PRICE NEGOTIATION IN DIFFERENTIATED PRODUCT MARKETS: THE CASE OF INSURED MORTGAGES IN CANADA.

JASON ALLEN
BANK OF CANADA

ROBERT CLARK
HEC MONTRÉAL

JEAN-FRANÇOIS HOUDE
UW-MADISON

PRELIMINARY AND INCOMPLETE
DO NOT CITE WITHOUT PERMISSION.

1. INTRODUCTION

Mortgage markets are decentralized. Contract terms for each borrower are determined through a search and negotiation process: borrowers search across different lender options and then bargain over a mortgage rate. The objective of this paper is to measure market power in this environment, and to quantify the importance of different market frictions that can lead to positive profit margins for lenders.

The mortgage industry has many institutional features which suggest it should be competitive: homogeneous contracts, negotiable rates, and, for a given consumer, common lending costs across lenders (due to securitization). As a result, even with a small number of competing lenders, informed borrowers can gather multiple quotes offering interest rates that reflect the expected cost of lending.

However, there is important heterogeneity in the ability of consumers to understand the subtleties of financial contracts, in their ability or willingness to negotiate and search

Date: This version: January 31, 2012.
Correspondence to Jason Allen: Bank of Canada, Ottawa, Ontario, K1A 0G9; Phone: (613) 782-8712; Email: jallen@bankofcanada.ca, Robert Clark: HEC Montréal, CIRANO and CIRPÉE, Montreal, Quebec; Phone: (514) 340-7034; Email: robert.clark@hec.ca, Jean-François Houde: University of Wisconsin-Madison, Madison, Wisconsin and CIRANO; Phone: (608) 262-3805; Email: houdejf@ssc.wisc.edu. This research has benefited from the financial support of the NSF (SES-1024840). We thank the Canadian Mortgage and Housing Corporation and Genworth for providing us with the data. The views in this paper are those of the authors and do not necessarily reflect those of the Bank of Canada. All errors are our own.

for multiple quotes, and also in their degree of loyalty to their main financial institutions. Survey evidence in Canada and in the United States reveals that, while some buyers get multiple quotes when shopping for their mortgage contract, nearly half only get one. Moreover, recent surveys suggest that 80% of consumers search for a quote at their main financial institution, and that the majority of these end up contracting with them. Even among homogeneous contracts consumers who remain with their home bank, or who gather few quotes may, as a result, be offered higher rates by lenders.

We propose and estimate a model to disentangle the different channels through which market power can arise for a given transaction in this environment. The first source of market power is search frictions. In our context search frictions do not arise because borrowers lack information, but rather because of the effort they must put forth when gathering multiple quotes. These frictions may induce profits for lenders since they permit them to price discriminate between consumers. To quantify search frictions the model we develop is sequential; consumers are initially matched with a home bank to obtain a mortgage quote, and can then decide, based on their search costs, whether or not to gather additional quotes from banks in their neighborhood.

The second source of market power is switching costs. In our context, switching costs might arise because consumers have a higher willingness-to-pay for their main financial institution than for competing lenders. This preference could stem from the fact that most lenders offer complementary services, and many consumers combine their deposit-taking, day-to-day banking, and loan transactions with the same financial institution. To the extent that there are costs of moving a checking account, investment account, etc., consumers may be willing to pay a higher price to stay with their home bank. Note that this home-bank premium could also come from other sources. For instance a consumer's main financial institution may be better equipped to provide a more competitive overall banking service, perhaps by reducing the fees on other products.

Finally, our model permits an idiosyncratic match value between consumers and lenders which represents a form of cost differentiation. Lenders can value a particular borrower differently, and so, for observationally equivalent consumers some lenders will be more competitive than others.

To shed light on these issues, we analyze detailed transaction-level data on a large set of approved mortgages in Canada between 1999 and 2001 and administered by either the Canadian Mortgage and Housing Corporation or Genworth Financial. These data provide information on features of the mortgage, household characteristics (including place of residence), and market-level characteristics. One advantage of our setting is that all of the mortgage contracts in our sample are insured. Since lenders are protected in the case of default and insurance qualifications and premiums are the same across lenders,

borrowers who qualify at one lender know they will also qualify at other lenders. The richness of the consumer data in combination with lender-level location data and survey data on the shopping habits of consumers allow us to empirically measure market power and distinguish between the three different sources of market power just described.

The key parameters of the model are the mean and variance of search costs and the home bank premium. Depending on the specification, we find that over the five year period of the contract the average search cost corresponds to an upfront sunk cost of between \$1, 047 and \$1, 590. We also estimate substantial amount of dispersion across consumers, which leads to significant amount of asymmetric information at the initial stage of the negotiation process. The home bank premium ranges from about \$759 to \$1, 617. In other words, consumers are willing to pay between \$759 and \$1, 617 upfront to stay with their home bank and avoid having to switch banks.

These two sets of parameters are mostly responsible for generating positive markups for lenders. The average markup is estimated to be 2.9%. The remaining parameters suggest that conditional on searching, consumers are able to extract most of the transaction surplus. Indeed, we estimate that lenders are close to be homogenous in costs, which leads to intense price competition for consumers that exhibit low search costs. The average markup is estimated to be 4.1% for searchers and 1.9% for searchers, but the distribution is much more skewed for searchers with close to 25% of them facing zero markup.

We use our model to simulate the effect of a merger between two of Canada's largest financial institutions. We find that the overall impact of the merger on mortgage rates is quite small (around $\$5.50$ a month). The effect is much larger for searchers and for consumers with a smaller set of lenders from which to choose. These results are consistent with the descriptive analysis presented in our companion paper measuring the impact of an actual merger that occurred in the Canadian mortgage market (Allen, Clark, and Houde (2011b)).

Since our model is one of search and negotiation in which rate dispersion is endogenously determined, we are related to two important literatures: search and bargaining. There is a small but growing empirical search literature, but it has mostly focused on posted-price markets and/or assumes exogenous price distributions. See for instance Sorensen (2000), Hortaçsu and Syverson (2004), Hong and Shum (2006), and De Los Santos, Hortacsu, and Wildenbeest (2011). There is also a growing empirical literature on the relationship between bargaining and price dispersion. This literature has mostly concentrated on health markets and markets for medical devices (see Dafny (2010), Grennan (2011), Capps, Dranove, and Satterthwaite (2003), Dranove, Satterthwaite, and Sfekas (2008), and Town and Vistnes (2001)), although more recently has looked at the market

for televisions (Crawford and Yurukoglu (2011)). One limitation of this literature is that it largely focuses on bilateral bargaining models. Specifically, the outside option of buyers is not determined as an equilibrium object that depends on offers they could expect to get from other sellers. This makes it difficult to study questions related to market structure.

While the industrial organization theory literature provides a number of relevant models for combining these two elements, to our knowledge, the only other empirical paper that combines them is Hall and Woodward (2010) which also studies the mortgage market. Hall and Woodward study the compensation (ie. the origination fee) paid to mortgage brokers in the United States, and model the potential benefits to consumers of gathering quotes from two brokers rather than just one. They model the negotiation process as an English auction in which the lowest-cost broker wins and pays the cost of the losing broker. Theoretically, our setup is closest to Armstrong and Zhou (2011), Wolinsky (1986), and Bester (1993). In Bester (1993) competing firms negotiate with consumers that can search across stores for better prices. Armstrong and Zhou (2011) develop a similar model and focus on the incumbency advantage that firms can develop with regard to their loyal customers. In Wolinsky (1986) consumers are motivated by more than just price. They search for a firm that will provide them with a suitable product, not just one with a low price.

In the labor literature, empirical models combining search and negotiation have been developed and estimated. Postel-Vinay and Robin (2002) for instance estimate a model in which firms that are differentiated in terms of their productivity make take-it-or-leave it wage offers to workers. Firms can adjust their offers depending on the characteristics of workers, and can make counter offers should their employees receive offers from competing firms.

The market for mortgages is not the only one in which prices are negotiated and consumers incur search costs to choose among a set of differentiated products. Other examples include markets for personal insurance, market for both new and used cars, and markets for other consumer loans. Despite its prevalence, this form of pricing has been largely ignored by researchers studying market power in differentiated products markets. This is potentially problematic since these markets do not fit the standard discrete-choice model used to evaluate market power. Specifically, consumers do not necessarily consider all available products, and the researcher has no knowledge of the distribution of rejected prices. Ignoring the actual pricing mechanism, however, can lead to an incomplete and biased analysis. To the extent that transaction prices reveal something about the valuation of consumers for the product that they choose, this can lead to a biased estimate of preferences. There have been two main approaches to solving this problem. In their study of the demand for new automobiles, Berry, Levinsohn, and Pakes (2004) ignore

transaction prices, abstracting away from the price setting mechanism actually used in the market. In contrast, in their analysis of sub-prime used car loans, Adams, Einav, and Levin (2009) assume monopoly pricing. We believe that the framework we have proposed here could be adapted to study any of the markets listed above. Moreover, although we focus on the home bank premium and switching costs, our framework can accommodate more general forms of differentiation. Our approach could, for instance, be applied in differentiated product markets with search and bargaining, such as the car market (see Berry, Levinsohn, and Pakes (2004) and Langer (2011)).

The paper is organized as follows. Section 2 presents details on the Canadian mortgage market, including market structure, contract types, and pricing strategies. Section 3 presents the model. Section 4 presents a description of the household-level data. Section 5 discusses the estimation strategy and Section 6 describes the empirical results. Section 7 concludes.

2. THE CANADIAN MORTGAGE MARKET

2.1. **Market structure.** The Canadian mortgage market is currently dominated by six national banks (Bank of Montreal, Bank of Nova Scotia, Banque Nationale, Canadian Imperial Bank of Commerce, Royal Bank Financial Group, and TD Bank Financial Group), a regional cooperative network (Desjardins in Quebec), and a provincially owned deposit-taking institution (Alberta's ATB Financial). Collectively, they control 90 per cent of assets in the banking industry and are called the "Big 8."

