Bank of England

Screening using a menu of contracts: a structural model of lending markets

Staff Working Paper No. 1,057

February 2024

Arthur Taburet, Alberto Polo and Quynh-Anh Vo

Staff Working Papers describe research in progress by the author(s) and are published to elicit comments and to further debate. Any views expressed are solely those of the author(s) and so cannot be taken to represent those of the Bank of England or to state Bank of England policy. This paper should therefore not be reported as representing the views of the Bank of England or members of the Monetary Policy Committee, Financial Policy Committee or Prudential Regulation Committee.

Bank of England

Staff Working Paper No. 1,057

Screening using a menu of contracts: a structural model of lending markets

Arthur Taburet,[(1)] Alberto Polo[(2)] and Quynh-Anh Vo[(3)]

Abstract

When lenders screen borrowers using a menu of contracts, they generate a contractual externality by making 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 borrowers' utility, is equivalent to an interest rate increase of 30 basis points (a 15% increase) on all loans.

Key words: Adverse selection, screening, structural model.

JEL classification: D82, G21, L13.

---

(1) Duke's Fuqua School of Business. Email: arthur.taburet@duke.edu
(2) Bank of England. Email: alberto.polo@bankofengland.co.uk
(3) Bank of England. Email: quynh-anh.vo@bankofengland.co.uk

The views expressed in this paper are those of the authors, and not necessarily those of the Bank of England or its committees. 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. Arthur Taburet would like to express his deepest gratitude to his advisors Alessandro Gavazza, Martin Oehmke, Daniel Paravisini and Ricardo Reis for their patient guidance, constant encouragement and invaluable advice. All errors are our own.

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

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

©2024 Bank of England
ISSN 1749-9135 (on-line)

1 Introduction

Menus of contracts are widely used in financial markets. 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 make borrowers 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 to 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 than high-default borrowers, which is not necessarily what would happen in the first best.

The theoretical literature has highlighted that pooling contracts cannot be offered in competitive markets even when pooling is a Pareto improvement over screening (see for instance, Rothschild and Stiglitz 1976 and Rosenthal and Weiss 1984). This is because a lender can take advantage of their competitors' pooling contract by introducing a contract with — let us say — an LTV just below the one offered by the competitors to steal the safer and more profitable borrowers.[1] This market failure emerges when lenders do not internalize how their screening strategies change the types of borrowers selecting competitors' products —– and thus the cost of lending via those products. Yet, how relevant this issue is in practice, and more generally, how adverse selection impacts contract terms and welfare, is still an open question (see Einav, Finkelstein, and Mahoney 2021 for a literature review). Quantifying the impact of adverse selection on contract terms requires determining what contracts would be offered if there were no adverse selection ("first best") or the ones offered by a social planner who internalizes that deviating from pooling may be inefficient ("second best"). This is challenging as 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 for default probabilities. 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 to 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 information about their default probabilities and their preferences over each contract characteristic. We identify and estimate the model parameters using administrative data on lenders' menus, borrowers'

1. In Rothschild and Stiglitz 1976 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.

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 choosing 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" changes the type of borrowers that choose another contract "B". Adverse selection can thus be recovered by comparing the default probability of groups of borrowers that 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 deliver three new empirical results. First, we find that the LTV ratio is used together with interest rates to screen borrowers along their default probability. Lenders set their LTV pricing schedule such that high-default borrowers chose 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[2] — have a 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 percent of borrowers (those with a lower default probability) choose contracts with an LTV between 70 and 85 percent. However, under perfect information, those borrowers, and most other borrowers, would have obtained an LTV above 85 percent and bought a bigger house. Thus, according to our model, contracts with an LTV between 70 and 85 percent 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 percent 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 that deviating from a pooling contract can be inefficient (second best), we isolate the effect

2. 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. (2020) for empirical evidence).

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.[3]

Our results show that screening is an important force in the UK mortgage market and that the associated contractual externality is costly. This suggests there is room for Pareto improving policy interventions. Examples of such policies — analysed in the theoretical companion paper Taburet (2022) — include lowering competition, increasing the capital requirement on low LTV mortgages in a low-competition environment, or banning the use of lower LTV products. These 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 in their individual specific menu that maximizes their utility. Following the industrial organisation literature (see Berry, Levinsohn, and Pakes 1995 and Crawford, Pavanini, and Schivardi 2018) we assume that borrowers' utilities are linear functions of contract characteristics (loan size, interest rate type, LTV, lender and fees) and estimate the contributory value for 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. We use our model to derive a discrete-continuous demand system (as in for instance Train 1986) composed of a mixed logit for the product choice and a linear regression for the loan demand. We also specify borrowers' default probabilities as a linear function of contract terms, as well as borrowers' observable and unobservable characteristics. On the supply side, we model lenders as heterogeneous multi-product firms offering differentiated menus of mortgages and competing on the number of contracts, their interest rates, LTV and fees.

We identify the model parameters using a three-step approach. First, we use, as in Nevo (2001), a revealed preference 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 that chose the same contract but have different average preferences. The variation in the average preference comes from changes over time in the

3. 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.

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 selecting a given contract are correlated with changes in unobservable characteristics. To address this identification threat, we instrument the average borrower type using product-specific risk weights and minimum capital requirements for contracts other than the one chosen. Risk weights are pre-determined 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 2018 and Robles-Garcia 2019). Our instrument is relevant as it affects the spread between interest rates and thus the type of borrower choosing 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 2018), the exclusion restriction requires that our cost shifter is not correlated with economic shocks affecting borrower types differently, and with changes in unobserved product characteristics, or acceptance and rejection rules based on characteristics unobserved by the econometrician only. It is plausible that the endogeneity from mortgage application rejections based on soft information observed by the lender but not the econometrician is not fully addressed, as lenders can update their acceptance and rejection criteria following any product cost shock. In that case, our results should be interpreted as a lower bound on adverse selection as lenders are likely to become stricter to mitigate the increase in the cost of lending.

Related literature This paper contributes to the empirical literature on adverse selection and the industrial organisation literature on credit markets.

A large empirical literature tests whether or not adverse selection and screening occur in practice. Seminal papers are Chiappori and Salanie 2000 for the positive correlation test approach and Einav, Finkelstein, and Cullen (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. Their identification relies on lenders that just started using menus (Hertzberg, Liberman, and Paravisini 2018) or the use of experimental data (Karlan and Zinman 2009). We contribute to the 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 as the literature has extensively documented plausibly exogenous variations in interest rates. We also implement our identification strategy within a structural model, which allows for answering a more comprehensive range of questions by doing counterfactual simulations. In particular, we are 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 the second best (when the contractual externality is internalized by lenders).

This paper also relates to the literature analysing consumers' and lenders' behaviours in retail financial markets. Our paper contributes to this literature by studying screening. To do so, we build on Benetton (2018) and Crawford, Pavanini, and Schivardi (2018) 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 modelling assumptions. In Section 3, we present the structural model and discuss its main assumptions. Section 4 discusses the identification strategy and estimation procedure. In Section 5, we analyse the estimation results and the counterfactual experiment outcomes are presented in Section 6. Finally, Section 7 concludes.

2 Institutional Setting, Data, and Motivating Evidence

This section describes the key institutional features of the market and the data used in this paper. It then provides suggestive evidence that screening is an important driver of the UK mortgage market contracts offering.

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 — such as that in the US, for instance — 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 two, three, or five). 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 percent of the outstanding loan — refinancing around the time when 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 choosing that contract. This is because minimal negotiation takes place between borrowers and lenders, and banks do not practice individual-based pricing.[4] 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) or online platforms.[5]

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 Figure 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.

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

Figure 1: Extract of the menu of contracts offered by HSBC in January 2023
Source: HSBC's website

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 2018, Robles-Garcia 2019) — that 90 percent 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 the borrowers choosing the contract.

4. The search platform Moneyfacts reports: "A personal Annual Percentage 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/.

5. This can be rationalized by the fixed cost of negotiation being high compared to the size of loans in the consumer market compared to the firm market.

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). As we do not observe loan applications or the criteria used by banks, we will build our empirical strategy considering 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 about 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 containing the credit events on 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 a tiny fraction of the house is repossessed conditional on default (e.g. according to the data from UK finance, in the third quarter of 2023, among 87930 homeowner mortgage in arrears, only 630 properties were taken into possession[6]).

In this paper, we focus on the years 2015 to the end of 2018. During this period, we observe for each mortgage origination 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.

The structural estimation uses 2018 data (See Table 2 in Appendix A for the 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 two, three, or five years. The average interest rate is 2.5 percentage 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 from the Moneyfacts database.

6. See https://www.ukfinance.org.uk/data-and-research/data/arrears-and-possessions

2.3 Motivating Evidence

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

Variation in product offering As shown in Figure 2, the number of products varies over time and across market participants. In particular, borrowers shopping for 90% LTV contracts faced on average two different options at each bank in 2010, six options in 2018. Menu sizes are larger at 75% LTV. Indeed, the average menu contains 6 alternative contracts at 75% LTV in 2010 and 16 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, 1500). A higher level of fee 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 percent of those products are offered by the average bank. These empirical facts are the motivation for making the number of products endogenous in the model.

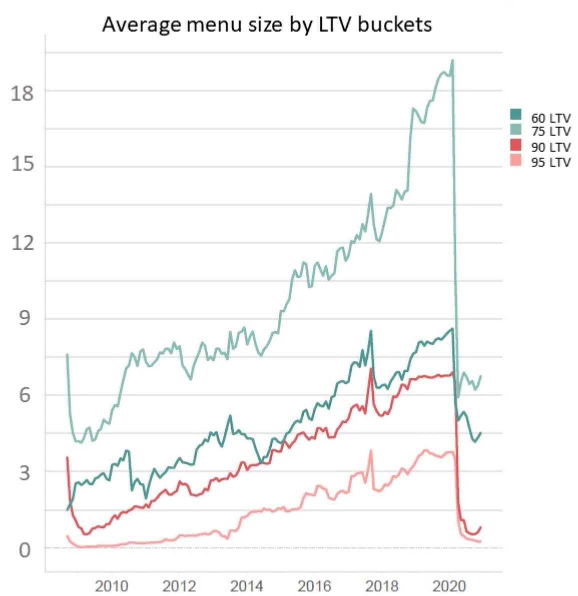

Figure 2: Average number of advertised mortgage products for Buy-to-Let, First-time-Buyer and Remortgage
Source: Moneyfacts and Bank of England's calculations

Sorting on observables As suggestive evidence that borrowers with different characteristics tend to select different products, we regress borrowers' observable characteristics on LTV dummies (see Table 1). We document that — compared to borrowers choosing

75% LTV contracts — borrowers choosing 95% LTV contracts are on average 1.5 years younger, earn £7,400 net less a year, and are 20 percent 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. As 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 behaviour. Making borrowers self-select (on observable characteristics) using LTV is thus feasible.

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

$$Default_i = \beta X_i + \epsilon_i \tag{1}$$

$Default_i$ is equal to 1 if borrower i has been in arrears by the end of 2019, andX_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 percent of the loans originated in 2018 had defaulted by 2020. The default probability on 85-95% LTV loans is 1.4 percent, while the average for 75-85% LTV loans is 0.8 percent.

