Price Discrimination and Mortgage Choice*

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

July 2022

Abstract

We characterize the large number of mortgage offers for which people qualify. Almost no one picks the cheapest option, nonetheless the one selected is not usually much more expensive. A few borrowers make very expensive choices. These expensive choices are most common when the menu they face has many expensive options, and are most likely for high loan-to-value and loan-to-income borrowers. Young people and first-time buyers are more prone to making expensive choices. The dispersion in the mortgage menu is consistent with banks attempting to price discriminate for some borrowers who might pick poorly while competing for others who might shop more effectively.

Keywords: Price discrimination, consumer choice, mortgages.

JEL Codes: G21, G51, G53, D12.

*We thank João Cocco, Patrick Coen, Robert Hall, Guido Menzio, David Miles, Daniel Paravisini, Morten Ravn, 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, and the London School of Economics 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.

[†]London School of Economics and Bank of England, w.j.coen@lse.ac.uk.

[‡]University of Chicago, Booth School of Business, National Bureau of Economic Research, Centre for Economic Policy Research, and Bank of England, 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 that 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. While customers rarely pick the cheapest products on offer, the cost implications are typically relatively small. Those that pick particularly expensive mortgages generally do so because they face poor menus, with large price dispersion and lots of expensive choices.

What drives these patterns of customer choices and price dispersion? 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 are more likely to be buying a house for the first time. Lenders thus price discriminate, offering menus with greater price dispersion to customers who may be less able to identify and avoid expensive options, or have fewer options to go elsewhere.

These patterns are consistent with the following view of the trade-off facing lenders competing for customers. On the one hand, they want to offer cheap mortgages to entice sophisticated customers who might comparison shop to borrow from their bank. To attract these customers they need to have some competitively priced options available. On the other hand, they 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 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 borrowers. In some cases banks suspect that the customers will not be in a position to shop around and hence 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 a costly mortgage.

Our paper contributes to a literature studying how customers choose complex financial products, and how this affects firms' supply of these products. Liu (2019) shows that customers are less cost-sensitive to products with fees than products without, and that lenders respond to cost shocks by adjusting fees but not rates. This is consistent with fees being a relatively non-salient price characteristic which lenders adjust to make profits. Iscenko

(2020) documents that 30% of UK mortgage customers pick mortgages that have higher prices in all price dimensions than other mortgages with the same non-price characteristics. Mysliwski and Rostom (2020) document significant mortgage price dispersion in the UK. They 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. Finally, Woodward and Hall (2012) find that customers would make large savings if they consulted more brokers.[1]

Our contribution is to set out a new channel by which firms take advantage of customers' imperfect choices. Firms adjust the extent of the price dispersion in the menu they offer across products in such a way that customers that one might expect to be more constrained or less financially savvy receive menus with greater price dispersion. In doing so, they induce these customers to choose expensive mortgages while not necessarily losing sales to customers who are, in fact, sophisticated or can shop elsewhere.

The closest paper 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 to theirs. Because their dataset lacks information on the identity of lenders, they have no concept of 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 - the 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 customers' ability to pick well from a menu from whether or not they borrowed at a bank with mostly cheap options. We find a role both for picking badly and for shopping at an expensive bank.[2]

We contribute to a second 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

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

[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 poor. Put differently, for people with very different incomes, we do not suppose the same level difference in the cost of a mortgage, e.g. £50 per month, is equivalent.

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.

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 set out the relevant institutional features of the UK mortgage market. In Section 3 we explain how we characterise 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 what makes 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, a loan-level administrative dataset capturing all newly issued mortgages in the UK. The data contain information that would be recorded by a financial institution at the time of taking out a mortgage. This includes information on the borrower characteristics, such as income and age; information on the property, such as its postcode and dwelling value; and loan details, such as the amount borrowed, the interest rate in the promotional period (initial rate), the length of the initial period, the mortgage term, and the issuing institution.

We merge these data with a secondary data source, MoneyFacts, which records the set of mortgages on offer in the UK at any given time. This enables us to construct the set of mortgages that were on offer to customers when they shopped, and to fill in two elements of the mortgage that are missing from the PSD in our sample period: upfront fees paid by the borrower, and the Standard Variable Rate (SVR) which the mortgage product resets to once the initial promotional period expires.[3] We match the two datasets together using a

[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) and Bracke and Tenreyro (2021). More recent vintages of

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 during our sample amortize over a period of 25 years. 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. A different rate might apply if a different fee were paid, or if the LTV limit differed, or if the amount borrowed was capped at a different level. The combinations of different fees, loan amounts and LTV limits means that most people qualify for many different mortgages, and these mortgages have different cost implications. Table A1 in the appendix shows an example of the kind of menu a borrower might face.

After having selected a mortgage, customers pay the initial promotional rate for a set period, after which the interest rate changes. Table A6 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 19% 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 (SVR), or reset rate, which fluctuates from month to month depending on macroeconomic conditions or the bank's own idiosyncrasies, for example their funding costs.[6] The bank has 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 reset rate.

the PSD have included 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.

[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 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.

Reset rates vary across lenders and over time. Figure A1 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 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. Of course many borrowers will plan to refinance the mortgage before it resets.

Summary statistics for our data are on shown in Table A1. 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.[7] First-time buyers constitute 40% of the sample. Most of the sample is younger borrowers, no doubt in part because older ones would be re-financing mortgages that have lower LTVs than we consider.

3 The choice problem

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

To define a customer's choice set we first identify all mortgages that were on offer when they were shopping around.[8] We then restrict this set to include only mortgages with the 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 an 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.

[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.

[8] 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 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.

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. A borrower needing to borrow at 80% LTV picking a mortgage available to borrowers with LTV of 85% would almost always involve paying a higher initial interest rate. 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.[9]

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.

The across-bank menu is informative about the gains from comparison shopping, and the role of competition between banks. The majority of borrowers use brokers to assist them in picking mortgages and hence will have compared offers at multiple lenders.[10] For any given individual, however, we cannot be sure whether that person did shop at different lenders and if so, which ones were considered. As we will see, however, banks differ in the average cost of the mortgages that they offer, so even choosing reasonably well at some banks for some types of loans can lead to a more expensive choice than picking an average loan from another bank that had a more attractive menu.

Having characterized the choice set, we need a metric by which to compare mortgages. 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.[11,12]

In the calculation, we assume that the reset rate remains constant, which is the assumption that is embedded in the initial monthly payment that the borrower will be given upon

[9] 91% of borrowers pick a mortgage with the lowest possible LTV cap for which they qualify at their chosen bank. See Tables A8 and A9 for the results with the larger choice set which can be compared to Tables A11 and A12.

[10] For example, Mysliwski and Rostom (2020) show that about two-thirds of UK home buyers use brokers.

[11] 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 A6) and most mortgages amortise over 25 years, this is consistent with customers refinancing once every 8 to 9 years.

[12] Calculating the payments over the first seven years only has two benefits. First, as a practical matter most borrowers do refinance at some point. 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.

signing the contract.[13] We use the seven year LIBOR rate to discount the payments. Hence the formula 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.[14]

3.1 Customer choices

Table A2 summarizes the size of the menus and describes the cost of the choices people make. The typical 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, we define a 'baseline mortgage' against which we evaluate customers' choices. We set this baseline to be the 15$^{\text{th}}$ percentile option in their menu, where options are ordered from cheapest to most expensive.

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 25$^{\text{th}}$ and 75$^{\text{th}}$ percentile of the within-bank choice set size is 11 and 23 respectively, so this cut off means that we are using the 2$^{\text{nd}}$ or 3$^{\text{rd}}$ cheapest mortgage for most people rather than the absolutely cheapest mortgage.

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

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

We then compute the amount of money a customer saves or leaves on the table relative to this baseline as a percentage of the customer's monthly income after tax.