The market was not always this concentrated. Until the early 1990s the Canadian residential-mortgage market also featured a large number of trust companies. Trusts make mortgage loans, funding them by issuing guaranteed investment certificates and accepting deposits. At the time the main difference between trusts and banks was that trusts were more lightly regulated with regards to reserve requirements. In particular, trusts did not have to hold reserves against mortgages, while chartered banks did. This provided trusts with a competitive advantage in the mortgage market due to lower cost of funding. Cross-ownership between the two types of institutions was not permitted until the 1992 revisions to the Bank Act. Following these revisions banks and trusts were granted almost identical powers, making them undifferentiated products from the point of view of consumers.

As a result of the Bank Act revisions and a series of bad residential and commercial loans that created solvency and liquidity issues for the trusts in the 1980s, Canadian chartered banks acquired the majority of trust companies over the course of the following decade. The merger wave led to the six largest banks controlling approximately 80 per cent of the mortgage market – almost double their 1980s market share. These mergers

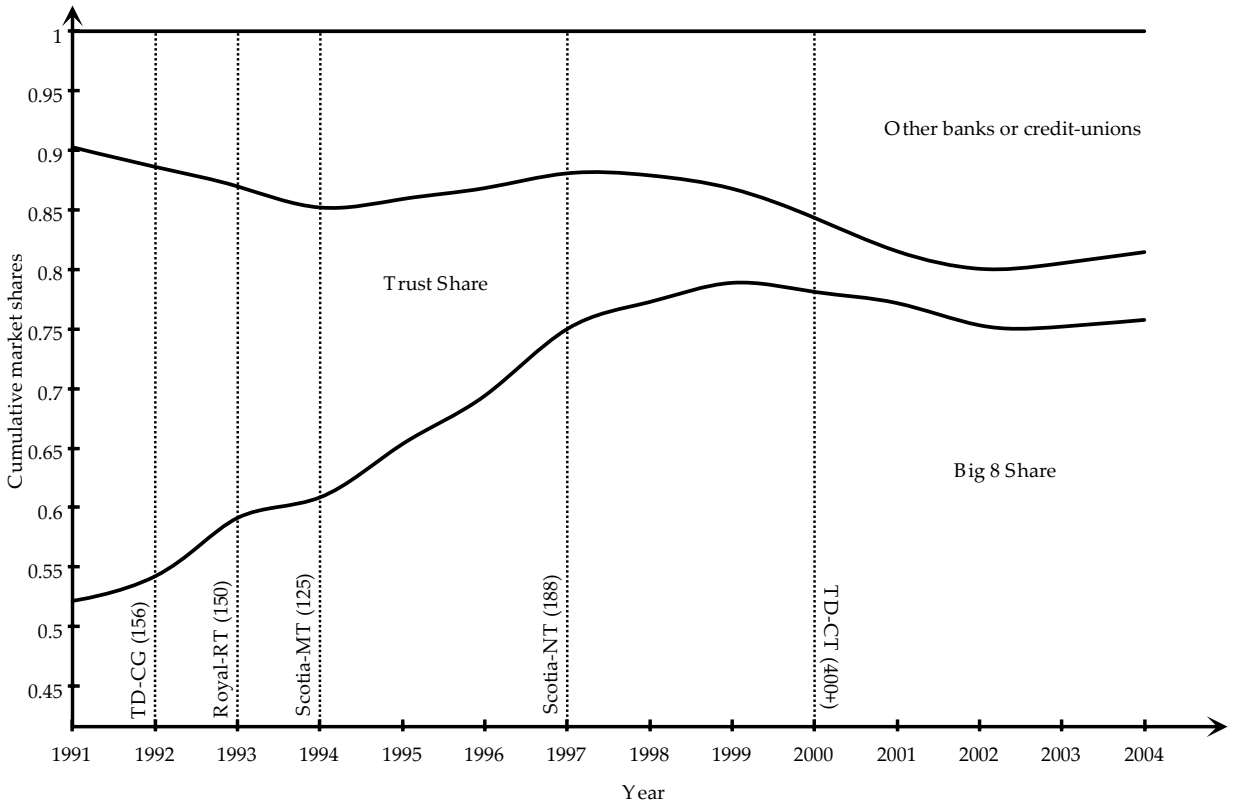


FIGURE 1. Evolution of the market share of financial institutions among new insured mortgage contracts (smoothed)

all resulted in significant expansion of the merged entity's branch network since in each case the Canadian Competition Bureau required little or no forced divestiture of branches. Figure 1 presents the evolution of the mortgage-market share of the main lending groups – The Big 8, Trusts, Credit Unions and other banks. Today, there are still many trusts operating in Canada, but they are small and their influence on the mortgage market is much less than it was prior to 2000.

2.2. **Mortgage contracts and mortgage insurance.** There are two types of mortgage contracts in Canada – conventional mortgages which are uninsured since they have a low loan-to-value ratio, and high loan-to-value mortgages, which require insurance (for the lifetime of the mortgage). Today, 85% of newly issued mortgages fall in the latter category. The primary insurer is the Canada Mortgage and Housing Corporation (CMHC), a crown corporation with an explicit backstop from the federal government. There are a number of private insurers as well, the only one in existence during our sample was Genworth Financial, which also has an explicit government of Canada guarantee, albeit

for 90 per cent. CMHC's market share during our sample averages around 80 per cent. Our analysis focuses on mortgages insured by CMHC or Genworth.

All insurers use the same strict guidelines for insuring mortgages. First, borrowers with less than 25% equity must purchase insurance.[1] Second, borrowers with monthly gross debt payments that are more than 32% of gross income or a total debt service ratio of 40% will almost certainly be rejected.[2] The mortgage insurers charge the lenders an insurance premium, ranging from 1.75 to 3.75 per cent of the value of the loan – lenders pass this premium onto borrowers. Insurance qualifications (and premiums) are common across lenders and based on the posted rate. Borrowers qualifying at one bank, therefore, know that they can qualify at other institutions, given that the lender is protected in case of default.

During our sample period, nearly all mortgage contracts were fixed rate, among which over 85 per cent had a 5 year term (the second most common term is 36 months). A 5 year fixed-rate mortgage contract must be renegotiated every five years, which in effect acts like an adjustable rate mortgage with a fixed time-frame to renegotiate. This has been the standard contract offered by Canadian banks since the late 1960's. Almost all contracts have 25 year amortization periods.

2.3. **Pricing and negotiation.** The large Canadian banks operate nationally and post prices that are common across the country on a weekly basis in both national and local newspapers, as well as online. There is little dispersion in posted prices, especially at the big banks: the coefficient of variation on posted rates for the Big six is close to zero (Allen and McVanel (2009)).

In contrast, as we discuss in further detail below, there is a significant amount of dispersion in transaction rates. In Allen, Clark, and Houde (2011a), for example, we document that the coefficient of variation in margins between 1999 and 2004 was 60%. Approximately 10% of borrowers pay the posted rate. The remainder receive a discount below the posted price. This comes about because borrowers can search for and negotiate better rates. One option for borrowers is to visit local branches and negotiate directly with branch managers who have the authority to offer borrowers discounts below the posted price under general guidelines from headquarters. Local branch managers compete against rival banks, but not against other branches of the same bank. Alternatively borrowers can hire brokers to search for the best rates on their behalf. Unlike in the United States, brokers in Canada have fiduciary duties. Brokers are compensated by lenders, but

[1]This is true during our sample. Today borrowers with less than 20% equity must purchase insurance.

[2]Gross debt service is defined as principal and interest payments on the home, property taxes, heating costs, annual site lease in case of leasehold, and 50 per cent of condominium fees. Total debt service is defined as all payments for housing and other debt.

TABLE 1. Summary statistics on shopping habits

	1999	2000	2001	2002	2003	2004
Contact main FI						80%
Number of quotes						1-2
Several rate offers	61%	56%	46%	57%	51%	
Loyalty to FI	57%	57%	48%	63%	63%	54%

Source: Canadian Association of Accredited Mortgage Professionals (CAAMP). Each entry calculates the average answer for new home-buyers.

"hired" by borrowers to gather the best quotes from multiple lenders.[3] The model that we use in this paper focuses only on branch-level transactions, and therefore excludes brokers.

Our data do not provide direct information on the number of quotes gathered by borrowers. However, survey evidence from the Canadian Association of Accredited Mortgage Professionals (CAAMP) revealed that on average borrowers negotiate with between one and two financial institutions when searching for a rate, and between 46% and 61% of first-time home buyers gather multiple quotes. Table 1 reproduces these statistics from an annual survey conducted by CAAMP. Notice that we will use these aggregate statistics on the fraction of consumers gathering more than one quote in the estimation model model (see section 5.2).

In 2004, 80% of new borrowers revealed that they contacted their main financial institution when shopping for their mortgage. Depending on the year, nearly 60% of new home-buyers remained loyal to their main institution. This loyalty rate is higher in our data-set since the survey includes broker transactions (which we exclude from our analysis), and focuses only on first-time buyers (we also consider former home owners who buy a new house). Indeed, from our contract-level data-set, we observe that only 35% of consumers dealing with brokers remain loyal to their home institution, while 73% of individual transactions are loyal consumers.

This loyal stems largely from the evolution of the banking system following the 1992 Bank Act revisions which led many Canadian households to treat their primary bank as a "one-stop shop", where they purchase the majority of their financial services. Another survey, the Canadian Financial Monitor (Ipsos-Reid), also characterizes the leading role played by the main institution of consumers; defined as the one with which borrowers conduct day-to-day banking activities. From Table 2 we see that 67 per cent of Canadian households have their mortgage at the same financial institution as their main checking

[3]Detailed survey evidence by Taddingstone in 2005 (MortgageBrokerReport@taddingstone.com) found that brokers on average contact 5.9 lenders for their clients, suggesting they do, in fact, assist in gathering multiple quotes.