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 percent increase in default probability; the default probability of a 5-year fixed rate contract is 40 percent lower than that of flexible rate contracts; the default probability of a zero fee contract is 30 percent lower than contracts with fees of 1,000; and borrowers whose income is one standard deviation lower (i.e. 16,000 less) are 16 percent more likely to default.

In Figure 3, we plot the share or mortgages that are in arrears as a function of the LTV at origination. Loans with an LTV bellow 75 percent are twice as less likely to be in arrears than loans with an LTV above 90 percent. This can for instance be due to the causal impact of borrowers type (i.e., adverse selection) or to the causal impact of contract terms (i.e., moral hazard or burden of payment).

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

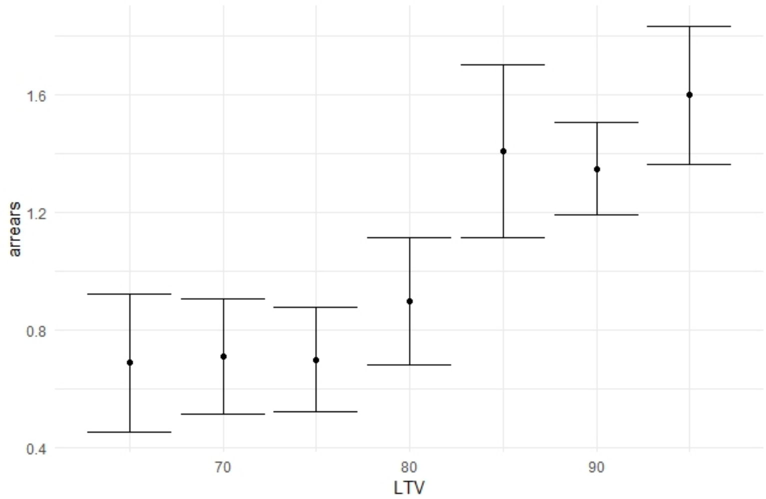

Figure 3: Percentage of mortgages in arrears by LTV at origination
Source: PSD001-PSD007

Need for a structural model To further understand the impact of screening on equilibrium quantities, one needs to compare the observed equilibrium contracts terms to a counterfactual in which there is no private information. Given the difficulty of finding the right counterfactual in the data, we build a structural framework and rely on simulations instead. The following sections discuss the model assumptions and our identification strategy. Our modeling approach and identification strategy also allow us to look at selection on unobservable borrowers' characteristics, take care of 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 read the data through the lens of the model of supply and demand described in this section. To simplify the notation, we drop the index t on the variables unless for variables with a different time index.

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 the UK institutional features, we define as a loan contract the object (L, X, r) where L is the loan size, X is a vector containing other contract characteristics (lender dummy, rate type, maximum LTV and fees) and r is the interest rate on the loan.

Following the IO 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 its content M_b) is endogenous.

Supply and demand Our model is based on the following maximization problems. Borrowers choose the bank and contract among their individual specific set that maximizes 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

$$(c_i, b_i) = \text{argmax}_{\{b \in B_i, c \in P_{ib}\}} \{ \underbrace{V_i(\overbrace{X_{cb}, r_{cb}}^{\text{contract terms and price}}, \overbrace{L_i(X_{cb}, r_{cb}, d_i(X_{cb}, r_{cb}))}^{\text{Loan demand}}, \overbrace{d_i(X_{cb}, r_{cb})}^{\text{Default probability}})}_{\text{Indirect utility}} + \underbrace{\varepsilon_{icb}}_{\text{Demand shock}} \} \tag{2}$$

$V_i(\cdot) + \varepsilon_{icb}$ is the borrower indirect utility. The demand shock ε has 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

$$M_b \subset \text{argmax}_{\{M_b, C_b, P_{ib}\}} \ \mathbb{E}[\sum_{i,c} 1_{\{(c_i,b_i)=(c,b)\}} \underbrace{NPV(r_{cb}, d_{icb}, \overbrace{mc_{cb}}^{\text{marginal cost of lending}})}_{\text{Expected NPV if i chooses cb}}] - \underbrace{F(M_b, M_{bt-1})}_{\text{Fixed cost of changing menu}} \tag{3}$$

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

This formulation is a general version of screening models such as Rothschild and Stiglitz (1976). Constraint (2) can be written as a participation constraint and an incentive compatibility constraint when the revelation principle is used. The fixed cost F is

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

required to match the number of contracts offered in the data. M_{bt-1} is the menu offered by bank b in the previous period.

Information set The expectation in Equation 3 is conditional on the lender information set. The information set contains competitors' contract terms and prices, 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 modelling is the most conservative as it is likely to lower the contractual externality when lenders partially internalize their competitors' behaviour as explained below.

3.2 Screening mechanism: A heuristic explanation

Before specifying key parametric assumptions and the main parameters of interest in the model, let us heuristically explain in this section the intuition underlying the screening that banks can do using a menu of contracts. For illustrative purpose, we focus on screening via the Loan-to-Values (LTVs) and we assume that there are only two borrower types which differ only in terms of the probability of experiencing negative income shocks. Therefore one type of borrowers has higher probability of default than the other. The credit market is perfectly competitive but is adversely selected: high default borrowers have a higher "willingness to pay" for loan size. This can be the result of, for instance, high default borrowers being less sensitive to interest rates because they expect to repay the loan less often (Stiglitz and Weiss 1981) or because they are less financially sophisticated and underestimate the cost of defaulting.

First best contracts If borrowers' default probabilities are observable and the marginal cost of lending (or regulatory constraints on the maximum loan size) is such that both borrower types buy the same house size under perfect information, the first best contracts have the same LTVs. Lenders break even on each contract and thus charge borrowers different interest rates. We illustrate this situation in Figure 4 by plotting on the LTV-interest rate plane the perfect information contracts (c_1, c_2), borrowers' indifference curves, and the break-even rates.

Screening contracts Under unobservable default probabilities, the first best contracts (c_1, c_2) are not incentive compatible. Indeed, as can be seen in Figure 4, by pretending to be low default type and choosing contract c_2, high default borrowers will get higher utility. To prevent this from happening, one option for banks is to reduce the LTV of the

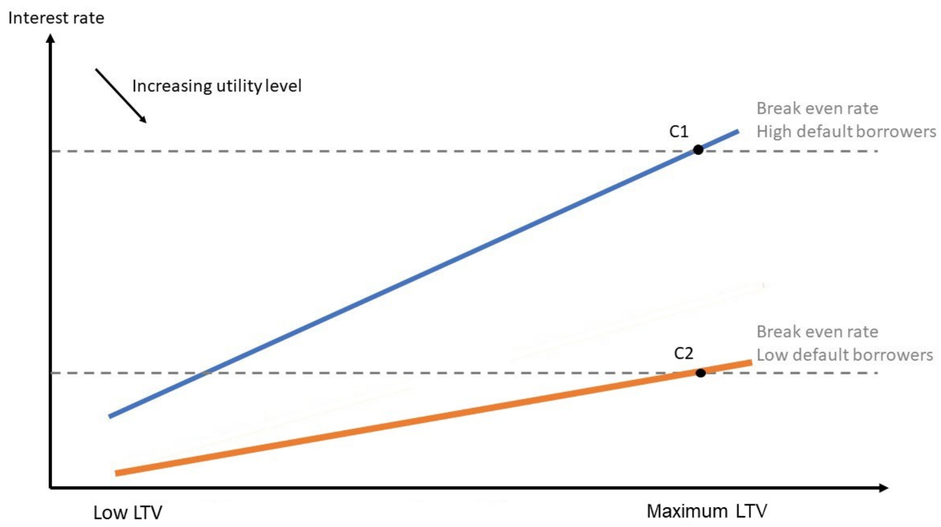

Figure 4: The perfect information, perfect competition contracts (c_1, c_2) are not incentive compatible.

contract offered to the low default type: the new contract lies on the same indifference curve of the high-default type as contract c_1 as illustrated in Figure 5. Screening works with this new menu as the high default type is indifferent between choosing the new contract c_2 and contract c_1 specifically designed for that type. The low default type also prefers its targeted contract c_2 than contract c_1. Note also that the screening leads to a downward distortion of the LTV offered to the low default type.

Contractual externality When the LTV distortions needed to screen borrowers are high, the pooling contract c_p (see Figure 5) is a Pareto improvement over screening. This happens because, under pooling contracts, high-default borrowers get a lower rate and low-default borrowers are less credit constrained.

Yet, pooling contracts cannot be offered in equilibrium. Indeed, if a lender prices their customers using the average default probability (pooling), a competitor can take advantage of the pooling contract by introducing a low rate-low credit constraint contract that will steal only the low default customers (so-called cream-skimming deviations, see Rothschild and Stiglitz 1976). We illustrate that situation in Figure 6: We can see that any contract in the profitable deviation region indicated in the figure is preferred by the low default type. All else equal, this deviation of competitors leaves the pooling lender with negative profits. A contractual externality thus exists when lenders do not internalize how their screening strategies change the types of borrowers selecting competitors' products — and thus the cost of lending via those products.

Overall, two additional remarks from the graphical analysis are worth mentioning. First, the banks' ability to screen depends on the heterogeneous slope of borrowers'

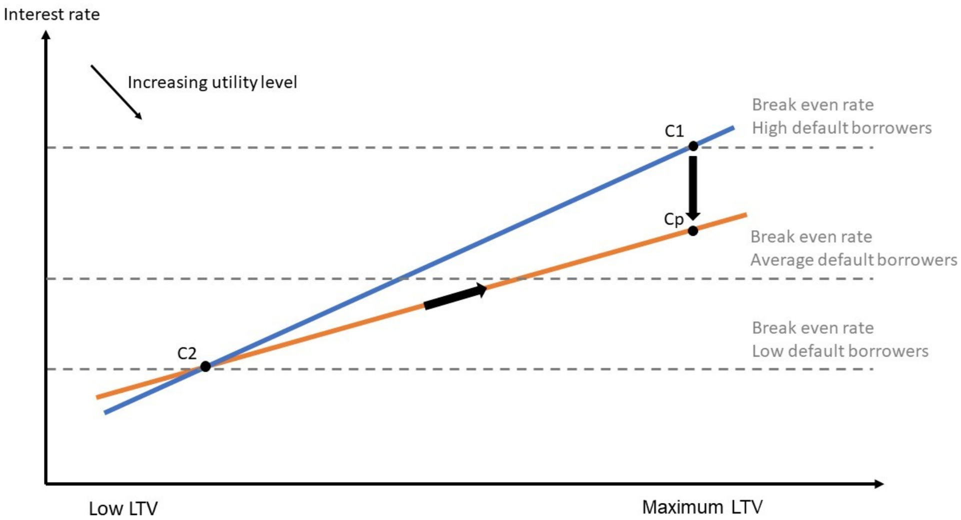

Figure 5: Screening may not be a equilibrium

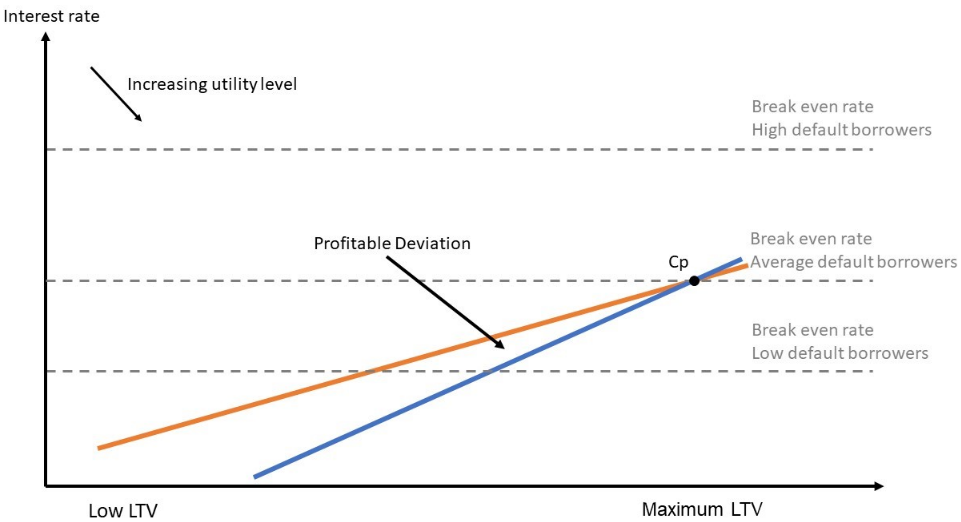

Figure 6: Pooling cannot not be a equilibrium

indifference curve. Once these slopes and the break-even rates are estimated, it is possible to back out the first-best optimal contracts. Second, lenders' incentives to deviate from pooling even when it is ex-post inefficient imply that a social planner with access to the same information as lenders could implement a Pareto improvement over the market equilibrium.

