Journal of Financial Economics 169 (2025) 104056

Contents lists available at ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/finec

Screening using a menu of contracts: A structural model for lending markets☆

Alberto Polo[a], Arthur Taburet[b,*], Quynh-Anh Vo[a]

[a] Bank of England, United Kingdom
[b] Duke University, United States of America

ARTICLE INFO

Dataset link: https://doi.org/10.7910/DVN/QQJOB4

Keywords:
Menu of contracts
Screening
Structural models

ABSTRACT

When lenders screen borrowers using a menu, they generate a contractual externality by rendering the composition of their competitors' borrowers worse. Using data from the UK mortgage market and a structural model of screening with endogenous menus, this paper quantifies the impact of asymmetric information on equilibrium contracts and welfare. Counterfactual simulations show that, because of the externality, there is too much screening along the loan-to-value dimension. The deadweight loss, expressed in borrower utility, is equivalent to an interest rate increase of 30 basis points (a 15% increase) on all loans.

1. Introduction

Menus of contracts are widely used in financial markets.[1] For instance, mortgage borrowers often have the choice between fixed or flexible interest rates, high or low loan-to-value (LTV) ratios, and different combinations of interest rates and fees. A leading explanation is that lenders offer menus to induce borrowers to reveal their private information through their choices (i.e., screening). By screening borrowers, lenders can restore perfect information pricing, but this may come at the cost of distortions in other contract terms. For example, if high LTV contracts are more valuable for high-default borrowers, lenders can make them self-select into a high interest rate high LTV contract. However, maintaining borrowers' incentives requires that low-default borrowers get a lower LTV contract than high-default borrowers, which is not necessarily what would happen in the first best.

The theoretical literature has demonstrated that the equilibrium in markets with adverse selection can be inefficient. For instance, Rothschild and Stiglitz (1976) shows that pooling contracts cannot be offered in competitive markets even when pooling is a Pareto improvement over screening. The reason is that a lender can take advantage of their competitors' pooling contract by introducing a contract with — for instance — an LTV just below the one offered by competitors to steal the safer and more profitable borrowers (i.e., "cream skimming").[2] This market failure emerges when lenders do not internalize how their screening strategies change the types of borrowers who select competitors' products and, thus, the cost of lending via those products.

Yet the practical relevance of the question and the impact of adverse selection on contract terms and welfare remains open (see Einav et al., 2021 for a literature review). Quantifying the impact of adverse selection on contract terms requires determining which contracts would be offered if there were no adverse selection ("first best") and which contracts would be offered by a social planner who internalizes the fact that deviating from pooling may be inefficient ("second best"). Answering this question is challenging since those situations are not directly observed in the data.

In this paper, we quantify the impact of asymmetric information on contract terms and welfare using the first structural model of screening. We use our structural model to simulate the menu of contracts that would be offered in the first and second-best cases. By comparing the simulated contracts with the ones in the data, we assess the extent to which contracts are distorted and quantify the welfare loss. To flexibly capture screening incentives, we develop a supply and demand model with imperfect competition and allow borrowers to have private

☆ We thank the editor, Toni Whited, for helpful comments. A. Taburet is deeply grateful to his advisors Alessandro Gavazza, Martin Oehmke, Daniel Paravisini and Ricardo Reis, for their patient guidance, constant encouragement, and invaluable advice. The paper uses data from the Financial Conduct Authority and the Bank of England that has been provided under a data-access agreement for external research purposes. The views expressed in this paper are those of the authors and do not represent those of the Bank of England or members of the Monetary Policy Committee, Financial Policy Committee, or Prudential Regulation Authority Board views of the Financial Conduct Authority. Any errors are our own.

* Corresponding author.
E-mail addresses: alberto.polo@bankofengland.co.uk (A. Polo), arthur.taburet@duke.edu (A. Taburet), quynh-anh.vo@bankofengland.co.uk (Q.-A. Vo).

[1] Example includes insurance (Handel et al., 2015), consumer loans (Hertzberg et al., 2018), credit cards (Nelson, 2025), and mortgages.

[2] In Rothschild and Stiglitz (1978) the pure strategy equilibrium does not exist in that case because there is also a profitable pooling deviation when all lenders screen. Papers such as Lester et al. (2019) characterize the mixed strategy equilibrium and show that lenders cannot cross-subsidize — and thus pool — when competition is high enough.

https://doi.org/10.1016/j.jfineco.2025.104056
Received 6 August 2023; Received in revised form 24 March 2025; Accepted 25 March 2025
0304-405X/

information about their default probabilities and their preferences over each contract characteristic. We identify and estimate model parameters using administrative data on lenders' menus, borrowers' contract choices, and defaults in the United Kingdom (UK) mortgage market for first-time buyers from 2015 to 2019.

A key challenge when identifying screening incentives is the following: Borrowers who choose different contracts can have different default probabilities because of the causal impact of contract terms (i.e., burden of payment or moral hazard) rather than borrowers' unobservable characteristics (i.e., adverse selection). We propose a novel research design to disentangle moral hazard from adverse selection. We leverage the idea that, everything else equal, changes in the price of a given contract "A" change the type of borrowers who choose another contract "B". Adverse selection can thus be recovered by comparing the default probability of groups of borrowers who chose the same contract "B", but self-selected differently because contract "A" was offered at a different price for each group. We show how to implement this idea formally within a structural model using an instrumental variable approach to exogenously shift the interest rate spread between contracts. The IV is based on contract-specific capital requirements that affect contract-specific lending costs.

We obtain three new empirical results. First, we find that the LTV ratio and interest rates are used together to screen. Lenders set their LTV pricing schedule such that high-default borrowers choose a higher LTV-higher interest rate contract relative to low-default borrowers. Screening works because high-default borrowers — who also tend to be less price elastic[3] — have higher willingness to pay for LTV. That is, high-default borrowers are more reluctant to provide a higher down payment for each pound they borrow (i.e., they have a higher marginal rate of substitution of interest rate for LTV). We also find that other contract characteristics (fees and the type of interest rate) are also used to screen.

Second, using counterfactual simulations, we show that maintaining incentives to self-select requires distorting contract terms away from their perfect information value. In the data, 50% of borrowers (those with a lower default probability) choose contracts with an LTV between 70% and 85%. However, under perfect information, those borrowers and most other borrowers as well would have obtained an LTV above 85% and bought a bigger house. Thus, according to our model, contracts with an LTV between 70% and 85% are introduced primarily to screen borrowers rather than to cater to their preferences. We also find that because of screening, the interest rate on 95% LTV loans is lower by 70 basis points (bps) relative to what those borrowers would have gotten under perfect information.

Finally, by comparing the menu in the data to the one offered when lenders internalize the fact that deviating from a pooling contract can be inefficient (second best), we isolate the effect of the contractual externality and show that there is excessive screening. A lower bound of the deadweight loss generated by this externality is equivalent to the utility loss caused by a 30 bps interest rate increase on all loans.[4]

Our results show that screening is an important force in the UK mortgage market and that the associated contractual externality is costly. This suggests that there is room for Pareto improving policy interventions. Examples of such policies — which are analyzed in a theoretical companion paper (Taburet, 2024) — include lowering competition, increasing the capital requirement for low LTV in a low-competition environment, or banning the use of lower LTV products. Those policies reduce the impact of the contractual externality by preventing cream-skimming deviations from occurring.

We derive our empirical results using a novel structural model that allows us to recover the correlations between borrowers' preferences and their default probabilities, lenders' unobservable costs of originating mortgages, and the fixed cost of changing the menu size.

On the demand side, borrowers choose the contract on their individual specific menu that maximizes their utility. Following the industrial organization literature (Berry et al., 1995; Crawford et al., 2018), we assume that borrowers' utilities are linear functions of contract characteristics (loan size, interest rate type, LTV, lender, and fees) and estimate of the contributory value of each. We allow those contributory values to be heterogeneous and to depend on borrowers' observable characteristics (income, age, location of the house); unobservable characteristics (e.g., risk aversion, financial sophistication, income volatility); and their expected default probability. Our model yields a discrete-continuous demand system (as in, for instance, Train (1986)) composed of a mixed logit model for the product choice and a linear regression model for the loan demand. We specify borrowers' default probabilities as a linear function of contract terms and borrowers' observable and unobservable characteristics.

On the supply side, we model lenders as heterogeneous multiproduct firms that offer differentiated menus of mortgages and compete based on the number of contracts, their interest rates, LTV, and fees. Contracts are exclusive. As such, our framework is adapted to mortgage markets but not to corporate and consumer loans, whereby firms and households borrow from multiple banks simultaneously (See Attar et al. (2011) for a theoretical framework with non-exclusivity).

We identify model parameters using a three-step approach. First, we use, as in Nevo (2001), a revealed preference approach to recover moments of the distribution of borrowers' ex ante unobservable preferences from contract product choice and loan size choice data. In the second step, we use the demand estimates to build a measure of the average preferences of borrowers conditional on contract choice (henceforth, the average borrower type). We use this measure in a default probability regression in which we compare the default of groups of borrowers who choose the same contract but have different average preferences. Variation in the average preference comes from changes over time in the characteristics of the menu offered to a particular group of borrowers. In the third step, we then use the demand and default parameters together with formulas derived from the lenders' profit maximization problem to back out the marginal costs of originating mortgage products and the fixed cost of changing menus.

In the second step, an endogeneity concern can arise if changes in the average borrower type that selects a given contract are correlated with changes in unobservable characteristics. To address this identification threat, we instrument average borrower type using product-specific risk weights and minimum capital requirements for contracts other than the one chosen. Risk weights are predetermined and vary over time across lenders and mortgages with different maximum LTVs. Minimum capital requirements vary over time and across lenders. Both have been extensively used as an instrument for interest rates (e.g., Aiyar et al., 2014; Benetton, 2021; Robles-Garcia, 2019). Our instrument is relevant because it affects the spread between interest rates and, thus, the type of borrower who chooses a given contract. We control for unobserved characteristics that are common among products (lender shocks) and those that are common across lenders (market shocks). Given the absence of individual-based pricing in the UK (see Benetton, 2021), the exclusion restriction requires that our cost shifter is not correlated with economic shocks that affects borrower types differently, changes in unobserved product characteristics, or acceptance and rejection rules based on characteristics unobserved by the econometrician. It is plausible that the endogeneity that stems from mortgage application rejections based on soft information observed by the lender but not the econometrician is not fully addressed, since lenders may update their acceptance and rejection criteria following

[3] The correlation between default and price elasticity is consistent with risky borrowers internalizing the probability that their application is rejected and thus behaving as if they had higher search costs (see Agarwal et al., 2024 for empirical evidence).

[4] Considering an average loan size of £200,000 and a 25-year maturity, this corresponds to a £25 monthly increase in borrowing expenses for all borrowers. In practice, this cost is borne by a third of borrowers and is thus equivalent to a £75 monthly increase.

any product cost shock. In that case, our results should be interpreted as a lower bound on adverse selection since lenders are likely to become stricter to mitigate the increase in the cost of lending.

This paper contributes to the empirical literature on adverse selection and the industrial organization literature on credit markets. A large empirical literature tests whether adverse selection and screening occur in practice. Seminal papers are Chiappori and Salanie (2000) for the positive correlation test approach and Einav et al. (2010) for the sufficient statistic approach. Our paper is closely related to a recent strand of the literature that focuses on disentangling moral hazard from adverse selection in credit markets. This literature uses reduced-form approaches. Identification relies on lenders that just started using menus (Hertzberg et al., 2018) or the use of experimental data (Karlan and Zinman, 2009). We contribute to this literature by showing that variation in interest rate spreads can be used to disentangle moral hazard from adverse selection. Our approach is thus applicable to a wide variety of setups, since the literature has extensively documented plausibly exogenous variations in interest rates. We also implement our identification strategy within a structural model, which enables us to answer a more comprehensive range of questions by performing counterfactual simulations. In particular, this is the first paper to quantify the impact of adverse selection on contract terms and welfare with respect to the first best (perfect information case) and second best (when the contractual externality is internalized by lenders).

This paper also relates to the literature that analyzes consumers' and lenders' behavior in retail financial markets. Our paper contributes to this literature by studying screening. To do so, we build on Benetton (2021) and Crawford et al. (2018) models and include endogenous mortgage product offering and screening.

The rest of the paper is structured as follows. In Section 2, we describe the institutional features of the UK mortgage market, outline the data used, and conduct a descriptive analysis to motivate the modeling assumptions. In Section 3, we present the structural model and discuss its main assumptions. We discuss the identification strategy and estimation procedure in Section 4. In Sections 5 and 6, we analyze the estimation results and counterfactual experiment outcomes. Section 7 concludes.

2. Institutional setting, data, and motivating evidence

This section describes the key institutional features of the market and the data used. It then provides suggestive evidence that screening is an important feature of the UK mortgage market.

2.1. Institutional setting

Market features: While mortgage markets are important credit markets in most countries, their institutional features vary (Campbell, 2013). The UK mortgage market differs from other mortgage markets, including the United States, along three dimensions.

First, lenders do not offer long-term fixed-rate contracts in the UK market. Instead, borrowers can fix the interest rate for a given number of years (typically 2, 3, or 5). After that period, the "teaser rate" is reset to a generally significantly higher and flexible "follow-on rate". Coupled with the fact that contracts feature high early repayment charges — which typically account for 5% or 10% of the outstanding loan — refinancing around the time the teaser rate period ends is very frequent in this market (Cloyne et al., 2019).

Second, the interest rate of a contract advertised by a given bank on its website or other platforms is the one paid by every borrower who chooses that contract. This is because minimal negotiation takes place between borrowers and lenders, and banks do not practice individual-based pricing.[5] However, while pricing is independent of borrowers' characteristics, banks may reject loan applications based on individual characteristics. This approach is common in other markets (credit cards, hedge funds) and online platforms.[6]

Finally, the UK mortgage market is very concentrated; The "big six" lenders account for approximately 75% of mortgage origination. The number of active banks is stable over time.

Loan contracts: As illustrated in Fig. 1, a borrower who is willing to take on a mortgage from a particular bank in the UK can choose from a menu of standardized loan contracts.

The pricing of those contracts is primarily based on product characteristics such as lender name, rate type, maximum LTV, and fees. Indeed, using a linear regression of rate on product characteristics, we show — consistent with other papers on the UK mortgage market (Benetton, 2021; Robles-Garcia, 2019) — that 90% of the price variation is explained by interacting time dummies with lender dummies, rate type, maximum LTV, and fees dummies. The remaining variation is independent of the characteristics of borrowers who choose the contract.

Conditional on those product characteristics, loan size and maturity choices do not impact interest rates.

While the contract pricing is independent of borrowers' characteristics, a bank can choose to reject a borrower's loan application based on their observable characteristics (e.g., income, age, credit score). Since we do not observe loan applications or the criteria used by banks, we will build our empirical strategy while accounting for this limitation.

2.2. Data

We use the Product Sales Database 001 (hereafter, PSD 001). The data are collected quarterly by the Financial Conduct Authority (FCA) and contain contract-level information on households' mortgage choices and detailed information on mortgage origination characteristics for the universe of residential mortgages in the UK. The dataset is available to restricted members of staff and associated researchers at the FCA or the Bank of England.

We merge the data with PSD 007, which contains the credit events for mortgages. We use arrears as a measure of default, which is defined as being 90 or more days delinquent on monthly payments. The loans are full recourse, but in practice, only 5% of the house is repossessed conditional on default, according to UK finance.

In this paper, we focus on the years 2015 to the end of 2018. During this period, for each mortgage origination, we observe details on the loan (interest rate, loan amount, initial fixed period, maturity, lender, fees); the borrower (income, age); and the property (value, location). We focus on the first-time buyer market to abstract from preexisting lending relationships between lender and borrower. In a given month, about 30% of the new mortgage contracts are first-time buyers. The other 70% are borrowers that remortgage and home movers (i.e. 35% remortgage and 35% home movers). First-time buyers are younger (31 years for FTB versus 42 for the other categories), have lower income (25% lower for first-time buyers), and have higher LTV than remortgages (average LTV of 75 versus 55 for remortgage and 65 for home mover).

[5] The search platform Moneyfacts reports: "A personal Annual%age Rate is what you will pay. For a mortgage this will be the same as the advertised APR, as with a mortgage you can either have it or you can't. If you can have the mortgage, the rate doesn't change depending on your credit score, which it may do with a credit card or a loan". See Leanne Macardle, "What is an APR?" Moneyfacts, https://moneyfacts.co.uk/guides/credit-cards/what-is-an-apr240211/.

[6] This can be rationalized by the fixed cost of negotiation's being high compared with the size of loans in the consumer market compared with the firm market.

90% Maximum Loan to Value (LTV)

Mortgage	Initial interest rate	Followed by a Variable Rate, currently	Booking fee
2 Year Fixed Fee Saver	5.59% fixed	6.29%	£0
2 Year Fixed Standard	5.34% fixed	6.29%	£999
5 Year Fixed Fee Saver	5.04% fixed	6.29%	£0

95% Maximum Loan to Value (LTV)

Mortgage	Initial interest rate	Followed by a Variable Rate, currently	Booking fee
2 Year Fixed Fee Saver	6.19% fixed	6.29%	£0
5 Year Fixed Fee Saver	5.75% fixed	6.29%	£0

Fig. 1. Extract of the menu of contracts offered by HSBC in January 2023.
Source: Source: HSBC website.

The structural estimation uses 2018 data (see Table 1 in Appendix A for data summary statistics), for which we also have Bank of England supervisory data about the risk weights and capital requirements. For that year, we observe 847,000 first-time buyer contracts, of which almost 90% are mortgages with initial fixed periods of 2, 3, or 5 years. The average interest rate is 2.5% points, and the average origination fee is £503. The average loan is almost £165,000 with an LTV of 80%, a loan-to-income of 4.6, and an average maturity of 29 years. Borrowers are, on average, 31 years old and have an annual income of £36,000.

We also supplement our analysis with data on the number of products from 2008 to 2022 (Moneyfacts database).