Figure A2 shows the distribution of customers' savings relative to the baseline 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 interpret this as the role of competition, which disciplines the banks and protects customers. The fact that customers can shop elsewhere prevents banks from offering many 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.

3.2 Expensive Choices

Nonetheless, there are some customers that do make expensive choices (Figure A2). 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%. 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 make these expensive choices.

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 baseline choice. Figure A3 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 in the sample.[15]

A customer can make an expensive choice for two reasons: choosing badly from a given menu, or when facing a menu with more poor choices making a more typical choice that is expensive. To separate these two effects, we define two variables. The first, *choice*, is simply the percentile rank of the choice the customer made, with a lower number representing a cheaper mortgage. The second, *bad tail*, 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 constant, bad tail is, by definition, 0. If

[15] From Figure A1, 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.

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.

Figure A4 shows how the qualities of both the choice and choice set combine to produce expensive choices. The chart plots the distribution of bad tail across customers. The vast majority of people - 87% - face a menu that precludes the possibility of making expensive choices at their bank; across banks, 42% face no bad choices. Competition protects these customers: no matter how poorly they pick, they can never lose much money.

As the quality of the menu deteriorates, the consequences of picking poorly increase. The shaded areas in the figure show the points in the distribution where it becomes possible to make an expensive choice. As is evident, there are some customers where the menus are littered with expensive options. For some of them, even picking the median product would represent an expensive choice.

Figure A5 plots the distribution of bad tail through time. Consistent with Figure A4, 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.

Our most remarkable finding about the time series patterns, however, is the strong influence the menu quality plays in driving expensive choices. In Figure A6, 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, suggesting that 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.

This leads to two questions worthy of consideration. First, what leads banks to price discriminate and give some customers menus with more expensive options? In other words, what determines the menu offering and who gets it? Second, for any given menu, what explains why some customers pick an expensive mortgage? We address each of these questions in turn.

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 can 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, race, or IQ), they can price discriminate by altering the characteristics of the mortgage contract presented to 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 A7 plots the average size of bad 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 (Figure A7). 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 doubles. 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. This also is somewhat expected. These borrowers would not necessarily sail through a mortgage approval 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. These customers 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 A8. 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 baseline mortgage at their chosen bank. Figure A8 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

[16]It is worth noting that just after the end of our sample, in 2015, 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) for a discussion of this policy and its consequences.

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%.[17] Second, the probability of going into arrears is very similar across higher and lower LTVs and LTIs.[18] This suggests default cannot be driving the kind of menu variation that we observe in Figure A7. Third, our measure of menu quality is relative, and thus relates to price *dispersion* - not *average* prices. While 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.

Adjusting the menu according to loan characteristics indirectly results in the menu banks offer differing by demographic characteristics. In Table A3, 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 can learn from the loan documents: whether the customer is young or old, a first-time buyer or not, and rich or poor. Table 3 reports the marginal effects that these variables have on mortgage choice.

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 statistically significant and 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, against a population average 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 are choosing the types of mortgage contracts where banks offer 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 pattern 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.[19]

[17]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.

[18]For instance, 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 A7 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 A7. We use PSD data from 2015 as this is the first available year for which arrears data is collected.

[19]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.

Table A4 shows how banks vary the menu across different customers in practice. 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. Across banks customers also face significant variation in the reset rates they face.

Table A5 relates the size of the bad tails that borrowers face to the dispersion in price. The key determinant of the size of the tail is the dispersion in initial rates across products. A customer whose initial rate dispersion is at the 75^{th} percentile of the sample distribution has a bad tail at the bank from which she borrowed that is 5 percentage points larger than a customer at the 25^{th} percentile. This difference is double the average size of bad tail in our sample. Taking Tables A4 and A5 together it is clear that banks predominantly rely on the initial rate to vary their menu offering.

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 a bad choice? Who makes bad choices? Were these bad choices driven by the menu the customer was offered, or some aspect of the choice they made?

5.1 Expensive Choice mechanisms

Table A6 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 last two 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.

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 A7 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 7 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 increases the chances of making an expensive choice, though the magnitude of this effect is modest.

Figure A9 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.[20] 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 A7, 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 expensive 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 cannot be the drivers of the main results in the paper. For instance, if we rerun the within-bank specification in Table A6 without controlling for multiple reset rates, the coefficients on the high and low fee are effectively identically to those we report in Table A6.

We now assess the drivers of expensive choices across banks. Table A8 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

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

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.

Figure A10 summarises the variation in cost that customers face within and across banks. For each bank in a customer's choice set we compute the cost of the mean and the 10$^{\text{th}}$ and 90$^{\text{th}}$ percentiles of the mortgages on offer at that bank as a percentage of the customer's income. Figure A10 plots these numbers relative to the cost of the 15$^{\text{th}}$ percentile mortgage in the customer's choice set, averaging across all customers (in red) and all those customers that face choice sets where at least 10% of the mortgages were bad choices (in blue). There is significant variation in the cost of mortgages both within and across banks. Making a bad choice entails shopping at the wrong bank *and* picking from the menu available at that bank. Notice also that at some of the banks, expensive choices are only possible by picking a mortgage that is worse than 90% of the ones for which the customer qualifies.

Tables A9 and A10 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 A9 summarises these two variables. The pattern in the table reinforces what we would expect from Figure A10: 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 banks the dispersion in cost is greater.

Table A10 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. A customer whose quality of choice of bank is at the 75$^{\text{th}}$ percentile of the distribution is 3 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 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 paying 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 A11 reports the marginal effects from regressions of making an expensive 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. 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.

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 is 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.

Table A12 reports the results of probit regressions of expensive choices on customer demographics. 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.

6 Discussion

Here is one stylized way to think about the price setting problem facing a lender. Suppose the lender assumes there are two types of customers: sophisticated customers and randomisers. Randomisers 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, because they are unaware of alternatives, or because they don't qualify for mortgages with other lenders. Because they prefer having a mortgage to not having one, they take the mortgage they randomly selected. 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 has to balance two considerations: providing cheap options in order to entice sophisticated customers to shop there, and to offer expensive options to profit from the randomisers. The menu on offer will therefore be one with price dispersion, with good options for sophisticated customers and bad options to profit from the randomiser. The more likely it is the person at their bank is a randomiser, the more they want to fill the mortgage menu with bad options.

This characterization is inspired by Menzio and Trachter (2018), who set out a model where 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.

The evidence in this paper 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 randomisers 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.

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 are constantly varying 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 five 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.

We know that there are many other circumstances where people must select between multiple outlets that are offering fairly similar products: cars, wedding venues, 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.

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*.

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.

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.

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.

Mysliwski, M. and Rostom, M. (2020). Value of information and the impact of mortgage intermediaries on lender competition and households' financial positions. Working paper.

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.

Appendices

A1 Sample formation

Our main data source is the Product Sales Database (PSD), a loan-level administrative dataset capturing all newly issued mortgaged in the UK. A typical set of choices that an individual would be presented with is shown in Table A1. As mentioned in the body of the paper, the combination of fees, loan-to-value options and initial interest rates leads to a wide set of choices for most people. The PSD contains information that would be recorded by a financial institution at the time of taking out a mortgage. In principle, this includes information on the borrower characteristics, such as income and age; information on the property, such as its post-code and value, and details about the loan such as the interest rate in the initial promotional period (initial rate), length of the promotional period, the mortgage term, and the issuing institution. However, in some cases information is missing. Table A2 shows the raw data between 2009 and 2014, after discarding observations for which no initial rate information is available.

The PSD, however, omits two crucial elements of the mortgage: upfront fees levied unto the borrower and the Standard Variable Rate (SVR) that the mortgage product resets to once the initial promotional period expires. To get this information we merge the PSD with a secondary data source, Moneyfacts, which records the fees and reset rates on all product offerings at a given point in time. We also merge in the minimum and maximum loan size available for each product. Most loans have no minimum loan requirement but a maximum loan requirement of £1million.