3.3 Key Parametric Assumptions

In this section, we present our parametric assumptions. The implications of our modelling assumptions are discussed in depth in Section 3.4.

3.3.1 Demand

We linearize the indirect utility, the loan demand and the default functions (V, L and d) around contract terms. 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.

The fact that the indirect utilities, the loan demand and the default function derive from the same maximization problem is captured by the use of correlated random coefficients. This is a generalization of Train (1986).

Formally, using the notation X_{cb} for the contract c characteristics in bank b and D_i for observable borrower characteristics, we have:[8]

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

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

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

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

$$with,\ \beta_i^x = \beta^x + \gamma^x D_i + PI_i^x,\ \alpha_i^x = \alpha^x + \tilde{\gamma}^x D_i + \tilde{P}I_i^x,\ x \in \{P, L\} \tag{8}$$

$$and\ (PI_i^x, \tilde{P}I_i^x, PI_i, \epsilon_{ic}^L,\ \epsilon_{ic}^d) \sim N(0, \Omega) \tag{9}$$

ε_{icb} contains deviations from the average borrowers' valuation of unobserved product characteristics. ε_{icb} is extreme value distributed, independent of the unobserved borrower characteristics (i.e., PI). We assume that $E[\varepsilon_{icb}|X_{cb}, r_{cb}, \beta_i, \alpha_i] = 0$ so that ε_{icb} represents the part of borrowers' demand that cannot be screened by banks when they use product characteristics (X_c, r_c) only. The potential identification threat caused by this assumption is discussed in Section 4.

8. Notice that we use the Industrial Organisation notation for the indexes. That is, as in for instance Benetton (2018), while we only observe one choice per borrower, we index all the possible alternatives that were available to the borrower.

(β_i^x, α_i^x) drive how borrower i values product characteristics and prices. We loosely refer to them as borrowers' preferences. They are a function of observable borrowers' characteristics (D_i) and unobserved borrower heterogeneity (PI). (β_i^P, α_i^P) captures the part of the valuation that is not a function of contract terms. Intuitively, how borrowers value contract terms might be a function of the default probabilities. For instance, risky borrowers might be less sensitive to prices if they expect that they won't be forced to repay the full face value of the loan upon default. In that case, α_i^P would be a decreasing function of default. In light of this example, borrower i valuation of characteristics (X) could be written in a more general form: $\beta_{icb}X$ with $\beta_{icb} := f_i(d_{icb}, X_{cb}, r_{cb}) \approx \beta_i + f_{cb}(X_{cb}, r_{cb})$. The elements that depend on contract terms (i.e., $f_{cb}(X_{cb}, r_{cb})$) are absorbed by the bank-product fixed effect ξ_{cb}. β_i can thus be interpreted as the part of borrowers' valuation that depends on unobservable borrower characteristics that may be correlated with default.[9]

ξ_{cb} is a product bank fixed effect. It captures the part of the average indirect utility that comes from unobserved (by the econometrician) contract characteristics.

ν^P are parameters capturing observable heterogeneity in borrowers' preferences.

PI_i^P is a random coefficient modeling unobserable heterogeneity in borrowers' preference. It is a key parameter for screening as it potentially contains information about borrower i's unobserved baseline default probability. β_i^P also contains borrowers' characteristics that are unobservable by the econometrician but observable by banks.[10]

The ratio $\frac{\beta_i^P}{\alpha_i^P}$ represents borrower i's willingness to pay for a characteristic. Indeed, if a bank proposes a new high LTV contract, borrower i would be happy to take it (i.e., its utility would increase by taking the contract) as long as the price increase is below the borrower's willingness to pay.

Key parameter for selected market In the spirit of the positive correlation literature (see Chiappori and Salanié 2002), we model adverse selection via the correlation between the random coefficients of the demand and default regressions $(PI^x, \tilde{P}I^x, PI)$. We denote $\rho_{x,y} := \rho_y^d \rho_{x,y}^d$ where ρ_y^d is the correlation between default and the y^{th} element of the private information component PI and $\rho_{x,y}^d$ is the correlation between the x^{th} preference parameter of $\Gamma := (\beta_i^P, \alpha_i^P, \beta_i^L, \alpha_i^L)$ and the y^{th} element of PI.

Given the assumption that the error term of the default regression is uncorrelated with observables and the private information component $(i.e., E[\epsilon_{icb}^d|X, r, PI] = 0)$, we have $\rho_x := \sum_y \rho_{x,y} \neq 0$, which implies that the market is a selection market with respect to the contract characteristic associated with preference parameter x. That is, borrowers

9. The logic behind our approach is as follows. The default probability is a function a of monthly repayment, the cost of defaulting and losing the house, the borrower's future income profile and the borrower's propensity to save. The loan size is an endogenous variable, so we replace it by its function defined in 6. We linearize the expression around the contract and borrowers' characteristics. Then, we explicitly acknowledge that the choice of contract and loan size depends on default in equation (7).

10. If the model is misspecified, this term includes the misspecification error terms as well.

that prefer characteristic (x) tend to be more (less) likely to default if $\rho_x > 0$ ($\rho_x < 0$). We denote $\rho_{x/r}$ when, instead of the preference parameter for product x, we use the willingness to pay for product characteristic x.

Key parameters for screening To capture the screening possibility, we allow the random coefficients to affect the slopes (via β_i^x) rather than the intercepts only (via ϵ_{icb}^x). This is because it is only when preferences (β_i^P) are heterogeneous that banks can influence the average characteristic of borrowers choosing a given product by changing their contract menus.[11] For instance, high default borrowers find it relatively more costly to provide a high level of down payments for each additional unit they borrow, then low LTV contracts attract unobservably safer borrowers and can be offered at a lower price. The following proposition states formally how lenders can screen in our setup.

Proposition: Test of screening and risk discrimination

It is possible to screen borrower a from borrower b with $\beta_{ax}^P \neq \beta_{bx}^P\ \forall(a \neq b)$, with β_{ix}^P the x^{th} element of vector β_i^P, using contract characteristics x and interest rate if and only if

$$\frac{\beta_{ax}^P}{\alpha_a} \neq \frac{\beta_{bx}^P}{\alpha_b} \forall(a \neq b). \tag{10}$$

We call this screening risk discrimination if $\rho_{x/r} \neq 0$, with $\rho_{x/r}$ defined above.

3.3.2 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 Benetton (2018) and Crawford, Pavanini, and Schivardi (2018), we consider that banks are risk neutral, that all borrowers refinance at the end of the teaser rate period[12] 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 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 equation 28), d is the default probability (defined in equation 1), r is the interest rate, f is the fixed rate period and *mc* is the marginal cost of lending. The derivation of the formula is in Appendix (F).

11. Given that banks do not offer a different price based on D_i in the UK, observable characteristics also drive the menu design.

12. 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 are held until maturity.

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). The fixed cost of designing a menu thus has the following form:

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

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

e_m^F is a cost shock. It 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.

F_b 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), λ is a scaling parameter that captures the cost or benefits of withdrawing a contract from the menu.

3.4 Discussion about the Model's Assumptions

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

3.4.1 Demand

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

Savings As we do not observe savings, we cannot explicitly model the constraints on the level of down payment (dp) a borrower can provide. We address this issue by modelling borrowers' choice of both LTV and the loan size and relying on a revealed preference approach to recover the demand parameters. Indeed, 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' maximisation problem (2), we allow for the menu available to each borrower (P_{ib}) to be different as a result of rejections of borrowers' applications for a particular contract. The modelling of the choice of product is general enough to encompass the case in which borrowers have or do not have perfect knowledge of which applications would be successful and which would not. We favour the perfect information case interpretation as this case can be justified by the heavy use of brokers in this market. The imperfect information case is discussed in Appendix (D.1).

Borrowers' participation in the mortgage market As shown in Andersen et al. (2021) and Benetton, Gavazza, and Surico (2021), borrowers' entry decision in the mortgage market is very inelastic to loan prices and characteristics.[13] Furthermore, Robles-Garcia (2019) and Benetton (2018) show that the level of competition is high in the UK mortgage market, making 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.

3.4.2 Supply

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

Collateral Our NPV parametrization 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 as we do not explicitly model the cost of default and instead rely on a revealed preference approach. However, it affects the 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, $\frac{L}{LTV}$ is the house value at the origination date, and δ is 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 $E[mc - min\{\delta \cdot \frac{1}{LTV}, r\}d|LTV]$. Following the literature (for instance, Benetton 2018 and Crawford, Pavanini, and Schivardi 2018), we do not identify δ and mc separately.

Static model of supply The supply model used in this paper is static, as at each period, lenders maximize the expected profits generated by current lending activities only.

13. 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.

This consideration is justified by the demand also being static. However, using the fixed cost function in the lenders' problem creates a dynamic relationship between current and past maximization problems. It makes the use of a dynamic model natural.

The following considerations can nonetheless justify the static supply approach. First, our static modelling 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 of our analysis. Indeed, the marginal costs are not affected as 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 where the counterfactual model would not be solvable with the current methods available.

Fixed cost The 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, capturing for instance marketing expenses or updates in softwares, but can also interpreted as non-monetary cost such as managerial frictions or collusion. As our model is static, the fixed cost may capture the impact of competitors' punishment strategy if the bank deviates from the current menu offering.