2.3. Motivating evidence

This section discusses descriptive patterns about banks' menus. We also provide suggestive evidence that screening is feasible in this market, because borrowers' (observable) characteristics are correlated with contract choices and default.

Variation in product offerings: As shown in Fig. 2, the number of products varies over time and across market participants. In particular, first-time buyers shopping for 90% LTV contracts faced on average two options at each bank in 2010 and six options in 2018. Menu sizes are larger at 75% LTV. Indeed, the average menu contains seven alternative contracts at 75% LTV in 2010 and 15 in 2018. Typically, in 2018 the average bank offers at 75% LTV, the option of fixing the rate for 0, 2, 3, or 5 years and proposes three levels of fees (0, 750, 1,500). A higher fee level is associated with a lower rate. Considering all combinations of fixed rates and fees for all LTV levels offered starting from 60% LTV (i.e., 60, 65, …, 90, 95), we find that, on average, only 40% of those products are offered by the average bank. This empirical result motivates the fact that the number of products needs to be endogenized in the model.

Sorting on observables: Because suggestive evidence shows that borrowers with different characteristics tend to select different products, we regress borrowers' observable characteristics on LTV dummies (see Table 2).

We document that compared with borrowers who choose 75% LTV contracts, borrowers who choose 95% LTV contracts are on average

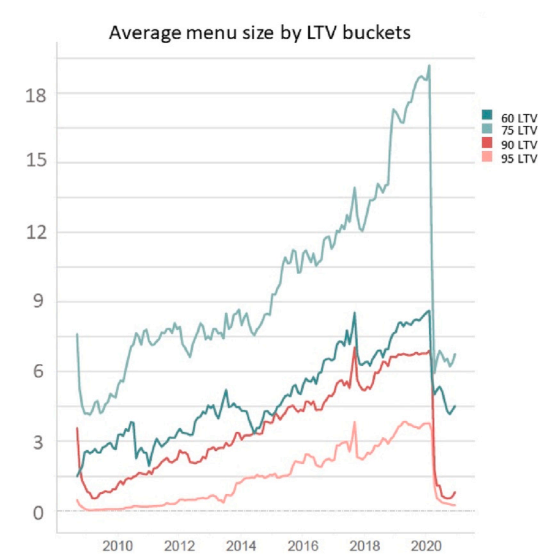

Fig. 2. Average number of advertised mortgage products for BtL, FtB and Remortgage.
Source: Moneyfacts and Bank of England's calculations.

1.5 years younger, earn 7,400 net pounds less a year, and are 20% more likely to be part of a couple.

This correlation between LTV and borrowers' characteristics can be the result of borrowers' self-selection or the fact that banks may decline the loan applications of riskier borrowers for a high LTV loan. Since banks generally offer high LTV loans only to safer borrowers, it is likely that the income and age gap between high and low LTV loans would be higher absent banks' rejection behavior. Causing borrowers to self-select (on observable characteristics) using LTV is thus feasible.

Sorting on default: Because suggestive evidence shows that borrowers who choose different products have different default behavior, we regress default on borrower and contract characteristics (see Table 3):

$$Default_i = \beta X_i + \epsilon_i \quad (1)$$

$Default_i$ is equal to 1 if borrower i has been in arrears by the end of 2019, and X_i includes borrower i's contract terms (lender, LTV, rate, fees, teaser period, mortgage term) and borrower i's characteristics (age, income, location of the house, number of applicants, month and year at which the loan has been originated).

We document that 1.2% of the loans originated in 2018 had defaulted by 2020. The default probability on 85%–95% LTV loans is 1.4%, while the average for 75%–85% LTV loans is 0.8%.

Using a baseline default of 1.2%, the regression of default on product and borrowers' characteristics implies that a 100 bps increase in rate is associated with a 50% increase in default probability; the default probability of a 5-year fixed rate contract is 40% lower than that of flexible rate contracts; the default probability of a zero fee contract is 30% lower than contracts with fees of £1,000; and borrowers whose income is one standard deviation lower (£16,000) are 16% more likely to default.

Those results, together with the one on borrowers' choice of contract — and given that pricing is independent of borrowers' income — provide suggestive evidence of adverse selection along the income dimension. Indeed, we document that low-income borrowers are more likely to choose high LTV contracts and are more likely to default.

Need for a structural model: To further understand the impact of screening on equilibrium quantities, we need to compare the observed equilibrium contracts' terms with a counterfactual in which there is no private information. Given the difficulty of finding the right counterfactual in the data, we instead build a structural framework that relies on simulations. The following sections discuss the model assumptions and our identification strategy. Our modeling approach and identification strategy also enable us to look at selection on unobservable borrowers' characteristics, address the bias generated by the rejection of mortgage applications, and disentangle moral hazard or burden of payment from adverse selection in the default regression.

3. General model setup

For each month t, we consider the data through the lens of the model of supply and demand described in this section. To simplify notation, we drop the time index t on variables except for the fixed cost section, for which the timing is relevant.

3.1. Overview of the model

There are two groups of agents: Borrowers and lenders. We also refer to the second group as banks. There are n borrowers indexed by i. There is a finite number of banks indexed by b $\in B$. The number of borrowers and lenders is exogenous.

Definition of contracts and products: Banks offer a menu of contracts. Based on UK institutional features, we define as a loan contract the object (L, X, r) in which L is the loan size, X is a vector that contains other contract characteristics (lender dummy, teaser rate period, maximum LTV, and fees), and r is the interest rate on the loan.

Using industrial organization literature vocabulary, we also refer to the vector of characteristics (X) as a product, r as the product price, and L as the product quantity. We index a product by the subscript c. We denote P_{ib} as the set of products (c) available to borrower i at bank b.[7] We denote by $M_{ib} := \{(X_{cb}, r_{cb})\}_{c\in P_{ib}}$ the menu of products offered to borrower i at bank b. We drop the b or i index in M and P to refer to the market menu ($M_i := \cup_b M_{ib}$ and $P_i := \cup_b P_{ib}$) or the bank menu ($M_b := \cup_i M_{ib}$ and $P_b := \cup_i P_{ib}$). $C_b := \mathbf{card}(P_b)$ is the number of products sold by bank b.

For each product $c \in P_b$, there exists a contract (L, X_c, r_c) for any loan amount $L \in [a, b]$. The menu of contracts (i.e., its size C_b and content M_b) is endogenous.

[7] Each combination of product characteristics (X) is a one-to-one mapping to a natural number.

Supply and demand: Our model is based on the following maximization problems. Borrower i chooses the bank and contract on its individual specific set that maximize its indirect utility. Lenders choose the menus of contracts they offer to maximize their expected profits. Lenders take competitors' contracts as given and know how borrowers select banks and contracts. They do not perfectly observe borrowers' characteristics but know their joint distribution. Formally, for each period t we have:

Borrower i : Contract c and lender b choice

$$(c_i, b_i) = \underset{\{b\in B_i, c\in P_{ib}\}}{\operatorname{argmax}} \{V_i(\overbrace{X_{cb}, r_{cb}}^{\text{contract terms and price}}, \overbrace{L_i(X_{cb}, r_{cb}, d_i(X_{cb}, r_{cb}))}^{\text{loan demand}}, \overbrace{d_i(X_{cb}, r_{cb})}^{\text{default probability}}) + \overbrace{\epsilon_{icb}}^{\text{demand shock}}\} \tag{2}$$

$V_i(\cdot) + \epsilon_{icb}$ is the borrower indirect utility. The demand shocks ϵ have a mean of zero. $L_i(\cdot)$ is the optimal loan size conditional on contract choice c. $d_i(\cdot)$ is the default probability conditional on contract and loan choice.

Lender b : Menu offering M

$$M_b \subset \underset{\{M\in\mathcal{F}, P_{ib}\}}{\operatorname{argmax}} \overbrace{\underbrace{\mathbb{E}\left[\sum_{i,c} \mathbf{1}_{\{(c_i,b_i)=(c,b)\}}\right]}_{Expected\ Demand} \underbrace{NPV(L_{icb}, r_{cb}, d_{icb}, \overbrace{mc_{cb}}^{\text{marginal cost of lending}})}_{\text{expected NPV if i chooses cb}}}^{\Pi_b(M,P_{ib})} - \underbrace{F(M, M_{bt-1})}_{\text{fixed cost of changing menu}} \tag{3}$$

where borrowers' choice of contract and bank (c_i, b_i) are given by equation (2)

This formulation is a general version of screening models such as that of Rothschild and Stiglitz (1978).

$\Pi_b(\cdot)$ is the expected gross margin, defined as the product of two terms. The first is $\mathbb{E}\left[\sum_{i,c} \mathbb{1}_{\{(c_i,b_i)=(c,b)\}}\right]$, the probability that the borrower i chooses the bank b and contract c. The second is $NPV(\cdot)$, the net present value of lending to borrower i conditional on their choice of contract. I use the shortcut notation L_{icb} and d_{icb} for the demand and default probability values when contract cb is chosen. mc_{cb} is the bank-product-specific marginal lending cost. $F(\cdot)$ is the fixed cost of designing a new menu. M_{bt-1} is the menu offered by bank b in the previous period. $\mathcal{F}$ is the feasible set of menus. To fit the empirical application, we allow the interest rate to be a continuous variable, but contract characteristics and the number of contracts are discrete variables.

Information set: The expectation in Eq. (3) is conditional on the lender information set. The information set contains competitors' menus of contracts and observable borrower characteristics.

Timing: Given that product characteristics are updated less frequently than interest rates, we follow Wollmann (2018) and assume that lenders play a two-stage game. They chose product characteristics first and then compete on interest rates. This modeling is the most conservative as it is likely to lower the contractual externality as lenders partially internalize their competitors' behavior.

3.2. Key parametric assumptions

In this section, we present our parametric assumptions. The implications of the modeling assumptions are discussed in depth in Section 3.3.

3.2.1. Demand

The following equations can be interpreted as linear approximations around contract terms of the indirect utility, the logarithm of the loan demand, and default functions (denoted V, $log(L)$ and d). As shown in Appendix C, the same formulas can also be obtained by specifying an indirect utility and using Roy's identity to derive the loan demand.

In the spirit of the positive correlation literature (see Chiappori and Salanie, 2000), we model adverse selection via the correlation between random variables in the demand default equations. We generalize their approach by allowing the random variables to affect the slopes (i.e., the demand sensitivity to contract terms) rather than only the intercepts. This generalization is needed for screening to be possible. A formal proof is provided in Appendix E; the intuition is the following. If high-default borrowers find it relatively more costly to provide a high level of down payment for each additional unit they borrow, low LTV contracts attract unobservably safer borrowers and can be offered at a lower price.

Borrower indirect utility: The indirect utility in Eq. (2) of borrowing via contract c at bank b is parametrized as a linear function of contract terms:[8]

$$V_{icb} := \beta_i^P X_{cb} - \alpha_i^P r_{cb} + \xi_{cb} \tag{4}$$

(β_i^P, α_i^P) drive how borrower i values product characteristics and interest rate; we loosely refer to these as borrowers' preferences. They are parametrized as a linear function of observable borrowers' characteristics (denoted D_i) and unobserved borrower heterogeneity (modeled by a random variable denoted PI_i^P). I describe the correlation structure between random variables more formally at the end of this section once the full demand system is presented.

ξ_{cb} is a product bank fixed effect. It captures the average borrower's indirect utility from contract characteristics unobserved by the econometrician.

Plugging Eq. (4) into Eq. (2) and assuming that the demand shocks ε_{icb} are extreme value distributed and independent of unobserved borrower characteristics, the probability that borrower i choose contract (c,b) is:[9]

$$Pr(i\ chooses\ (c,b)|\alpha_i^P \beta_i^P, \xi_{cb}) = \frac{exp(V_{icb})}{\sum_{x\in B, y\in\{P_{ib}\}} exp(V_{iyx})} \tag{5}$$

Notice that the logit functional form of Eq. (5) implies that any borrower fixed effect in the indirect utility cancels each other.

Loan demand: The loan demand is parametrized as:

$$ln(L_{icb}) = \beta_i^L X_{cb} - \alpha_i^L r_{cb} + v^L D_i + \sigma_L \epsilon_{icb}^L \tag{6}$$

(β_i^L, α_i^L) parameterize the loan size semi-elasticities. They are linear functions of observable borrowers' characteristics and unobserved borrower heterogeneity (modeled by a random variable denoted PI_i^L).

The vector v^L parameterizes the sensitivity of the loan demand to the borrower's characteristics.

σ_L is the standard deviation of the demand shock ϵ_{icb}^L.

Default probabilities: The default probabilities are specified as:

$$d_{icb} = \beta^d X_{cb} + \alpha^d r + v^d D_i + \rho PI_i + \sigma_d \epsilon_{icb}^d \tag{7}$$

The parameters β^d and α^d capture the impact of contact terms X and prices r on default probabilities. They can be interpreted as moral hazard or burdens of payment.

v^d parameterizes the sensitivity of default to borrowers' characteristics.

ρ is the key parameter of interest. This vector captures the average impact of borrowers' unobservable characteristics $PI_i := (PI_i^P, PI_i^L)$ on default probabilities. When the first element of ρ differs from zero, lenders can use contract terms to screen borrowers on their default probabilities.

$\sigma_d \epsilon_{icb}^d$ is a random variable that represents the part of default probabilities that is not captured by the other variables.

Unobservable heterogeneity: The key innovation of the demand system is to allow borrower unobservable characteristics — modeled by a random variable $PI := (PI_i^P, PI_i^L)$ — to jointly determine demand and default. Formally, the random variables $(\beta_i^P, \beta_i^L, \alpha_i^P, \alpha_i^L)$ driving the demand elasticities are linear functions of observable (D_i) and unobservable heterogeneity (P_i):

$$\beta_i^x = \beta^x + \gamma^x D_i + \bar{PI}_i^x,\ \ \alpha_i^x = \alpha^x + \tilde{\gamma}^x D_i + \tilde{PI}_i^x \tag{8}$$

$$(PI_i, \epsilon_{ic}^L,\ \epsilon_{ic}^d) \sim N(0, \Omega) \tag{9}$$

$$with\ PI_i := (PI_i^P, PI_i^L)\ and\ PI_i^x := (\bar{PI}_i^x, \tilde{PI}_i^x)_x,\ \ x \in \{P, L\} \tag{10}$$

(β^x, α^x) drive the average demand sensitivities.

$(\gamma^x, \tilde{\gamma}^x)$ drive the observable heterogeneity in demand sensitivities.

PI is a random variable capturing the part of the demand sensitivities (β_i^x, α_i^x) that is not captured by the observable characteristics.

Ω allows the random variables in the demand and default equations to be correlated. It captures the fact that (V, L, d) derive from the same maximization problem. This random variable approach helps to mitigate the selection on unobservables when bringing the model to the data. It is a generalization of Train (1986).

3.2.2. Supply

NPV: We follow the standard literature assumptions to build an approximation of the net present value of the cash flow associated with a mortgage. As in Benetton (2021) and Crawford et al. (2018), we assume that banks are risk neutral, that all borrowers refinance at the end of the teaser rate period[10], and that lenders do not forecast the probability of default in each period, but consider an average expected probability of default. The Net Present Value (NPV) of lending is thus well approximated by:

$$NPV_{icb} \approx L_{icb} \cdot [(1 - d_{icb})r_{cb} - mc_{cb}]f_{cb} \tag{11}$$

where L_{icb} is borrower i's loan demand conditional on choosing contract c at bank b (defined in Eq. (6)), d is the default probability (defined in Eq. (7)), r is the interest rate, f is the fixed rate period and mc is the marginal cost of lending.

Derivation of the formula is provided in Appendix F.

Fixed cost: Given that price changes are more common than product introduction and withdrawal, we consider that only changes in product characteristics affect the fixed cost as in Wollmann (2018). Thus, the fixed cost of designing a menu thus takes the following form:

$$F(m, M_{bt-1}) = \frac{\tilde{F}_b(m, M_{bt-1})}{\beta^F} + \beta^F e_m^F \tag{12}$$

$$\tilde{F}(m, M_{bt-1}) := \sum_{c\in P_{bt}} \theta' X_{cb}[\underbrace{\mathbf{1}_{c\in P_{bt}, c\notin P_{bt-1}}}_{Inclusion} + \lambda \underbrace{\mathbf{1}_{c\in P_{bt-1}, c\notin P_{bt}}}_{Exclusion}] \tag{13}$$

e_m^F is a cost shock that is independent across products and extreme value distributed. β^F is the variance of this cost. We scale down the fixed cost F_b by β^F for notational convenience in the estimation section, but this is without loss of generality.

$\tilde{F}$ is the non-random cost of changing the menu. We use the same functional form as Wollmann (2018), in which $\theta' X$ is the cost of introducing a new contract with characteristics X (i.e., the origination fee, LTV, and fixed-rate period), and λ is a scaling parameter that captures the cost or benefits of withdrawing a contract from the menu.

[8] Notice that we use IO notation for the indexes. That is, as in — for instance, Benetton (2021) — while we only observe one choice per borrower, we index all possible alternatives that were available to the borrower.

[9] The parameterization of the indirect utility is robust to the use of borrower and time fixed effect as they would disappear in Eq. (5).

[10] Given the high level of refinancing at the end of the initial period, it is unreasonable to assume that lenders compute the present value as if all mortgages were held until maturity.

3.3. Discussion of the model's assumptions

Any model simplifies reality to focus on a given economic phenomenon. For instance, we do not endogenize the house price upon default or model dynamic considerations in order to be able to model screening incentives in more detail. Our counterfactual simulations thus consider that those elements — as well as unobserved product characteristics (captured by product-lender fixed effects) — remain constant.

3.3.1. Demand

In this section, we discuss how our assumptions affect the interpretation of demand parameters.