TABLE 2. Distribution of financial services between main and secondary institutions

Account	Main FI	Second FI	All other FI
Mortgage (all)	67.4%	10.9%	21.7%[a]
Mortgage (no broker)	70.3%	10.8%	18.9%
Mortgage (broker)	37.3%	30.6%	32.1%
Loan	55.8%	9.6%	34.6%
Credit card	77.9%	20.7%	1.4%
GIC or term deposit	72.8%	15.8%	11.4%
Bonds, t-bills, GI's	45.3%	7.8%	46.9%
Mutual funds	38.8%	7.2%	54.0%

Source: Canadian Finance Monitor survey conducted by Ipsos-Reid, between 1999 and 2007.

account. In addition, 55 per cent of household loans, 78 per cent of credit cards, 73 per cent of term deposits, 45 per cent of bonds/guaranteed investments and 39 per cent of mutual funds are held at the same financial institution as the households main checking account.

3. MODEL

We propose a sequential model in which consumers with heterogeneous search costs are initially matched with their home bank to obtain a quote, and then decide whether to keep searching by gathering multiple quotes from the remaining lenders in their local market. The initial stage is analogous to a bargaining model with incomplete information. The home bank makes a take-it-or-leave-it initial offer without knowing the cost for the consumer of gathering multiple quotes, and tries to screen consumers who are unlikely to search. This is the first source of market power for large network institutions: banks with a large consumer base have an incumbency or first-mover advantage, and are more likely to transact with consumers with high search costs.

We consider a second source of market power, arising from differentiation across lenders. Differentiation arises first because consumers might favor their home bank: everything else being equal, consumers have a higher willingness-to-pay for their home bank relative to competing financial institutions. This "home-bank" premium induces loyalty or inertia in consumers' financial service choices. This is motivated by the fact that most lenders are multi-product firms, and a large fraction of consumers combine their deposit-taking, day-to-day banking and lending transactions with the same institution. Moreover, to the extent that consumers face switching costs to transact with a bank other than their home

bank, some consumers will be willing to pay more to avoid switching financial institutions.

The model also includes an idiosyncratic match value between borrowers and lenders, which introduces an additional source of differentiation. In particular, lenders in the model have heterogenous evaluations of the value of dealing with an observationally equivalent consumer.

We describe the model in detail in the next three subsections. First, we present the notation, and formally define the timing of the model. Then, we describe the search decision of consumers, and the process generating the initial quote. Finally, we introduce the second stage of the model in which banks are allowed to compete for consumers.

3.1. **Timing and payoffs.** The timing of the model is as follows. First, consumers obtain an initial quote P^0 from their home bank h. If the offer is rejected, consumers organize a multilateral negotiation game between a set of banks in their neighborhood denoted $\mathcal{N}_i$. We model the multilateral negotiation process as a simultaneous Bertrand-Nash game among lenders in $\mathcal{N}_i$, in which consumers choose the highest utility option rather than the lowest offer.

The simultaneous assumption in the second stage allows us to abstract from considerations related to the order of arrival of competing offers. We believe it is a more accurate description of the market than a model with sequential offers. In practice, banks are able to lower their initial offer if consumers receive a lower price quote from a competing bank.

We assume the following payoff structure for consumers and firms, respectively:

$$\text{Consumers:} \qquad U_{ij} = \theta_{ij} - P_{ij}, \tag{1}$$

$$\text{Firms:} \qquad \Pi_{ij} = P_{ij} - C_{ij}, \tag{2}$$

where $P_{ij} = r_{ij} \times L_i$ is the monthly payment on a loan of size L_i offered by bank j. Throughout the paper, we assume that the loan size and the downpayment are predetermined at the beginning of the initial negotiation stage. Consumers and banks are therefore assumed to negotiate solely on the interest rate.

The willingness to pay of consumer i for bank j is determined by the home-bank premium. Let E_{ij} be a dummy variable equal to one if consumer i has prior experience dealing with bank j, and zero otherwise. The willingness to pay function is expressed as:

$$\theta_{ij} = \begin{cases} \lambda & \text{If } E_{ij} = 1, \\ 0 & \text{Otherwise.} \end{cases} \tag{3}$$

This expression can easily be enriched, for instance by allowing λ to vary across consumers based on observed and unobserved attributes, or by introducing other forms of differentiation across lenders.

The cost term measures the direct lending costs for the bank, and the future benefits associated with selling complementary services to consumer i. Both components are related to variables affecting the risk of default, and the risk of loan pre-payment over contract length. While lenders are fully insured against default risk, the event of default certainly implies additional transaction costs to lenders that lowers the value of lending to risky borrowers. The pre-payment risk is perhaps more relevant in our context, since consumers are allowed to reimburse 15% of their mortgage every year without penalty.[4] Since we do not observe the performance of the contract along these two dimensions, we use a reduced-form expression to approximate the net present value of the contract. In particular, we model C_{ij} as a function of observed consumer and firms fixed-effects, denoted by Z_{ij}, an unobserved attribute ϵ_i that symmetrically affects all lenders, and an idiosyncratic match value u_{ij}:

$$C_{ij} = L_i \times (Z_{ij}\gamma + \epsilon_i - u_{ij}) = L_i \times (c_{ij} - u_{ij}). \tag{4}$$

Note that we use small-case letters to identify variables measured in terms of a hundred dollar loan. The idiosyncratic component of firms' profits comes from several sources: branch manager compensation, idiosyncratic evaluation of future revenues, and idiosyncratic evaluation of pre-payment or default risks. Importantly we assume that C_{ij} is observed by all parties at the beginning the negotiation process.

Finally, the transaction surplus from a (ij) match is equal to:

$$V_{ij} = \theta_{ij} - C_{ij}. \tag{5}$$

3.2. **Search decision and initial quote.** Consumers choose to search for additional quotes by weighting the value of accepting P^0, or paying a sunk cost κ_i in order to lower their monthly payment. The benefit of gathering quotes, denoted W_i, originates from generating competition between lenders. It is observed by consumers and their home bank at the beginning of the negotiation process. We therefore use the terminology "search" to describe the action of shopping for extra quotes, rather than acquiring information.

The search decision of consumers is defined by a threshold function, which yields a search probability that is increasing in the value of the initial offer and the outside option

[4]On average borrowers pre-pay an additional 1% of their mortgage every year.

of consumers, and decreasing in the home-bank premium:

$$\begin{aligned}\Pr\left(\text{Reject}|P^0, W_i\right) &= \Pr\left(\kappa_i < W_i - \left(\theta_{ih} - P^0\right)\right) \\ &= H(P^0|W_i). \end{aligned} \tag{6}$$

Bank h chooses P^0 to maximize its expected profit. As described in Section 2.3, lenders do not commit on a fixed interest rate, and are open to haggling with consumers based on their outside options. Therefore, we assume that the initial quote can be revised downwards in the second stage if consumers are willing to gather better offers. Our Bertrand-Nash assumption implies that bank h will offer a more competitive quote in the second stage of the game if it can provide a higher surplus than the outside option (i.e. $V_{ih} > W_i$). If this is the case, the second stage quote, denoted by P^*, will be such that the consumer is indifferent between h and the next best alternative:

$$\theta_{ih} - P^* = W_i \Leftrightarrow P^* = \theta_{ih} - W_i.$$

Anticipating this outcome, the home bank faces the following maximization problem:

$$\max_{P^0} \quad (P^0 - C_{ih})\left(1 - H(P^0|W_i)\right) + 1(V_{ih} > W_i)H(P^0|W_i)\underbrace{(P^* - C_{ih})}_{=V_{ih}-W_i},$$

which yields the following implicit representation of the initial-quote markup:

$$P_i^0 - C_{ih} = \underbrace{1(V_{ih} > W_i)(V_{ih} - W_i)}_{\text{Differentiation}} + \underbrace{\frac{1 - H(P_i^0|W_i)}{h(P_i^0|W_i)}}_{\text{Search cost}} \tag{7}$$

where $h(P_i^0|W_i) = \partial H/\partial P^0$. In words, the previous expression decomposes the home-bank markup into a component coming from differentiation across lenders, and the unobserved search cost of consumers.

In order to estimate the model, we assume that κ_i is distributed according to an exponential distribution with translation parameter $\bar{\kappa}$ and variance σ_κ.[5] The translation parameter corresponds to the common sunk search cost, and the variance term measures the importance of asymmetric information. The exponential distribution has a constant hazard-ratio, which yields the following piece-wise linear expression for the optimal initial quote:

$$P_{ih}^0 = \begin{cases} \theta_{ih} - W_i + \sigma_\kappa & \text{If } V_{ih} > W_i, \\ C_{ih} + \sigma_\kappa & \text{Otherwise.} \end{cases} \tag{8}$$

[5]The translated exponential cumulative distribution function takes the following form: $\Pr(\kappa_i < x) = 1 - \exp\left(-\frac{1}{\sigma_\kappa}(x - \bar{\kappa})\right)$, where $\sigma_\kappa > \bar{\kappa} \geq 0$.

Finally, by substituting this expression for the initial quote into equation 6, we can characterize the equilibrium search probability conditional on the value of the outside option W_i:

$$H(W_i) = 1 - \exp\left(-\frac{1}{\sigma_\kappa}(\max\{W_i - V_{ih}, 0\} + \sigma_\kappa - \bar{\kappa})\right). \tag{9}$$

The previous expression implies a lower bound on the rejection probability, given by $\bar{H} = 1 - \exp\left(-\frac{\sigma_\kappa - \bar{\kappa}}{\sigma_\kappa}\right)$. This probability is associated with the case in which the home bank is guaranteed to retain the consumer (i.e. $V_{ih} > W_i$). As the value of the outside option increases, the monthly payment that consumers can hope to obtain gets larger, and the search probability increases towards one.

3.3. **Competition stage.** Conditional on rejecting P^0, all lenders in the choice-set $\mathcal{N}_i$ compete for the contract (including bank h). We model this competition as a simultaneous Bertrand-Nash pricing game in which consumers choose the highest utility option.