4 Identification and Estimation

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

For notational convenience, we collect all the parameters into the vector $\Theta := (\Theta^D, \Theta^d, \Theta^S)$ where $\Theta^D := (\Theta^P, \Theta^L)$ denotes the demand parameters related to the product demand

(Θ^P) and the loan demand (Θ^L). Θ^d contains the default parameters (β^d, ν^d, ρ^d) and Θ^S the supply ones (mc, F_b). The elements of Θ^P and Θ^L are defined in the relevant sections. Each following section — demand (section 4.1.1), default (section 4.1.2), and supply (section 4.1.3) — focuses on the identification and estimation of their 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 that value high LTV contracts the most also tend to compare less intensively products across banks, lenders can use a menu to extract more surplus from them.

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

Identification challenges for bank and contract choice There are two classic challenges related to the demand estimation. The first one is that interest rates are endogeneous. 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 which 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 available to borrowers is likely to lead to a downward bias in willingness to pay for LTV.

We use an instrumental variable approach together with bank and product fixed effects to deal with the unobserved product characteristics. The 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., branch network or customer service). Following in Benetton 2018 and Robles-Garcia 2019, we use product-specific risk-weights as cost shifters. Risk weights drive the minimum amount of capital banks must have when lending with a given product and, thus, the lending cost. They vary across lenders and over time. For the largest banks, risk weights come from an Internal Rating-Based (IRB) model. While the choice of model is endogenous, they have to be approved by the central bank before being used. As the approval process features some delay, it is unlikely that current unobservable shocks are correlated with pre-determined risk weights. Given the absence of individual-based pricing in the UK (see Benetton 2018), 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.

The consideration set bias is dealt with using a sufficient set approach, as in Crawford, Griffith, Iaria, et al. 2016. This approach shows that taking a subset of the menu for which banks' rejection is independent on 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 a downward bias of the WTP LTV estimates, our main results about the LTV distortion level and the cost of those distortions should be interpreted with caution as a lower bound to the true effect. We construct the choice set the following way, as in Benetton, Gavazza, and Surico 2021 and Robles-Garcia 2019. We build sets based on the product sold in the same month in the same geographical regions. The geographical restriction affects mostly 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 do similar restrictions for small lenders. This captures consideration bias in the search or lender rejection. Furthermore, we assume 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 cut-off values. The rationale for these restrictions is based on lenders' most common set of affordability criteria.

Identification challenges loan amount There are two econometric challenges.

The first one comes from the fact that 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 comes from a selection bias. It arises 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 coefficients to be correlated. The seminal paper Train 1986 is an extreme version of this approach as it assumes that the coefficients are perfectly correlated.

4.1.2 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). The default parameters are identified and estimated using the cross-sectional variation and the variation in the month of the mortgage origination.

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 $\hat{\beta}^P_{gcb}$. This is formally defined as:

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

$\mathcal{I}^E_P$ denotes the econometrician information set, β^P_i is defined in equation (4).[14]

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 $\hat{\beta}^P_{gcb}$ contains no information about moral hazard or burden of payment.[15]

We use the index c^- to emphasize that the average borrower type selecting 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, therefore changing the average type of borrower choosing low LTV contracts. It is this source of variation — i.e., changes in the outside options c^-— that identifies the screening for default parameter ρ^d in the default regression.

The identification of the coefficients driving moral hazard (α^d) and adverse selection (ρ) thus comes from two different sources of 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 the interest rate r_{cb} — while keeping the interest rate spreads constants — keep incentives to choose a given contract unchanged ($\hat{\beta}^P_{gcbc^-}$ does not vary) but changes the burden of payment of the borrower ($\alpha^d r_{cb}$). In contrast,

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

15. 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, Liberman, and Paravisini (2018).

variations in the spread between contracts c and other contracts — keeping r_{cb} constant — will only change the type of borrower getting contract c ($\beta^d r_{cb}$).

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

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

To limit this omitted variable concern, we use bank fixed effects, 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 $\hat{\beta}^P_{gcbc^-}$. The instrument is based on risk weights as for the product choice regression. The difference with product choice regression is that the risk weights of the 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 of those products, thus changing the type of borrower choosing product c even when the characteristics of product c did not change. Formally, we instrument $\hat{\beta}^P_{gcb}$ by replacing the interest rates by the capital requirements in equation (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, the main bias likely comes from a correlation of risk weights with acceptance and rejection rules. Such rules are lender choice variables. 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 as the economic literature argued that they are quite sticky (Agarwal et al. 2020). Yet, to be conservative, our results can be interpreted as a lower bound on adverse selection as 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, period in which the PSD data feature less information about contract characteristics.

This approach is thus outside the scope of our paper.

4.1.3 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 For notational convenience, let us rewrite the lender maximization problem (3) as follows:

$$\max_{M_b \in \mathcal{F}, P_{ib}} \Pi_b(M, P_i) - F(M_b, M_{b,t-1}) \tag{15}$$

where $\Pi_b(M, P_i)$ is the expected gross margin for a given market menu $M := (M_b, M_{b^-})$ and acceptance and rejection rule P_i offered by all banks. M_{b^-} denotes the menus offered by other lenders than b. $\mathcal{F}$ is the feasible set of menus. We allow the interest rate to be a continuous variable, but contract characteristics and the number of contracts are discrete variables. The optimality conditions, written from the econometrician information set, are thus:

$$\partial_{r_{cb}} \Pi_b(M, P_i) = 0, \text{ for all rates } r_{cb} \tag{16}$$

$$\Pr((M_b)|M_{b^-}, \Theta) = \Pr(M_b \in \text{argmax}_{m \in \mathcal{F}, P_{ib}} \{\Pi((m, M_{b^-}), P_i) - F(m, M_{t-1})\}|M_{b^-}, \Theta) \tag{17}$$

$$\iff \Pr((M_b)|M_{b^-}, \Theta) = \frac{\exp(\Pi((M_b, M_{b^-}), P_i) - F(M_b, M_{b,t-1}))}{\sum_{m \in \mathcal{F}} \exp(\Pi((m, M_{b^-}), P_i) - F(m, M_{b,t-1}))} \tag{18}$$

The first order condition with respect to interest rate (16) — analyzed formally in Appendix E — states that the interest rates can be written as a markup over the effective marginal cost. As in, for instance, Crawford, Pavanini, and Schivardi (2018), the markup is larger when the demand elasticity is low and when the burden of payment channel is low. Indeed, when elasticity is low, lenders can increase interest rates without losing customers. However, when the payment burden channel is large, increasing rates lead to larger default probabilities, providing incentives to keep interest rates low. As in Rothschild and Stiglitz (1976), interest rates, together with product characteristics, are used to maintain borrowers' incentives to self-select. As a result some contracts feature an asymmetric information premium while others get a discount (i.e., an information rent).

Equation (18) states that the menu M_b is more likely to be offered if it is the best response to menus offered by other lenders (M_b^-). 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 the identification of the parameter (λ) capturing the cost of benefits of withdrawing a product from the menu.

Contrary to Wollmann 2018, we assume that the error terms in the fixed cost function 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 equation (16) is the marginal cost of lending. As in the seminal paper Berry, Levinsohn, and Pakes 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 are recovered using equation (16), we can construct an estimate of the gross margin function $\hat{\Pi}(M_b, P_i)$. We then use it in equation (18) to recover the fixed cost.

The marginal cost equation are recovered by inverting equation (16) for each bank-product without making any identification assumptions. As a result, there are no identification challenges, but the 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 3.4.2.

Instead, the fixed cost estimation relies on the classic assumption that the error terms (e_m^F) are uncorrelated to 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 18.

An omitted variable bias if, for instance, high LTV products are often associated with higher marketing expenses. This would tend to bias the cost of high LTV products upwardly. To mitigate those issues, we use the 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, Levinsohn, and Pakes 1995)

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, Crawford, Griffith, Iaria, 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 upward bias the cost of introducing this product. To mitigate that issue, we do the estimation only at product introduction and product exclusion periods and calculate counterfactual profits in the equation using the menu from the previous period. As a robustness check, we also do the estimation considering as a set of potential products the combinations of the most common values for the characteristics of the existing products in the market.[16]

4.2 Estimation

The demand coefficient in the logit model are estimated separately for each consideration set as in Benetton 2018.

The joint estimation of the demand, default and supply parameters is computationally demanding as it would require iterating on the estimate of the average preference $E[\beta_i^P|\mathcal{I}_P^E, i\ choose\ cb]$ for each Θ^P. For this reason, we estimate each equation (5,6,7,16, 18) separately using GMM (see for instance Nevo 2001 for the mixed logit procedure) and calculate the 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 equation 6 is conditional on the product choice. The default moment built from equation 7, is conditional on the choice of product and loan size. The supply parameters are conditional on the demand parameters. The correlations between the random coefficients 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 the following. Given a consistent estimate for Θ^P — taken from the product demand estimation — we construct a consistent estimate for $E[\beta_i^P|\mathcal{I}_P^E, i\ choose\ cb])$ using Bayes' rule and the estimated preferences coefficients of equations (5,6).[17] To lower the computational burden of calculating the conditional random coefficients in $E[\beta_i^P|\mathcal{I}_P^E, i\ choose\ cb])$, 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 $E[\beta_i^P|\mathcal{I}_P^E, i\ choose\ cb])$ as a dependent variable in the default regression.

16. 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, 1500) and buckets of LTV from 60 to 95 by increasing levels of 5 percent. We only consider one product introduction for each market segment considered.

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

$$\hat{E}[\beta_i^P|\mathcal{I}_P^E, i\ choose\ cb] = \int \beta^P \frac{Prob(\text{i chooses cb}\,|\mathcal{I}_P^E, \hat{\Theta}^P, \beta_i^P)}{Prob(\text{i chooses cb}\,|\mathcal{I}_P^E, \hat{\Theta}^P)} dF(\beta^P; \hat{\Omega}^P) \tag{19}$$

$Prob(\text{i chooses cb}\,|\mathcal{I}_P^E, \Theta^P, \beta_i^P)$ is defined in equation (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 $F(\beta^P; \hat{\Theta}^P)$.

The marginal cost estimates are recovered by inverting the equation (16). We get an estimate the gross margin using the demand and marginal cost into 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 deviation 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 (18) using the estimated gross margin.

5 Estimation Results

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

5.1 Demand Results

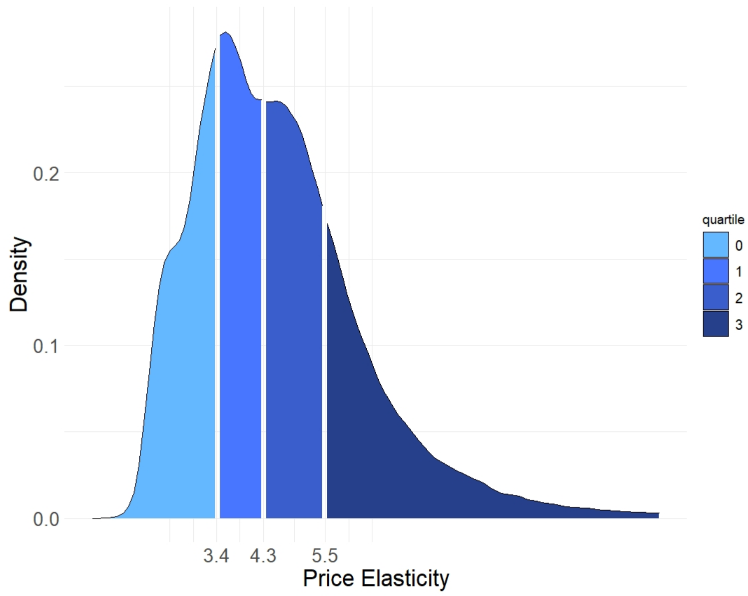

Figure 7: Distribution of price elasticity for the discrete choice regression for the full population

Discrete choice: The demand parameter coefficients are reported in Table 5.