We match the two datasets together using a matching algorithm that uses the name of the bank, the product type, the initial interest 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. 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 A3 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.[21] 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 after these filters are applied are shown in Table A4.

Although there are many small mortgage lenders, most of the mortgage market is domi-

[21] 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.

nated by a few top players. Given our interest in contrasting choices within the set of loans made by an individual lender, we want to have enough loans every month to make meaningful comparisons. So we eliminate the peripheral lenders and focus on six large mortgage providers. Table A5 shows the loan characteristics when we concentrate on the six lenders that make the most loans in the sample. Focusing on these lenders only shifts the sample from just over 2 million loans to just under 1.6 million loans.

In our analysis, we only keep households with loan amounts less than £1million, 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 A6. 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.[22] For the across-bank analysis we further 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 decide which mortgages households are compared against, we first need a set of all possible mortgages that a household could have chosen from. We do this by creating a dataset that contains unique product-level observations, that are arranged by month, six LTV buckets (65%, 70%, 75%, 80%, 85%, 90% and 95%), the minimum and maximum loan sizes allowable for each product and the product type (2, 3, 5yr FRM and 2yr ARM). Our final sample of mortgage choices is then compared against each eligible observation in this product-level dataset.

The date observed 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 house sale. The initial rates and fees associated with these completion dates will therefore be from offers that were available 3 to 5 months ago. However, the observed reset rate will be the SVR at the time of completion, not at the time of the mortgage offer. We therefore lag the reset rate by 4 months to ensure it matches the quoted SVR at around the time the mortgage offer was made and when the borrower was shopping, 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 loan amount; 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

[22] 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 common one.

where the max LTV is 85%.[23]

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

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

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

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

[23] 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.

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%.[24]

Figures A1 to A4 and Tables A7 to A9 replicate the key charts and figures 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 A7). 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 A1).

Expanding the choice set significantly increases the size of the bad tail (Figure A2). 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 A3). 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.

Figure A4 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

[24] 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.

- 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 A8 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 11). 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 11. 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 A9 shows the likelihood of expensive choices by demographic. The results are similar to the main results in Table 12 - 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.

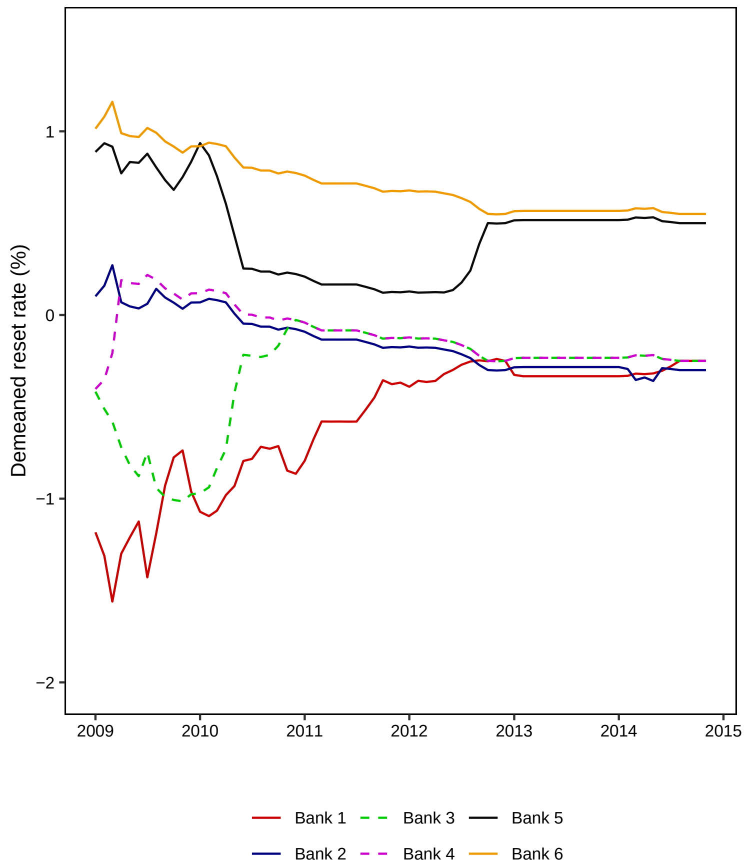
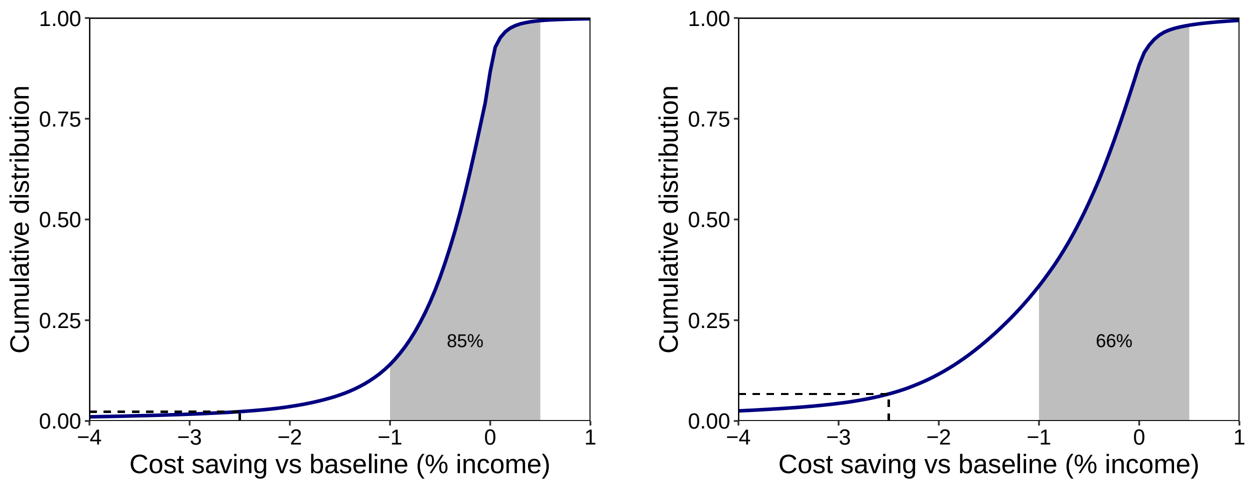


Figure A1: Across-bank variation in reset rates

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 (standard variable rate) across each of the products they offer in a given month. We then demean these by the simple average of the reset rate across banks. The demeaned series are plotted.



(a) Within each bank's choice set

(b) Across all banks' choice sets

Figure A2: Cost savings of chosen mortgage vs baseline (% of net income)

Note: These figures plot the distribution of the amount a customer saves relative to a benchmark 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). We take the difference between that cost and the cost of a baseline mortgage, taken to be the 15^{th} percentile mortgage in a customer's choice set (where mortgages are ordered from cheapest to most expensive). We divide this difference 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 panel) and of mortgages on offer across banks (right panel). 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 the portion of each sample where the choice represents a big mistake (defined as costing more than 2.5% of net income relative to the baseline mortgage).