Savings: Since we do not observe savings, we cannot explicitly model constraints on the level of down payment (dp) a borrower can provide. We address this issue by modeling borrowers' choice of both LTV and the loan size and relying on a revealed preference approach to recover the demand parameters. Using the definition of LTV, we get: $LTV := \frac{L}{dp+L} \Leftrightarrow dp = L \cdot \frac{1-LTV}{LTV}$. In the situation in which a borrower is constrained by their savings (s_i) when selecting their level of down payment, their loan demand function is $L_i(LTV) = s_i \frac{LTV}{1-LTV}$, where s_i is a parameter to be recovered using choice data. Our specification of the demand allows us to capture this situation.

Rejection of mortgage application: In borrowers' maximization problem (2), we allow for the menu available to each borrower (P_{ib}) to differ as a result of rejections of borrowers' applications for a particular contract. Our modeling of the choice of product is general enough to encompass the case in which borrowers have or do not have perfect knowledge regarding which applications would be successful and which would not. We favor the perfect information case interpretation, since this case can be justified by the heavy use of brokers in this market. The imperfect information case is discussed in Appendix D.

Borrowers' participation in the mortgage market: As shown by Benetton et al. (2021), borrowers' entry decision in the mortgage market is very inelastic to loan prices and characteristics.[11] Furthermore, Robles-Garcia (2019) and Benetton (2021) show that the level of competition is high in the UK mortgage market, which makes it unlikely that banks will be able to extract the full surplus from borrowers. This motivates the assumption of taking borrowers' participation as given and the use of a static demand model.

3.3.2. Supply

In this section, we discuss how our assumptions affect interpretation of the supply parameters.

Collateral: Our NPV parameterization is derived in Appendix F from a model in which banks do not recover anything following borrowers' default. This assumption does not affect the demand estimation, because we do not explicitly model the cost of default and instead use a revealed preference approach. However, it affects interpretation of the marginal cost of lending parameter that is recovered in the estimation section. To provide intuition for how to interpret the results given our assumption about collateral, let us introduce the following notation. Upon default, the mortgage originator can seize the lender's house and get $min\{\delta \cdot \frac{L}{LTV}, rL\}$. L is the loan size, r the interest rate, $\frac{L}{LTV}$ the house value at the origination date, and δ the ratio of the house price upon default over the one at origination. Default happens with probability d. If δ is not equal to zero, the estimated marginal cost we recover will capture the average loss given default conditional on LTV $\mathbb{E}\left[mc - min\{\delta \cdot \frac{1}{LTV}, r\} d | LTV\right]$. Following the literature (for instance, Benetton (2021), Crawford et al. (2018)), we do not identify δ and mc separately.

[11] They estimate the entry decision in regular time, as opposed to a financial crisis, but it seems that even during the COVID-19 crisis, the number of borrowers did not drop on average.

Static model of supply: The supply model used in this paper is static, since at each period lenders maximize the expected profits generated by current lending activities only. This consideration is justified by demand's also being static. However, using the fixed cost function in the lenders' problem creates a dynamic relationship between current and past maximization problems, which makes the use of a dynamic model natural.

The following considerations can nonetheless justify the static supply approach. First, our static modeling can be written as the hurdle rate approach, which is a good approximation of firms' product-offering decisions according to recent surveys (see Wollmann, 2018). The hurdle rate approach assumes that firms choose to offer a set of products such that, for any other feasible set, the expected ratio of the added profits to added sunk costs does not exceed a set number (the hurdle rate).

Second, the only parameter affected by a dynamic modeling approach is the fixed cost function, which is not an object of interest in our analysis. Indeed, the marginal costs are not affected, because they are identified from a model optimality condition that depends on the number of products being fixed. The counterfactual experiment is not affected by the use of the static model as long as the relationship between current and expected profits in the counterfactual experiment remains the same as in the data. The static estimation affects the economic interpretation of the size of the fixed cost. As a complementary approach, we show in Appendix B how methods used in the dynamic demand estimation literature could be used in a dynamic version of our model to estimate the supply parameters. However, the dynamic estimation increases the computational burden of counterfactual experiments to the point at which the counterfactual model would not be solvable with the methods currently available.

Fixed cost: Fixed costs are needed to rationalize the fact that banks do not offer a continuum of products despite the large heterogeneity in preferences. They can be interpreted as monetary costs that capture, for instance, marketing expenses or updates in software, but can also be interpreted as nonmonetary costs such as managerial frictions or collusion. Since our model is static, the fixed cost may capture the impact of competitors' punishment strategy if the bank deviates from the current menu offering (see, for instance, Dou et al. (2022) for a game theoretical modeling of dynamic collusion).

4. Identification and estimation

We use product choice and loan size data to recover demand parameters, and default data to recover default probabilities. Once the demand parameters are estimated, we use the lender model optimality conditions together with data on the menus offered and estimated demand parameters to recover the supply parameters.

For notational convenience, we collect all parameters in the vector $\Theta := (\Theta^D, \Theta^d, \Theta^S)$ where $\Theta^D := (\Theta^P, \Theta^L))$ denotes demand parameters related to product demand (Θ^P) and loan demand (Θ^L). Θ^d contains the default parameters (β^d, ν^d, ρ^d) and Θ^S the supply ones (mc, F). The elements of Θ^P and Θ^L are defined in the relevant sections. Each of the following sections — demand (Section 4.1.1), default (Section 4.1.3), and supply (Section 4.1.4) — focuses on identification and estimation of its respective Θ element.

4.1. Identification

4.1.1. Step 1: Demand

In this first step, we use contract choice data to identify and estimate borrowers' heterogeneous demand elasticities. Those elasticities capture banks' ability to screen borrowers along their outside options. For instance, if borrowers who value high LTV contracts the most also tend to less intensively compare products across banks, lenders can use a menu to extract more surplus from them.

The demand parameters (Θ^P, Θ^L) that govern the choice of contract (the mixed logit equation (5)) and optimal loan choice (linear regression (6)) are identified using the cross section for a given month. However, we estimate the model using both time and cross-sectional variation.

4.1.2. Identification of the product choice and loan demand parameters

Identification challenges for bank and contract choice: There are two classic challenges related to demand estimation. The first is that interest rates are endogenous. In particular, interest rates are likely to be correlated with unobserved product characteristics such as product-specific marketing expenses. The second identification threat comes from unobserved lender loan application rejection criteria that affect the borrower-specific choice set (P_i). For instance, it may be that some borrowers did not choose a higher LTV contract because they were unable to rather than because it was too expensive. As a result, using a larger choice set in the logit regression than the one offered to borrowers is likely to lead to downward bias in willingness to pay for LTV.

We use an instrumental variable approach together with bank and product fixed effects to address the unobserved product characteristics. Fixed effects control for unobserved product characteristics that are common across banks (e.g., market segment-specific advertising) or common across products of the same bank (e.g., the branch network or customer service).

Following Benetton (2021) and Robles-Garcia (2019), we use product-specific risk weights as cost shifters. The instrument is relevant for the following reasons. Lenders fund their lending via deposits and their own capital (equity). Typically, deposits are a cheaper funding source because of the convenience yield of holding deposits. Risk weights affect the amount of capital requirements a lender must hold when lending via a particular type of contract. They thus affect the cost of lending. Under imperfect competition, this cost is partially passed through the contract interest rates (this holds empirically) and thus affects the spread between interest rates. It is likely for this reason that our instrument is relevant.

Risk weights vary across lenders and over time. For the largest banks, risk weights come from an Internal Ratings-Based (IRB) model. While the choice of model is endogenous, it must be approved by the central bank before being used. Since the approval process entails some delay, it is unlikely that current unobservable shocks are correlated with predetermined risk weights. Given the absence of individual-based pricing in the UK (see Benetton, 2021), the exclusion restriction requires that our cost shifter is not correlated with unobserved bank-product-specific unobservable characteristics. This restriction is violated if lenders react to cost by changing unobservable product characteristics. However, given that observable contract characteristics other than interest rates (e.g., reset rates, pre-payment penalties) are relatively constant over time for any given bank, it is unlikely that the time series variation in risk weights is highly correlated with changes in unobservable product characteristics.

Consideration set bias is dealt with using a sufficient set approach, as in Crawford et al. (2021). This approach shows that taking a subset of the menu for which banks' rejection is independent of variables unobserved by the econometrician restores the consistency of the estimates. The choice of subset is subject to the econometrician's judgment. Since a failure of the sufficient set correction would lead to downward bias of the WTP LTV estimates, our main results for the LTV distortion level and the cost of those distortions should be interpreted with caution as a lower bound of the true effect. We construct the choice set as follows, as in Benetton et al. (2021) and Robles-Garcia (2019). We build sets based on the product sold in the same month in the same geographic region. The geographic restriction mostly affects building societies and smaller banks, because they often have limited coverage across regions. The time restriction accounts for the entry and exit of products. We then further restrict the choice set by considering products with LTV just above and just below the one actually chosen. In addition, if a borrower got a loan from a large bank (top 8), we restrict his choice set to large banks. We impose similar restrictions for small lenders. This captures consideration bias in the search or lender rejection. Furthermore, we assume that a household will not qualify for a product if it has a larger loan-to-income ratio, or if they are older than any of the cutoff values. The rationale for these restrictions is based on lenders' most common set of affordability criteria.

Identification challenges regarding loan amount: There are two econometric challenges.

The first one arises from the fact that the interest rate can be correlated with unobservable bank-product characteristics. We deal with this using the same instrumental variable approach as in the product choice estimation.

The second challenge arises from selection bias. It occurs if, for instance, borrowers with a high unobservable propensity to borrow are also more likely to compare products more intensively and thus end up choosing lower-rate contracts. This bias is mitigated by allowing random variables in the demand and default regressions ($PI_i, \epsilon^L_{ic}, \epsilon^d_{ic}$) to be correlated. Train (1986) is an extreme version of this approach because it assumes that the coefficients are perfectly correlated.

4.1.3. Step 2: Default probabilities

In this second step, we use contract default data together with our demand estimates to identify and estimate adverse selection (ρ) and moral hazard parameters (α^d).

Default parameters are identified and estimated using cross -sectional variation and variation in the mortgage origination month.

Research design: In the default regression (7), some product or borrower characteristics may be unobservable by the econometrician. In particular, the borrower's private information PI_i is unobservable. However, given the use of menus in this market, we can construct a measure of the average borrower type conditional on product choice from our demand estimates. We denote it $\hat{\beta}^P_{icb}$. This is formally defined as

$$\hat{\beta}^P_{icbc^-} := \hat{\mathbb{E}}[PI^P_i | \mathcal{I}^E_P, i\ choose\ cb] \tag{14}$$

$\mathcal{I}^E_P$ denotes the econometrician information set and β^P_i is defined in Eq. (4).[12]

Because of the specification of preferences β^P_i and the use of bank-product fixed effects (ξ_{cb}), β^P_i is uncorrelated with observable and unobservable contract characteristics (X^o_{cb}, X^u_{cb}). As a result, the coefficient $\hat{\beta}^P_{icb}$ contains no information on moral hazard or burden of payment.[13]

We use the index c^- to emphasize that the average borrower type who selects product c depends on the other contracts c^- offered. To gain intuition, let us consider a situation in which high-default borrowers tend to choose high LTV contracts. If the price of high LTV contracts increases, some borrowers will substitute to lower LTV contracts, which therefore changes the average type of borrower who chooses low LTV contracts. It is this source of variation — i.e., changes in outside options c^- — that identifies the screening for default parameter ρ in the default regression.

Identification of the coefficients that drive moral hazard (α^d) and adverse selection (ρ^d) thus comes from two different sources of

[12] The average type can be recovered as follows. $\mathbb{E}[\beta^P_i | \mathcal{I}^E_P, i\ choose\ c] = \int \beta^P \frac{Prob(\text{i chooses cb}|\mathcal{I}^E_P,\Theta^P,\beta^P_i)}{Prob(\text{i chooses cb}|\mathcal{I}^E_P,\Theta^P)} dF(\beta^P;\Omega^P)$ and $Prob(\text{i chooses cb}|\mathcal{I}^E_P,\Theta^P,\beta^P_i)$, given by Eq. (5), depends on the spread between contracts only ($\frac{exp(\beta X_c - \alpha r_c)}{\sum_x exp(\beta X_x - \alpha r_x)} = \frac{1}{\sum_x exp((\beta X_x - \alpha r_x)-(\beta X_c - \alpha r_c))}$). $Prob(\text{i chooses cb}|\mathcal{I}^E_P,\Theta^P)$ is given by integrating $Prob(\text{i chooses cb}|\mathcal{I}^E_P,\Theta^P,\beta^P_i)$ over β^P.

[13] This last statement is conditional on interpreting the β^P_i coefficient as coming from a first-degree approximation of borrowers' valuation of contract characteristics ($\beta^P_{icb} \approx \beta^P_i + f(X_{cb})$), or assuming that the causal impact of contract terms is homogeneous across agents—as in, for instance, Hertzberg et al. (2018).

variation—for instance, changes in the interest rate of product c and changes in the interest rate spread between product c and its close substitutes (c^-). Variations in interest rate r_{cb} — while keeping interest rate spreads constant — keep incentives to choose a given contract unchanged ($\hat{\beta}^P_{icbc^-}$ does not vary), but change the burden of payment for the borrower ($\alpha^d r_{cb}$). In contrast, variations in the spread between contracts c and other contracts — keeping r_{cb} constant — will only change the type of borrower who gets contract c ($\beta^d r_{cb}$).

Identification challenges: Let us now formally discuss the identification challenges. As in the product choice regression, we might worry that unobserved product characteristics are correlated with interest rates. We mitigate this concern using the same cost shifter (capital requirement) to instrument the for interest rate.

One potential identification challenge is associated with adverse selection coefficients ρ^d. This arises if changes in outside options — for instance, the interest rates for contract c^- — are correlated with changes in contract c characteristics that we cannot control for or to changes in other variables that affect default, such as sector specific labor shocks.

To limit this omitted variable concern, we use bank fixed effects and product fixed effects and control for the mortgage origination date. Our empirical strategy thus controls for differences across acceptance and rejection rules that are common among products (lender shocks) and differences across products that are common across lenders (market shocks).

We also use a new instrument for our measure of borrower average type $\hat{\beta}^P_{icbc^-}$. The instrument is based on risk weights, as in the product choice regression. The difference with product choice regression is that the risk weights of products that were not chosen are also used as instruments. The instrument is relevant, because changes in the cost of producing products other than product c are passed through the interest rate for those products, and thus change the type of borrower who chooses product c even when the characteristics of product c did not change. Formally, we instrument $\hat{\beta}^P_{icb}$ by replacing interest rates with capital requirements in Eq. (14).

The instrument exclusion restrictions are that changes in capital requirements of other contracts than contract c are uncorrelated with unobserved bank-product specific characteristics of contract c or any bank-product specific shocks that also affect default. The instrument faces the same limitation as the one discussed in the product demand Section 4.1.1. Given the comprehensive set of contract characteristics, we observe that the main bias likely stems from a correlation of risk weights with acceptance and rejection rules. Such rules are lender choice variables, and as such, they may react to any product-specific cost shocks. Any instrumental variable would thus face this caveat, but this may be less of an issue for rejection rules because the economic literature argues that they are quite sticky (Agarwal et al., 2024). Yet, to be conservative, our results can be interpreted as a lower bound on adverse selection since lenders are likely to become stricter to mitigate the increase in the cost of lending. An alternative IV approach could exploit the timing of a bank-specific internal rate-based approval as an exogenous variation in the interest rate spread between products, assuming that acceptance and rejection rules take time to react to that change. However, the internal rate-based model mostly happens around 2010—A period in which the PSD data feature less information about contract characteristics. This approach is thus outside the scope of our paper.

4.1.4. Step 3: Supply

In this third step, we use menu data together with our demand and default estimates to identify and estimate the marginal costs of lending and the fixed cost of designing a new product.

Conditional on the demand and default parameters' being identified and estimated, the supply parameters are identified and estimated using the cross-sectional variation.

Identification: The optimality conditions of the lender problem (3), written from the econometrician information set, are:

$$\partial_{r_{cb}} \Pi_b(M_b, P_i) = 0, \text{ for all rates } r_{cb} \quad (15)$$

$$\Pr(M_b|\Theta) = \Pr(M_b \in \underset{\{m \in F\}}{\operatorname{argmax}}\{\Pi(m, P_i) - F(m, M_{b,t-1})\}|\Theta) \quad (16)$$

$$\iff \Pr(M_b|\Theta) = \frac{\exp(\beta^F \Pi(M_b, P_i) - \tilde{F}(M_b, M_{b,t-1}))}{\sum_{m\in F} \exp(\beta^F \Pi(m, P_i) - \tilde{F}(m, M_{b,t-1}))} \quad (17)$$

Θ is the set of parameters. All other elements are defined below Eqs. (3) and (13).

The first-order condition with respect to interest rate (15) — analyzed formally in Appendix H — states that the interest rates can be written as a markup over the effective marginal cost. As in, for instance, Crawford et al. (2018), the markup is larger when the demand elasticity and default elasticity (i.e., the burden of payment) are low. Indeed, when elasticity is low, lenders can increase interest rates without losing customers. However, when the payment burden channel is large, increasing rates leads to larger default probabilities, which provides incentives to keep interest rates low. As in Rothschild and Stiglitz (1978) some contracts feature an asymmetric information premium while others offer a discount (i.e., an information rent). The premium and discounts are decrease and increase in interest rates relative to the perfect information case. They are used to maintain borrowers' incentives to self-select.

Eq. (17) states that the menu M_b is more likely to be offered if it is the best response to menus offered by other lenders. Given that the gross margin function is increasing and concave in the number of products, the fixed cost is such that any additional product introduction beyond what we observe in the data must generate less revenue than the fixed cost of introducing the said product. Using this condition for all banks, we can point-identify the fixed cost parameters using a standard logit model argument. A similar argument holds for identification of the parameter (λ) that captures the cost of the benefits of withdrawing a product from the menu.