There is substantial heterogeneity in the interest rate elasticity, mainly driven by

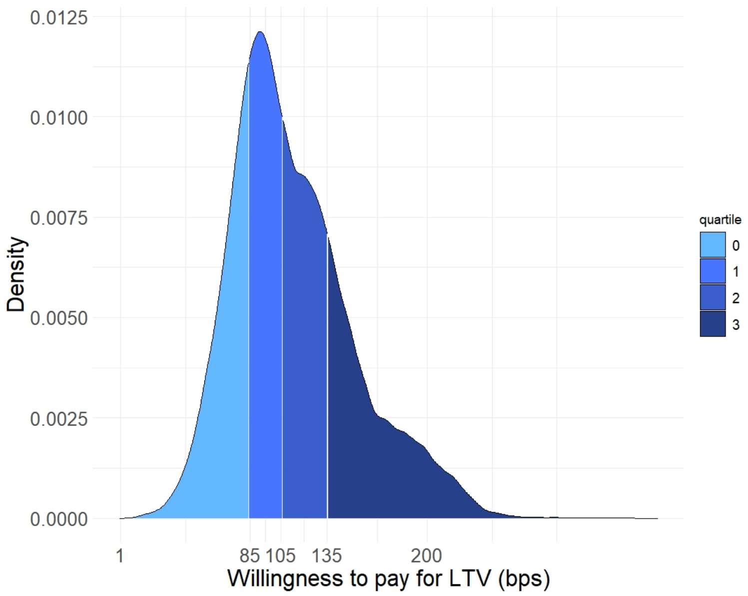

Figure 8: Distribution of WTP for LTV for the full population

income.[18] The corresponding average own-product demand elasticity is equal to 2.6, 3.6 and 5.1 for borrowers shopping 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%. Figure 7 represents the price elasticity distribution for the whole borrower population.

The estimate implies that borrowers with higher income are more sensitive to rates. It can be rationalized by, for instance, search costs as in Agarwal et al. (2020). Borrowers with higher income are more likely to be accepted into 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 the next section, default is indeed correlated with income.

The average LTV coefficient is 0.17, meaning that the average borrower likes high LTV loans. Contrary to the interest rate case, the heterogeneity is mainly driven by the random coefficient rather than income. This coefficient is significant at the 0.1 level[19].

18. The other source of heterogeneity coming from the observable heterogeneity and the random coefficients term are non-significant (statistically and economically).

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

The first quartile of the distribution is 0.13 while the third quartile is 0.21. However, when considering only the 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 for the positive coefficient results is that borrowers do not like to make down payments as 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 percent LTV increase. Figure 8 represents 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 coefficient 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 having different degrees of risk aversion or expectations about the future economic path. This implies that some borrowers prefer a fixed rate while 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 willingness to pay of -30, 8 and 50 bps for a one-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 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-pound decrease in fees.

Loan demand: The loan coefficients are all significant and reported in Table (8). We find that high LTV increases the amount borrowed by 15 percent. For the teaser rate, we find that increasing the teaser rate by 0.8 percent. 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 percent more. Borrowers with a unobserved preference for an LTV that is one standard deviation higher borrow, on average, 1.3 percent 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

The default parameter coefficients are significant and reported in Table 9.

We document positive selection along the LTV random coefficient. For a given level of income and other observable characteristics, borrowers that 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 willing to get high leverage and thus a bigger house when they know that 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 the buy the same house size but have different amounts of savings due to their different income levels. As in the UK mortgage market, there is no individual-based pricing, this correlation between observable characteristics and default drives adverse selection.

As mentioned in the demand section, longer teaser rates period 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 preferring higher teaser rates are thus likely to be more risk averse or 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 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 percent less likely to default. The results suggest that the risk aversion channel dominates. This interpretation is also consistent with the loan regression results showing that those customers tend to borrow more. Indeed, 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 The 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 percent, this implies an average fair price of between 220 and 231 bps. The 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 one. 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 The results are reported in Table 11. We find that the average fixed costs of introducing a new product are about (£ 16 M) per product or 2% of current profits. Around 30% of the fixed cost is recovered after the withdrawal of an existing product. Those numbers are comparable to Wollmann (2018), which analyses the car industry. The estimates are the ones implied by the model to justify that banks offer a discrete number of products. The sunk cost includes monetary costs such as marketing expenses, updates of the menu on all lending platforms, and changes in risk weights calculations. They also include non-monetary costs such as within-firm managing frictions. Given recent papers showing that collusion is large in banking markets (**dou2022cost**, **bruguestaxation**), our preferred interpretation for the fixed cost is collusion. Indeed, as our model is static, the fixed cost also capture the impact of competitors' punishment strategy if the bank deviates from the current menu offering.[20]

6 Counterfactual Analysis

In section 6.1, we use simulations to provide a measure of product distortions relative to the perfect information benchmark. In section 6.2, we calculate the cost of the contractual externality.

One of the standard issues in industrial organization models of product or firm entry with fixed cost is the multiplicity of equilibria, as explained, for example, in Eizenberg (2014). Another standard issue is that, in screening models, the equilibrium is difficult to characterise even in simple cases such as Rothschild and Stiglitz (1976) because the equilibrium may not exist in pure strategy.

To mitigate these concerns, we develop two well-behaved benchmarks to analyse the contractual externality: the perfect information benchmark - which features a closed-

20. Alternatively, the high cost of introducing a new contract could be caused by demand misspecification. For instance, the cost may capture the fact the unobserved bank characteristics may be related to borrowers' valuation of the size of the menu being offered to them. If borrowers dislike having a large menu because it looks complicated, then our cost function would be large to capture this channel. We do not find evidence of this channel when augmenting the demand with the number of products offered by banks.

The fixed cost estimates are also influenced by the choice of modelling the competition as a two-stage or simultaneous game. In the paper, we assumed that lenders chose product characteristics first and then compete on interest rates. The fixed costs are higher by a factor of two when allowing for a simultaneous 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.

form solution - and the social planner benchmark. We discuss those benchmarks in the following sections.

6.1 Product and Interest Rate Distortions

6.1.1 Conceptual Framework

As explained in Section 3.2, perfect information contracts may not be incentive compatible. To maintain incentives to self-select, lenders can lower the LTV of the contract designed for the low default borrowers. Alternatively, lenders can also lower the interest rate spread so that high default borrowers are indifferent between the high LTV contract and the lower rate contract. Therefore, screening will generally lead to some distortions of contracts offered by banks relative to the first best contracts.

We characterize the product distortions by comparing the menu offered in the data to the one offered under perfect information. In the perfect information case, it can be shown[21] that lenders set the product characteristics such that it maximizes the surplus generated by the loan. Formally, the lender increases the characteristic X, up to the point that the marginal cost of lending equals the borrower's willingness to pay:

$$\partial_X \frac{\beta_i^P X}{\alpha_i^P} = \partial_X \frac{mc_{icb}}{1 - d_{icb}} \tag{20}$$

We abstract away from the fixed cost for the product distortion to overcome the multiple equilibria problem mentioned in Eizenberg 2014. We use a conservative approach instead. If the result gives a characteristic X in between the discrete values observed in the data (e.g., an LTV of 83 while only LTVs of 70, 75, 80, etc are observed in the data) we set the counterfactual characteristic X to its discrete value closer to the data equilibrium value (e.g., 85 if the LTV of that borrower was for instance 90 in the data, and 80 if its LTV was 75).

We characterize the internet rate distortion by using the model 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). The formula is presented formally in Appendix E. The different components of the formula are functions of the model parameters and the data and do not require simulations. As a result, there is no equilibrium multiplicity problem.

21. Using the change of variable $V_{icb} = \beta_i^P X_{cb} - \alpha_i^P r_{cb} + \xi_{cb}$ and maximizing over V_{icb} instead or r_{icb} we get the desired first order conditions.

6.1.2 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 a high willingness to pay 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 percent of borrowers shopping between 70 and 95% LTV would get a 85-95% LTV product under perfect information-perfect competition (see table 12). This finding suggests that products below 85% LTV are introduced to screen rather than to cater to borrowers' heterogeneous preferences. We exclude borrowers shopping below 65% LTV as they constitute less than 10 per cent of the loans originated, and the data quality is lower for that sub-sample.[22] 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 observable heterogeneity and estimating the coefficient separately for each sufficient set.[23] 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. Indeed, one may be worried that this sign results from banks selecting good borrowers into 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 one hundred times larger in absolute value to imply that 10% of borrowers get offered lower than 90% LTV products. Given the standard errors of 2.810^{-6} and the average coefficient of $-3.9 \cdot 10^{-5}$ on the LTV coefficient, this situation is not likely.

As summarized in Table 12 in the appendix, we find that the product distortions when it comes to fees and teaser rates are milder. Indeed, 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 percent of the 80+ LTV borrowers). In addition, this result highly depends on how the marginal costs of lending vary with fees and teaser period. As the marginal costs are estimated for products with fees ranging from 0 to 1500 and teaser rate from 0 to 7, the product introduction results are highly dependent on our extrapolation of the marginal cost function. We find that the distribution of borrowers would shift

22. Including them would imply that LTV between 50 and 75 would be introduced but would account for less than 5 percent of the market shares.

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

towards lower-fee products and more flexible rate contracts. This is the result of interest rate distortions. Those distortions are analyzed in the next section.

6.1.3 Interest Rate Distortions: Results

The results on the interest rate decomposition are summarized in Table 15, Table 14 and Figure 9. Doing this decomposition, we find that the average fair price is 231 bps, the markup is about 116 bps while 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 while the average information rent is -30. These difference across LTV are mainly due to the fact that lower LTV loans are chosen by borrowers that are more price elastic on average. As a result banks have less able to apply large interest rate 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 to 75% LTV products. This is due to the fact that high LTV products are also more expensive to produce, implying 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 fees products are chosen by more price elastic borrowers. Under perfect information those borrowers would thus get a lower markup (see mark up columns in Tables 15 and 14). To be able to extract more surplus from other borrowers, banks make high fee product 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, as a result more price elastic borrowers are pushed to high fees products when they exist. This creates incentives introduce more high fees products relative to the first best in order to implement the screening.