Firms are willing to "bid" for consumer i as long as they can earn positive profits: $P_{ih} \geq C_{ij}$. An offer equal to C_{ih} yields the highest utility that each bank can provide, and is equal to the transaction surplus V_{ij}. If bank j is the highest surplus lender, it can match the offer of the second-highest surplus bank and still make positive profits. The unique Nash equilibrium transaction price is therefore a quote P_i^* such that the consumer is indifferent between the highest-surplus bank, and the zero-profit offer from the second-highest surplus bank:

$$P_i^* = \theta_{ij} - \max_{k \neq j} V_{ik}, \text{ if } V_{ij} = \max_{k \in \mathcal{N}_i} V_{ik} = V_{(1)}. \tag{10}$$

This expression implies that the transaction price will include a positive markup that depends on the importance of product differentiation between lenders, and heterogeneity in match values. For instance, lenders competing with the home-bank will on average have to discount the contract by a larger margin in order to attract new customers.

Finally, the value of shopping corresponds the surplus generated by the second-highest lender in consumer i's choice-set:

$$W_i = V_{(2)} = \max_{k: V_{ik} < V_{(1)}} V_{ik}. \tag{11}$$

This is the value of the outside option determining the outcome of the first-stage negotiation.

3.4. **Discussion of model assumptions.**

Inelastic housing demand. We model the choice of lenders as a discrete choice, abstracting away from the possibility that consumers can decrease their downpayment or buy a larger house when receiving larger discounts. In principle this assumption could be relaxed by

modeling the decision as a discrete-continuous choice problem, adapting the framework developed Dubin and McFadden (1984) and Hanemann (1984).

We decided against this approach mostly because we observe a discrete distribution of loan-to-value ratios. The insurance premium charged by the two insurance companies has a piece-wise linear form, with kinks at every increase of 10%. As a result consumers tend to bunch around the kinks of the premium schedule (see Figure 2 (a) below), which discretizes the downpayment decision. This prevents us from using Roy's identity to derive a smooth loan-size demand function, complicating the analysis. Moreover, the discrete nature of the LTV distribution suggests that loan size demand is locally inelastic: for small changes in interest rate offers consumers choose a fixed loan-size in order to avoid paying an extra amount in insurance premium.

Posted interest rates. Posted rates do not enter the model described above. In practice, banks post a common interest rate that is adjusted weekly to reflect changes in the cost of funding. This rate can be thought of as an upper bound on the monthly payment that branch managers across the country can offer, since overage is illegal in Canada. This could in principle constrain the equilibrium interest rates, and affect the lending decisions of banks. For instance, consumers should not qualify for a loan if the cost of lending exceeds the revenue evaluated at the current posted rate. Similarly, high cost borrowers could be constrained to borrow at the posted rate if their outside option was less attractive than the posted-rate at their home bank.

We abstract from these considerations in the current paper because they would add unnecessary complexities to the likelihood function, without providing extra benefits in terms of our analysis. First, very few consumers are actually paying the posted interest rate. We estimate that less than 10% of borrowers pay a rate that is within 10 basis points of the current posted interest rate.[6] Second, for nearly every week in our sample, there is no dispersion in posted interest rates across the 12 largest lenders. Therefore, it is unlikely that the posted rate is used to attract new customers; at least not during our sample period. Third, banks have an incentive to post an artificially high interest rate that is not binding. Indeed, the pre-payment penalty is calculated as a fraction of the interest payments remaining on the contract, evaluated at the posted rate valid at the signature date, rather than the transaction interest rate. Banks therefore have an incentive to raise the posted rate, in order to reduce their pre-payment risk.

[6]We observe the posted rate with error, since we do not observe the date at which the contract was negotiated, but the closing date on the sale of the house. As a result about 5% of borrowers pay an interest rate above the rate posted by their lender.

Complete versus incomplete information. The model also assumes that the value of consumers' outside options is known by consumers and banks (i.e. u_{ij} is observed to all parties). This assumption greatly simplifies our analysis, by providing analytical expressions for both stages of the game. The incomplete version of the model, in which the outcome of the competition stage is privately observed is perhaps more intuitive, but is significantly more complicated to estimate. Moreover, it exhibits very similar empirical predictions.

To see this, notice that the Bertrand pricing game is strategically equivalent to a descending auction in which the consumer sequentially lowers the asking price, and openly announces a bid preference formula given by his/her willingness to pay function. Therefore, the transaction price is invariant to our information assumption.

However, our information assumption does affect the outcome of the first-stage negotiation. Under incomplete information, both lenders and borrowers are uncertain about the value of the competitive outcome, but observe the value of the home-bank transaction V_h. The value of search then becomes a function of two random variables: the surplus form the first and second best options among $\mathcal{N}_i \backslash h$ banks. This affects the expression for the search probability, and modifies the profit maximization problem of the home bank.

Although the optimal initial offer does not have a simple analytical expression, it shares similar properties with the full-information optimal quote. In particular, for low values of the winning probability $F_1(V_{ih})$, the initial quote is equal to its complete information counterpart (i.e. $P_i^0 = C_{ih} + \sigma_\kappa$). As V_{ih} gets larger, the premium over C_{ih} increases non-linearly, and converges to a finite constant function of the expected consumer surplus from the highest surplus option (excluding the home bank). In this region of V_{ih} the two models differ. Empirically, the incomplete-information model is less flexible than the full-information model, since it imposes a strict lower bound on the size of monthly payments. A detailed analysis of the incomplete information model, including its corresponding likelihood function is available upon request.

4. DATA

4.1. **Mortgage contracts and sample selection.** Our main data-set is a sample of insured contracts from the Canada Mortgage and Housing Corporation (CMHC) and Genworth Financial between January 1999 and December 2001. We focus on this window for three reasons. First, between 1992 and 1999, the market has been transiting from markets with a larger fraction of posted-price transactions and loans originated by trust companies, to a decentralized market dominated large multi-product lenders. Our model is a better description of the latter period. Second, between August 2002 and September 2003,

TABLE 3. Summary statistics on contractual characteristics in the full sample

	Distribution observations	
	Number	Fraction
New home purchase	74,155	0.939
Fixed-rate term	75,576	0.957
5 Years term	68,469	0.867
25 Years amortization	72,891	0.923
Non-broker transaction	47,429	0.755
Missing values (broker, fico, residential status)	20,000	0.253
Total sample size	29,279	0.419

TD/Canada Trust experienced with a new pricing scheme based on a "no-haggle" principle. Understanding the consequences of this experimentation is beyond the scope of this paper. Finally, the 1999-2001 period also includes the TD-Canada Trust merger, which produces useful variation in the choice-set of consumers.

We obtained a 10% random sample of all contracts from CMHC, and the full set of contracts originated by the 12 largest lenders from Genworth Financial. We further sample from the Genworth contracts to match their annual market share, which by 2004 was approximately 30%. Both data-sets contain information on 20 household/mortgage characteristics, including all of the financial characteristics of the contract (i.e. rate, loan size, house price, debt-ratio, risk-type), and some demographic characteristics (e.g. income, prior relationship with the bank, residential status, dwelling type). Table 10 in the Appendix lists all of the variables included in data-set. In addition, we observe the location of the purchased house up to the forward sortation area (or FSA).[7]

We restrict our sample to contracts with homogenous terms. In particular, from the original sample we select contracts that have the following characteristics: (i) 25 years amortization period, (ii) 5 year fixed-rate term, (iii) newly issued mortgages (i.e. excluding refinancing and renewal), (iii) contracts that were negotiated individually (i.e. without a broker), (iv) contracts without missing values for key attributes (e.g. credit score, broker and residential status). The 5 year fixed-rate mortgage contract must be renegotiated every five years, which in effect acts like an adjustable rate mortgage with a fixed time-frame to renegotiate. This contract type has traditionally been the most popular in Canada, and its market share has been stable over the sample period. Table 3 illustrates the breakdown of the full sample according to those characteristics. The final sample includes slightly more than fifty thousand observations, or 42% of the initial sample. Most of this drop originates from omitting broker transactions, which represent 25% of newly

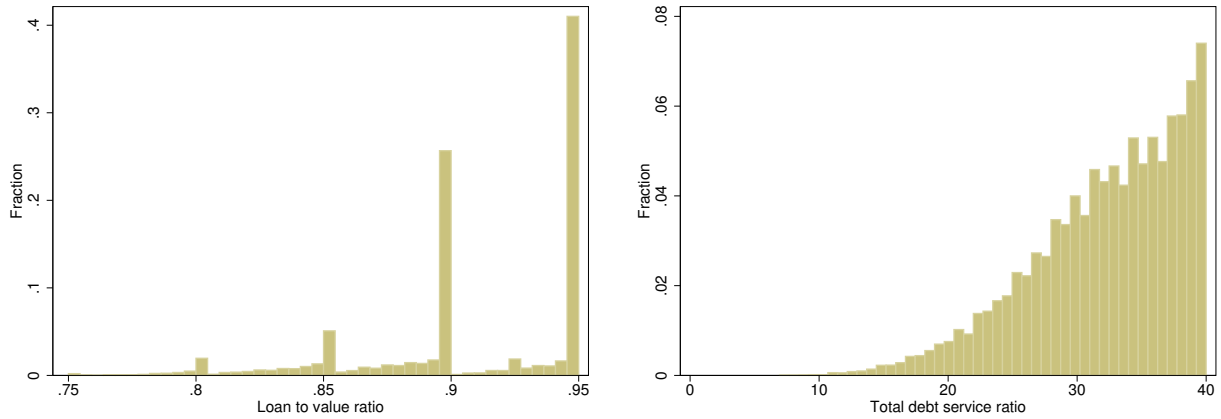
[7]The FSA is the first half of a postal code. We observe nearly 1,300 FSA in the sample. While the average forward sortation area (FSA) has a radius of 7.6 kilometers, the median is much lower at 2.6 kilometers.