Contrary to Wollmann (2018) and, for instance, Pakes et al. (2021), we assume that the error terms in the fixed cost function (F) are extreme value distributed. This parametric assumption allows us to write the moment inequalities (derived using the best response function approach as in Wollmann, 2018) as a logit model. This parametric assumption allows us to point-identify the parameters using classing logit model arguments.

Identification challenges: Given demand and default estimates and data on menus, the only unknown in Eq. (15) is the marginal cost of lending. As in the seminal paper by Berry et al. (1995), we thus recover the bank product-specific lending cost by inverting the first-order condition with respect to interest rates. Once the marginal cost is recovered using Eq. (15), we can construct an estimate of the gross margin function $\hat{\Pi}(M_b, P_i)$. We then use it in Eq. (17) to recover the fixed cost.

The marginal cost equation is recovered by inverting equation (15) for each bank-product without making any identification assumptions. As a result, there are no identification challenges, but interpretation of the marginal cost coefficient changes depending on the model used. We discuss this point extensively — as well as the fixed cost interpretation — in Section 3.3.2.

Instead, the fixed cost estimation relies on the classic assumption that the error terms (e^F_m) are uncorrelated with observable characteristics. Those error terms can be interpreted as both unobserved fixed cost heterogeneity and growth margin misspecification. The latter occurs because we use an estimate $\hat{\Pi}$ instead of the true Π in the fixed cost regression (17).

Omitted variable bias emerges if, for instance, high LTV products are often associated with higher marketing expenses. This would tend to bias upward the cost of high LTV products. To mitigate these issues, we use product-fixed effects from the demand regressions as dependent variables. The reasoning is that those fixed effects can be interpreted as unobservable product-bank characteristics (see, for instance, Berry et al. (1995) or Granja (2021)).

The fixed cost identification also relies on an assumption about the set of alternative menus that were considered by the lender (i.e., $\mathcal{F}$). This issue is common to the demand estimation, and has been analyzed in the consideration set literature (see, e.g., Crawford et al. (2021)). For instance, wrongly including a highly profitable product that is not being offered because of regulations or that is mistakenly not considered by the banks will bias the cost of introducing this product upward. To mitigate this issue and the computational burden, we perform the estimation only at product introduction and in product exclusion periods and calculate counterfactual profits in the equation using the menu from the previous period. As a robustness check, we also conduct the estimation considering as a set of potential products, combinations of the most common values for the characteristics of existing products in the market.[14]

4.2. Estimation

The demand coefficient in the logit model is estimated separately for each consideration set, as in Benetton (2021).

Joint estimation of the demand, default and supply parameters is computationally demanding, because it would require iterating on the estimate of average preference $\mathbb{E}[\beta_i^P|\mathcal{I}_P^E, i\ choose\ cb]$ for each Θ^P. For this reason, we estimate each equation ((5), (6), (7), (15), (17)) separately using GMM (see, for instance, Nevo (2001) for the mixed logit procedure) and calculate standard errors using a bootstrap method.

We condition the moments on the information gathered from previous steps. That is, the loan choice moment built from Eq. (6) is conditional on the product choice. The default moment built from Eq. (7), is conditional on the choice of product and loan size. The supply parameters are conditional on the demand parameters. The correlations between random variables are recovered by constructing a consistent estimate of their average value conditional on product choice (and loan choice for the default regression) and using this value as a dependent variable. For instance, the procedure for the default regression is as follows. Given a consistent estimate for Θ^P — taken from the product demand estimation — we construct a consistent estimate for $\mathbb{E}[\beta_i^P|\mathcal{I}_P^E, i\ choose\ cb]$ using Bayes' rule and the estimated preferences coefficients of Eqs. ((5), (6)).[15] To lower the computational burden of calculating conditional random variables in $\mathbb{E}[\beta_i^P|\mathcal{I}_P^E, i\ choose\ cb]$, we approximate the equation using a linearized version of the logit model in the spirit of Salanié and Wolak (2019). We then use our estimate of $\mathbb{E}[\beta_i^P|\mathcal{I}_P^E, i\ choose\ cb]$ as a dependent variable in the default regression.

Marginal cost estimates are recovered by inverting equation (15). We obtain an estimate the gross margin using demand and marginal cost in the gross margin function $\Pi(M_b, P_i)$.

We then simulate the equilibrium gross margin in the second stage of the game for all product deviations in the feasible set $\mathcal{F}$. Because product characteristics are fixed in that stage, the gross margin calculation is similar to standard IO setups. We can thus use (Morrow and Skerlos, 2011) contraction mapping to recover the equilibrium interest rate. As such, calculating lenders' growth margins for each deviation does not feature any multiple equilibrium issues.

We then estimate the logit equation (17) using the estimated gross margin.

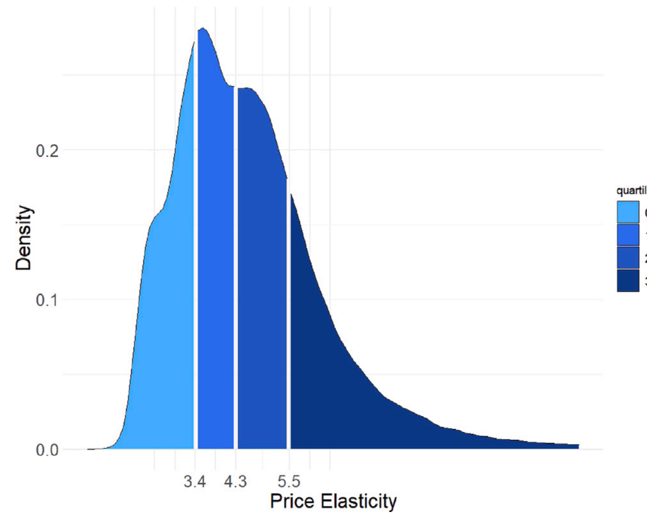

Fig. 3. Distribution of price elasticity for the discrete choice regression for the full population.

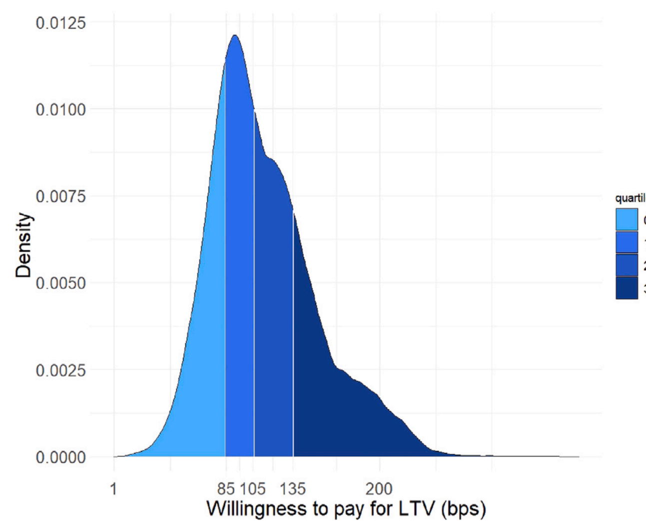

Fig. 4. Distribution of WTP for LTV for the full population.

5. Estimation results

This section presents estimation results for demand, default, and supply parameters. The implied interest rate and product distortions are studied in Section 6.1.

5.1. Demand results

Discrete choice: Demand parameter coefficients are reported in Table 5.

[14] We limit the feasible set to a combination of products with teaser rates of 0, 2, 3 or 5 years, three potential levels of fees (0, 750, 1,500) and buckets of LTV from 60% to 95% by increasing levels of 5%. We only consider one product introduction for each market segment considered. Each time, I chose the product withdrawal or introduction that led to the higher gross margin. When calculated this way, the fixed cots are a bit higher (£25M instead of £16M)

[15] We construct a consistent estimate of $\mathbb{E}[\beta_i^P|\mathcal{I}_P^E, i\ choose\ cb]$ by using Bayes' rule and the estimated preferences coefficients of equation ((5), (6)) to get

$$\hat{\mathbb{E}}[\beta_i^P|\mathcal{I}_P^E, i\ choose\ cb] = \int \beta^P \frac{Prob(\text{i chooses cb}|\mathcal{I}_P^E, \hat{\Theta}^P, \beta_i^P)}{Prob(\text{i chooses cb}|\mathcal{I}_P^E, \hat{\Theta}^P)} dF(\beta^P; \hat{\Omega}^P) \quad (18)$$

$Prob(\text{i chooses cb}|\mathcal{I}_P^E, \Theta^P, \beta_i^P)$ is defined in Eq. (5). $Prob(\text{i chooses cb}|\mathcal{I}_P^E, \hat{\Theta}^P)$ is given by integrating $Prob(\text{i chooses cb}|\mathcal{I}_P^E, \hat{\Theta}^P, \beta_i^P)$ over β^P using the cumulative distribution function $F(\beta^P; \hat{\Theta}^P)$.

There is substantial heterogeneity in the interest rate elasticity, mainly driven by income.[16] The corresponding average own-product demand elasticity is equal to 2.6, 3.6 and 5.1 for borrowers shopping for a 70%–85% LTV loan that are in the first, second, and third quartile of the income distribution (see Table 7). This result implies that on average, a 1% increase in the interest rate decreases the market share of the mortgage by 3.6% for 70%–85% LTV shoppers. Looking at the market share of low-income borrowers only (the first quartile of the distribution), we see that a 1% increase in the interest rate decreases the market share by 2.6%. Fig. 3 depicts the price elasticity distribution for the whole borrower population.

The estimate implies that borrowers with higher income are more sensitive to rates. This can be rationalized by, for instance, search costs as in Agarwal et al. (2024). Borrowers with higher income are more likely to be accepted for any loan contract and thus have more incentives to search intensively. The correlation between income and price elasticity can also be related to the fact that income could be a proxy for other variables such as financial sophistication. Alternatively, this correlation can be rationalized by the direct effect of default probabilities: Borrowers who are more likely to default are also less likely to repay the full face value of the debt and thus end up being less price elastic. As shown in the motivating evidence and in Section 5.2, default is indeed correlated with income.

The average LTV coefficient is 0.17, implying an elasticity of about 8, which means that the average borrower likes high LTV loans. In contrast to the interest rate case, the heterogeneity is mainly driven by the random variable rather than income. This coefficient is significant at the 0.1 level.[17] The first quartile of the distribution is 0.13 and the third quartile is 0.21. However, when considering only observable heterogeneity, we find that the lower quartile of the distribution has an average of 0.16 and the third quartile's average is 0.18. One interpretation of the positive coefficient results is that borrowers do not like to make down payments, because they may be credit constrained. Combining the two coefficients' estimates, we find that 70%–85% LTV shoppers in the first, second and third quantile are, respectively, willing to pay ($\frac{\beta}{\alpha}$) up to 7, 10, and 14 bps for a 1% LTV increase. Fig. 4 depicts the distribution of WTP for LTV for the whole population.

We also find substantial heterogeneity for the teaser rate parameter. The heterogeneity is driven by the random variable term rather than income. This is the only product characteristic that is valued positively by certain borrowers and negatively by others. Fixing rates for a longer period provides a hedge against interest rate increases when borrowers refinance their loan. The interest rate risk, and thus the benefit of fixing rates, can be a result of future changes in borrowers' credit risk or variation in lenders' cost of lending. Consequently, the teaser rate coefficients can be rationalized by borrowers who have different degrees of risk aversion or expectations about the future economic path. This implies that some borrowers prefer a fixed rate and others prefer a flexible rate. Borrowers in the first, second, and third quantile have a coefficient of −0.4, 0.1, and 0.9. Those coefficients imply a WTP of −30, 8, and 50 bps for a 1-year increase in the teaser rate.

The average borrower dislikes fees. There is no observable and unobservable heterogeneity for that coefficient, given the other coefficient for heterogeneity. Borrowers have an average coefficient of $-7 \cdot 10^{-4}$. Those coefficients imply a willingness to pay of 32, 43, and 60 bps for a £1,000 decrease in fees.

Loan demand: Loan coefficients are all significant and reported in Table 8. We find that high LTV increases the amount borrowed by 15%. For the teaser rate, we find that increasing the teaser rate by 0.8%. We further document that borrowers with a high unobserved preference for LTV or a fixed rate also have a higher propensity to borrow. Indeed, borrowers with an unobserved preference for a fixed rate that is one standard deviation higher borrow, on average, 20% more. Borrowers with an unobserved preference for an LTV that is one standard deviation higher borrow, on average, 1.3% more. If those borrowers are also profitable, this creates incentives for banks to create a menu to extract more surplus from them.

5.2. Default results

Default parameter coefficients are significant and reported in Table 9.

We document positive selection along the LTV random variable. For a given level of income and other observable characteristics, borrowers who have an unobserved propensity to choose high LTV products (high $\hat{e}_{LTV}$) that are one standard deviation above the average of the $\hat{e}_{LTV}$ distribution also have a baseline default probability that is twice as low relative to the average borrower (assuming the average is 1.2%). The positive selection along the $\hat{e}_{LTV}$ dimension can be the result of borrowers who are willing to obtain high leverage and thus a bigger house when they know they are less likely to pay the cost of defaulting. This effect goes in the other direction relative to the income effect.

We also document that low-income borrowers are more likely to default and are also more likely to choose a high LTV loan. The latter can be rationalized by a model in which borrowers want to buy the same house size but have different amounts of savings due to their different income levels. In the UK mortgage market, because there is no individual-based pricing, this correlation between observable characteristics and default drives adverse selection.

The correlation of the random variable associated with unobserved LTV preferences implies positive selection, and the one with income implies adverse selection. If the positive selection effect dominated, screening along the LTV dimension would not be possible as screening requires a positive correlation between borrowers' willingness to pay and the cost of offering them the product (see, for instance, Rothschild and Stiglitz (1978)). In the counterfactual section (Section 6), we show that contracts are distorted along the LTV dimension, which implies that adverse selection dominates.[18] The net effect is exactly given by looking at the difference between the perfect information benchmark and the data menus (done in Section 6.1).

As explained in the demand section, longer teaser rates hedge borrowers against changes in interest rates. Variation in future rates can be a result of, for instance, general economic conditions or borrower-specific credit risk changes. Borrowers who prefer higher teaser rates are thus likely to be more risk averse or to see their credit score decrease (and thus their refinancing rate goes up). Those two channels imply opposite predictions regarding adverse or advantageous selection. Indeed, theoretically, borrowers who are highly risk averse are less likely to default. In contrast, private information about a credit risk interpretation will likely lead to adverse selection along the teaser rate dimension. Indeed, borrowers with private information about their credit risk's being likely to go up over time are more likely to fix their contract terms. Those borrowers are also more likely to default.

Our estimates imply mild positive selection along the teaser rate dimension. Indeed, borrowers who are one standard deviation above the mean are 2% less likely to default. These results suggest that the risk aversion channel dominates. This interpretation is also consistent with the loan regression results, which show that those customers tend

[16] The other source of heterogeneity arising from the observable heterogeneity and the random variable terms are non-significant (statistically and economically).

[17] The income interaction term is not significant and has almost no impact on the parameter.

[18] Indeed, if high LTV contracts are more valuable for high-default borrowers, lenders can make them self-select into a high interest rate-high LTV contract. However, maintaining borrowers' incentives requires that low-default borrowers get a lower LTV contract than high-default borrowers, which is not necessarily what would happen in the first best.

to borrow more: Those borrowers are less likely to lose their house and thus benefit more from each extra unit of house bought. However, the fact that the teaser rate coefficients are low may be a result of both channels being present.

5.3. Marginal costs and fixed cost results

Marginal costs: Results are reported in Table 10. We find that the average marginal cost is 220 bps. Scaled up by a default probability between 0% and 5%, this implies an average fair price of between 220 and 231 bps. Marginal costs are increasing in LTV in a convex fashion. While the average marginal cost increases by 10 bps between 70% and 80% LTV loans, it increases by 110 bps between 90% and 95% LTV loans. Longer teaser rate products are more expensive to produce. One year longer costs 4 bps at a low level but 14 bps per year above the fifth level. Finally, higher fee products are associated with lower marginal costs. A 500£ fee increase is associated with a marginal costs decrease of 10 bps starting from a zero fee product. This decrease is even bigger for higher fee products.

Fixed costs: Results are reported in Table 11. The fixed cost is obtained by taking the ratio of θ over β in Table 11.

We find that the average fixed costs of introducing a new product are between £40,000 and £180,000 per product, depending on the assumption about refinancing. This represents about 2% of current profits. Those costs are equivalent to an interest rate decrease of 0.2 bps on each contract in a given market segment and represent 0.1% of the revenue generated. Around 30% of the fixed cost is recovered after the withdrawal of an existing product. As a comparison, Wollmann (2018) finds fixed costs in between $5 and $25 million for the car industry. The difference makes economic sense since the fixed costs of setting up or dismantling a production line for trucks is likely to be much higher than adding or scraping a mortgage product.

The difference in the estimates depends on assumptions about whether the borrower stays at the bank or refinances the contract elsewhere.[19] The conservative estimate (£40,000) is obtained assuming profits are zero after the teaser rate period (i.e. f is set to be the teaser rate period in Eq. (11)). For the less conservative estimate (£180,000), we assume the lender can extract as much surplus with the refinancing contract as the first-time buyer contract and set the number of periods to the loan's maturity (i.e., f in Eq. (11) is set to the loan maturity, which is on average 30 years).

The estimates are those implied by the model to justify that banks offer a discrete number of products and should not be interpreted as a monetary cost. The sunk cost includes monetary costs such as marketing expenses, updates of the menu on all lending platforms, and changes in risk-weight calculations. They also include nonmonetary costs such as within-firm managing frictions.

Tables 12 and 13, 14 provide additional robustness checks. Tables 12 and 13 show that the size of the fixed cost is mainly driven by the 90–95 LTV market segment that features a lower number of products and is also the most profitable. Estimates using only the 90–95 market segments yield fixed costs that are 10 times higher.

Table 14 shows that the fixed costs are multiplied by four when allowing for a simultaneous instead of a two-stage game; the intuition is that it is more profitable to deviate when competitors keep interest rates than when they optimally adapt their price following the deviation.