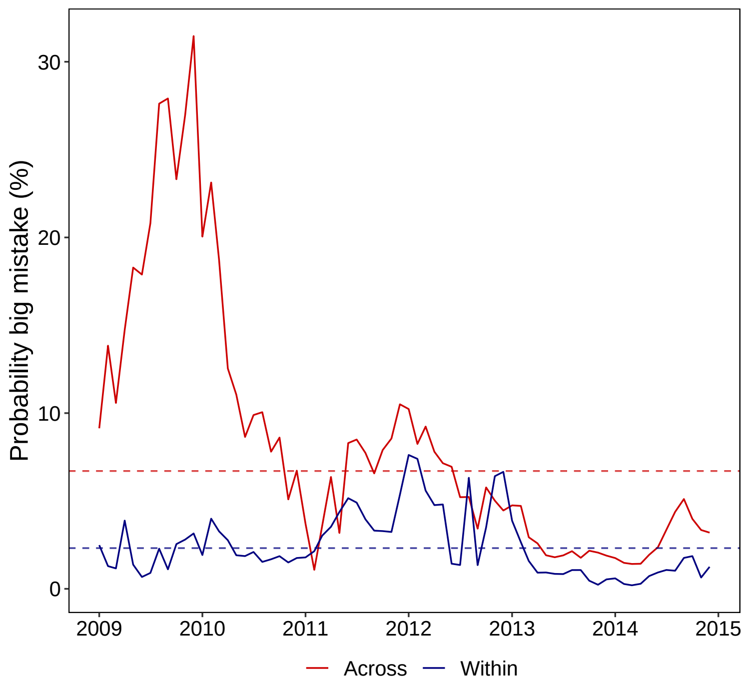
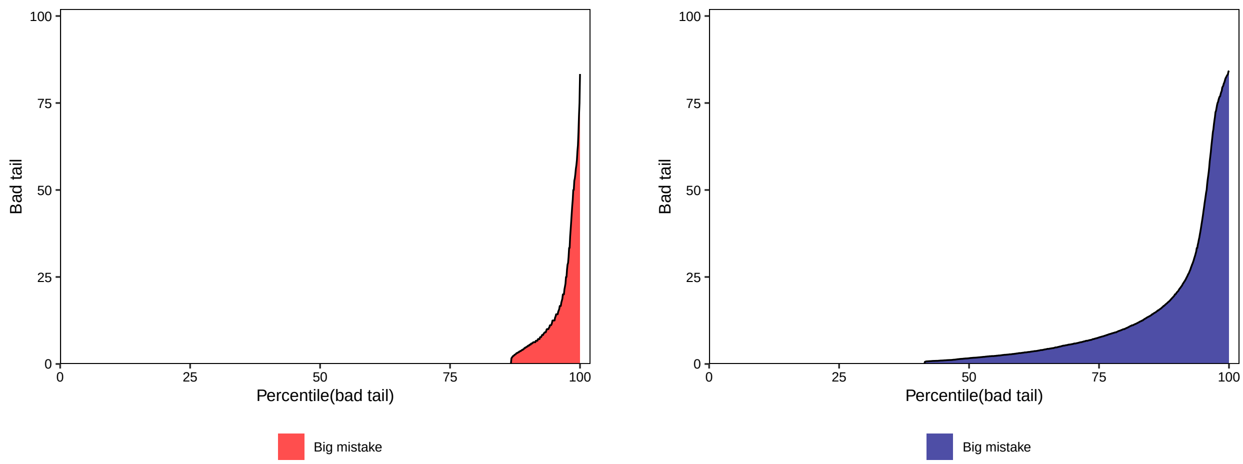


Figure A3: Big mistakes through time

Note: This figure plots the percentage of customers that make big mistakes each month, at the bank where they shopped and across banks. A customer makes a big mistake if the mortgage they choose costs them at least 2.5% of their monthly income more than the 15^{th} percentile in their choice set, where the choice set consists only of mortgages on offer at the bank that gave the customer their mortgage (blue line) and of mortgages on offer across banks (red line). The horizontal lines plot the means over the sample period for each comparison.

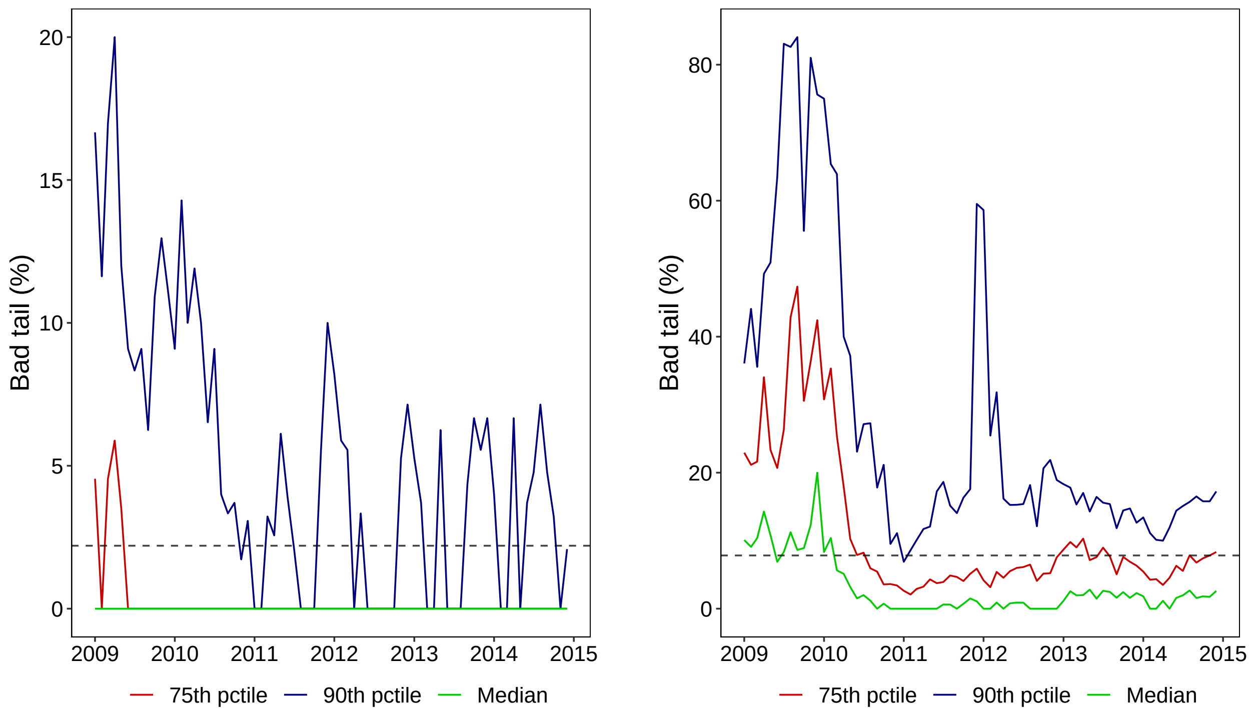


(a) Within each bank's choice set

(b) Across all banks' choice sets

Figure A4: Distributions of bad tail

Note: These figures summarize the distribution of the fraction of a customer's choice set that would represent a big mistake if chosen, both within and across banks. A mortgage represents a big mistake if choosing it would cost a customer at least 2.5% of their monthly income more than the 15^{th} percentile in their choice set. The variable *bad tail* computes the fraction of the mortgages in a customer's choice set that would represent a big mistake. The x-axis plots the percentile of bad tail across customers, and the shaded area identifies the set of choices that would represent a big mistake if chosen. The left panel takes the choice set to consist only of mortgages on offer at the bank that gave the customer their mortgage, whilst the right panel takes the choice set across banks.