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

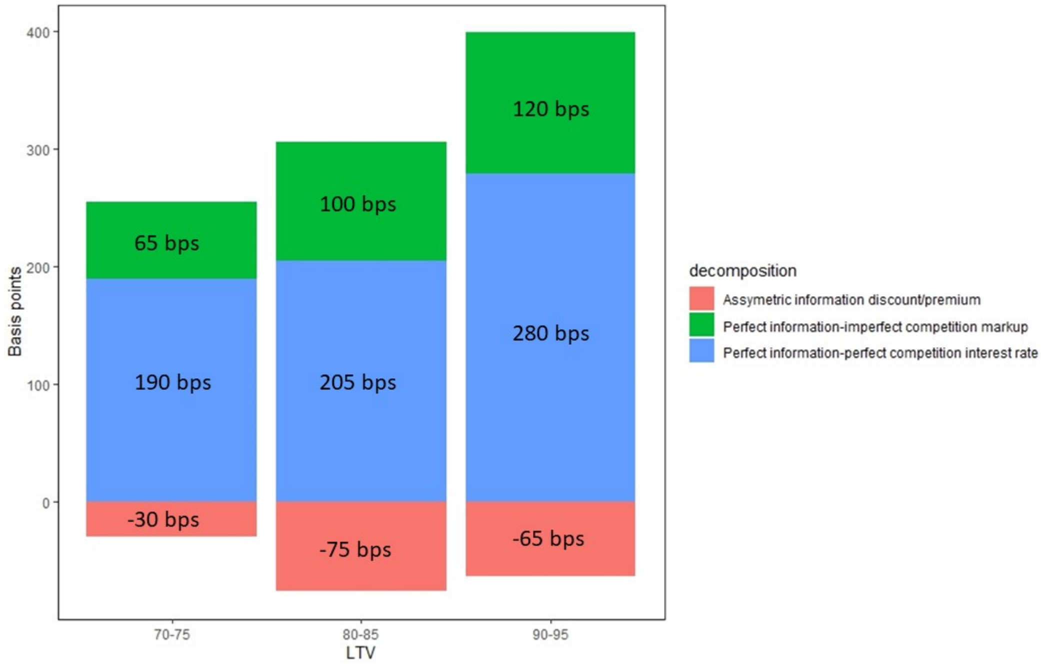

Figure 9: Interest rate decomposition by LTV

6.1.4 Summary of the Results and Economic Interpretation

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

Those 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 basis point 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) to 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 are 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 Contractual Externality

6.2.1 Conceptual Framework

To capture the contractual externality explained in Section 3.2, we solve for the contract under the following specification. 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 contractual externality by preventing an increase in welfare generated by a better allocation of borrowers to cheaper banks.

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 1976. 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, Hendel, and Whinston (2015) that rely on Riley (1979) equilibrium 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.[24] We present the formal specification in Appendix G.

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 to one achievable in a benchmark in which the contractual externality is internalized.

6.2.2 Summary of the Results and Economic Interpretation

As illustrated by Figure 10, the counterfactual simulation shows that the social planner could do a Pareto improvement by pooling more borrowers at higher LTV. Low-default borrowers are better off because they can buy a larger house. High-default borrowers

24. 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.

Figure 10: Data and social planner simulation distribution of the equilibrium interest rate and LTV distribution.

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 contractual 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.

This finding suggests there is room for Pareto improving policy interventions. As shown in the theoretical companion paper Taburet (2022), lowering competition, increasing the capital requirement on 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 risk of lending via higher LTV. This friction would then lead to too much leverage in the mortgage market instead of too little leverage. As a result, a policy Policy interventions should consider both frictions before implementing a low LTV ban.

6.3 Ban on High LTV Products

Limits on LTV are becoming increasingly popular. Indeed, according to the IMF's Global Macroprudential Policy Instruments (GMPI) database, 47 countries have introduced limits on LTVs. While those policies are used as part of the macroprudential policy toolkit, LTV limits also have an effect on the market equilibrium by restricting banks' ability to screen using LTV.

Indeed, by doing so, borrowers shopping at high LTV will be forced to move to lower LTV loans. Banks thus have to pool borrowers with different price elasticities and default probabilities or introduce new products in order to sort borrowers. To assess the impact of those policies, we solve for the situation in which the banks cannot change their menu offers and the situation in which the product offering is endogenous.

Solving the model Given the difficulties of solving for more than one endogenous characteristic using the first order condition approach (Einav, Finkelstein, and Mahoney (2021)), the numerical exercise is based on discretizing products' characteristics and using a contraction mapping to solve for rates using the interest rate first-order conditions for a given menu offering. Instead of looking at all the possible menu offering combinations, which would be too computationally demanding,footnoteIndeed, even restricting ourselves to 10 potential products of 6 banks, the potential equilibriums to compute are greater than 10^6. , we use an algorithm proposed by Lee and Pakes (2009). The idea is to start from a given equilibrium, change a fundamental parameter and allow a first bank to optimally choose which products to enter or exit, taking other banks' offers as given and knowing what the interest rate equilibrium will be.[25] We compute the new equilibrium prices using a classic contraction mapping. Then, we allow a second bank to best respond to the new equilibrium. The program cycles through the banks, continually updating the offerings until an entire cycle is complete and no firm wishes to deviate.

Fixed products scenario The average rate for 80-90 products increases from 244 bps to 255 bps. Using the interest rate decomposition we find that the average markup for 80-90 products goes from 33 bps to 48 bps. This is because borrowers who previously shopped at 95% LTV are, on average, less price elastic and more likely to default. After the LTV ban, they substitute for a lower LTV. The average price elasticity and default probability of borrowers shopping at lower thus increase leading to a price increase.

The average information rent decreases from 66 bps to 58 bps implying either that banks pool more borrowers or that the incentive compatibility constraints are easier to maintain. Using the structural model, we find that the average cost of the LTV ban is equivalent to a 10 bps interest rate increase for all borrowers.

Endogenous products scenario Allowing for product entry increases the average price from 244 bps to 283 bps and expands the choice set. This is a 30 bps increase relative to the fixed product scenario. While allowing for endogenous products could have disciplined prices by increasing competition in market segments with high markups, we find that the opposite result holds because that endogenous products allows banks to extract more surplus from high WTP borrowers. In particular, we find that the products introduced by banks following the high LTV ban are the ones that are more likely to be chosen by the new borrowers that are less price elastic: 90% LTV products, low fees, and longer teaser rates.

25. We could also consider that other banks do not change their rate

The number of products increases for two reasons. The first reason is that the number of borrowers shopping at a given LTV range increases, and the price elasticity decreases. As a result, the expected profit for any given product increases due to the market size and the markup effect; thus, it is more likely that the fixed cost becomes lower than the potential profits. This product introduction effect lowers mark-ups. However, as discussed in Tirole (1988), the existence of fixed costs can lead to too much product being offered. This happens because lenders do not internalize the business stealing effect (cannibalization) of their product introduction on competitors. As a result, competitors tend to offer too many products. Including product introduction and exclusion thus also allows for this effect to be present.

The second effect comes from incentives to screen borrowers. As the preference heterogeneity of borrowers shopping at lower LTV increases, banks have incentives to increase the number of products to screen borrowers. As discussed in section 3.2, because of the contractual externality, banks may create too many products (i.e., screen borrowers) even when the social planner would not do so.

The overall effect of product introduction on welfare in thus theoretically ambiguous. Using the structural model, we find that, compared to the situation without the ban, welfare decreases by 30 bps. This result implies that product introduction is, in our case, detrimental to borrowers' welfare as it allows banks to extract more surplus from high WTP borrowers and pushes other borrowers towards products with distorted characteristics. Not considering product introduction thus underestimates the negative impact of an LTV ban by a factor of three.

7 Conclusion

The main contribution of this paper is to provide 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 the first structural model of screening with endogenous menus of contracts.

To identify and estimate the model, we make several contributions. First, we develop a new identification strategy to test whether screening for default probability is possible. Along the way, we discuss how to adapt classic structural models to 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 accept to sell borrowers some products (rejection of loan applications), and the market is likely to feature adverse or positive selection. The second contribution is to propose a new set of tools to analyse 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 analysed by the literature in the context of multiple endogenous variables — we propose a new complementary approach. We first use perfect information, well-behaved model, as a benchmark to analyze product distortions. Second, we use a "sufficient statistic approach" to decompose the equilibrium interest rates into a fair price, a perfect information markup and an asymmetric information premium or discount. Finally, we propose a social planner benchmark to deliver a measure of the cost of coordination problems related to screening. The third contribution is to estimate the impact of policies affecting incentives to screen using the classic structural approach and discuss why their impact on contract terms is theoretically ambiguous.

In addition, our paper touches on several topics that we think are exciting avenues for future research. First, although not at the centre of our analysis, we document that the banking market features a large fixed cost of introducing products (£ 16 M per product or 2 percent of current profits). That results is comparable to the one of Wollmann (2018) for the car industry. Given that introducing a new product in credit markets does not require — contrary to the car industry — any new machine or raw material expenses, that result may imply large managerial frictions or collusion between banks. However, given the static nature of our supply model, our estimated fixed cost should not be taken at face value. We believe using a dynamic approach like the one explored in Appendix (B) instead of the static one used in this paper could help provide better estimates of those fixed costs. In turn, this would help in designing better models and policies in credit markets. Second, although acceptance and rejections are important drivers of the market equilibrium, those thresholds are unobserved in most data sets. We deal with this limitation by using a sufficient set approach in this paper, but, we believe that using a structural approach to back out those rules is also an interesting avenue for research.[26]

References

Agarwal, Sumit, John Grigsby, Ali Hortaçsu, Gregor Matvos, Amit Seru, and Vincent Yao. 2020. *Searching for approval.* Technical report. National Bureau of Economic Research.

Aiyar, Shekhar, Charles W Calomiris, John Hooley, Yevgeniya Korniyenko, and Tomasz Wieladek. 2014. "The international transmission of bank capital requirements: Evidence from the UK." *Journal of Financial Economics* 113 (3): 368–382.

Andersen, Steffen, Cristian Badarinza, Lu Liu, Julie Marx, and Tarun Ramadorai. 2021. "Reference dependence in the housing market." *Available at SSRN 3396506.*

26. We do not compute it in our estimation due to the computational burden.

Arcidiacono, Peter, and Robert A Miller. 2011. "Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity." *Econometrica* 79 (6): 1823–1867.

Benetton, Matteo. 2018. *Leverage regulation and market structure: A structural model of the uk mortgage market.* Technical report. Working Paper.

Benetton, Matteo, Alessandro Gavazza, and Paolo Surico. 2021. "Mortgage pricing and monetary policy."

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile prices in market equilibrium." *Econometrica: Journal of the Econometric Society,* 841–890.

Campbell, John Y. 2013. "Mortgage market design." *Review of finance* 17 (1): 1–33.

Chiappori, Pierre-Andre, and Bernard Salanié. 2002. "Testing contract theory: A survey of some recent work."

Chiappori, Pierre-André, and Bernard Salanie. 2000. "Testing for asymmetric information in insurance markets." *Journal of political Economy* 108 (1): 56–78.

Cloyne, James, Kilian Huber, Ethan Ilzetzki, and Henrik Kleven. 2019. "The effect of house prices on household borrowing: a new approach." *American Economic Review* 109 (6): 2104–36.

Crawford, Gregory S, Rachel Griffith, Alessandro Iaria, et al. 2016. *Demand estimation with unobserved choice set heterogeneity.* Technical report. CEPR Discussion Papers.

———. 2021. "Demand estimation with unobserved choice set heterogeneity."

Crawford, Gregory S, Nicola Pavanini, and Fabiano Schivardi. 2018. "Asymmetric information and imperfect competition in lending markets." *American Economic Review* 108 (7): 1659–1701.

Einav, Liran, Amy Finkelstein, and Mark R Cullen. 2010. "Estimating welfare in insurance markets using variation in prices." *The quarterly journal of economics* 125 (3): 877–921.