6. Counterfactual analysis

In the following sections, we use simulations to provide a measure of product distortions relative to the perfect information benchmark (Section 6.1) and to the social planner benchmark (Section 6.2). The later allows us to calculate the cost of the contractual externality.

We propose two well-behaved benchmarks to mitigate the two key challenges related to computing the counterfactuals. The first challenge stems from the multiplicity of equilibria in models of products or firms entry with fixed cost (see, for instance, Eizenberg (2014)).[20] The second one is the difficulty of characterizing the equilibrium in competitive screening models because the equilibrium may not exist in pure strategy (see, for instance, Rothschild and Stiglitz, 1978 or Azevedo and Gottlieb, 2017).

6.1. Product and interest rate distortions

This section compares perfect information benchmarks equilibrium contracts to the data to quantify the interest rate and product distortions.

6.1.1. Conceptual framework

As well known in the contract theory literature (see, for instance, Rothschild and Stiglitz, 1978), perfect information contracts may not be incentive compatible. To illustrate this point, let us consider that the first-best contracts consist of contracts with the same LTV but with different interest rates to reflect borrowers' heterogeneous default probabilities. If lenders were to offer the first-best contracts in a menu, all borrowers would choose the lower-rate contract.

Thus, to maintain incentives to self-select, lenders must distort contract terms away from their first best value. If high LTV contracts are more valuable for high-default borrowers, lenders can distort interest rates and LTVs to make the low-rate contract less attractive to high-default borrowers.

One option is to lower the LTV on the low-rate contract. This works because choosing a lower LTV is relatively more costly to high-default borrowers, given that they have a larger willingness to pay. We call this channel product distortions.

Another option is to decrease the interest rate of the high-rate contract—or, more generally, lower the interest rate spreads between contracts. We call this channel interest rate distortions or cross -subsidies. In the extreme situation in which only interest rate distortions are used, this leads to a single pooling contract.

6.1.2. Perfect information benchmark

We characterize the product and interest rate distortions using different approaches.

We characterize product distortions by comparing the menu offered in the data with the one offered under perfect information. We consider that under perfect information, all the random variables except the demand shocks are observable by the lender. The demand shocks allow for imperfect competition as Taburet (2024). The model is presented in Appendix G.

Because the fixed cost estimates heavily depend on the modeling assumptions (see Section 5.3), we abstract away from them and use a conservative approach instead. We first construct a continuous marginal cost function by interpolating our marginal cost estimates using a spline interpolation. Then, we solve for the optimal contract characteristics for each borrower using the model first-order conditions (see Appendix G). If the result gives a contract characteristic in between

[19] This assumption impacts the estimated profits by the number of periods (as in Eq. (33) in Appendix F).

[20] A multiple equilibrium problem arises when moving away from the two above-mentioned benchmarks to, for instance, simulate the impact of a change in competition or a ban in high LTV. The literature uses Lee and Pakes (2009) algorithm as in Wollmann (2018). However, the algorithm is known to be non-stable in some instances. The logit functional form developed in this paper allows to assign a probability to each potential equilibrium. This approach is feasible and could be a good alternative to Lee and Pakes (2009) algorithm when the set of players and strategy is small enough.

the discrete values observed in the data (e.g., an LTV of 83 while only LTVs of 75, 80 and 85 are observed in the data) we set the counterfactual characteristic to its discrete value closer to the data equilibrium value (e.g., 85 if the borrower had chosen 85 in the data, or 80 if he had chosen 75 or 80).

The interest rate distortions are measured holding the other contract terms constant, and are recovered using a decomposition of the imperfect information structural model described in Appendix H. Formally, we use the imperfect information structural model's first-order conditions to decompose the interest rate into a perfect information perfect competition price, a perfect information markup, and an asymmetric information discount or premium (i.e., the amount of cross-subsidy generated by adverse selection). Again, perfect information refers to the situation in which all the random variables except the logit demand shocks are observable.

Holding the contract characteristic constant is a relevant metric for describing the distortions, and it also allows the results to be independent of the fixed cost value. In addition, the different components of the formula are functions of model parameters and the data and do not require additional simulations. As a result, there is no equilibrium multiplicity problem or computational burden. The cost of this approach is that we can only recover the average perfect information interest rate for borrowers who chose the same contract under imperfect information.

6.1.3. Product distortions: results

Our results imply that maintaining borrowers' incentives to self-select requires distorting contract terms away from their perfect information value. Because high-default-low price elastic borrowers have high WTP for LTV, low-default-high price elastic borrowers get a lower LTV and thus a lower house size under imperfect information.

We find that more than 90% of borrowers shopping between 70 and 95% LTV would get an 85%–95% LTV product under perfect information-perfect competition (see Table 15). This finding suggests that products below 85% LTV are introduced to screen rather than to cater to borrowers' heterogeneous preferences. We exclude borrowers shopping for below 65% LTV, because they constitute less than 10% of the loans originated, and the data quality is lower for that subsample.[21] Our benchmark does not endogenize house prices and does not feature any risk associated with having a portfolio composed of high-leveraged loans only. The results should be thus interpreted as a comparative static, holding those elements constant.

Our results are robust to the use of models with observable heterogeneity and estimating the coefficient separately for each sufficient set.[22] The amount of product distortion relative to the perfect information situation is accentuated when moving away from perfect competition. Finally, the result is robust to changing the fact that a higher LTV decreases default. One might be worry that this sign results from banks that select good borrowers for high LTV loans based on soft information not observable by the econometrician. However, the LTV coefficient of the default regression would need to be positive and 100 times larger in absolute value to imply that 10% of borrowers get offered lower than 90% LTV products. Given the standard error of $2.8 \cdot 10^{-6}$ and the average coefficient of $-3.9 \cdot 10^{-5}$ on the LTV coefficient, this situation is not likely.

As summarized in Table 15, we find distortions in fees and teaser rate periods. The model implies that more products should be offered—in particular, higher fee products (more than £1500) and longer teaser rate periods (longer than 7 years). The share of the population that would like to get them is low (below 20% of the 80+ LTV borrowers). In addition, this result highly depends on how the marginal costs of lending vary with fees and the teaser period. Since marginal costs are estimated for products with fees ranging from 0 to 1500 and teaser rates from 0 to 7, product introduction results are highly dependent on our extrapolation of the marginal cost function. We find that the distribution of borrowers would shift toward lower-fee products and more flexible rate contracts. This is the result of interest rate distortions. Those distortions are analyzed in Section 6.1.3.

6.1.4. Interest rate distortions: Results

Results for the interest rate decomposition are summarized in Table 18, Table 17, and Fig. 5. Doing this decomposition, we find that the average fair price is 231 bps, the markup is about 116 bps, and the average information rent is −70 bps for high LTV loans (above 80). For loans with LTV between 70 and 80, the average fair price is 202 bps, the markup is about 60 bps, and the average information rent is −30. These differences across LTV are mainly due to the fact that lower LTV loans are chosen by borrowers who are more price elastic on average. As a result, banks are less able to apply large interest rates or large information rents. The impact of default is mild when explaining the interest rate level. For instance, the difference between the effective marginal cost and the marginal cost is on average less than 5 bps (and less than 10 bps when we scale up all default probabilities by 5 to take into account that the estimated default probabilities may underestimate banks' true default expectations). However, even a mild difference in default can lead to big product distortions when the screening device is not very effective.

Looking at the differences in the average information rent between different products, we find that high LTV products (95% LTV) earn low information rents (5 bps) compared with 75% LTV products. This is because high LTV products are also more expensive to produce, which implies that the information rent need not be large. This result is also consistent with the fact that banks maintain incentives to self-select by distorting the LTV rather than rates. Contrarily, we find that lower fee contracts and longer rate contracts get a substantial information rent. This can be explained by the fact that high-fee products are chosen by more price elastic borrowers. Under perfect information those borrowers would thus get a lower markup (see markup columns in Tables 17 and 18). To be able to extract more surplus from other borrowers, banks make high-fee products relatively more expensive than what they should be. This is consistent with the product distortion and the shift in the low-fee products category observed under perfect information: Banks increase rates in low-fee products to extract more surplus from the low price elastic borrowers, and as a result more price elastic borrowers are pushed to high-fee products when they exist. This creates incentives to introduce more high-fee products relative to the first best in order to implement screening.

Longer teaser rate products are more expensive to produce and are chosen by less price elastic borrowers. Under perfect information, those borrowers would get a higher markup. Those products also benefit from an information rent.

6.1.5. Summary of results and economic interpretation

Our estimates imply that, in the perfect information case, borrowers in the first and last WTP quartile of the LTV distribution would get contracts with similar LTVs — respectively, 85% and 95% — and be charged different prices because of their heterogeneous price elasticity and default probabilities. As a result, a menu composed of perfect information contracts cannot be offered under imperfect information, since high default-low price elastic borrowers would be tempted to choose lower rate contracts. This creates incentives to decrease the interest rate on high LTV contracts (i.e., an asymmetric information discount, which is also called information rent in monopoly models) and increase the interest rate on low LTV contracts (i.e., an asymmetric

[21] Including them would imply that LTV between 50% and 75% would be introduced but would account for less than 5% of market shares.

[22] As the unobservable heterogeneity uses a normal random variable, there is always a mass of borrowers with very low WTP for any characteristics. However, borrowers who will choose lower than 90% LTV in the heterogeneity case account for less than 5% of the population.

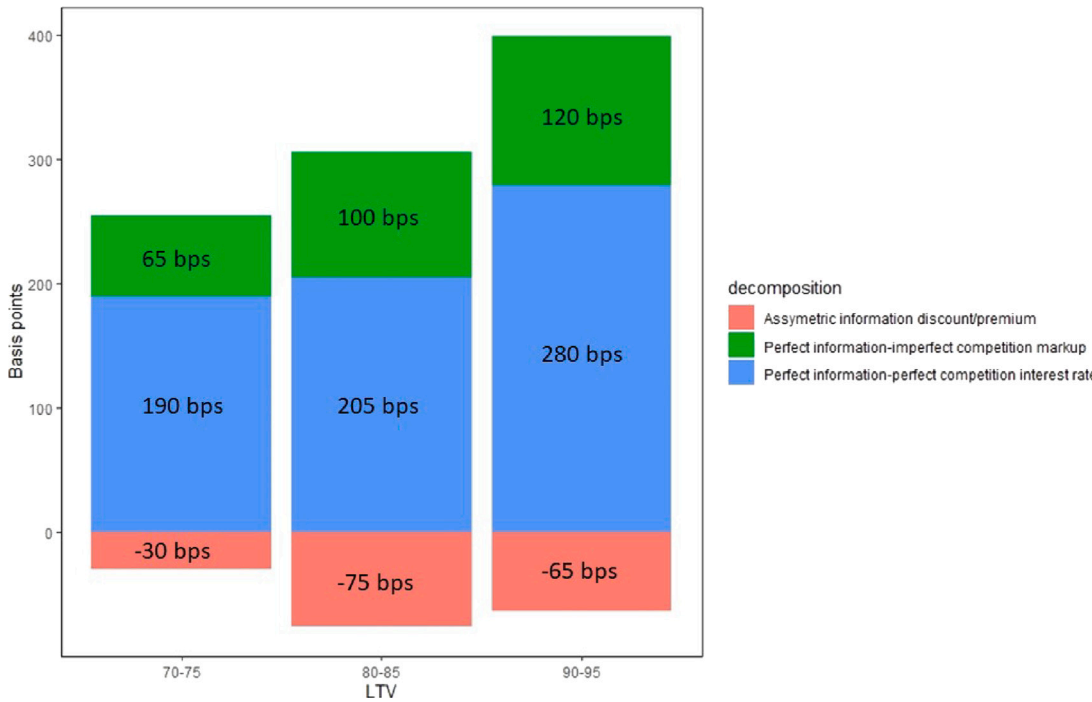

Fig. 5. Interest rate decomposition by LTV.

information premium) relative to the perfect information case. As a complementary incentive, lenders also introduce LTV contracts that are lower than 85%. Since high default-low price elastic borrowers are more reluctant to provide higher down payments for each loan unit, low LTV contracts attract unobservably safer borrowers and can be offered at a lower price.

These results imply that welfare is lower relative to the perfect information-perfect competition case. The overall loss in borrowers' utility in the current data is equivalent to the loss in utility following a 100 bps interest rate increase on all loans.

The perfect information-imperfect competition case is not a natural benchmark to study welfare, given that asymmetric information and imperfect competition interact. Removing one friction can thus increase the other. For instance, by removing asymmetric information, lenders are able to set a higher interest rate (70 bps) for high LTV contracts without the fear of borrowers' substituting to a lower LTV contract designed to attract safer borrowers.

Reducing the level of asymmetric information or allowing lenders to price borrowers on all observable characteristics, such as ethnicity, gender, disability, or religious beliefs, may not be feasible or desirable. As a result, it is also relevant to look at how far the product offered is from the second best (i.e., the menus offered by an informationally constrained social planner). This is the purpose of the following section.

6.2. Quantitative analysis of the screening externality

This section compares social planner benchmarks equilibrium contracts to the data to quantify the contractual externality.

6.2.1. Conceptual framework

Screening allows lenders to (partially) restore perfect information pricing at the cost of contract term distortions relative to the perfect information case. Yet, due to a contractual externality, socially optimal menus may not be offered (see Rothschild and Stiglitz, 1978, Dasgupta and Maskin (1986) or Taburet (2024) for characterization of the equilibrium). The friction emerges because lenders do not internalize how their screening strategies change the types of borrowers who select their competitors' products—and thus the cost of lending via those products.

To illustrate this point, let us consider a perfectly competitive market in which screening is achieved by making high-default borrowers to self-select a high-rate contract because lower-interest-rate contracts contain features that are relatively more costly for them, such as a low LTV. When the LTV distortions needed to screen borrowers are high, pooling at least some types of borrowers is a Pareto improvement over screening. This happens because with pooling contracts, high-default borrowers get a lower rate and low-default borrowers are less credit constrained. Yet, if a lender prices its customers using the average default probability (pooling), a competitor can use the lender pooling strategy to introduce contracts that will steal only low-cost customers (cream skimming) by offering a low rate-low credit constraint contract.

6.2.2. Social planner benchmark:

To capture the contractual externality, we solve for the contract under the following specification (See Appendix I for the formal model description). We fix the customers of each bank and look at whether a Pareto-improving menu exists. Fixing the market share eliminates the externality by preventing borrowers from moving from one bank to another. It also allows us to focus exclusively on the screening externality by preventing an increase in welfare generated by a better allocation of borrowers to more productive banks. Because we shut down competition, we add a participation constraint (PC) to prevent the lender from extracting too much surplus from borrowers. The constraint ensures that the new contracts offered are Pareto improvements over the contract observed in the data.

The benefit of setting the social planner benchmark this way is that it becomes similar to the textbook monopolistic screening model, which does not feature the equilibrium problems in competitive models such as, for instance, Rothschild and Stiglitz (1978). This allows us to overcome the issues related to solving for a potentially mixed-strategy equilibrium (see, for instance, Lester et al. (2019)) or the need to use equilibrium refinements to solve screening models (see, for instance, Handel et al. (2015), who rely on Riley (1979) equilibrium

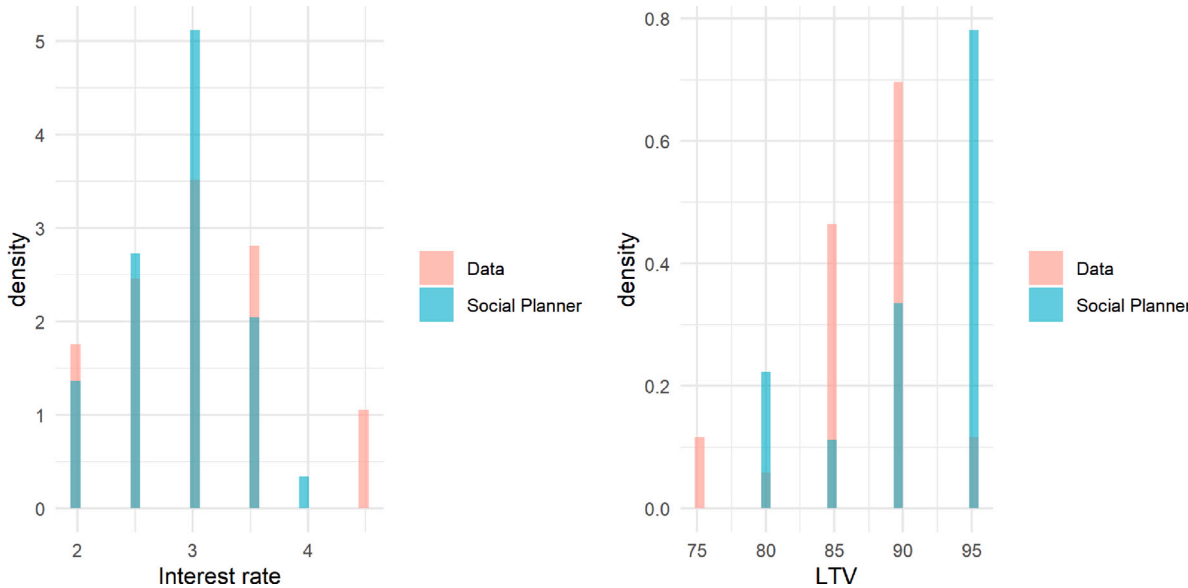

Fig. 6. Data and social planner simulation distribution of the equilibrium interest rate and LTV distribution.

concepts, which forces the screening equilibrium to occur). The multiplicity arising from the fixed cost is also mitigated as we can solve each bank problem separately for each menu in the bank feasible set.[23]

We define social welfare as the sum of firms' profits plus the sum of borrowers' utility expressed in monetary terms. We measure the cost of the screening externality by comparing the utilitarian social welfare level implied by our structural model with one achievable in a benchmark in which the contractual externality is internalized.

