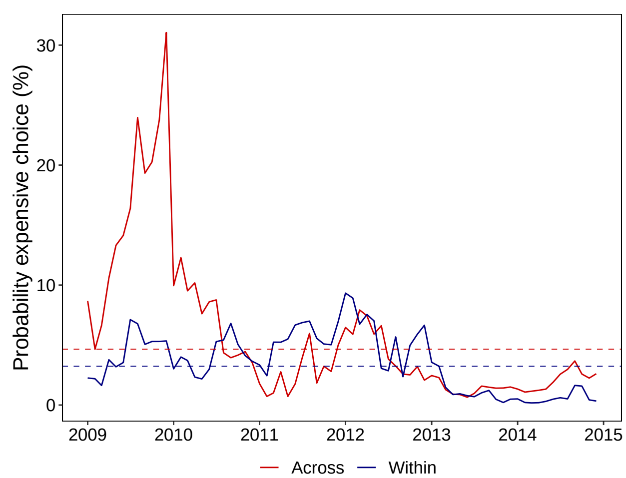
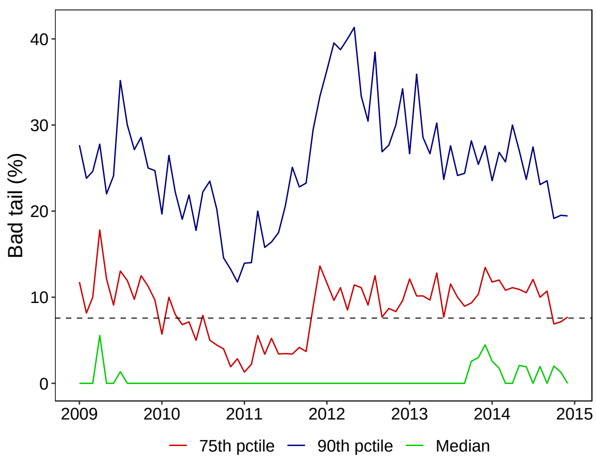
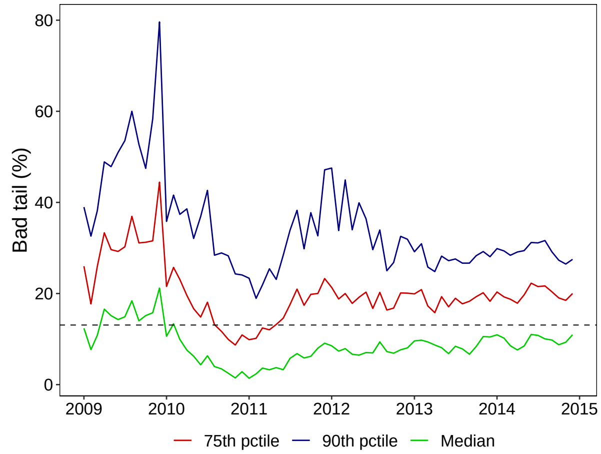


Figure A6: Expensive choices through time: alternative choice set

Note: This figure plots the percentage of customers that make expensive choices each month, at the bank where they shopped and across banks. Expensive choices are defined in Section 3. The blue line takes the choice set to consist only of mortgages on offer at the bank that gave the customer their mortgage, and the red line includes mortgages on offer across all banks. The horizontal lines plot the means over the sample period for each comparison. In contrast to the main analysis, a customer's choice set is taken to include mortgages with maximum loan-to-value *greater than or equal to* the customer's loan-to-value.