Einav, Liran, Amy Finkelstein, and Neale Mahoney. 2021. "The IO of selection markets." In *Handbook of Industrial Organization,* 5:389–426. 1. Elsevier.

Eizenberg, Alon. 2014. "Upstream innovation and product variety in the U.S. home PC market." *Review of Economic Studies* 81 (3): 1003–1045.

Handel, Ben, Igal Hendel, and Michael D Whinston. 2015. "Equilibria in health exchanges: Adverse selection versus reclassification risk." *Econometrica* 83 (4): 1261–1313.

Hertzberg, Andrew, Andres Liberman, and Daniel Paravisini. 2018. "Screening on loan terms: evidence from maturity choice in consumer credit." *The Review of Financial Studies* 31 (9): 3532–3567.

Karlan, Dean, and Jonathan Zinman. 2009. "Observing unobservables: Identifying information asymmetries with a consumer credit field experiment." *Econometrica* 77 (6): 1993–2008.

Lee, Robin S, and Ariel Pakes. 2009. "Multiple equilibria and selection by learning in an applied setting." *Economics Letters* 104 (1): 13–16.

Lester, Benjamin, Ali Shourideh, Venky Venkateswaran, and Ariel Zetlin-Jones. 2019. "Screening and adverse selection in frictional markets." *Journal of Political Economy* 127 (1): 338–377.

Morrow, W Ross, and Steven J Skerlos. 2011. "Fixed-point approaches to computing Bertrand-Nash equilibrium prices under mixed-logit demand." *Operations research* 59 (2): 328–345.

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

Riley, John G. 1979. "Informational equilibrium." *Econometrica: Journal of the Econometric Society,* 331–359.

Robles-Garcia, Claudia. 2019. *Competition and incentives in mortgage markets: The role of brokers.* Technical report. Working paper.

Rosenthal, Robert W, and Andrew Weiss. 1984. "Mixed-strategy equilibrium in a market with asymmetric information." *The Review of Economic Studies* 51 (2): 333–342.

Rothschild, Michael, and Joseph Stiglitz. 1976. "Equilibrium in competitive insurance markets: An essay on the economics of imperfect information." *The Quarterly Journal of Economics.*

Salanié, Bernard, and Frank A Wolak. 2019. *Fast," robust", and approximately correct: estimating mixed demand systems.* Technical report. National Bureau of Economic Research.

Stiglitz, Joseph E, and Andrew Weiss. 1981. "Credit rationing in markets with imperfect information." *The American economic review* 71 (3): 393–410.

Taburet, Arthur. 2022. "Screening using a menu of contracts: theory and policy implications for lending markets." *Working Paper.*

Tirole, Jean. 1988. *The theory of industrial organization.* MIT press.

Train, Kenneth. 1986. *Qualitative choice analysis: Theory, econometrics, and an application to automobile demand.* Vol. 10. MIT press.

Wollmann, Thomas G. 2018. "Trucks without bailouts: Equilibrium product characteristics for commercial vehicles." *American Economic Review* 108 (6): 1364–1406.

A Tables

A.1 Descriptive statistics

Table 1: Regression LTV on borrowers' characteristics

Variable	Age	Yearly net income	Number of borrowers	Self employed
Bellow 60% LTV	33***	39,855	1.35***	0.085***
60-70% LTV	-0.7***	-82	0.04***	0.01***
70-75% LTV	-1.5***	3675***	0.007***	-0.005***
75-80% LTV	-1.3***	1793*	0.11***	0.006***
80-85% LTV	-1.7***	1941**	0.16***	0.007***
85-90% LTV	-2.4***	-2716***	0.22***	-0.024***
95+ ltv	-2.7***	-3842***	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 2: Summary Statistics for 2018

Variable	Mean	SD	Min	Max
Loan Characteristics:				
Max LTV (percent)	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 (percent)	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	279,379			

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

	Mortgage Holiday by 2021	Arrears by 2020 (Origination: 2018)
Interest (in percent)	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

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)

Source: PSD001 + Moneyfact

A.2 Estimation Results

Table 5: Mixed logit (Origination: 2018)

	85 + LTV loans	70-85% LTV loans
Interest rate (percent)	-54	-7.1
	(50)	(44)
LTV (percent)	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 coefficient Fixed rate period	250***	100***
	(48)	(27)
Standard deviation random coefficient 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 coefficient	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 (per cent)	**LTV (per cent)**	**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 (percent)	WTP teaser rate (year)	WTP fees (pounds)
	85+ loans			
Observable heterogeneity only				
First quartile	28	1	-6.5	-6.6$\cdot10^{-3}$
Second quartile	39	1.4	-4.9	-5$\cdot10^{-3}$
Third quartile	54	1.9	-3.7	-3.8$\cdot10^{-3}$
Observable and unobservable heterogeneity				
First quartile	28	0.5	-1.1	-6.6$\cdot10^{-3}$
Second quartile	39	1.3	-4.4	-5$\cdot10^{-3}$
Third quartile	54	2.4	5.3	-3.8$\cdot10^{-3}$
	70-85 loans			
Observable heterogeneity only				
First quartile	26	0.8	0.7	-6$\cdot10^{-3}$
Second quartile	36	1	0.9	-4.3$\cdot10^{-3}$
Third quartile	51	1.4	1.2	-3.2$\cdot10^{-3}$
Observable and unobservable heterogeneity				
First quartile	26	0.7	-3	-6$\cdot10^{-3}$
Second quartile	36	1	0.8	-4.3$\cdot10^{-3}$
Third quartile	51	1.4	5	-3.2$\cdot10^{-3}$

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 (percent)	-52^{***}	-52^{***}
	(3.9)	(3.9)
LTV (percent)	0.9^{***}	0.8^{***}
	(0.2)	(0.4)
LTV=95 (percent)	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 effect	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 percent)	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)
$1_{LTV<85}$×LTV (percent)	12***	14***
	(2.8)	(0.1)
$1_{LTV>85}$×LTV (percent)	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

	$\tilde{X}_c$=1 (1)
Profits (β)	4.47^{***}
	$(6.37 \cdot 10^{-2})$
Nbr of Product included (θ)	$7.8 \cdot 10^{7***}$
	$(5.04 \cdot 10^{3})$
Nbr of Product excluded ($\theta \cdot \lambda$)	$-2.4 \cdot 10^{7***}$
	$(5.04 \cdot 10^{3})$
Bank fixed effect	No
Time fixed effect	No
Observations	61

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

A.3 Counterfactual Results

Table 12: Product distortion (80+ LTV loans)

	Ideal LTV (percent)	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 13: 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 14: 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 15: Intest 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

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

B Menu Adjustment Costs: Dynamic approach

I want to estimate:

$$Pr(d_{jt}(M, \tilde{M})),\ with\ d_{jt}(M, \tilde{M}) := 1_{\{V_{jt}(M) - sc(M, \tilde{M})) \geqslant V_{jt}(\tilde{M}) + e_{Mtj} - e_{\tilde{M}tj}\}}$$
$$and\ V(M_{t-1}, (e_t)) = max_{M \in \mathcal{M}_j} \underbrace{\Pi_j(M) - sc_M + \beta E[V(M, (e_{t+1}))]}_{v(M, M_{t-1})} + e_{Mtj}$$

With $(e_{Mtj} - e_{\tilde{M}tj})$ are iid and EVD I get:

$$Pr(d_{jt}(M, \tilde{M})) = \frac{exp(u_{\tilde{M}}(M_{t-1}, M_t))}{1 + \sum_{m \neq \tilde{M}} exp(u_{\tilde{M}}(M_{t-1}, m))} \tag{21}$$

with:

$$\begin{aligned} u_{\tilde{M}}(M, M_{t-1}) &:= \Pi(M) - \Pi(\tilde{M}) - [sc(M, M_{t-1}) - sc(\tilde{M}, M_{t-1})] \\ &- \beta[\underbrace{log(Pr(\tilde{M}|M)) - log(Pr(\tilde{M}|\tilde{M}))}_{Observable\ in\ data}] \end{aligned} \tag{22}$$

The last term $(log(Pr(\tilde{M}|M)) - log(Pr(\tilde{M}|\tilde{M})))$ comes from using the EV assumption and rewriting the value function as in Arcidiacono and Miller (2011) (cf. **Proof**)

I parametrize:

$$sc(M', M) = \sum_c \theta' \tilde{X}_c [\underbrace{\mathbf{1}_{c \in M', c \notin M}}_{Inclusion} + \underbrace{\lambda \mathbf{1}_{c \in M, c \notin M'}}_{Exclusion}] \tag{23}$$

Proof:

$$
\begin{aligned}
v(M, M_{t-1}) =&\Pi(M) - sc(M, M_{t-1}) + \beta[log(\sum_m exp(v(m, M))) + cst] \\
=&\Pi(M) - sc(M, M_{t-1}) + \beta[log(v(\tilde{M}, M)) - log(Pr(\tilde{M}|M)) + cst] \qquad (24)
\end{aligned}
$$

and noting that, as in (Arcidiacono and Miller (2011)):

$$
\begin{aligned}
&u_{\tilde{M}}(M, M_{t-1}) := v(M, M_{t-1}) - v(\tilde{M}, M_{t-1}) \\
&= \Pi(M) - \Pi(\tilde{M}) - [sc(M, M_{t-1}) - sc(\tilde{M}, M_{t-1})] \\
&- \beta[\underbrace{log(Pr(\tilde{M}|M)) - log(Pr(\tilde{M}|\tilde{M}))}_{Observable\ in\ data}] \qquad (25)
\end{aligned}
$$

C Demand with Roy's Identity

Guided by the micro foundation presented in appendix D.2, we parametrize the indirect utility derived at the optimal borrowing amount given the loan characteristics X and price r as

$$
U_i(L_i(X, r); X, r) := A_i(X)\frac{L_i(X, r)}{LTV} + V_i(Y_i), \qquad (26)
$$

where Y_i is the income of borrower i, A_i is a function of the 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 parametrization 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 Roys' 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 (D.3), we show a parametrization 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 $L_i(c, b) := L(X_{cb}, r_{cb})$:

$$
\begin{aligned}
&V_i(c, b) = \overbrace{\beta_{icb}X_{cb} - \alpha_{icb}r_{cb} + \xi_{cb}}^{\tilde{u}_i(c,b)} + \sigma_{icb}^{-1}\varepsilon_{icb} \qquad (27) \\
&ln(L_i(c, b)) = \tilde{\beta}_{icb}X_{cb} - \tilde{\alpha}_{icb}r_{cb} + \nu D_i + e_{icb}^L \qquad (28) \\
&with\ (\beta_{icb}, \alpha_{icb}, \sigma_i^{-1}, \tilde{\beta}_{icb}, \tilde{\alpha}_{icb}, e_{cb}^L)\ correlated,
\end{aligned}
$$

where u_i is a monotonic transformation of the indirect utility U_i defined in equation

(26).[27]

D Model Extensions and Micro-foundations

D.1 Imperfect Information about acceptance and rejections

When borrowers do not observe acceptance and rejection rules, denoting p_{icb} the probability of being accepted, the utility they derive from a contract $c \in C$ is:

$$p_{icb}u(cb) + (1 - p_{icb})\beta[E_\varepsilon[V(c)] - cost] \tag{29}$$

$$V(c) = max_{\{x \in C \backslash c\}}[p_{ix}u(x) + (1 - p_{ix})\beta E_\varepsilon[max_{\{x \in C \backslash c\}} V((c, x)) - cost] \tag{30}$$

$V(x)$ is the expected utility after being rejected from the contracts present in vector x. Since rejections are observed by other banks, the probability of being accepted in another contract may be lower upon rejections. Assuming that borrowers get a new extreme value draw after each rejection, once can calculate V in a closed form manner. To ease computational burden, one can assume that the probability of being accepted after the first rejection i 0 and replace $V(c)$ by an outside option that is borrower specific $\bar{u}_i$.