6.2.3. *Summary of results and economic interpretation*

As illustrated by Fig. 6, the counterfactual simulation shows that the social planner could achieve a Pareto improvement by pooling more borrowers at higher LTV. Low-default borrowers are better off because they can buy a larger house, and high-default borrowers benefit from being pooled by getting a lower interest rate. Lenders are also better off because lower LTV distortions imply that the surplus generated by the lending activity is larger, and they are thus able to extract more surplus and increase their profits.

We find that despite the low spread between defaults, the cost of the screening externality is quite large. The deadweight loss associated with the externality is equivalent to the loss in borrowers' utility following a 32 bps increase in interest rates for all contracts.

The results are robust to using the fixed cost approach or the more conservative one that abstracts from the fixed costs (See Appendix I for a description of the two approaches). The similarity comes from the optimal menu mainly involving the removal of some products to pool borrowers, making the fixed costs less relevant. Fig. 7 reports the equilibrium distributions. The differences come from the lower inclusion of products in the 95 LTV market segments with the fixed cost approach.

This finding suggests that there is room for Pareto to improve policy interventions. As shown in the theoretical companion paper by Taburet (2024), lowering competition, increasing the capital requirement for low LTV in a low-competition environment, or banning the use of lower LTV products could reduce the impact of the contractual externality by preventing cream-skimming deviations to occur. However, our model focuses on asymmetric information distortions and does not explicitly model other frictions. For instance, deposit insurance could lead banks to underestimate the lending risk via higher LTV. This friction would then lead to too much leverage in the mortgage market instead of too little leverage. As a result, policy interventions should consider both frictions before implementing a low LTV ban.

This paper focuses on first-time buyers to reduce concern that lenders have superior information than their competitors and the econometrician. For the following reasons, I expect the screening externality to be lower in the remortgaging market. First, lenders learn about their customers over time, so I expect the information asymmetry to be lower in the population of mortgage borrowers. In addition, borrowers may have higher switching costs in the refinancing market. The combination of lower competition and lower asymmetric information is likely to lower the contractual externality

7. Conclusion

This paper provides the first analysis of product and price distortions in the context of credit markets in which menus of contracts are used. We do so by developing and identifying a novel structural model of screening. Thanks to the tools developed in this paper, We provide the first quantification of the contractual externality and show that there is room for policy interventions.

We develop a novel identification strategy to test whether screening for default probability is possible. Along the way, we discuss how to adapt classic structural models for the banking market. Those changes are guided by the fact that financial markets are not a classic IO market in many regards. For instance, contrary to a traditional IO market, the quantity (loan size) of products being sold to a given borrower may be limited by sellers, sellers may not agree to sell borrowers some products (i.e., rejection of loan applications), and the market is likely to feature adverse or positive selection.

The paper's second innovation is to propose a new set of tools to analyze the impact of screening on product and price distortions. Instead of using the classic counterfactual analysis — for which the technical properties (equilibrium uniqueness) have not been fully analyzed by the literature in the context of multiple endogenous variables — we propose a new, complementary approach. We first use perfect information and a well-behaved model as a benchmark to analyze product distortions. Second, we use a sufficient statistic approach to decompose equilibrium interest rates into a fair price, a perfect information markup, and an asymmetric information premium or discount.

[23] To lower the computational burden, we restrict the feasible set for LTV to contract just above and below the one currently offered to the borrower.

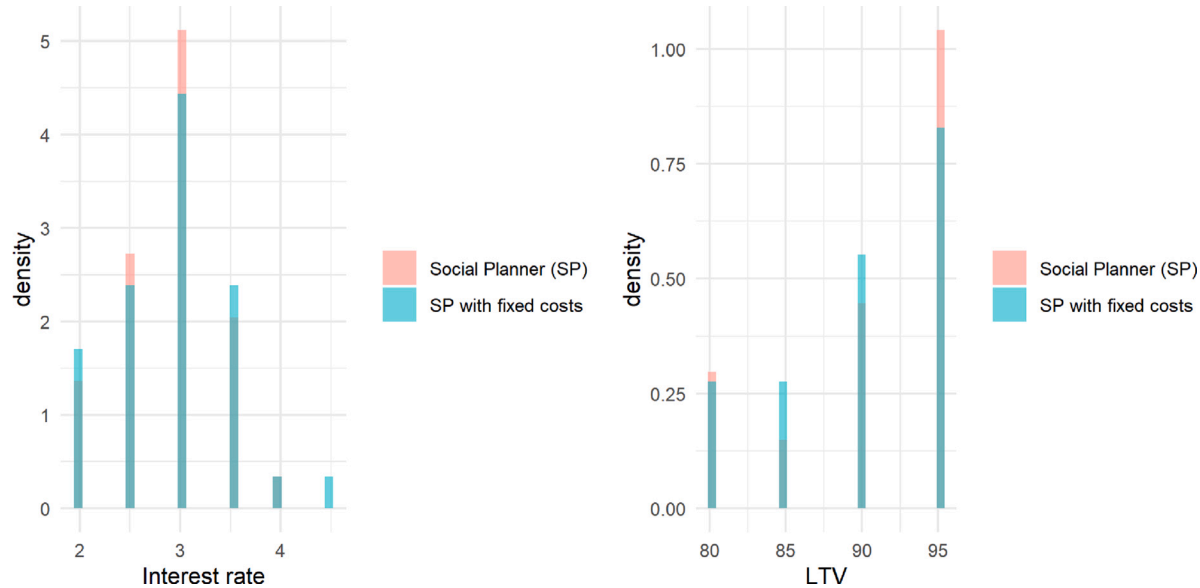

Fig. 7. Social planner distribution of the equilibrium interest rate and LTV distribution with two different approaches for the fixed costs.

Table 1
Summary statistics for 2018.

Variable	Mean	SD	Min	Max
Loan Characteristics:				
Max LTV (%)	82.5	10.8	50	95
Teaser rate period (years)	3.3	1.6	0	7
Maturity (years)	29.7	5.7	8	40
Fees (£)	503	631	0	2610
Rate (%)	2.5	0.8	1.1	8
Loan amount (£ 1000)	164	129	35	864
Borrower Characteristics:				
Household income (£ 1000)	36	16	25	944
Loan applicants	1.56	0.5	1	2
Age (years)	31	7	18	75
Loan to Income	4.6	1.2	1.1	6.1
N	847,379			

CRediT authorship contribution statement

Alberto Polo: Resources. **Arthur Taburet:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Quynh-Anh Vo:** Resources.

Declaration of competing interest

Arthur Taburet declared that he has no financial and personal relationships with other people or organizations that could inappropriately influence or bias my work.

Appendix A. Tables

A.1. Descriptive statistics

See Tables 1–4.

A.2. Estimation results

See Tables 5–14.

A.3. Counterfactual results

Counterfactuals are presented in Appendices G and H (see Table 16).

Appendix B. Menu adjustment costs: Dynamic approach

Let us denote M_{t-1} the menu offered by a bank in the previous period. The probability of observing menu M in period t when menu M_{t-1} was offered in the previous period is:

$$Pr(d_{jt}(M, M_{t-1})),\ with\ d_{jt}(M, M_{t-1}) := \mathbb{1}_{\{V_{jt}(M) - F(M, M_{t-1}) \geq V_{jt}(M_{t-1}) + e_{Mtj} - e_{M_{t-1}tj}\}}$$

F is the fixed cost of changing the menu. $V_{jt}(M)$ is the value function of the bank j at time t when offering menu M. (e) are error terms. Formally, the value function is:

$$V_{jt}(M_{t-1}) = \max_{M \in \mathcal{M}_j} \underbrace{\Pi_j(M) - F(M, M_{t-1}) + \beta E[V_{jt+1}(M)]}_{v(M, M_{t-1})} + e_{Mtj}$$

Where $\Pi_j(M)$ is the lender's gross margin and β is a discount factor.

With (e_{Mtj}) being iid and extreme value distributed, the probability $Pr(d_{jt}(M, M_{t-1}))$ is:

$$Pr(d_{jt}(M, M_{t-1})) = \frac{exp(u_{\tilde{M}}(M, M_{t-1}))}{1 + \sum_{m \neq \tilde{M}} exp(u_{\tilde{M}}(m, M_{t-1}))} \tag{19}$$

with:

$$u_{\tilde{M}}(M, M_{t-1}) := \Pi(M) - \Pi(\tilde{M}) - [F(M, M_{t-1}) - F(\tilde{M}, M_{t-1})] - \beta \underbrace{[log(Pr(d_{jt}(\tilde{M}, M))) - log(Pr(d_{jt}(\tilde{M}, \tilde{M})))]}_{Observable\ in\ data} \tag{20}$$

The term $(log(Pr(d_{jt}(\tilde{M}, M))) - log(Pr(d_{jt}(\tilde{M}, \tilde{M}))))$ comes from using the shock being extreme value distributed and rewriting the value function as in Arcidiacono and Miller (2011).

Appendix C. Demand with Roy's identity

This section derives the demand equations by specifying an indirect utility and using Roy's identity to derive the loan demand.

Table 2
Regression LTV on borrowers' characteristics.

Variable	Age	Yearly net income	Number of borrowers	Self employed
Below 60% LTV	33***	39,855	1.35***	0.085***
60%–70% LTV	−0.7***	−82	0.04***	0.01***
70%–75% LTV	−1.5***	3,675***	0.007***	−0.005***
75%–80% LTV	−1.3***	1,793*	0.11***	0.006***
80%–85% LTV	−1.7***	1,941**	0.16***	0.007***
85%–90% LTV	−2.4***	−2,716***	0.22***	−0.024***
95+ LTV	−2.7***	−3,842***	0.28***	−0.06***
N	1,077,291	1,077,291	1,077,291	1,077,291

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

Table 3
Mortgage holiday take up and arrears. A mortgage holiday is a payment deferral (up to 6 months).

	Mortgage holiday by 2021	Arrears by 2020 (Origination: 2018)
Interest (in%)	12***	5.8***
LTV> 90	−3.5	−1.4***
Fixed rate period (years)		−0.9***
Lender fees		$3.7 \cdot 10^{-3}$***
Income		$-1.2 \cdot 10^{-4}$***
Nb applicants		−3.9***
Age		$6.7 \cdot 10^{-2}$*
LTI		−1.4***
Time fixed effect	No	Yes
Bank fixed effect	Yes	Yes
Region fixed effect	No	Yes
Mean	26%	1.2%
Observations	53	279,379

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.
Notes: All the coefficients of the first column need to be rescaled by $\cdot 10^{-2}$ and by $\cdot 10^{-3}$ in the second column.

Table 4
Most common product characteristics.
Source: PSD001 + Moneyfact.

Variable	2019	2021
	high LTV (95)	
Average number of products (rounded)	8	0–2
Fixed rate period (years)	(5,3,2,0)	5-year more likely
Average lender fees (rounded)	(0, 750)	high fees more likely
	medium LTV (75–85)	
Average number of products (rounded)	12	16
Fixed rate period (years)	(5,3,2,0)	(5,3,2,0) + longer fixed rates
Average lender fees (rounded)	(0, 750, 1450)	(0, 750, 1450)

Table 5
Mixed logit (Origination: 2018).

	85 + LTV loans	70%–85% LTV loans
Interest rate (%)	−54	−7.1
	(50)	(44)
LTV (%)	23***	21***
	(1.2)	(5)
Fixed rate period (years)	−78*	−18
	(40)	(19)
Lender fees (pounds)	$-9 \cdot 10^{-2}$***	$-7 \cdot 10^{-2}$***
	(1.610^{-2})	$(5 \cdot 10^{-3})$
Interest rate ×Yearly Net Income (pounds)	$-4.5 \cdot 10^{-3}$***	$-3.2 \cdot 10^{-3}$***
	$(1.1 \cdot 10^{-5})$	$(1.7 \cdot 10^{-5})$
Standard deviation random variable Fixed rate period	250***	100***
	(48)	(27)
Standard deviation random variable LTV	24***	4.8***
	(2.7)	$(2 \cdot 10^{-1})$
Region-Age-Nb applicants interaction terms for all product characteristics	Yes	Yes
Interest rate- Fixed rate period-fees random variable	Yes	Yes
Observations	279,379	230,680

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.
Notes: All the coefficients of the first column need to be rescaled by $\cdot 10^{-2}$.

Table 6
Coefficient heterogeneity.

	Interest rate (%)	LTV (%)	Teaser rate (year)	Fees (pounds)
		85+ loans		
Observable heterogeneity only				
First quartile	−11	2.3	−7.8	$-8 \cdot 10^{-3}$
Second quartile	−8.6	2.3	−7.8	$-8 \cdot 10^{-3}$
Third quartile	−6.3	2.3	−7.8	$-8 \cdot 10^{-3}$
Observable and unobservable heterogeneity				
First quartile	−11	1.5	−2.4	$-8 \cdot 10^{-3}$
Second quartile	−8.6	2.3	−7.8	$-8 \cdot 10^{-3}$
Third quartile	−6.3	3	9.2	$-8 \cdot 10^{-3}$
		70–85 loans		
Observable heterogeneity only				
First quartile	−23	1.6	1.5	$-7 \cdot 10^{-3}$
Second quartile	−19	1.7	1.5	$-7 \cdot 10^{-3}$
Third quartile	−15	1.8	1.5	$-7 \cdot 10^{-3}$
Observable and unobservable heterogeneity				
First quartile	−23	1.3	−4.3	$-7 \cdot 10^{-3}$
Second quartile	−19	1.7	1.5	$-7 \cdot 10^{-3}$
Third quartile	−15	2.1	9.1	$-7 \cdot 10^{-3}$

Notes: All the coefficients of the first column need to be rescaled by $\cdot 10^{-1}$.

Table 7
WTP and elasticity heterogeneity.

	Price elasticity	WTP LTV (%)	WTP teaser rate (year)	WTP fees (pounds)
		85+ loans		
Observable heterogeneity only				
First quartile	28	1	−6.5	$-6.6 \cdot 10^{-3}$
Second quartile	39	1.4	−4.9	$-5 \cdot 10^{-3}$
Third quartile	54	1.9	−3.7	$-3.8 \cdot 10^{-3}$
Observable and unobservable heterogeneity				
First quartile	28	0.5	−1.1	$-6.6 \cdot 10^{-3}$
Second quartile	39	1.3	−4.4	$-5 \cdot 10^{-3}$
Third quartile	54	2.4	5.3	$-3.8 \cdot 10^{-3}$
		70-85 loans		
Observable heterogeneity only				
First quartile	26	0.8	0.7	$-6 \cdot 10^{-3}$
Second quartile	36	1	0.9	$-4.3 \cdot 10^{-3}$
Third quartile	51	1.4	1.2	$-3.2 \cdot 10^{-3}$
Observable and unobservable heterogeneity				
First quartile	26	0.7	−3	$-6 \cdot 10^{-3}$
Second quartile	36	1	0.8	$-4.3 \cdot 10^{-3}$
Third quartile	51	1.4	5	$-3.2 \cdot 10^{-3}$

Notes: All the coefficients of the first column need to be rescaled by $\cdot 10^{-1}$.

Table 8
Loan demand (Origination: 2018).

	log(Loan size)	log(Loan size)
Interest rate (%)	-520^{***}	-520^{***}
	(3.9)	(3.9)
LTV (%)	0.9^{***}	0.8^{***}
	(0.2)	(0.4)
LTV=95 (%)	76^{***}	150^{***}
	(7.3)	(21)
Fixed rate period (years)	-1.7^{*}	-8.5^{***}
	$(9 \cdot 10^{-4})$	$(2.4 \cdot 10^{-3})$
Lender fees (pounds)	$6.5 \cdot 10^{-2***}$	$6.9 \cdot 10^{-2***}$
	$(1.6 \cdot 10^{-3})$	$(1.6 \cdot 10^{-3})$
log(Income) (pounds)	800^{***}	800^{***}
	(4)	(4)
unobserved WTP fixed rate: $\hat{e}_{TR}$ (mean 0 sd normalized to 1)		200^{***}
		(2.3)
unobserved WTP LTV: $\hat{e}_{LTV}$ (mean 0 sd normalized to 1)		80^{***}
		(13)
Lender, Region, time fixed effects	Yes	Yes
Borrowers' characteristics control	Yes	Yes
Borrowers' WTP interaction terms	Yes	Yes
R^2	0.76	0.77
Observations	279,379	279,379

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.
Notes: All the coefficients of the first column need to be rescaled by $\cdot 10^{-3}$.

Table 9
Default regression (mortgage originated in 2018).

	Arrears by 2020	Arrears by 2020
Interest (in %)	5.8***	3.9***
	(0.4)	(0.34)
LTV	−0.2***	0.12
	(2.7)	(2.1)
Fixed rate period (years)	−0.1 ***	−0.6 10*
	(2.7)	(0.16)
Lender fees (in thousands)	3.7	4
	(0.7)	(0.6)
Income (in thousands)	−0.12***	−0.24***
	$(1.1 \cdot 10^{-2})$	$(1.6 \cdot 10^{-2})$
Nb applicants	−3.9***	−3.1***
	(0.3)	(0.2)
Age	0.067*	0.07*
	$(1.1 \cdot 10^{-2})$	$(1.9 \cdot 10^{-2})$
$\hat{e}_{LTV}$ (sd normalized to 1)		−5.4***
		(9.4)
$\hat{e}_{TR}$ (sd normalized to 1)		−0.2***
		(5.1)
Time fixed effect	Yes	Yes
Lender fixed effect	Yes	Yes
Region fixed effect	Yes	Yes
Macroeconomics controls (monthly GDP)	Yes	Yes
Control for loan size	Yes	Yes
Mean	1.2%	1.2%
Observations	279,379	279,379

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.
Notes: All the coefficients of the first column need to be rescaled by $\cdot 10^{-3}$.

Table 10
Marginal costs regression and interest rate (LTV > 70).