TABLE 4. Summary statistics on mortgage contracts in the selected sample

	N	Mean	SD	Min	Median	Max
Loan ($ X1000K)	29,279	1.34	.533	.405	1.26	3
Income ($ X1000K)	29,279	.682	.259	.161	.647	1.98
Monthly payment ($ 1000K)	29,279	.9742	.3801	.2688	.9206	2.439
Total debt service ratio (TDS)	29,279	.3222	.05798	.0692	.332	.4
Other debt ($ X1000K)	29,279	.8595	.5206	.000046	.7568	4.758
LTV	29,279	.91	.0439	.75	.9	.95
LTV=95%	29,279	.38	.485	0	0	1
FICO (mid-point)	29,279	.666	.0734	.5	.7	.75
Switchers	18,692	.268	.443	0	0	1
Renters	29,279	.546	.498	0	1	1
Parents	29,279	.0672	.25	0	0	1

Sample: 5-year fixed-rate contracts issued between 1999 and 2001. Contracts negotiated through brokers are excluded. The sample also excludes top and bottom 1% of the loan size distribution.

FIGURE 2. Loan to Value and Total Debt Service Ratios: 1999-2001



(A) Loan to value

(B) Total debt service

issued mortgages. We also drop a large number of contracts with missing characteristics. Most of these missing values are concentrated in first six months of 1999, when CMHC and Genworth started collecting additional information on broker transactions and the residential status on households (i.e. new home owner).

Table 4 describes the main financial and demographic characteristics of the borrowers in our sample, where we trim the top and bottom 1% of observations in terms of income, loan-size, and interest-rate premium. The resulting sample corresponds to a fairly symmetric distribution of income and loan size. The average loan size is nearly $\$140,000$

which is twice the average annual household income. The total debt service (TDS) ratio is capped at 40%, but most consumers are not constrained by this maximum. Figure 2 (b) illustrates the distribution of TDS in our sample. From this variable we construct a measure of the total other monthly debt payments subtracting the mortgage payments from the total debt services (e.g. credit card debts, car loans). On average households monthly debt payments other than the mortgage are $\$862$.

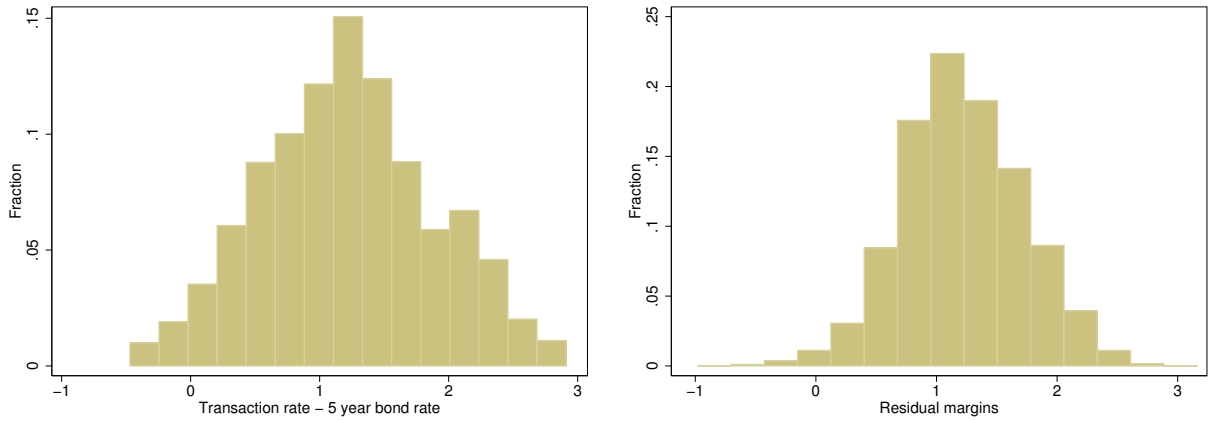
The loan-to-value (LTV) variable shows that many consumers are constrained by the minimum down-payment of 5% imposed by the government guidelines. Nearly 40% of households invest the minimum, and the average loan-to-value is 91%. Figure 2 (a) plots the distribution of the LTV ratio. LTV ratios are highly localized around 90 and 95, and to a lesser extent 75, 80, and 85. The clustering comes about because the insurance premium schedule is discrete, and there are only a small number of price-quantity pairs. Moreover, the vast majority of households in our data (i.e. 96%) roll-over the insurance premium into the initial mortgage loan. The loan size measure that we use includes the insurance premium for those households.

The variable labeled "switchers" is a dummy variable equal to one if the duration of the prior relationship with the mortgage lender is zero. 73% of households choose a lender with which they already have a prior financial relationship.[8] The fraction of switchers is significantly larger for new home-buyers (i.e. formerly renters or living with their parents), and for contracts negotiated through a broker. Notice that this variable is missing for nearly 10,000 contracts. Those contracts were either insured by Genworth (which does not report the prior experience variable), or issued by Royal Bank (for which the variable is mis-measured).

The fact that transaction interest rates are negotiated rather than posted induces a substantial amount of dispersion. Figures 4a and 4b illustrate this dispersion by plotting the distribution of retail margins in the sample. We measure margins using the 5-year bond-rate as a proxy for marginal cost. The average transaction rate is 1.2 percentage point above the 5-year bond rate, and exhibits substantial dispersion. Importantly, a large share of the dispersion is left unexplained when we control for a rich set of covariates: financial characteristics, week fixed effects, lend/province fixed-effects, lender/year fixed-effects, and location fixed-effects. These covariates explain 43% of the total variance of observed margins. Figure 4b shows the histogram of the residual dispersion in margins, scaled up using the unconditional average margin. The standard-deviation of retail margins is

[8]Note that due to data limitations we do not measure the switcher variable for contracts issued by Genworth, and for one financial institution. The fraction of switchers is measures using only the remaining contracts.

FIGURE 3. Dispersion of retail margins between 1999-2001



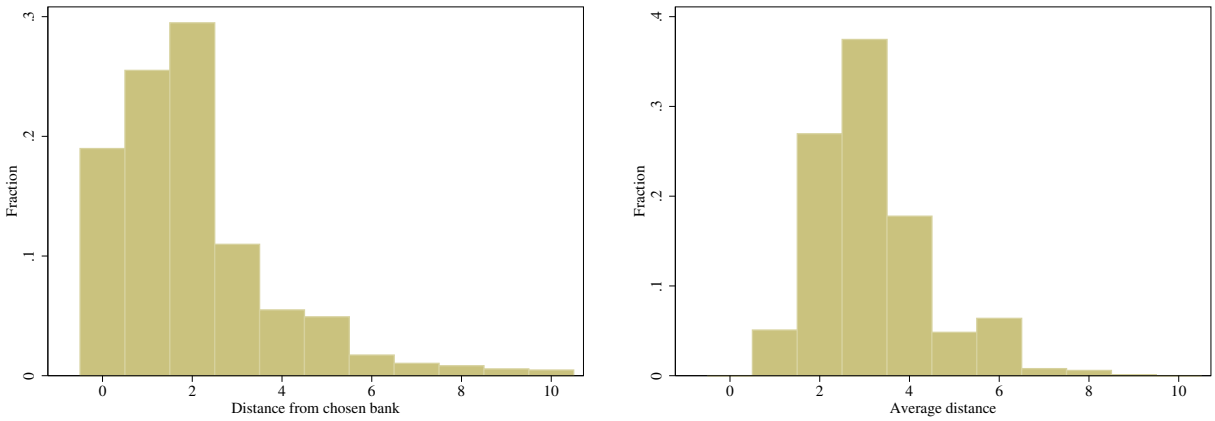
(A) Margins (B) Residual margins

equal to 66 basis points, while the residual margins has a standard-deviation of 50 basis points.

4.2. **Local markets and lender information.** Our main data-set contains the lender information for 10 lenders during our sample period. The remaining lenders are coded as "Other Bank", "Other credit union", and "Other trusts". While the "Other trusts" category corresponds to less than 2% of contracts in our sample, the other two categories represent a sizable share of contracts in some regions. We assign the contracts issued by Trusts to generic "Other lender" category. The credit-union market is fragmented, and we do not attempt to impute the missing lender information (i.e. this would amount to estimate a large number of credit-union fixed effects). Instead, consumers transacting with an "Other credit-union" are assumed to deal with the same "Other lender" category, which shares common characteristics across the country. The "Other Bank" category is different and includes mostly two institutions: Laurentian Bank is mostly present in Quebec and Eastern Ontario, while HSBC is present mostly in British Colombia and Ontario. We exploit this geographic segmentation and assign the "Other banks" customers to HSBC or Laurentian based on their relative presence in the local market around each home location. After performing this imputation, consumers face at most 13 lending options: Alberta Treasure Bank (ATB), Bank of Montreal (BMO), Bank of Nova Scotia (NS), Canada Trust (CT), Canadian Imperial Bank of Commerce (CIBC), Desjardins, Laurentian, National Bank of Canada (NBC), Royal Bank of Canada (RBC), Vancity, and Other Lender.

Not all consumers have access to every option, because of the uneven distribution of branches across local markets. We exploit this variation by assuming that consumers shop for their mortgage locally, in a neighborhood around the location of their new house (e.g.

FIGURE 4. Distribution of minimum distances between banks and consumers



(A) Distance to chosen lender

(B) Distance to other lenders

municipality). To implement this, we match the new house location with the postal code associated with each financial institution's branches. The branch location data is available annually, and comes from Micromedia-ProQuest; a provider of commercial address information in Canada. The information relative to the location of each house is coarser than the location of branches. Therefore, we assume that each house is located in the center of its FSA, and calculate a somewhat large euclidian distance radius of 10 KM around it to define the borrower's maximum choice-set. Formally, a lender is part of consumer i maximum choice-set if it has a branch located within less than 10KM of the house location. We use this definition to measure the relative presence of each lender (i.e. number of branches in choice-set), and the number of lenders within each choice-set (i.e. number of lender with at least one branch).