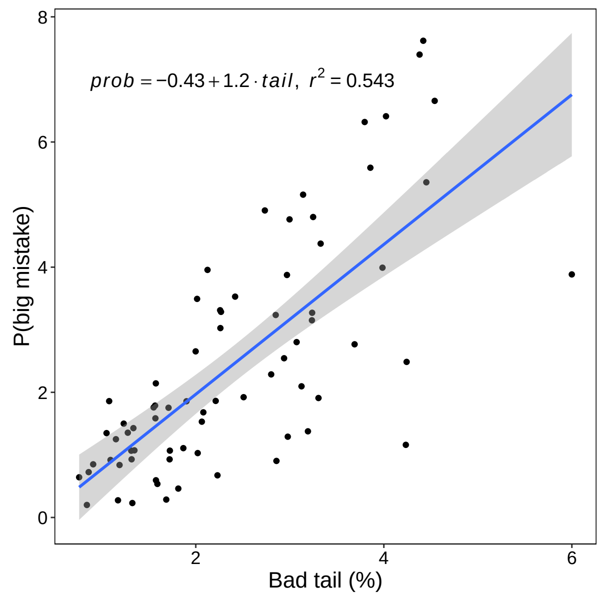


(a) Within each bank's choice set

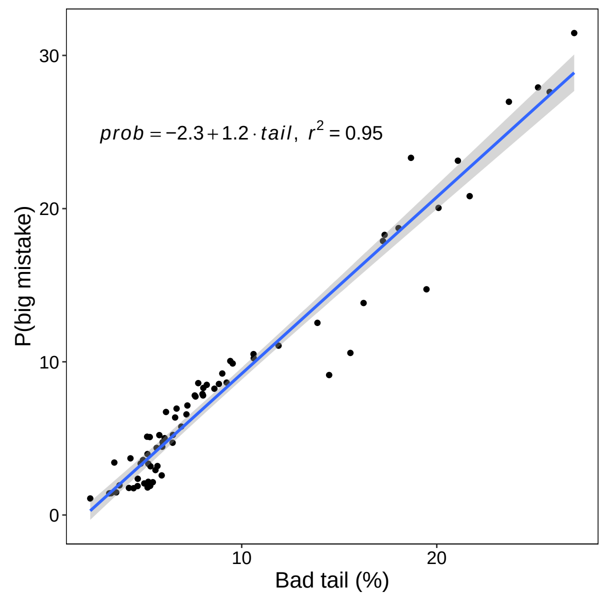
(b) Across all banks' choice sets

Figure A5: Bad tail through time

Note: These figures summarize the distribution of the fraction of a customer's choice set that would represent a big mistake if chosen, both within and across banks, through time. A mortgage represents a big mistake if choosing it would cost a customer at least 2.5% of their monthly income more than the 15^{th} percentile in their choice set. The variable *bad tail* computes the fraction of the mortgages in a customer's choice set that would represent a big mistake. 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.



(a) Within each bank's choice set



(b) Across all banks' choice sets

Figure A6: Big mistakes and bad tails

Note: These figures summarize the relationship between the frequency of big mistakes and the average quality of customers' choice sets in a given month. A mortgage represents a big mistake if choosing it would cost a customer at least 2.5% of their monthly income more than the 15^{th} percentile in their choice set. The variable *bad tail* computes the fraction of the mortgages in a customer's choice set that would represent a big mistake. The figures plot the percentage of customers that make big mistakes 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 a big mistake on the size of *bad tail*, with equation displayed in each panel. The shaded area represents the 95% confidence interval.

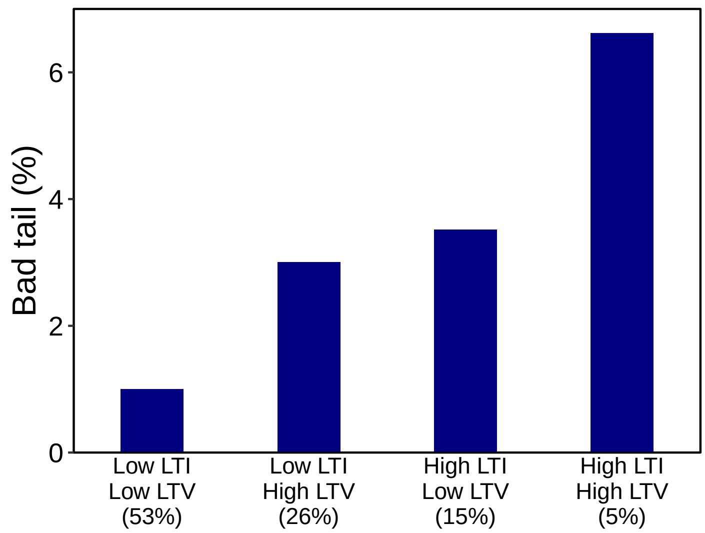


Figure A7: Bad tail by loan-to-value and loan-to-income ratio

Note: This figure summarizes the average quality of customers' choice sets by combinations of their LTV and LTI ratios. A mortgage represents a big mistake if choosing it would cost a customer at least 2.5% of their monthly income more than the 15^{th} percentile in the choice set at the bank where they shopped. The variable *bad tail* computes the fraction of the mortgages in a customer's choice set that would represent a big mistake. 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 below the x-axis are the percentages of the sample in each bin.

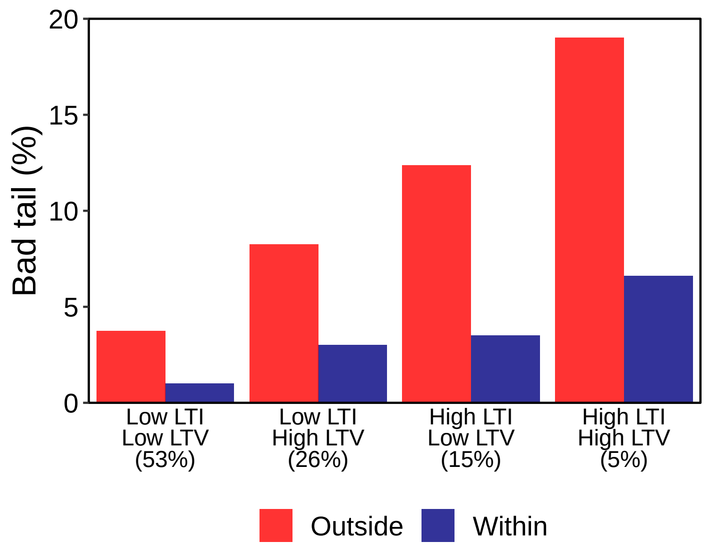


Figure A8: Tails within and outside banks by loan-to-value and loan-to-income ratio

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. A mortgage represents a big mistake if choosing it would cost a customer at least 2.5% of their monthly income more than the 15^{th} percentile in the choice set at the bank where they shopped. The variable *bad tail* computes the fraction of the mortgages in a customer's choice set that would represent a big mistake. 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 a big mistake, relative to the same baseline 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 below the x-axis are the percentages of the sample in each bin.

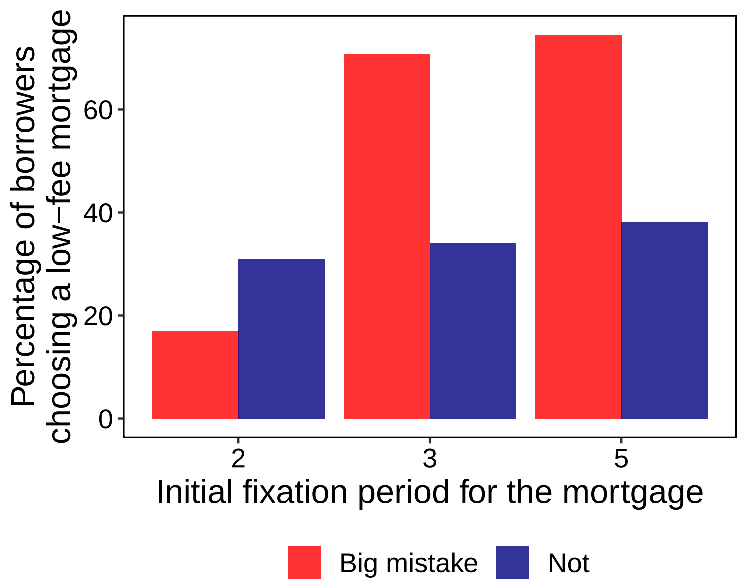


Figure A9: Big mistakes by choice of fee

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. A mortgage represents a big mistake if choosing it would cost a customer at least 2.5% of their monthly income more than the 15^{th} percentile in the choice set at the bank where they shopped.