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

$$p_{icb}u_i(cb) + (1 - p_{icb})\bar{u}_i \tag{31}$$

D.2 Micro-foundation borrowers ' utility mortgage market

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

The assumptions about borrowers' utility function are made for tractability and do not impact the qualitative results.

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

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

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.

D.2.1 Indirect utility functional form micro foundation

Toy Model consume in period 1, default and loose the house in period 2

$$
\begin{aligned}
u(C^*, H^*) :=& max_{\{C,L\}}\ \mu C_1 + \overbrace{(1 - \frac{\delta}{2}\frac{r}{Y_2}L)}^{survival\ probability}\ [\frac{\phi}{P_H}\frac{L}{ltv} + \mu C_2] \\
& pC_1 + (1 - ltv)\frac{L}{ltv} = Y_1 \\
& pC_2 = Y_2 - rL
\end{aligned}
$$

$\frac{\delta}{2}\frac{r}{Y_2}\frac{L}{ltv}$ represents the fact that you are more likely to default as you leverage This implies:

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

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:

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

D.3 Derivation of the Demand system

Borrowers maximize:

$$
max_c u(L_{ci}, c) = max_c A_{ic}\frac{L_{ci}}{ltv} + V(Y)
$$

A_c captures that default or consumption trade-off depends on contracts c features

$$
\begin{aligned}
max_c u(L_c, c) =& max_c ln(A_{ic}) + ln(L_c) - ltv(ltv) \\
& ln(A_{ic}) = \tilde{\beta}_i X_c + \sigma_i \varepsilon_{ic}
\end{aligned}
$$

From Roy's Identity (A_{ic} $doesn't\ vary\ with\ Y, r$):

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

Integrating with respect to DF_i (loan discount factor):

$$ln(L_c) = ln(ltv) + \gamma \frac{V_Y}{A_{ic}} DF_i + cst$$
$$with: \ cst := \beta_i X_c + \epsilon_i, with\ (cstr)$$

set $\frac{DF_i}{A_{ic}} = \nu D_i + (\beta_1 X_c + \gamma_c + \beta_2 X_i + \gamma_i) r_c$

In the regression, allow for some element of A_i to be proxied by income. That way the income element of β_i^b need not be equal to the one in β^L

$$Pr(i\ choose\ c) = \frac{exp(\beta_i^b X_c - \alpha_i^b r_c)}{\sum_j exp(\beta_i^b X_j - \alpha_i^b r_c)} \tag{32}$$
$$ln(L_{ci}) = \alpha_i^L r_c + \nu D_i + \beta_i^L X_c + \sigma_\epsilon \epsilon_i \tag{33}$$

with β_i^d, α_i^b correlated with α_i^L, β_i^L.

E First order conditions with respect to interest rates

The First order conditions with respect to interest rate yield:

$$r_c = \{ \overbrace{\frac{mc}{1 - E[d|bc]}}^{Fair\ price} + \overbrace{\frac{E[\Phi_c]}{E[-\Phi_c']} (\overbrace{\frac{1 - E[d|bc] + \beta_r^d r}{1 - E[d|bc]}}^{Burden\ of\ Payment})}^{PI\ mark-up} + \overbrace{\sum_{j \neq c} \frac{E[\Phi_j']}{E[-\Phi_c']} \frac{\tilde{\pi}_c}{1 - E[d|bc]}}^{AI\ discount/premium} \} \frac{1 - E[d|bc]}{1 - E[d'|bc]} \tag{34}$$

Where $\Phi_c := \sum_i \frac{exp(u_{ic})}{\sum_x exp(u_{ix})} L_{ic}$ is the expected amount lent, $\tilde{\pi}_c := (1 - E[d|cb]) r_c - mc_c$ is the expected profit on each loan unit given that the borrower choose the contract c at bank b.

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 the contract c at bank b ($E[d|bc] := \beta^d X_c + \alpha^d r_c + \rho \frac{E[\Phi_c \beta_i]}{E[\Phi_c]}$). It 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 choosing each contracts ($E[d|bc]\forall cb$). $\frac{E[\Phi_j]}{E[-\Phi'_j]}$ is the impact of borrowers product elasticity (i.e., competition). $\frac{(\beta^d_r r}{1-E[d|bc]})$ 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 $\sum_{j\neq c} \frac{E[\Phi'_j]}{E[-\Phi'_c]} \frac{\tilde{\pi}_j}{1-E[d|bc]}$ is the equivalent of the information rent in the textbook principal agent model.

The ratio $\frac{1-E[d|bc]}{1-E[d'|bc]}$ in which $E[d'|bc] := \beta^d X_c + \alpha^d r_c + \rho\frac{E[\Phi'_c \beta_i]}{E[\Phi'_c]}$ is scale up the three terms by taking into account the fact that changes in r impact the type of borrowers choosing a given contract.[28]

F Derivation 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:

$$C = \frac{Lr(1+r)^T}{(1+r)^T - 1} \tag{35}$$

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):

$$D = \frac{Lmc(1+mc)^T}{(1+mc)^T - 1} \tag{36}$$

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

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

$$L\sum_{k=1}^{F} \delta^k [\frac{r(r+1)^T}{(r+1)^T - 1} - \frac{mc(mc+1)^T}{(mc+1)^T - 1}] + \gamma b \sum_{k=F+1}^{T} \delta^k [\frac{R(R+1)^{T-F}}{(R+1)^{T-F} - 1} - \frac{mc(mc+1)^{T-F}}{(mc+1)^{T-F} - 1}] \tag{37}$$

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, Pavanini, and Schivardi (2018), assuming that banks consider the average default instead of the probability of defaulting in each period, for a constant discount rate ($\delta < 0$), denoting d a dummy equal to 1 if borrower default, the present

28. When the number of product 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, $\Phi' \approx \Phi(\tilde{\beta}_r + 1)\Phi \approx \Phi$.

value of lending up period F is:

$$C \cdot E[(1-d)] \cdot \sum_{k=1}^{F} \delta^k = Lr\frac{(1+r)^T}{(1+r)^T - 1} \cdot E[(1-d)] \cdot \frac{1-\delta^F}{1-\delta}\delta \tag{38}$$

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:

$$PV \approx L \cdot \{E[(1-d)]r\frac{\delta}{1-\delta} + \gamma E[(1-d)]R\frac{1-\delta^{T-F}}{1-\delta}\delta^F - [\frac{\delta}{1-\delta} + \gamma\frac{1-\delta^{T-F}}{1-\delta}\delta^F]mc\} \tag{39}$$

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

$$PV \approx L \cdot [E[(1-d)]rF + \gamma RE[(1-d)](T-F) - [F + \gamma(T-F)]mc] \tag{40}$$

We further assume as in Benetton (2018) 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 borrower remortgage). We can thus also abstract from the discount rate if $\delta < 1$ as it is constant across mortgages, we thus get:

$$NPV_{icb} := L \cdot [E[(1-d)]r - mc] \text{ when } \delta < 1 \tag{41}$$

The above expression comes implies that banks do care about fixing the interest rate except from its impact on the cost of lending (mc), default (d) or on demand (L). This result comes from the assumption that $\delta^F \approx 0$. It may be problematic as 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 $\delta^F \approx 0$ 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 like the 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:

$$NPV_{icb} := L \cdot [(1-d)r - mc]F \text{ when } \delta = 1 \tag{42}$$

Alternative approach:

Without using Crawford, Pavanini, and Schivardi (2018) assumption about default, the expression for the annuity would be would be, using d as the per period default probability:

$$C\sum_{k=1}^{t}((1-d)\delta)^t = Lr((1-d)\delta)\frac{(1+r)^T}{(1+r)^T-1}\frac{1-((1-d)\delta)^t}{1-((1-d)\delta)} \tag{43}$$

Using the same approximations as in Benetton (2018), $\frac{(1+r)^T}{(1+r)^T-1} \approx 1$ and $\partial_r\gamma = 0$, the expression for the NPV becomes:

$$NPV_{icb} := L\cdot[(1-d)\delta\frac{1-((1-d)\delta)^F}{1-\delta+d\delta}r - mc\frac{1-\delta^F}{1-\delta}] \text{ when } \delta < 1 \tag{44}$$

$$NPV_{icb} := L\cdot[(1-d)\frac{1-(1-d)^F}{d}r - mc\cdot F] \text{ when } \delta = 1 \tag{45}$$

Here again, as 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 the interpretation of the mc as discussed in 3.4.2. Moreover, when d is small as in our empirical application and δ equal to 1, the expression becomes the same as in Crawford, Pavanini, and Schivardi (2018):

$$NPV_{icb} \underset{d\to 0}{\sim} L\cdot[(1-d)r - mc]\cdot F,\ \text{when } \delta = 1 \tag{46}$$

G Contractual externality model

Formally, lender problem is defined as:

$$max_{C_{bt},M_{bt}\in\mathcal{F}^{C_{bt}},P_{ibt}}\sum_i n_i\sum_{c=1}^{C}\overbrace{\Pr(\text{i chooses c}\,|\,\text{i chooses b})}^{IC}NPV_{ic} - F(M_{bt},M_{bt-1}) \tag{47}$$
$$s.t.\ \forall i\ E[\max_c u_{ic}+\epsilon] \geqslant E[\max_c \bar{u}_i+\epsilon]\ (PC)$$

$\Pr(\text{i chooses c}\,|\,\text{i chooses b}) := \frac{exp(u_{ic})}{\sum_{x\in[\![1,C]\!]}exp(u_{ix})}$ captures how borrowers i make their choice of contract when having only access to bank b contracts. We use this demand instead of the one used in the structural model ($\frac{exp(u_{ic})}{\sum_{x\in B}exp(u_{ix})}$) to shut down the intensive margin (i.e., competition) channel.

$\Pr(\text{i chooses b}) := \frac{\sum_{x\in[\![1;C]\!]}exp(u_{ix})}{\sum_{x\in[\![1;C]\!]}exp(u_{ix})}$

$E[max_c u_{ic}+\epsilon] \geqslant E[max_c\bar{u}_i+\epsilon] \iff \sum_{c=1}^{C}exp(u_{ic}) \geqslant Cexp(\bar{u}_i) \iff E_c[exp(u_{ic})] \geqslant exp(\bar{u}_i)$ 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.

NPV_{ic} is the net present value of lending to borrower i via contract c. It is formally defined in Appendix F as the amount lent to borrower i multiplied by the expected revenues generated by each lending unit minus the unit cost of lending via contract c.