Figures 4 illustrate the distribution of minimum distances between each house's FSA centroid and the closest branch of each lender. On average consumers transact with banks that tend to be located close to their house. The average minimum Euclidian distance is nearly 2 KM for the chosen institution, and above three for the other lenders. In fact the distributions indicate that 80% of consumers transact with a bank that has a branch within 2 KM of their new house, while only 30% of consumers have an average distance to competing lenders lower or equal to 2 KM. This feature reflects the fact consumers tend to choose lenders with large networks of branches.

In Table 5 we measure the average network size of the chosen institution relative to the average size of others present in the same neighborhood (i.e. relative network size). On average consumers transact with lenders that are nearly 60% larger than their competitors in terms of branches; the median is smaller at 28%.

TABLE 5. Descriptive statistics on local market structure

	Mean	**Min**	**P25**	**P50**	**P75**	**Max**
Nb. contracts	455	11	29	169	410	4288
Nb. FIs (in 10 KM)	6.09	2	5.18	6.12	7.03	8.12
HHI-Branch (in 10 KM)	2240	1527	1874	2089	2325	5370
C1-Contract	41.4	21.6	29.2	36.8	48.5	90
HHI-Contract	1304	338	517	762	1424	7300
Relative network size	1.58	.831	1.11	1.28	1.52	10.6

Markets are defined as census-divisions (130 obs.). Sample excludes market with less than 10 contracts between 1999 and 2001.

The remaining variables in Table 5 measure the level of concentration aggregated at the census-division level. On average each consumer faces six lenders within 10 KM. Most of these banks have a relatively small presence, as indicated by the large Herfindahl-Hirschman index, calculated using the distribution of branches within 10 KM of each contract (i.e. both the mean and median are above two thousands). The C1 and HHI-contract measures also suggest a lack of competition. On average, the top lender in each region controls 41.4% of contracts. The HHI-contract variable suggests a somewhat lower level of concentration, although this variable is subject to measurement error due to the small sample size in some regions. This difference nonetheless suggests that, although the top lender in each region has a disproportionately large market share, the remaining contracts are distributed more uniformly across other banks.

5. ESTIMATION METHOD

In this section we describe the steps we take to estimate the model parameters. We begin by describing the functional form assumptions imposed on consumers and lenders unobserved attributes. Then we derive the likelihood function induced by the model, and we discuss the sources of identification in the final subsection.

5.1. **Distributional assumptions.**

Our baseline model has four sources of randomness beyond observed financial and demographic characteristics: (i) consumer choice-set, (ii) identity of banks with prior experience and origin of the first quote, (iii) common unobserved profit shock ϵ_i, and (iv) idiosyncratic match values u_{ij}.

Consumer choice-sets. As discussed in section 4.2, we assume that consumers shop locally for their new mortgage. In our baseline specification, we assume that all consumers consider lenders located within a 10 KM euclidian distance around the center of their postal code area (i.e. FSA).

We also consider a richer econometric specification in which a fraction μ of consumers only consider dominant banks in their local area, while the remaining "sophisticated" consumers consider the full set of banks in $\mathcal{N}_i$. We identify dominant lenders by computing the cumulative distribution of branches in each local market: dominant lenders are defined as the largest banks in $\mathcal{N}_i$ controlling at least 75% of branches. Recall that the distribution of branches in highly skewed for most consumers, and the largest bank in each local market controls on average 40% of branches. The average number of lenders in the restricted choice-set is equal to 3, compare to 6 in the full set.

Distribution of main financial institutions. The identity of home banks is partially observed when consumers transact with a bank which they have at least one month of experience, and consumers are assumed to have experience with at most one bank. For the consumers that switch institutions, the identity of the bank with prior experience in unknown (i.e. we only know it is not the chosen lender). Moreover, this variable is absent for the contracts insured by Genworth Financial.

We assume that E_{ij} is a multinomial random variable with a probability distribution ψ_{ij}. This distribution is a function of the location of consumers, and income group. We estimate this probability distribution separately using a survey of consumer finances performed by-yearly (Ipsos-Reid) which identifies the main financial institution of consumers. This data-set surveys nearly $12,000$ households per year in all the regions of the country. We group the data into six years, ten regions, and four income categories. Within these sub-samples we estimate the probability of a consumer choosing one of the twelve largest lenders as their main financial institution. We denote this estimated probability by $\psi_j(X_i)$, where X_i identifies consumer i's group. This probability corresponds to the density of positive experience level given the year, income, and location of borrower i.

In addition, it is possible that consumers have no prior experience with lenders in their choice-set. For instance, a bank might not be present in the new residential neighborhood of consumers. As a result, the identity of the first offer (i.e. h_i) is not always equal to the "home" bank, which means that we must integrate out two possibilities when evaluating the likelihood: (i) receiving an initial quote from the home bank (i.e. $E_{ij} = 1$), and (ii) receiving an initial quote from a bank with no prior experience (i.e. $E_{ij} = 0$). In the latter case, we assume that the matching probability is proportional to the branch network share of bank j, denoted by s_{ij}. Formally the probability of the pair (h, E) is:

$$\Pr\left(h, E | X_i, \mathcal{N}_i\right) = \begin{cases} 1(h \in \mathcal{N}_i)\hat{\psi}_h(X_i) & \text{If } E_{ih} = 1, \\ \sum_{k \notin \mathcal{N}_i} \hat{\psi}_k(X_i) s_{ih} & \text{If } \sum_{k \in \mathcal{N}_i} E_{ik} = 0, \\ 0 & \text{Otherwise.} \end{cases} \tag{12}$$

In words, when possible, the initial offer comes from a bank with prior experience of consumer i, otherwise, it is randomly sampled from the set of available options.

In the estimation of the model, we also allow a fraction η of consumers to have zero valuation for their home bank (i.e. $\theta_{ij} = 0$ for all j). For those consumers, the initial matching is random, and is solely determined by the distribution of branch share in $\mathcal{N}_i$.

Unobserved profit shocks. The common unobserved lending cost ϵ_i is normally distributed with mean zero and variance σ_ϵ^2, and the bank-specific idiosyncratic match values u_{ij} are independently distributed according to a type-1 extreme-value (EV) distribution with location and scale parameters $(0, \sigma_u)$. As a result, the surplus V_{ij} is also distributed according to a type-1 extreme-value distribution with location $\xi_{ij} = (\theta_{ij} - C_{ij})/L_i$ and scale σ_u. Let $F(v; \xi_{ij}, \sigma_u)$ and $f(v; \xi_{ij}, \sigma_u)$ denote the CDF and PDF of V_{ij}/Li.

The EV distribution assumption leads to analytical expressions for the distribution functions of the first and second-order statistics, and has often been used to model asymmetric value distributions in auction settings (see for instance Brannan and Froeb (2000)). To simplify the notation, we use the term $\mathcal{N}_i$ to denote the number and identity of lenders present in the choice-set of consumer i (i.e. vector of ξ_{ij}). Also, the notation $\mathcal{N}_i \backslash j$ identifies the choice-set of consumer i excluding option j. The distribution of the highest surplus in consumer i's choice-set is directly obtained from the extreme-value functional form:

$$F_1(v; \mathcal{N}_i) \equiv F(v; \xi_{i,\max}, \sigma_u), \quad \text{where } \xi_{i,\max} = \sigma_u \log \left(\sum_{j \in \mathcal{N}_i} \exp\left(\xi_{ij}/\sigma_u\right) \right). \tag{13}$$

This leads to the familiar multinomial logit form for the probability that bank j offers the highest surplus:

$$\rho_{ij} = \Pr\left(V_{ij} = \max_{k \in \mathcal{N}_i} \{V_{i,k}\} \right) = \frac{\exp\left(\sigma_u \xi_{ij}\right)}{\sum_{k \in \mathcal{N}_i} \exp\left(\sigma_u \xi_{ik}\right)} = \frac{\partial \xi_{i,\max}}{\partial \xi_{ij}}. \tag{14}$$

The second-order statistics of the V's distribution can also be derived analytically from $F_1(\cdot)$:

$$\begin{aligned} F_2(v | V_{ij} = V_{(1)}; \mathcal{N}_i) &= \frac{1}{\rho_{ij}} \Big(F_1(v; \mathcal{N}_i \backslash j) + (\rho_{ij} - 1) F_1(v; \mathcal{N}_i) \Big) \\ F_2(v; \mathcal{N}_i) &= \sum_{j \in \mathcal{N}_i} F_1(v; \mathcal{N}_i \backslash j) + F_1(v; \mathcal{N}_i) \sum_{j \in \mathcal{N}_i} (\rho_{ij} - 1) \\ &= \sum_{j \in \mathcal{N}_i} F_1(v; \mathcal{N}_i \backslash j) + (1 - N_i) F_1(v; \mathcal{N}_i), \end{aligned}$$

where $N_i = |\mathcal{N}_i|$. The densities $f_2(v | V_{ij} = V_{(1)}; \mathcal{N}_i)$ and $f_2(v; \mathcal{N}_i)$ are defined analogously.

5.2. **Likelihood function.** We estimate the model by maximum likelihood. In order to derive the likelihood function, we first consider the likelihood contribution of an individual i, first conditioning on Z_i, ϵ_i, and E_i which groups all the relevant information to calculate θ_{ij} and c_{ij}, the choice-set of consumers, as well as the model parameter vector β and the identity of the bank issuing the first quote h_i. After describing the likelihood contribution conditional on $\mathcal{I}_i = (\mathcal{N}_i, Z_i, \epsilon_i, E_i, h_i)$, we discuss the integration of the model unobservables. We do so for the baseline specification in which consumers have homogenous choice sets (i.e. $\mu = 0$), and all consumers have positive home bank premium (i.e. $\eta = 0$).