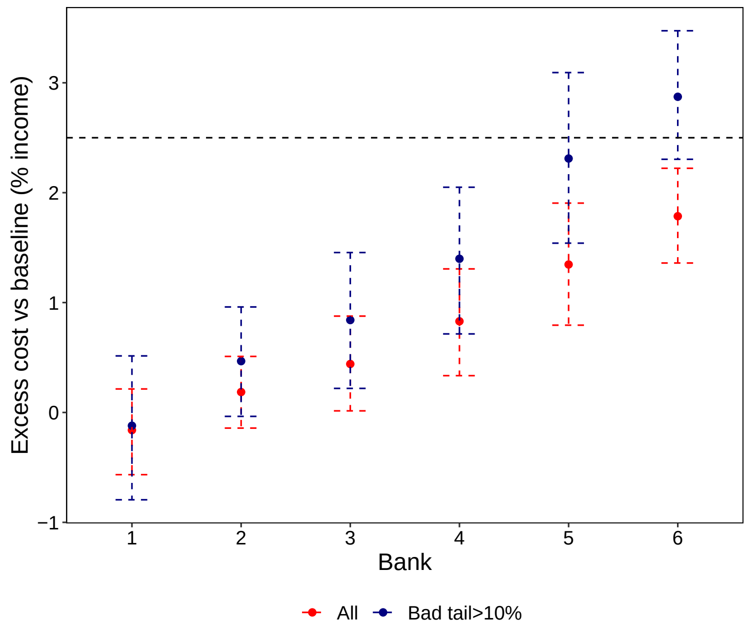


Figure A10: Decomposing price dispersion across banks

Note: This figure summarizes the quality of customers' choice sets within and across banks. For each bank in a customer's choice set we compute the mean cost of a mortgage at that bank, together with the 10^{th} and 90^{th} percentiles, as a percentage of the customer's income. We subtract from these figures the cost of the 15^{th} percentile mortgage in their choice set, and then average across all customers that qualify for mortgages at all 6 banks. We plot this for all customers (in red) and for all customers who faced choice sets where at least 10% of the mortgages represent big mistakes, defined as costing at least 2.5% of the customer's monthly income more than the 15^{th} percentile in their choice set (in blue). Choices above the dashed horizontal line represent big mistakes.

Table A1: 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 A2: 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 summarises the distribution of this variable, along with the size of the choice set, across the sample.

Table A3: Probit regressions of high loan-to-value and high loan-to-income on borrower types

	Dependent variable:		
	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 tope of each column in the table. 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 customers have income in the lower tertile of the net income distribution whilst rich customers have income in the upper tertile of this 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. *indicates the coefficient is different from zero at the 10 percent level of significance, **indicates the coefficient is different from zero at the 5 percent level of significance, ***indicates the coefficient is different from zero at the 1 percent level of significance. The pseudo R-squared is McFadden's R-squared, equal to 1 minus the ratio of the regression's log likelihood to the log likelihood of a regression with only an intercept and no covariates.

Table A4: 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 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, with fees measured in 1000s of pounds, and the reset rate and initial rate measured in percentage points. 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 A5: Determinants of the size of the bad tail for each borrower

	Dependent variable:	
	Bad tail within	Bad tail across
	(1)	(2)
Fee dispersion (£000s)	0.103***	1.960***
	(0.011)	(0.035)
Rate dispersion (pp)	7.440***	13.900***
	(0.011)	(0.014)
Reset rate dispersion (pp)	0.657***	3.250***
	(0.023)	(0.022)
Bank dummies	Yes	No
Product dummies	Yes	Yes
Mean dependent variable	2.2	7.83
R-squared	0.42	0.6
Observations	894901	883459

Note: This table reports coefficients from OLS regressions with the size of borrowers' bad 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, with fees measured in 1000s of pounds, and the reset rate and initial rate measured in percentage points. The dependent variable measures the percentage of a customer's choice set that would represent a big mistake, where a big mistake 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 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. 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. *indicates the coefficient is different from zero at the 10 percent level of significance, **indicates the coefficient is different from zero at the 5 percent level of significance, ***indicates the coefficient is different from zero at the 1 percent level of significance.

Table A6: Probit regressions of big mistakes on choices of price components

	Dependent variable:	
	Big mistake within	Big mistake 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	0.32
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 in the table. 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. A big mistake 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 a big mistake. 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. 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. *indicates the coefficient is different from zero at the 10 percent level of significance, **indicates the coefficient is different from zero at the 5 percent level of significance, ***indicates the coefficient is different from zero at the 1 percent level of significance. The pseudo R-squared is McFadden's R-squared, equal to 1 minus the ratio of the regression's log likelihood to the log likelihood of a regression with only an intercept and no covariates.

Table A7: Determinants of within-bank mistakes

	Dependent variable:	
	High rate within	Big mistake 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 in the table. 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. A big mistake 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 dependent variable in the first column is a dummy variable equal to 1 if the customer picks a product with a high initial rate. The dependent variable in the second column is a dummy variable equal to 1 if the customer makes a big mistake. The choice set consists only of mortgages on offer at the bank that granted the customer their mortgage. 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. *indicates the coefficient is different from zero at the 10 percent level of significance, **indicates the coefficient is different from zero at the 5 percent level of significance, ***indicates the coefficient is different from zero at the 1 percent level of significance. The pseudo R-squared is McFadden's R-squared, equal to 1 minus the ratio of the regression's log likelihood to the log likelihood of a regression with only an intercept and no covariates.

Table A8: Mistakes within and across banks

	Big mistake across	Not
Big mistake within	1.2	1.1
Not	5.5	92.2

Note: This table shows the distribution of mistakes 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. A big mistake is defined as a choice that costs a customer at least 2.5% of their income more than the 15$^{\text{th}}$ percentile in their choice set. The entries in the table show the percentages of the sample in each of the four possible combinations of mistakes and non-mistakes.

Table A9: Choice quality within and across banks

	25$^{\text{th}}$ pctile	Median	75$^{\text{th}}$ pctile	90$^{\text{th}}$ pctile
Cost difference vs baseline 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 variable *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 variable *cost difference within bank* measures the difference in cost between the mortgage the customer chose and the 15$^{\text{th}}$ percentile mortgage on offer at their bank. The entries in table show the values of these variables at different points (the 25$^{\text{th}}$ percentile, median, 75$^{\text{th}}$ percentile and 90$^{\text{th}}$ percentile) in the customer distribution.

Table A10: Determinants of across-bank mistakes

	Dependent variable:				
	Big mistake across				
	(1)	(2)	(3)	(4)	(5)
Cost difference vs. best bank			0.023***		0.029***
			(0.0002)		(0.0002)
Cost difference within bank				0.016***	0.023***
				(0.0002)	(0.0002)
Bad tail		0.304***	0.224***	0.279***	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 a big mistake across banks. A big mistake is defined as a choice that costs a customer at least 2.5% of their income more than the 15^{th} percentile in their across-bank choice set. The variable *bad tail* measures the percentage of a customer's across-bank choice set that would represent a big mistake. The variable *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 variable *cost difference within bank* measures the difference in cost between the mortgage the customer chose and the 15^{th} percentile mortgage on offer at their bank. 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. *indicates the coefficient is different from zero at the 10 percent level of significance, **indicates the coefficient is different from zero at the 5 percent level of significance, ***indicates the coefficient is different from zero at the 1 percent level of significance. The pseudo R-squared is McFadden's R-squared, equal to 1 minus the ratio of the regression's log likelihood to the log likelihood of a regression with only an intercept and no covariates.