	Marginal costs	Interest rates
Intercept	85***	12***
	(20)	(8.69)
$\mathbb{1}_{LTV<85}$×LTV (%)	12***	14***
	(2.8)	(0.1)
$\mathbb{1}_{LTV>85}$×LTV (%)	18***	20***
	(1.5)	(0.1)
95% LTV (dummy)	98***	120***
	(91)	(2.1)
Fixed rate period (years)	$40 \cdot 10$**	44**
	(10)	$(5.3 \cdot 10^{-1})$
High Fixed rate period (≥5)	180***	23010***
	$(5 \cdot 10^{-2})$	$(1.6 \cdot 10^{-3})$
Lender fees (pounds)	−0.2***	−0.4***
	$(1.8 \cdot 10^{-2})$	$(5.7 \cdot 10^{-4})$
High fees (1000–1500)	−100*	−130***
	(40)	(2.7)
Bank fixed effect	Yes	Yes
Average	2.12	2.42
N	278	647,433
R^2	0.88	0.76

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.
Notes: All the coefficients of the first column need to be rescaled by $\cdot 10^{-2}$.

Table 11
Fixed cost results: all market segments.

	NPV using teaser rate	NPV using maturity
Profits (β)	2.3***	4.47***
	$(1.4 \cdot 10^{-2})$	$(6.37 \cdot 10^{-2})$
Nbr of Product included (θ)	$8.8 \cdot 10^{4***}$	$8.3 \cdot 10^{5***}$
	$(3.6 \cdot 10)$	$(4.5 \cdot 10)$
Nbr of Product excluded ($\theta \cdot \lambda$)	$-1.4 \cdot 10^{4***}$	$-2.4 \cdot 10^{5***}$
	$(2.6 \cdot 10)$	$(2.1 \cdot 10)$
Bank fixed effect	No	No
Time fixed effect	No	No

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.
Note: In the first column of Table 11, the NPV is calculated using the teaser rate period as f in Eq. (33), in the second column the maturity is used instead.

Table 12
Fixed cost results: above 90 LTV.

	NPV using teaser rate
Profits (β)	3.7***
	$(4.9 \cdot 10^{-1})$
Nbr of Product included (θ)	$1.1 \cdot 10^{6***}$
	$(9 \cdot 10^{4})$
Nbr of Product excluded ($\theta \cdot \lambda$)	$-1.8 \cdot 10^{5***}$
	$(5 \cdot 10^{4})$
Bank fixed effect	No
Time fixed effect	No

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 13
Fixed cost results: below 90 LTV.

	NPV using teaser rate
Profits (β)	3.8***
	$(1.6 \cdot 10^{-2})$
Nbr of Product included (θ)	$4.2 \cdot 10^{3***}$
	$(3.3 \cdot 10^{1})$
Nbr of Product excluded ($\theta \cdot \lambda$)	$-2.7 \cdot 10^{3***}$
	$(1.5 \cdot 10^{1})$
Bank fixed effect	No
Time fixed effect	No

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 14
Fixed cost results: simultaneous game.

	NPV using teaser rate
Profits (β)	6.1***
	$(7.9 \cdot 10^{-2})$
Nbr of Product included (θ)	$9.6 \cdot 10^{6***}$
	$(7.3 \cdot 10^{4})$
Nbr of Product excluded ($\theta \cdot \lambda$)	$-8.1 \cdot 10^{5***}$
	$(5.8 \cdot 10^{3})$
Bank fixed effect	No
Time fixed effect	No

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

C.1. Main

Guided by the micro foundation presented in Appendix C.3, we parameterize the indirect utility derived at the optimal borrowing amount given loan characteristics X and price r as

$$U_i(L_i(X,r);X,r) := A_i(X)\frac{L_i(X,r)}{LTV} + V_i(Y_i), \tag{21}$$

where Y_i is the income of borrower i, A_i is a function of product characteristics X, V_i is a function of income, and L_i is the optimal loan size as a function of product characteristics X and rate r. LTV is the loan-to-value of the contract, so $\frac{L_i(X,r)}{LTV}$ is the house price.

This parameterization is a generalized version of Train (1986). The main departure from Train (1986) is that we allow A_i to be a general function that varies with products' and borrowers' characteristics instead of a constant.

Using Roy's identity, the optimal loan size should satisfy $L_i(X,r) = -\frac{\partial_r U(L,X,r)}{\partial_Y U(L,X,r)}$. In Appendix C.2, we show a parameterization of the function (A) that leads to the following demand system. We index by c a product (X_{cb}, r_{cb}) offered by bank b and relabel $L_{icb} := L(X_{cb}, r_{cb})$:

$$V_{icb} = \overbrace{\beta^P_{icb}X_{cb} - \alpha^P_{icb}r_{cb} + \xi_{cb}}^{u_i(c,b)} + \sigma^{-1}_{icb}\varepsilon_{icb} \tag{22}$$

$$ln(L_{icb}) = \beta^L_{icb}X_{cb} - \alpha^L_{icb}r_{cb} + \nu D_i + e^L_{icb} \tag{23}$$

$$with\ (\beta^P_{icb}, \alpha^P_{icb}, \sigma^{-1}_i, \beta^L_{icb}, \alpha^L_{icb}, e^L_{cb})\ correlated,$$

Table 15
Product distortion (80+ LTV loans).

	Ideal LTV (%)	Ideal teaser rate (year)	Ideal Fees (pounds)
Observable heterogeneity only (perfect information+perfect competition)			
First quartile	95	0	0
Second quartile	95	0	0
Third quartile	95	0	500
Observable and unobservable heterogeneity (perfect information+perfect competition)			
First quartile	90	0	0
Second quartile	95	2	0
Third quartile	95	5–7	500
Product choice distribution (data)			
First quartile	85	2	0
Second quartile	90	2	500
Third quartile	95	5–7	1000

Table 16
LTV distortion perfect competition-perfect information benchmark (70+ LTV loans).

Decile	10%	20%	30%	40%	50%	60%	70%	80%	90%
Product choice distribution (data)									
	75	75	80	85	90	90	90	90	95
Benchmark-implied distribution (observable heterogeneity)									
	90	90	95	95	95	95	95	95	95
Benchmark-implied distribution (observable + unobservable heterogeneity)									
	85–90	90	90	90	95	95	95	95	95

Table 17
Interest rate decomposition (70-80+ LTV loans).

	Fair price (bps)	Perfect information mark-up (bps)	Asymmetric Information discount/premium (bps)
LTV	2^{***}	$1 \cdot 10^{-1}$	$2 \cdot 10^{-1*}$
fees (500)	-16^{***}	-9^{***}	6^{***}
fees (1000)	-29^{***}	-20^{***}	13^{***}
fees (1500)	-35^{***}	-30^{***}	17^{***}
teaser rate period (2 years)	-40^{***}	−8	0
teaser rate period (5 years)	-20^{*}	−4	-10^{**}
teaser rate period (7 years)	7	10	-20^{**}
Average	202	65	−30

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 18
Interest rate decomposition (80+ LTV loans).

	Fair Price (bps)	Perfect information mark-up (bps)	Asymmetric Information discount/premium (bps)
LTV	12^{***}	$8 \cdot 10^{-2}$	2^{***}
fees (500)	-12^{***}	-19^{***}	20^{***}
fees (1000)	-35^{***}	-46^{***}	41^{***}
fees (1500)	-46^{***}	-55^{***}	45^{***}
teaser rate period (2 years)	3	15^{***}	-16^{***}
teaser rate period (5 years)	15^{***}	35^{***}	-31^{***}
teaser rate period (7 years)	27^{**}	43^{***}	-40^{***}
Average	231	116	−68

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

where $u_i(c, b)$ is a monotonic transformation of the indirect utility U_i defined in Eq. (21).[24]

[24] $V_i(Y_i)$ is not present, since $argmax_c U_i(L^*, X_c, r_c) = argmax_c U_i(L^*, X_c, r_c) - V_i(Y_i)$. For those who are skeptical about the discrete-continuous approach, we could end up with the same functional form by assuming that borrower i chooses product c and the optimal loan size $L_i(X_c, r_c)$:

$$max_{c \in M_{ib}} u_i(L_i(X_{cb}, r_{cb}), X_{cb}, r_{cb}) + \sigma_i^{-1} \epsilon_{icb}$$

and make the assumption that $L_i(X_c, r_c)$ and $u_i(L_i^*(X_c, r_c), X_c, r_c)$ are linear in contract terms.

C.2. Derivation of the demand system

Indexing each combination of contract characteristic and prices (X, r) by c, borrower i maximization problem is:

$$\max_c u(L_c, c) = \max_c A_{ic} \frac{L_c}{LTV_c} + V(Y_i)$$

A_c captures the fact that default or consumption trade-off depends on contract c features.

The problem is equivalent to:

$$\max_c u(L_c, c) = \max_c ln(A_{ic}) + ln(L_c) - log(LTV_c)$$

We parameterize $ln(A_{ic}) = \beta_i^L X_c + \sigma_i \epsilon_{ic}$. From Roy's identity:

$$\frac{L}{LTV} = \gamma^{-1}\frac{[\partial_{DF_i}\{\frac{L}{LTV}\}]A_{ic}}{V_Y(Y)}$$

Integrating with respect to DF_i (loan discount factor):

$$ln(L_c) = ln(LTV) + \gamma\frac{V_Y}{A_{ic}}DF_i + cst$$

$$with:\ cst := \beta_i X_c + \epsilon_i, with\ (cst\ r)$$

setting $\frac{DF_i}{A_{ic}}$ to be a linear function of contract terms and borrowers' characteristics and allowing for some element of A_i to be proxied by the income, we get the demand system used in this paper.

C.3. Micro-foundation for borrowers' utility mortgage market

In this section, I micro-found the borrowers' indirect utility function used in the main text of the paper.

Assumptions regarding borrowers' utility function are made for tractability and do not impact the qualitative results.

Toy Model: Consume in period 1, default and lose the house in period 2

$$u(C^*, H^*) := \max_{\{C,L\}}\ \mu\ \overbrace{C_1}^{Consumption} + \overbrace{(1-\frac{\delta}{2}\frac{r}{Y_2}L)}^{survival\ probability}\ [\phi\overbrace{\frac{L}{P_H LTV}}^{House\ size} + \mu C_2]$$

$$pC_1 + (1-LTV)\frac{L}{LTV} = Y_1$$

$$pC_2 = Y_2 - rL$$

p is the price of consumption and P_H is the house price. μ is the marginal utility of consumption. ϕ the marginal utility of housing. $\frac{\delta}{2}\frac{r}{Y_2}\frac{L}{LTV}$ represents the fact that the borrower is more likely to default when he has a higher leverage. δ parameterizes the default sensitivity to the interest rate r. L is the loan size and LTV is the Loan-to-Value.

The optimal loan size is:

$$H^* = \frac{L^*}{LTV} = \frac{\overbrace{\frac{\phi}{P_H}}^{bigger\ house} - \overbrace{\frac{\mu}{p}(1-LTV)}^{lower\ consumption\ period\ 1} - \overbrace{\mu r}^{lower\ consumption\ period\ 2}}{\underbrace{(\frac{\phi_c}{P_H} - \frac{\mu r}{p})\delta\frac{r}{Y_2}}_{Higher\ default}}$$

Thus:

$$V(Y_1, H^*) := u(C^*, H^*) = \frac{\mu}{p}[Y_1 + Y_2] + H^*\{(\frac{\phi_c}{P_H} - \mu r\cdot ltv)[\frac{\frac{\phi_c}{P_H} - \frac{\mu}{p}r + \frac{\mu}{p}(1-ltv)}{2(\frac{\phi_c}{P_H} - \frac{\mu r}{p})}] - \frac{\mu}{p}\frac{\delta r}{2ltv}\}$$

Without consumption in period 2 this yields the functional form used in Appendix C.2:

$$V(Y_1, H^*) := u(C^*, H^*) = \frac{\mu}{p}Y_1 + H^*[\frac{\frac{\phi_c}{P_H} + \frac{\mu}{p}(1-ltv)}{2}]$$

Appendix D. Model extensions: Imperfect information about acceptance and rejections

When borrowers do not observe acceptance and rejection rules the utility they derive from applying to a contract $c \in C$ is:

$$p_{icR}u(c) + (1-p_{icR})\beta[E_\epsilon[U(R)] - cost] \quad (24)$$

$$where:\ U(R) = \max_{\{x\in C\setminus R\}}[p_{ixR'}u(x) + (1-p_{ixR'})]\beta E_\epsilon[U(R') - cost], with\ R' := \{x \cup R\} \quad (25)$$

p_{icR} is the probability of being accepted to contract c conditional on the rejection history R. $u(c)$ is the utility of getting contract c. β is a discount rate. $U(R)$ is the expected utility after being rejected from the contracts present in vector R. $cost$ is the cost ob being rejected and filling another application.

Since rejections are observed by other banks, the probability of being accepted for another contract may be lower upon rejections. Assuming that borrowers get a new extreme value drawn after each rejection, we can calculate U in a closed-form manner. To ease the computational burden, we can assume that the probability of being accepted after the first rejection is 0 and replace $U(\cdot)$ with an outside option that is borrower-specific $\bar{u}_i$.

Assuming that the term $p_{icb}[\sigma_i^{-1}\epsilon_{ibc} - \bar{u}_i] + \bar{u}_i$ is extreme value distributed with a variance $\bar{\sigma}_i^{-1}$, the new model thus becomes equivalent to the perfect information case with all utility parameters scaled by p_{icR}.

Appendix E. Proof screening

Let us start by considering that the demand parameters β_i^P in Eq. (4) can take only two values, β_{Li}^P and β_{Hi}^P, with $\beta_{Li}^P < \beta_{Hi}^P$; preferences along other dimensions are continuously distributed and independent of contract characteristics. We can then generalize the proof to any number of values.

Using Bayes' rule:

$$Pr(\beta_i^P = \beta_{ji}^P | choose\ contract\ b) = \frac{Pr(choose\ contract\ b|\beta_i^P = \beta_{ji}^P)Pr(\beta_i^P = \beta_{ji}^P)}{\sum_{m\in\{L,H\}} Pr(choose\ contract\ b|\beta_i^P = \beta_{mi}^P)Pr(\beta_i^P = \beta_{mi}^P)}$$

We start by offering contract A with characteristics (X,r) that would be accepted by both borrowers and then offer another contract B with characteristics $X + \Delta X, r + \omega\Delta X_c$ with $\omega > \min_j\{\frac{\beta_{ij}^P}{\alpha^P}\}$. When $\Delta X = 0$, contract A has the following market share: $Pr(V_{iA} + \epsilon_{iA} > V_{iB} + \epsilon_{iB})$ where $V_{ic} + \epsilon_{ic}$ is the borrower utility when getting contract c, defined is Eq. (2). Denoting f the pdf of $V_{iA} + \epsilon_{iA} - V_{iB} - \epsilon_{iB}$, by increasing ΔX the probability of contract B being chosen by the high willingness to pay (WTP) borrower increases by approximately $\frac{\frac{\beta_{Hi}}{\alpha} - \omega}{\sigma}f(0)$, while for the other it decreases by $\frac{\frac{\beta_{Li}}{\alpha} - \omega}{\sigma}f(0)$.

We can generalize this proof by first separating the lowest WTP from all other borrowers, then the second-lowest WTP from all other borrowers. When there are no bounds on r and on X, or when $\sigma \to 0$, borrowers can be almost perfectly screened $\forall\epsilon, \exists(\Delta X_m, \omega_m)_m : \forall j, \exists m : Pr(\beta_i^P = \beta_{ij}^P | choose\ contract\ m) > 1 - \epsilon$.

Appendix F. Derivation of the present value of lending

Given a loan size L, a maturity T, and a per period compound interest rate r, the per period mortgage repayment C is given by the annuity formula:

$$C = \frac{Lr(1+r)^T}{(1+r)^T - 1} \quad (26)$$

Similarly, we can express the bank cost of lending an amount L as a constant rate (mc) and write it as an annuity to make it comparable to the interest rate (r):

$$D = \frac{Lmc(1+mc)^T}{(1+mc)^T - 1} \quad (27)$$

The marginal cost includes, among others, the interest rate banks need to pay on their deposits.

Using δ as the discount rate, the present value of lending the amount L, abstracting from default, can thus be written:

$$L\sum_{k=1}^{f}\delta^k[\frac{r(r+1)^T}{(r+1)^T - 1} - \frac{mc(mc+1)^T}{(mc+1)^T - 1}]$$

$$+\gamma b\sum_{k=f+1}^{T}\delta^k[\frac{R(R+1)^{T-f}}{(R+1)^{T-f}-1}-\frac{mc(mc+1)^{T-f}}{(mc+1)^{T-f}-1}] \quad (28)$$

R is the reset rate and b is the remaining balance at the end of the teaser rate period. F is the fixed rate period, T is the maturity of the loan, γ is the share of people not refinancing, and mc is the marginal cost of lending.

As in Crawford et al. (2018), assuming that banks consider the average default instead of the probability of defaulting in each period, for a constant discount rate ($\delta < 0$) and denoting d a dummy equal to 1 if borrower defaults, the present value of lending up to period f is

$$C\cdot E[(1-d)]\cdot\sum_{k=1}^{f}\delta^k = Lr\frac{(1+r)^T}{(1+r)^T-1}\cdot E[(1-d)]\cdot\frac{1-\delta^f}{1-\delta}\delta \quad (29)$$

When T and f are large, $\frac{(1+r)^T}{(1+r)^T-1}\approx 1$ and $\delta^f\approx 0$, the net present value of lending, is thus

$$PV\approx L\cdot\{E[(1-d)]r\frac{\delta}{1-\delta}+\gamma E[(1-d)]R\frac{1-\delta^{T-f}}{1-\delta}\delta^f - [\frac{\delta}{1-\delta}+\gamma\frac{1-\delta^{T-f}}{1-\delta}\delta^f]mc\} \quad (30)$$

With ($\delta = 1$), the expression is instead

$$PV\approx L\cdot[E[(1-d)]rf+\gamma RE[(1-d)](T-f)-[f+\gamma(T-f)]mc] \quad (31)$$

We further assume, as in Benetton (2021), that $\partial_r\gamma = 0$ so that it does not enter inside the FoC of r_c and set γ_c to 0 (i.e., all borrowers remortgage). We can thus also abstract from the discount rate if $\delta < 1$, since it is constant across mortgages, and we thus get

$$NPV_{icb} := L\cdot[E[(1-d)]r-mc] \text{ when } \delta<1 \quad (32)$$

The above expression implies that banks do care about fixing the interest rate, except for its impact on the cost of lending (mc), default (d) or demand (L). This result comes from the assumption that $\delta^f\approx 0$. This may be problematic, for a given demand, interest rate, default, and marginal cost, profits are likely to be increasing in f as the loan generates annuities for a longer period.