We use the following notation. With a slight abuse of notation, we use cap-letters to refer to random variables, and small-case letters to refer to the realizations of consumer i. We will also remove the conditioning $(\mathcal{I}_i, \beta)$ whenever possible, since it is common to all probabilities.

The endogenous outcomes of the model are: the chosen lender and transaction price (b_i, p_i), as well as the selling mechanism $M_i = \{\mathbf{c}, \mathbf{n}\}$ (i.e. **c**ompetition versus **n**egotiation). The observed prices are either generated from consumers accepting the initial quote (i.e. $M_i = \mathbf{n}$), or accepting the competitive offer (i.e. $M_i = \mathbf{c}$). Importantly, only the latter case is feasible if $B_i \neq h_i$, while both cases have positive likelihood if $B_i = h_i$. We derive the likelihood contribution for the loyal case first, and then discuss the case of switchers.

Case 1: Loyal consumers ($B_i = h_i$). The main obstacle in evaluating the likelihood function is that we do not observe the selling mechanism. Since we do not observe M_i, the likelihood contribution of loyal consumers is:

$$L^l(p_i, B_i = h_i) = l(p_i, B_i = h_i, M_i = \mathbf{c}) + l(p_i, B_i = h_i, M_i = \mathbf{n}). \tag{15}$$

Three random variables determine the observed outcomes: (i) surplus generated by the home bank V_{h_i}, (ii) the surplus generated by the second highest-option, and (iii) the search cost κ_i.

Recall that for loyal consumers who choose to gather extra quotes, the price reveals perfectly the value of $V_{(2)} = \theta_{i,b_i} - P_i$. To construct the likelihood we consider first the

joint probability of three discrete outcomes: $P_i < p, B_i = h_i, M_i = \mathbf{c}$.

$$\begin{aligned}
&\Pr(P_i < p, B_i = h_i, M_i = \mathbf{c}) \\
&= \int \Pr(P_i < p, B_i = h_i | M_i = \mathbf{c}, v_h) \Pr(M_i = \mathbf{c}|v_h) f(v_h; \xi_{h_i}, \sigma_u) dv_h \\
&= \int \Pr(\theta_{h_i} - p < \max_{j \neq h_i} V_j < v_h) \Pr(M_i = \mathbf{c}|v_h) f(v_h; \xi_{h_i}, \sigma_u) dv_h \\
&= \int_{\theta_{h_i} - p}^{\infty} \left[F_1(v_h; \mathcal{N}_i \backslash h_i) - F_1(\theta_{h_i} - p; \mathcal{N}_i \backslash h_i) \right] \bar{H} f(v_h; \xi_{h_i}, \sigma_u) dv_h,
\end{aligned}$$

where $F_1(v_h; \mathcal{N}_i \backslash h_i)$ is the CDF of the first-order statistics of surpluses excluding the winning banks h_i.

The likelihood contribution is obtained by differentiating the previous joint probability with respect to p, and evaluating it at the observed transaction price p_i:

$$l(p_i, B_i = h_i, M_i = \mathbf{c}) \quad = \quad \int_{\theta_{h_i} - p_i}^{\infty} f_1(\theta_{h_i} - p_i; \mathcal{N} \backslash h_i) \bar{H} f(v_h; \xi_{h_i}, \sigma_u) dv_h. \tag{16}$$

From equation 8, we know that the initial quote is linearly increasing in C_{i,h_i} if $V_{i,h_i} < V_{(1)}$, and linearly decreasing in $V_{(2)}$ otherwise. Depending on the case, the transaction price of loyal consumers choosing the negotiation mechanism reveals V_{ih} or $V_{(2)}$. Moreover, in the first case the acceptance probability is a function of $V_{(2)} - V_{ih}$, while it is constant when the home bank offers the highest surplus. Using these two results from the model, we can write the conditional likelihood as the sum of two independent events:

$$\begin{aligned}
l(p_i, B_i = h_i, M_i = \mathbf{n}) &= f_1(\theta_{h_i} + \sigma_\kappa - p_i; \mathcal{N}_i \backslash h_i)(1 - \bar{H}) \left[1 - F(\theta_{h_i} + \sigma_\kappa - p_i; \xi_{h_i}, \sigma_u) \right] \\
&+ f(\theta_{h_i} + \sigma_\kappa - p_i; \xi_{h_i}, \sigma_u) \Pr\left(M_i = \mathbf{n} \middle| V_{(1)} > \theta_{h_i} + \sigma_\kappa - p_i \right) \left[1 - F_1(\theta_{h_i} + \sigma_\kappa - p_i; \mathcal{N}_i \backslash h_i) \right].
\end{aligned} \tag{17}$$

The conditional acceptance probability included in the the second term of the previous expression is given by:

$$\begin{aligned}
\Pr\left(M_i = \mathbf{n} \middle| V_{(1 \backslash h_i)} > v_h \right) = \Bigg[&\int_{v_h}^{\infty} \exp\left(-\frac{1}{\sigma_\kappa} (v_{(2)} - v_h + \sigma_\kappa - \bar{\kappa}) \right) f_2(v_{(2)} | \mathcal{N}_i \backslash h_i) dv_{(2)} \\
&+ \bar{H} \left[F_2(v_h | \mathcal{N}_i \backslash h_i) - F_1(v_h | \mathcal{N}_i \backslash h_i) \right] \Bigg] \Big/ \left(1 - F_1(v_h | \mathcal{N}_i \backslash h_i) \right).
\end{aligned}$$

Case 2: Switching consumers ($B_i \neq h_i$). The density of prices for switchers reflects the distribution of the second-highest surplus option since the bank b_i cannot be the outcome of the negotiation mechanism: $p_i = \theta_{i,b_i} - \max\left\{ V_{h_i}, V_{(1)} \right\}$, where $V_{(1)} = \max_{j \neq b_i, h_i} V_j$ and $\theta_{i,b_i} = 0$ (abusing notation slightly). Let $\mathcal{N}_i \backslash (b_i, h_i)$ denote the choice-set excluding the home bank

and the choice. Importantly V_{i,b_i}, V_{i,h_i}, and $V_{(1)}$ are three independent extreme-value random variables. We proceed as before by integrating out the value of V_{h_i}, and then integrating out the value of $V_{(1)}$ (note: the order of integration is different from the first to the second term):

$$\begin{aligned}
&\Pr(P_i < p, B_i = b_i, M_i = \mathbf{c}) \\
&= \int_{-p}^{\infty} \int_{-\infty}^{v_h} \Pr(V_{i,b_i} > v_h) \bar{H} f_1\left(v_{(1)}; \mathcal{N}_i \backslash (b_i, h_i)\right) f(v_h; \xi_{h_i}, \sigma_u)\, dv_{(1)} dv_h \\
&\quad + \int_{-p}^{\infty} \int_{-\infty}^{v_{(1)}} \Pr(V_{i,b_i} > v_{(1)}) H(v_h, v_{(1)}) f(v_h; \xi_{h_i}, \sigma_u) f_1\left(v_{(1)}; \mathcal{N} \backslash (b_i, h_i)\right) dv_h dv_{(1)} \\
&= \int_{-p}^{\infty} \Pr(V_{i,b_i} > v_h) \Pr(V_{(1)} < v_h) \bar{H}^c f(v_h; \xi_{h_i}, \sigma_u) dv_h \\
&\quad + \int_{-p}^{\infty} \Pr(V_{i,b_i} > v_{(1)}) \left[\int_{-\infty}^{v_{(1)}} H(v_h, v_{(1)}) f_1\left(v_{(1)}; \mathcal{N}_i \backslash (b_i, h_i)\right) dv_h \right] f_1(v_{(1)}; \mathcal{N}_i \backslash (b_i, h_i)) dv_{(1)}.
\end{aligned}$$

Differentiating the previous expression with respect to $p = p_i$ gives us the likelihood of switchers:

$$\begin{aligned}
L^s(p_i, B_i = b_i | \mathcal{I}_i, \beta) = (1 - F(-p_i; \xi_{b_i}, \sigma_u)) \Bigg[& f(-p_i; \xi_{h_i}, \sigma_u) F_1(-p_i; \mathcal{N}_i \backslash (h_i, b_i)) \bar{H} \\
& + f_1(-p_i; \mathcal{N}_i \backslash (h_i, b_i)) \int_{-\infty}^{-p_i} \left(1 - \exp\left(-\frac{1}{\sigma_i}(-p_i - v_h + \sigma_i - \bar{\kappa})\right)\right) f(v_h; \xi_{h_i}, \sigma_u) dv_h \Bigg].
\end{aligned} \tag{18}$$

Integration of other unobservables. The likelihood function is evaluated by integrating out three other unobservables: h_i, E_i and ϵ_i. The common lending profit shock ϵ_i is distributed according to a normal distribution with common variance σ_ϵ. We integrate it out using a quadrature approximation. The identity of the home bank, and the choice-set heterogeneity is then integrated out by summing the possible combinations.

The likelihood contribution of a contract i can therefore be written as:

$$L(b_i, p_i | Z_i, \beta) = \int \sum_{h,E} \Pr(h, E | X_i) \begin{pmatrix} 1(b_i = h) L^l(b_i, p_i | \mathcal{I}_i) \\ +1(b_i \neq h_i) L^s(b_i, p_i | \mathcal{I}_i) \end{pmatrix} \phi(\epsilon_i; \sigma_\epsilon) d\epsilon_i. \tag{19}$$

Aggregate likelihood function. The aggregate likelihood function sums over the n observed contracts, and incorporates additional external survey information on search effort. We use the results of the annual survey conducted by CAAMP and presented in Table 1 to match the aggregate probability of gathering more than one quote. More specifically, we

augment the data with the extra aggregate moment that 54% of consumers gather more than one quote. On average around $1,000$ consumers are surveyed each year.[9]

Using the model and the observed new-home buyers characteristics we calculate for each year the unconditional probability of rejecting the initial quote; integrating over V_{ih}, ϵ_i and the identity of the home bank. Let $\bar{H}_t(\beta)$ denotes this function, and $\hat{H}_t$ the survey estimate for year t. Since the number of observations used to calculate $\hat{H}_t$ is large (i.e $N_2 \approx 1,000$), we use the normal approximation to the binomial distribution to construct the likelihood from this second source of data. That is:

$$L^2(\hat{H}_t|\beta) = \phi\left(N_2\hat{H}_t; N_2\bar{H}_t(\beta), N_2\bar{H}_t(\beta)(1-\bar{H}_t(\beta))\right), \tag{20}$$

where $\phi(y;\mu,\sigma)$ is the normal density with mean and variance μ and σ.