Table A11: Big mistakes and loan characteristics

	Dependent variable:			
	Big mistake within		Big mistake across	
	(1)	(2)	(3)	(4)
High LTV & High LTI	0.093***	−0.002***	0.228***	0.023***
	(0.002)	(0.0003)	(0.002)	(0.001)
High LTV & Low LTI	0.044***	0.0001	0.076***	0.001
	(0.001)	(0.0002)	(0.001)	(0.0005)
Low LTV & High LTI	0.025***	0.002***	0.087***	0.008***
	(0.001)	(0.0003)	(0.001)	(0.0005)
Bad tail		0.118***		0.299***
		(0.001)		(0.001)
Bank dummies	Yes	Yes	No	No
Product dummies	Yes	Yes	Yes	Yes
Pseudo R-squared	0.35	0.69	0.15	0.57
Mean dependent variable	0.023	0.023	0.067	0.067
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 in the table. High (low) LTI customers have loan-to-income above (below) 4. High (low) LTV customers have loan-to-value above (below) 85%. A big mistake is defined as a choice that costs a customer at least 2.5% of their income more than the 15$^{\text{th}}$ percentile in their choice set. The variable *bad tail* measures the percentage of a customer's choice set that would represent a big mistake. 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. 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. *indicates the coefficient is different from zero at the 10 percent level of significance, **indicates the coefficient is different from zero at the 5 percent level of significance, ***indicates the coefficient is different from zero at the 1 percent level of significance. The pseudo R-squared is McFadden's R-squared, equal to 1 minus the ratio of the regression's log likelihood to the log likelihood of a regression with only an intercept and no covariates.

Table A12: Big mistakes and borrower characteristics

	Dependent variable:			
	Big mistake within		Big mistake across	
	(1)	(2)	(3)	(4)
Young	0.005***	0.001***	0.018***	0.005***
	(0.0004)	(0.0002)	(0.001)	(0.0004)
Old	−0.008***	−0.0003	−0.031***	−0.006***
	(0.0004)	(0.0003)	(0.001)	(0.001)
First-time buyer	0.006***	−0.0003	0.005***	−0.005***
	(0.0004)	(0.0002)	(0.001)	(0.0004)
Poor	0.0005	0.001***	0.003***	0.001**
	(0.0004)	(0.0002)	(0.001)	(0.0004)
Rich	−0.0001	−0.001***	−0.006***	−0.006***
	(0.0003)	(0.0002)	(0.001)	(0.0004)
Bad tail		0.117***		0.303***
		(0.001)		(0.001)
Bank dummies	Yes	Yes	No	No
Product dummies	Yes	Yes	Yes	Yes
Pseudo R-squared	0.3	0.69	0.09	0.56
Mean dependent variable	0.023	0.023	0.067	0.067
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 in the table. 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. A big mistake 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 a big mistake. 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. 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. *indicates the coefficient is different from zero at the 10 percent level of significance, **indicates the coefficient is different from zero at the 5 percent level of significance, ***indicates the coefficient is different from zero at the 1 percent level of significance. The pseudo R-squared is McFadden's R-squared, equal to 1 minus the ratio of the regression's log likelihood to the log likelihood of a regression with only an intercept and no covariates.

Table A1: Example of products on offer

	Initial rate (%)	Fee (£)	Reset rate (%)	Max LTV (%)	Max loan (£000)
Bank A	2.19	2,260	3.99	80	1,000
Bank A	2.49	1,260	3.99	80	1,000
Bank A	3.54	0	3.99	80	1,000
Bank A	3.69	1,260	3.99	80	1,000
Bank A	3.34	1,260	3.99	85	1,000
Bank A	3.79	0	3.99	85	1,000
Bank B	3.39	995	4.79	80	1,000
Bank B	3.49	0	4.79	80	1,000
Bank B	4.05	995	4.79	85	1,000
Bank B	4.14	0	4.79	85	1,000

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

Table A2: 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 A3: 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 A4: 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 A5: 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 A6: 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 A7: Characteristics of choice sets and choices made

	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 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 summarises the distribution of this variable, along with 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 loan-to-value *greater than or equal to* the customer's loan-to-value.

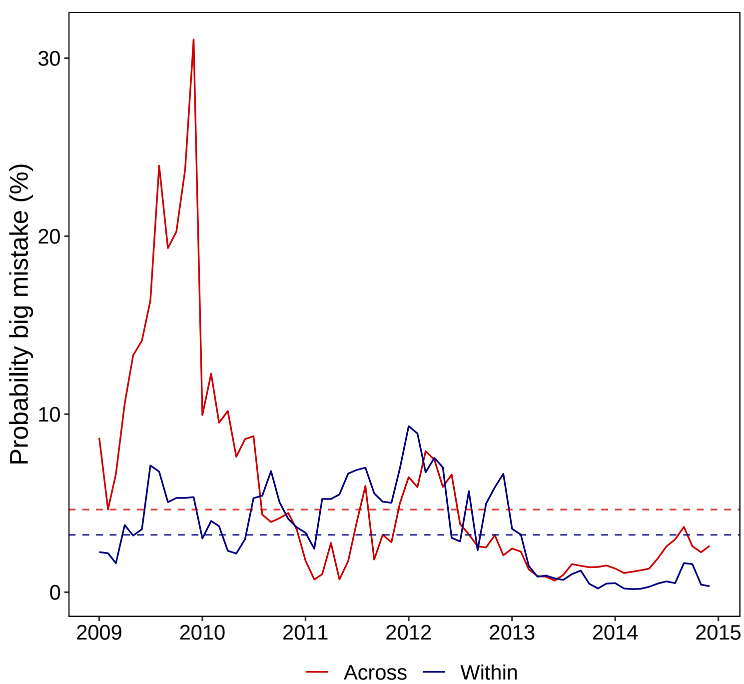
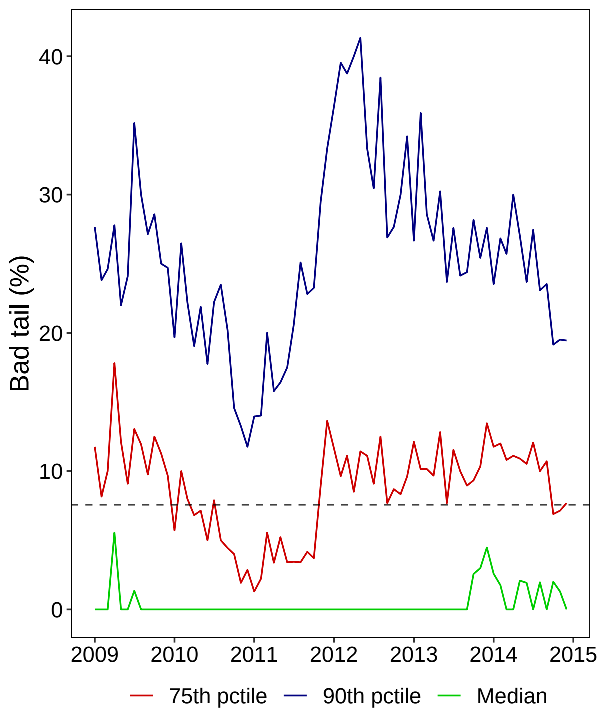
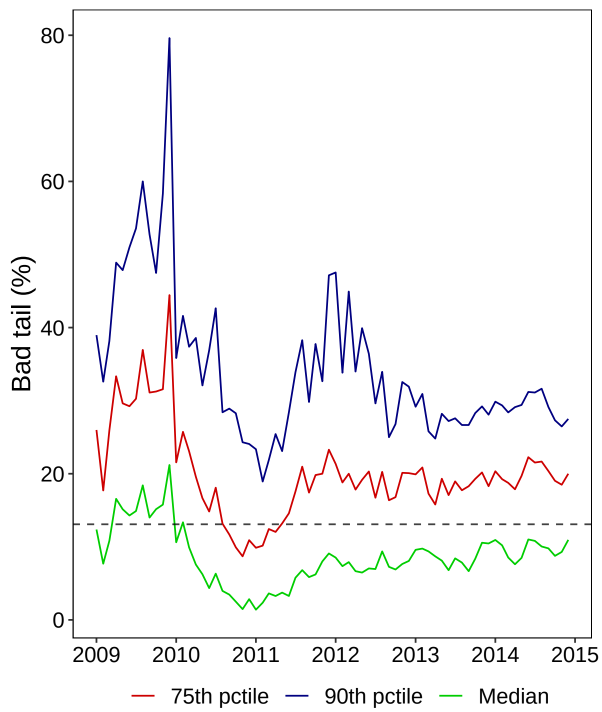


Figure A1: Big mistakes through time

Note: This figure plots the percentage of customers that make big mistakes each month, at the bank where they shopped and across banks. A customer makes a big mistake if the mortgage they choose costs them at least 2.5% of their monthly income more than the 15^{th} percentile in their choice set, where the choice set consists only of mortgages on offer at the bank that gave the customer their mortgage (blue line) and of mortgages on offer across banks (red line). 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.