Relaxing the assumption that $\delta^f\approx 0$ would, however, require an assumption about the discount rate used (for instance, the bond of or deposit rates) or the use of non-standard approaches such as integrating over one. This last method is too computationally demanding for our set-up. We thus go with the first approach and assume that $\delta = 1$. We get

$$NPV_{icb} := L\cdot[(1-d)r-mc]f \text{ when } \delta=1 \quad (33)$$

Alternative approach

Without using Crawford et al. (2018) assumption about default, the expression for the annuity would be, using d as the per period default probability:

$$C\sum_{k=1}^{t}((1-d)\delta)^t = Lr((1-d)\delta)\frac{(1+r)^T}{(1+r)^T-1}\frac{1-((1-d)\delta)^t}{1-((1-d)\delta)} \quad (34)$$

Using the same approximations as in Benetton (2021) — $\frac{(1+r)^T}{(1+r)^T-1}\approx 1$ and $\partial_r\gamma = 0$ — the expression for the NPV becomes

$$NPV_{icb} := L\cdot[(1-d)\delta\frac{1-((1-d)\delta)^f}{1-\delta+d\delta}r - mc\frac{1-\delta^f}{1-\delta}] \text{ when } \delta<1 \quad (35)$$

$$NPV_{icb} := L\cdot[(1-d)\frac{1-(1-d)^f}{d}r - mc\cdot f] \text{ when } \delta=1 \quad (36)$$

Here again, since the discount rate is not observable, the NPV would require estimating both the discount rate δ and the marginal cost mc. In a low rate environment, the discount factor can be approximated by 1. Changing the definition of the NPV will impact interpretation of the mc as discussed in Section 3.3.2. Moreover, when d is small, as in our empirical application, and δ is equal to 1, the expression becomes the same as in Crawford et al. (2018):

$$NPV_{icb} \underset{d\to 0}{\sim} L\cdot[(1-d)r-mc]\cdot f,\ \text{when } \delta=1 \quad (37)$$

Appendix G. Perfect information model

Under perfect information, a lender designs a contract for each borrower to maximize static profits, knowing the borrower survival probability function ($\theta_i(X,r) := 1-d_i(X,r)$, with d defined in Eq. (7)), the choice of bank function ($\Phi(X,r) := \frac{exp(V_i(X,r))}{\sum_{b\in B}exp(V_i(X_b,r_b))}$), with the utility $V_i(X,r)$ defined in Eq. (4)) and loan demand function—$L_i(X,r)$, defined in Eq. (6). Formally, the problem is:

$$\max_{\{X,r\}}\Phi(V_i(X,r))[\theta_i(X,r)r-mc(X)]L_i(X,r) \quad (38)$$

Where $mc(X)$ is the marginal cost of lending function, recovered by interpolating our point estimates.

Using the change of variable $\alpha_i^P u_i = \beta_i^P X - \alpha_i^P r + \xi \iff r = \frac{\beta_i^P}{\alpha_i^P}X + \frac{\xi}{\alpha_i^P} - u_i$, we have that:

$$\max_{\{X,u_i\}}\Phi(u_i)[\theta_i(X,u_i)[\frac{\beta_i^P}{\alpha_i^P}X+\frac{\xi}{\alpha_i^P}-u_i]-mc(X)]L_i(X,u_i) \quad (39)$$

Denoting $\tilde\beta := \frac{L_X}{L}$, the first order conditions imply that the lender chooses characteristics X to maximize the social surplus:

$$\theta_X r + \theta\frac{\beta_i^P}{\alpha_i^P} - mc' + \tilde\beta[\theta r - mc] = 0 \quad (40)$$

Eq. (40) can be solved for any value of promised utility u_i. In particular, under perfect competition, lenders break even so that $\theta r = mc$, which implies that:

$$\frac{\beta_i^P}{\alpha_i^P} = \frac{mc}{\theta}[\frac{mc_X}{mc}-\frac{\theta_X}{\theta}] \quad (41)$$

A similar expression yields when lenders can use non-linear pricing so that the lender set L to the first best value as in Taburet (2024).

Using the results from Table 9, we neglect the default elasticity with respect to LTV. Using the notation $d := 1-\theta$, the expression becomes:

$$\underbrace{\frac{\beta_i^P}{\alpha_i^P}}_{willingness\ to\ pay} = \underbrace{\frac{mc_{LTV}}{1-d_i}}_{effective\ marginal\ cost} \quad (42)$$

This expression states that the LTV should increase until the borrower's willingness to pay equals the effective marginal cost of increasing the LTV.

Appendix H. First-order conditions with respect to interest rates

First-order conditions with respect to interest rate of the model presented in Section 3 yield:

$$r_{cb} = \{\overbrace{\frac{mc_b}{1-E[d|bc]}}^{Fair\ price} + \overbrace{\frac{E[\Phi_{cb}]}{E[-\Phi'_{cb}]}(\overbrace{\frac{1-E[d|bc]+\beta_r^d r}{1-E[d|bc]}}^{Burden\ of\ Payment})}^{Perfect\ information\ mark-up} + \overbrace{\sum_{j\neq c}\frac{E[\Phi'_j]}{E[-\Phi'_{cb}]}\frac{\tilde\pi_{cb}}{1-E[d|bc]}}^{AI\ discount/premium}\}\frac{1-E[d|bc]}{1-E[d'|bc]} \quad (43)$$

where $\Phi_{cb} := \sum_i\frac{exp(V_{icb})}{\sum_x exp(V_{ix})}L_{icb}$ is the expected amount lent and $\tilde\pi_{cb} := (1-E[d|cb])r_{cb} - mc_{cb}$ is the expected profit on each loan unit given that the borrower chooses contract c at bank b. $E[d|cb]$ is the expected default conditional on contract and bank choice ($E[d|bc] := \beta^d X_{cb} + \alpha^d r_{cb} + \rho\frac{E[\Phi_{cb}\beta_i]}{E[\Phi_{cb}]}$).

The first term $\frac{mc}{1-E[d|bc]}$ is the pricing at which banks break even given the expected default probability of borrowers' choosing contract c at bank b. This is the marginal cost scaled up by the survival probability.

The second term is $\frac{E[\Phi_j]}{E[-\Phi'_j]}(\frac{1-E[d|bc]+\beta_r^d r}{1-E[d|bc]})$ is the pricing set by banks above the fair price if they could observe the average default probability of the type of borrowers who choose each contract ($E[d|bc]\forall cb$). $\frac{E[\Phi_j]}{E[-\Phi'_j]}$ is the impact of borrowers' product elasticity (i.e., competition). $\frac{\beta_r^d r}{1-E[d|bc]}$ accounts for the burden of payment: When increasing r, borrowers are more likely to default ($\beta^d < 0$); this creates incentives to lower the mark-up.

The last term $\sum_{j\neq c}\frac{E[\Phi'_j]}{E[-\Phi'_{cb}]}\frac{\tilde{\pi}_j}{1-E[d|bc]}$ is the equivalent of the information rent in the textbook principal–agent model. It ensures that borrowers self-select. Under perfect information, menus are not offered, so this term would be equal to zero.

The ratio $\frac{1-E[d|bc]}{1-E[d'|bc]}$, in which $E[d'|bc] := \beta^d X_{cb} + \alpha^d r_{cb} + \rho\frac{E[\Phi'_{cb}\beta_i]}{E[\Phi'_{cb}]}$, scales up the three terms by taking into account the fact that changes in r impact the type of borrowers who choose a given contract.[25]

For the equilibrium observed in the data, all the elements in Eq. (43) are observable (e.g., contract terms) or estimated (e.g., demand and default elasticities, marginal cost).

Appendix I. Screening externality model

Formally, lender b problem is defined as a monopoly problem:

$$\max_{M_{bt}\in\mathcal{F},P_{ibt}} \sum_i \sum_{c=1}^{C} \overbrace{\tilde{\Phi}(V_i(X_c,r_c))}^{Demand} NPV_{ic} \tag{44}$$
$$s.t.\ \forall i\ E[\max_c V_{ic}+\epsilon] \geq E[\max_c \bar{V}_i+\epsilon]\ (PC)$$

We drop the b subscript for notational convenience. The key difference with the model used in the estimation comes from the demand $\Phi(V_i(X_c,r_c))$ and the constraint (PC).

The NPV term is the same as the main model and is described in Eq. (11). It is the net present value of lending to borrower i via contract c.

We restrict borrower consideration set to the set of contracts offered by bank b. The demand is thus $\Phi(V_i(X_c,r_c)) := n_i\frac{exp(V_{ic})}{\sum_{x\in[\![1,C]\!]}exp(V_{ix})}$. It captures how borrowers i make their choice of contract when they only have access to bank b contracts (indexed by $c\in[\![1,C]\!]$). We use this demand instead of the one used in the structural model ($\frac{exp(V_{ic})}{\sum_{x\in B}exp(V_{ix})}$) to shut down the intensive margin (i.e., competition) channel which drives the contractual externality.

Because we shut down competition, we add a participation constraint (PC) to prevent the lender from extracting too much surplus from borrowers. The constraint ensures that the new contracts offered are Pareto improvements over the contract observed in the data.

Formally, the participation constraint (PC) is:

$$E[\max_c V_{ic}+\epsilon] \geq E[\max_c \bar{V}_i+\epsilon] \tag{45}$$

$$\iff \sum_{c=1}^{C} exp(V_{ic}) \geq C exp(\bar{V}_i) \iff E_c[exp(V_{ic})] \geq exp(\bar{V}_i) \tag{46}$$

Eq. (45) states that borrower i's expected utility should be at least as big as what they got under the competitive equilibrium if they chose bank b (denoted $\bar{u}_i$). $\bar{u}_i$ is the utility the borrower derive from the contract they got in the data.[26]

[25] When the number of products in the market is large and the loan rate elasticity is low ($\tilde{\beta}_r$ low), $E[d|bc]$ and $E[d'|bc]$ are relatively close to each other. Indeed, $\Phi' \approx \Phi(\tilde{\beta}_r+1)\Phi \approx \Phi$.

[26] Ideally, we would have used a nested demand system where the borrower chooses first the bank and then the contract within the menu. This way, the error terms ϵ in Eq. (45) would have had the same economic interpretation as in the structural model. However, this approach would have been too computationally demanding. I refer the interested reader to Taburet (2024), which proposes an alternative approach to calculate the contractual externality.

Because the fixed cost estimates heavily depend on the modeling assumptions, we abstract away from them and use a conservative approach instead. As in the estimation, we limit the feasible set to a combination of products with teaser rates of 0, 2, 3 or 5 years, three potential levels of fees (0, 750, 1,500) and buckets of LTV from 60% to 95% by increasing levels of 5%. In the estimation (see Section 4.1.2), we assumed that a given borrower can only get access to a contract with an LTV just below and above the one they get in the data. In the counterfactual, we consider adding or removing products that were not in the choice set, but we keep the constraint of the LTV being just above or below the one chosen in the data.

Without any more restrictions, we end up with menus that can be composed of up to 48 products. This implies that there exist $C_{10}^{48} = 6,540,715,896$ different potential menus with 10 products. To overcome the computational burden, we hold the teaser rate period and fee level and focus on changes in LTVs only. For instance, let us consider a menu composed of two contracts $\{(90,0,5),(95,500,5)\}$ where the first number is the LTV, the second is the fees, and the last is the teaser rate period. If the borrower chose the second one, I consider that $(95,0,5)$ and $(90,500,5)$ are potential products to include. I do not consider LTV above 95 because they are never observed in the data, and I assume that 85 LTV products would not be preferred or available because they are 10% below the contract they chose.

As a robustness check, I also solve the model with fixed costs. For consistency, I consider the same feasible set as described in the previous paragraph. I use the conservative estimates of the fixed costs for the simulation.

Data availability

The reference data is available at https://doi.org/10.7910/DVN/QQJOB4.

References

Agarwal, S., Grigsby, J., Hortaçsu, A., Matvos, G., Seru, A., Yao, V., 2024. Searching for approval. Econometrica 92 (4), 1195–1231.

Aiyar, S., Calomiris, C.W., Hooley, J., Korniyenko, Y., Wieladek, T., 2014. The international transmission of bank capital requirements: Evidence from the UK. J. Financ. Econ. 113 (3), 368–382.

Arcidiacono, P., Miller, R.A., 2011. Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity. Econometrica 79 (6), 1823–1867.

Attar, A., Mariotti, T., Salanié, F., 2011. Nonexclusive competition in the market for lemons. Econometrica 79 (6), 1869–1918.

Azevedo, E.M., Gottlieb, D., 2017. Perfect competition in markets with adverse selection. Econometrica 85 (1), 67–105.

Benetton, M., 2021. Leverage regulation and market structure: A structural model of the UK mortgage market. J. Financ. 76 (6), 2997–3053.

Benetton, M., Gavazza, A., Surico, P., 2021. Mortgage pricing and monetary policy. Working paper, CEPR Discussion Paper No. DP16456.

Berry, S., Levinsohn, J., Pakes, A., 1995. Automobile prices in market equilibrium. Econometrica: J. Econ. Soc. 841–890.

Campbell, J.Y., 2013. Mortgage market design. Rev. Financ. 17 (1), 1–33.

Chiappori, P.-A., Salanie, B., 2000. Testing for asymmetric information in insurance markets. J. Political Econ. 108 (1), 56–78.

Cloyne, J., Huber, K., Ilzetzki, E., Kleven, H., 2019. The effect of house prices on household borrowing: a new approach. Am. Econ. Rev. 109 (6), 2104–2136.

Crawford, G.S., Griffith, R., Iaria, A., 2021. A survey of preference estimation with unobserved choice set heterogeneity. J. Econometrics 222 (1), 4–43.

Crawford, G.S., Pavanini, N., Schivardi, F., 2018. Asymmetric information and imperfect competition in lending markets. Am. Econ. Rev. 108 (7), 1659–1701.

Dasgupta, P., Maskin, E., 1986. The existence of equilibrium in discontinuous economic games, I: Theory. Rev. Econ. Stud. 53 (1), 1–26.

Dou, W.W., Wang, W., Wang, W., 2022. The cost of intermediary market power for distressed borrowers. Available at SSRN 4163319.

Einav, L., Finkelstein, A., Cullen, M.R., 2010. Estimating welfare in insurance markets using variation in prices. Q. J. Econ. 125 (3), 877–921.

Einav, L., Finkelstein, A., Mahoney, N., 2021. The IO of selection markets. In: Handbook of Industrial Organization, vol. 5, (no. 1), Elsevier, pp. 389–426.

Eizenberg, A., 2014. Upstream innovation and product variety in the us home pc market. Rev. Econ. Stud. 81 (3), 1003–1045.

Granja, J., 2021. Regulation and service provision in dynamic oligopoly: Evidence from mobile telecommunications. Working paper.

Handel, B., Hendel, I., Whinston, M.D., 2015. Equilibria in health exchanges: Adverse selection versus reclassification risk. Econometrica 83 (4), 1261–1313.

Hertzberg, A., Liberman, A., Paravisini, D., 2018. Screening on loan terms: evidence from maturity choice in consumer credit. Rev. Financ. Stud. 31 (9), 3532–3567.

Karlan, D., Zinman, J., 2009. Observing unobservables: Identifying information asymmetries with a consumer credit field experiment. Econometrica 77 (6), 1993–2008.

Lee, R.S., Pakes, A., 2009. Multiple equilibria and selection by learning in an applied setting. Econom. Lett. 104 (1), 13–16.

Lester, B., Shourideh, A., Venkateswaran, V., Zetlin-Jones, A., 2019. Screening and adverse selection in frictional markets. J. Political Econ. 127 (1), 338–377.

Morrow, W.R., Skerlos, S.J., 2011. Fixed-point approaches to computing Bertrand–Nash equilibrium prices under mixed-logit demand. Oper. Res. 59 (2), 328–345.

Nelson, S., 2025. Private information and price regulation in the us credit card market. Econometrica (Forthcoming).

Nevo, A., 2001. Measuring market power in the ready-to-eat cereal industry. Econometrica 69 (2), 307–342.

Pakes, A., Porter, J.R., Shepard, M., Calder-Wang, S., 2021. Unobserved heterogeneity, state dependence, and health plan choices. Working paper, National Bureau of Economic Research.

Riley, J.G., 1979. Informational equilibrium. Econometrica: J. Econ. Soc. 331–359.

Robles-Garcia, C., 2019. Competition and incentives in mortgage markets: The role of brokers. Working paper.

Rothschild, M., Stiglitz, J., 1976. Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. Q. J. Econ. 90 (4), 629–649.

Rothschild, M., Stiglitz, J., 1978. Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. In: Uncertainty in Economics. Elsevier, pp. 257–280.

Salanié, B., Wolak, F.A., 2019. Fast, ‘‘robust’’, and approximately correct: estimating mixed demand systems. Working paper, National Bureau of Economic Research.

Taburet, A., 2024. Screening using a menu of contracts in imperfectly competitive markets with adverse selection. Working Paper.

Train, K., 1986. Qualitative Choice Analysis: Theory, Econometrics, and an Application to Automobile Demand, vol. 10, MIT Press.

Wollmann, T.G., 2018. Trucks without bailouts: Equilibrium product characteristics for commercial vehicles. Am. Econ. Rev. 108 (6), 1364–1406.