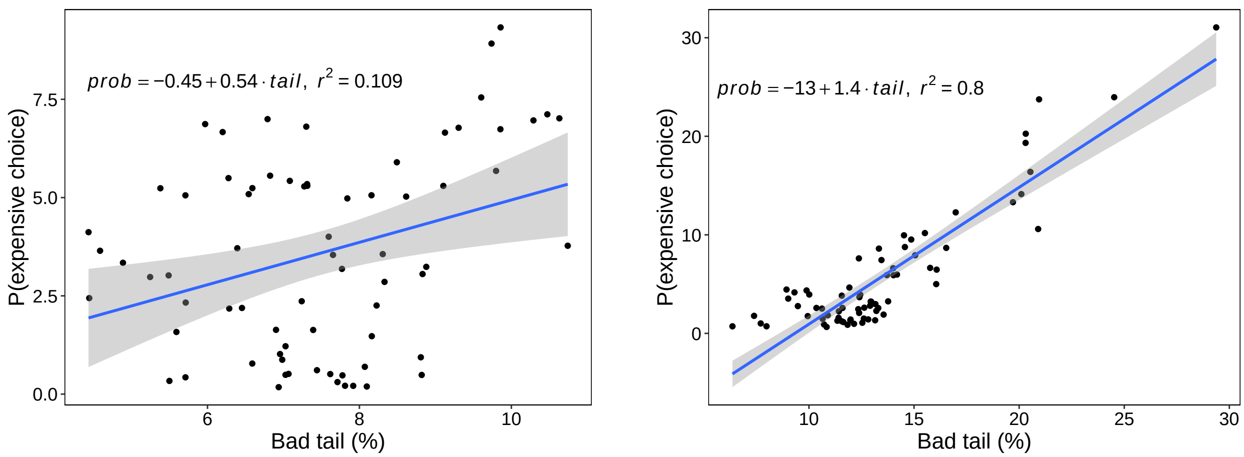
(a) Within each bank's choice set



(b) Across all banks' choice sets

Figure A7: Bad tail through time: alternative choice set

Note: These figures summarize the distribution of the fraction of a customer's choice set that would represent an expensive choice if chosen, both within and across banks, through time. The *bad tail* computes the fraction of the mortgages in a customer's choice set that would represent an expensive choice, where an expensive choice is defined in Section 3. The figures plot percentiles of the distribution of *bad tail* through time for the median (green) 75^{th} percentile (red) and 90^{th} percentile (blue), where the choice set consists only of mortgages on offer at the bank that gave the customer their mortgage (left panel) and of mortgages on offer across banks (right panel). The horizontal lines plot the means over the sample period In contrast to the main analysis, a customer's choice set is taken to include mortgages with maximum LTV *greater than or equal to* the customer's LTV.



(a) Within each bank's choice set

(b) Across all banks' choice sets

Figure A8: Expensive choices and bad tails: alternative choice set

Note: These figures summarize the relationship between the frequency of expensive choices and the average quality of customers' choice sets in a given month. The *bad tail* computes the fraction of the mortgages in a customer's choice set that would represent an expensive choice, where an expensive choice is defined in Section 3. The figures plot the percentage of customers that make expensive choices in a month against the average size of *bad tail* in that month, where the choice set consists only of mortgages on offer at the bank that gave the customer their mortgage (left panel) and of mortgages on offer across banks (right panel). The blue line shows a linear regression of the probability of an expensive choice on the size of *bad tail*, with equation displayed in each panel. The shaded area represents the 95% confidence interval. In contrast to the main analysis, a customer's choice set is taken to include mortgages with maximum LTV *greater than or equal to* the customer's LTV.

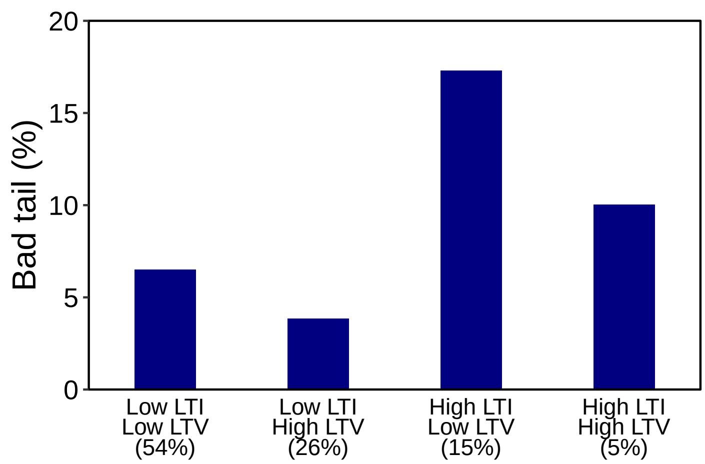


Figure A9: Bad tail by loan-to-value and loan-to-income ratio: alternative choice set

Note: This figure summarizes the average quality of customers' choice sets by combinations of their LTV and LTI ratios. The *bad tail* computes the fraction of the mortgages in a customer's choice set that would represent an expensive choice, where expensive choices are defined in Section 3. This figure plots the average of *bad tail* according to a customer's LTV and LTI. High LTV is defined as LTV$>85\%$, and low LTV as LTV$<85\%$. High LTI is defined as LTI>4, and low LTI as LTI<4. The numbers in parentheses are the percentages of the sample in each bin. In contrast to the main analysis, a customer's choice set is taken to include mortgages with maximum LTV *greater than or equal to* the customer's LTV.