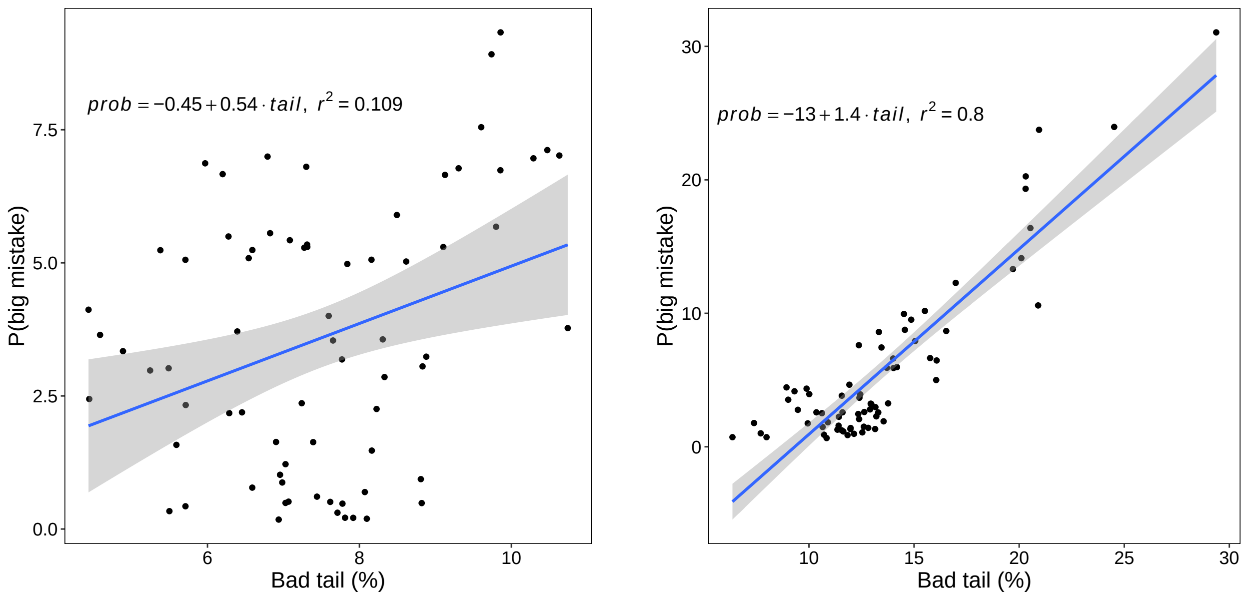
(a) Within each bank's choice set



(b) Across all banks' choice sets

Figure A2: Bad tail through time

Note: These figures summarize the distribution of the fraction of a customer's choice set that would represent a big mistake if chosen, both within and across banks, through time. A mortgage represents a big mistake if choosing it would cost a customer at least 2.5% of their monthly income more than the 15$^{\text{th}}$ percentile in their choice set. The variable *bad tail* computes the fraction of the mortgages in a customer's choice set that would represent a big mistake. 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, 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 A3: Big mistakes and bad tails

Note: These figures summarize the relationship between the frequency of big mistakes and the average quality of customers' choice sets in a given month. A mortgage represents a big mistake if choosing it would cost a customer at least 2.5% of their monthly income more than the 15$^{\text{th}}$ percentile in their choice set. The variable *bad tail* computes the fraction of the mortgages in a customer's choice set that would represent a big mistake. The figures plot the percentage of customers that make big mistakes 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 a big mistake 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 loan-to-value *greater than or equal to* the customer's loan-to-value.

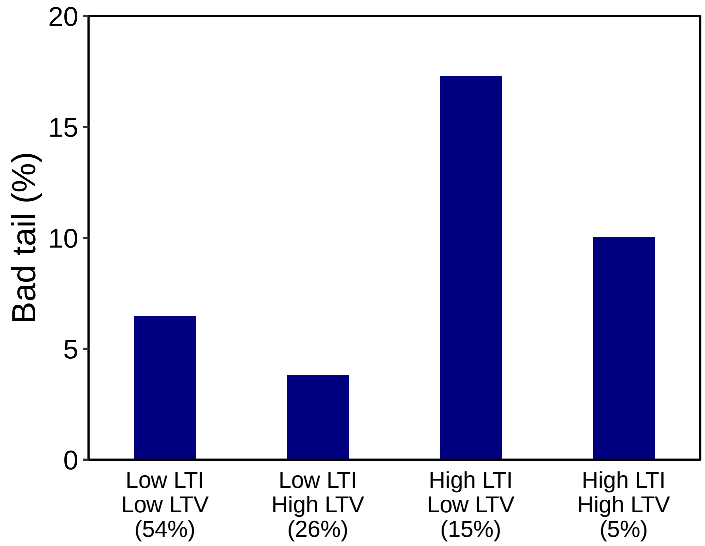


Figure A4: Bad tail by loan-to-value and loan-to-income ratio

Note: This figure summarizes the average quality of customers' choice sets by combinations of their LTV and LTI ratios. A mortgage represents a big mistake if choosing it would cost a customer at least 2.5% of their monthly income more than the 15^{th} percentile in the choice set at the bank where they shopped. The variable *bad tail* computes the fraction of the mortgages in a customer's choice set that would represent a big mistake. This figure plots the average of *bad tail* according to a customer's LTV and LTI. High LTV defined as LTV$>$ 85%, and low LTV as LTV$<$ 85%. High LTI defined as LTI$>$ 4, and low LTI as LTI$<$ 4. The numbers in parentheses below the x-axis 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 loan-to-value *greater than or equal to* the customer's loan-to-value.

Table A8: Mistakes by loan characteristic: marginal effects

	Dependent variable:			
	Big mistake within		Big mistake 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 in the table. High (low) LTI customers have loan-to-income above (below) 4. High (low) LTV customers have loan-to-value above (below) 85%. A big mistake 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 a big mistake. 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. 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. *indicates the coefficient is different from zero at the 10 percent level of significance, **indicates the coefficient is different from zero at the 5 percent level of significance, ***indicates the coefficient is different from zero at the 1 percent level of significance. The pseudo R-squared is McFadden's R-squared, equal to 1 minus the ratio of the regression's log likelihood to the log likelihood of a regression with only an intercept and no covariates. 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.

Table A9: Mistakes by demographic: marginal effects

	Dependent variable:			
	Big mistake within		Big mistake 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 in the table. 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. A big mistake is defined as a choice that costs a customer at least 2.5% of their income more than the 15$^{\text{th}}$ percentile in their choice set. The variable *bad tail* measures the percentage of a customer's choice set that would represent a big mistake. 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. 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. *indicates the coefficient is different from zero at the 10 percent level of significance, **indicates the coefficient is different from zero at the 5 percent level of significance, ***indicates the coefficient is different from zero at the 1 percent level of significance. The pseudo R-squared is McFadden's R-squared, equal to 1 minus the ratio of the regression's log likelihood to the log likelihood of a regression with only an intercept and no covariates. 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.