The log-likelihood function combines both contract-level and aggregate information:

$$\mathcal{L}(\mathbf{Y}|\mathbf{Z},\beta) = \sum_i \log L(b_i,p_i|Z_i,\beta) + \sum_t \log L^2(\hat{H}_t|\beta), \tag{21}$$

where $(\mathbf{Y},\mathbf{Z})$ denotes the vectors of observed outcomes and covariates.

5.3. **Identification.** The model includes four groups of parameters: (i) consumer observed heterogeneity (i.e. γ), (ii) unobserved cost heterogeneity (i.e. σ_u and σ_ϵ), (iii) search cost ($\bar{\kappa}$ and σ_κ), and (iv) home-bank loyalty (i.e. λ).

Although we estimate the model by maximum-likelihood, it is useful to consider the empirical moments contained in our main data-sets. The contract data include information on market share, and conditional price distributions. For instance, we can measure the reduced-form relationship between average prices and number of lenders in consumers' choice-sets, or other borrower-specific attributes. Similarly, we measure the fraction of switchers, along with the premium that loyal consumers pay above switchers. Finally, we augmented the data with the aggregate fraction of consumers who gather more than one quote.

Intuitively, the cost parameters can be identified from the sample of switching consumers. Under the timing assumption of the model, most switchers are consumers who reject the initial quote, and set-up the Bertrand game. The transaction price therefore reflects the second-order statistic of the value distribution. This conditional price distribution can therefore be used to identify the contribution of observed consumer characteristics.

The residual dispersion can be explained by u or ϵ (i.e. common versus idiosyncratic). To tell the difference between the two, we exploit variation in the size of consumers'

[9]Recall that this statistic from the CAAMP refers only to first-time home buyers, while our data-set includes fraction of previous home owners. Therefore, when we match the model prediction with this aggregate statistic, we use only the new-home owners from our data-set.

choice-sets. Indeed, the number of lenders directly affects the distribution of the second-order statistic through the value of σ_u. The "steepness" of the reduced-form relationship between transaction rates and number of lenders therefore identifies the relative importance of σ_u and σ_ϵ.

The data exhibits three sources of variation in the choice-set of consumers. First, consumers living in urban areas tend to face a richer choice-set than consumers living is small cities. Second, nearly 50% of consumers were directly affected by the merger between Canada Trust and Toronto Dominion in 2000, and effectively lost one lender. The third source of variation comes from difference in the shape of the branch distribution across markets. Indeed, we estimate a specification in which a fraction μ of consumers only consider "dominant" lenders, with a cumulate branch share greater than 75%. In this specification, the model has two parameters to explain the relationship between the structure of local markets and the distribution of prices.

The three remaining parameters are identified from differences in the price distribution across switching and loyal consumers, as well as from the relative fraction of switchers and searchers. Intuitively the task is to tell the difference between two competing interpretations for the observed consumer loyalty: high switching cost (or home-bank premium), and/or high search cost.

Using the model specification, we know that the equilibrium search probability is a function of the ratio $\frac{\sigma_\kappa - \bar{\kappa}}{\sigma_\kappa}$ (see equation 9). The observed aggregate fraction of consumers gathering more than one quote therefore pins down the ratio of the two search cost parameters. The level of the private information component and the home-bank premium are separately identified from the observed price difference between loyal and switching consumers, and the fraction of switchers. Indeed, we observe that 54% of consumers search in the population, and more than 70% of consumers remain loyal. The difference between those two fractions suggests a sizeable home-bank premium. Finally, the model specification imposes strong restrictions on the relationship between loan size, and the probability of searching and switching. The value of shopping is increasing in the loan size, but the home bank premium and the search cost are invariant to L_i. Therefore, the model implies that consumers financing larger loans are more likely to incur the sunk shopping cost.

6. Estimation results

6.1. **Reduced form relationships.** Before introducing the estimates of the structural model we first provide empirical evidence describing the reduced-form relationships discussed in section 5.3. Specifically, we want to quantify the relationship between interest rates

TABLE 6. Margin and switching probability regression results

VARIABLES	(1) Margin	(2) Margin	(3) Switcher	(4) Switcher
Annual income (X 100K)	-0.277^a	-0.282^a	0.087^a	0.088^a
	(0.043)	(0.043)	(0.033)	(0.033)
Loan size (X 100K)	0.084^a	0.088^a	-0.054^a	-0.054^a
	(0.022)	(0.022)	(0.016)	(0.016)
Loan/Income	-0.220^a	-0.222^a	0.051^a	0.051^a
	(0.015)	(0.015)	(0.012)	(0.012)
$LTV = 0.95$	0.141^a	0.140^a	0.027^a	0.027^a
	(0.011)	(0.011)	(0.008)	(0.008)
FICO (mid-point)	-0.763^a	-0.764^a	-0.124^a	-0.125^a
	(0.042)	(0.042)	(0.031)	(0.031)
Switcher	-0.090^a	-0.089^a		
	(0.010)	(0.010)		
Renter	-0.013	-0.013	0.080^a	0.080^a
	(0.008)	(0.008)	(0.006)	(0.006)
Living w/ parents	-0.052^a	-0.052^a	0.040^a	0.039^a
	(0.013)	(0.013)	(0.008)	(0.008)
Relative network	0.025^a	0.033^a	-0.009^a	-0.012^a
	(0.005)	(0.005)	(0.003)	(0.003)
Concentration level (C1)	0.277^a		0.023	
	(0.051)		(0.041)	
Nb. FIs > 7		-0.045^a		0.033^a
		(0.009)		(0.007)
Observations	29,279	29,279	24,015	24,015
R-squared	0.405	0.405	0.286	0.287
Period	99-01	99-01	99-01	99-01

and market structure, switching behavior, and financial factors, and measure the importance of unobserved heterogeneity. Allen, Clark, and Houde (2011a) provide a detailed description of mortgage discounting in Canada. Here we focus on a substantially smaller set of contracts, and the results are similar.

Columns 1 and 2 in Table 6 present the results of linear regressions of transaction margins, measured by subtracting the 5 year bond rate from the transaction interest rate, on financial and demographic characteristics of borrowers, as well as market structure controls. The relationships between the financial characteristics of consumers and retail margins are mostly consistent with an interpretation that low risk and wealthy consumers represent lower lending costs, and therefore tend to pay lower rates on average. The loan sizes and FICO scores of consumers are particularly strong predictors of the observed

transaction interest rates. Similarly, financially constrained consumers pay on average a premium equal to 14 basis points. Notice also that the marginal effect of income is positive and statistically different from zero for most observed contracts, except for richer households with relatively small loans. Therefore, conditional on loan size, richer households tend to pay more.

Although we do not take a stand on the specific channel that explains these relationships in the cost function of banks, it is reasonable that standard risk factors are correlated with transaction rates, even in a setting in which lenders are fully insured. On the one hand, lenders can incur transaction costs in the event of default, therefore lowering the expected revenue from risky borrowers. On the other hand, most of these characteristics are also associated with the expected revenues generated from complementary services offered by banks, including other loans and saving accounts. For instance, the income effect result is consistent with the fact that richer households are more likely to pre-pay their mortgage, which reduces the expected revenue for lenders.

Some of these reduced-form relationships are also captured in the model through the search decision. As we discussed above, consumers financing larger loans are more likely to search (and pay lower rates), since the sunk cost is invariant to the level of monthly payment. Although it is not currently in the model, it is also conceivable that richer households have a higher value of time, and therefore higher search costs on average. These two interpretations are also consistent with the fact that the marginal effect of loan size on the switching probability is positive, while the marginal effect of income is negative for most observed contracts (see columns 3 and 4). While we do not observe searches in the data, switching decisions are correlated with searching decisions in the model.

The coefficient associated with the "switcher" dummy variable implies that loyal consumers pay on average nearly 9 basis points above the rate paid by switchers. Similarly, consumers dealing with large-network institutions pay more on average for their mortgage. These relationships are captured in the model by the search cost and home-bank premium.

Finally, we control for two measures of the structure of local markets surrounding each consumer: a dummy equal to one when the number of lenders is greater than seven, and the branch share of the largest lender (C1). Both measures suggest that more competitive local markets are associated with lower rates. We included these measures, instead of the number of lenders for instance, because the relationship is non-linear. The presence of dominant banks matters more than just the number of firms. This suggest that our specification with heterogeneous choice-sets will fit the data better.

Finally, we point out that the R-squared of the rate regression is around 0.4 indicating that even after controlling for a wide range of financial, demographic, and market characteristics, there is still a significant amount of unexplained variation in margins.

6.2. **Preference parameter estimates.** Table 7 presents the maximum likelihood estimates for the key parameters. To reduce the computational burden, the model is estimated on a 3,000 random sample. We do not report the parameter estimates for the cost-function (i.e. γ), which includes financial characteristics, bank fixed-effects, and year/region fixed-effects. The baseline specification includes 39 parameters. The price coefficient is normalized to one and monthly payments are measured in hundreds of dollars. The scale of the parameters translates into $\$100$ of monthly expenses for the life of the contract (i.e. 5 years).

We report the results of four specifications. The first column corresponds to our baseline specification: the choice-sets are a deterministic function of the location of individuals, and consumers have a homogenous valuation for their home bank (i.e. λ). Columns (2) and (3) relax these assumptions, first introducing a fraction μ of consumers facing a smaller choice-set (i.e. only dominant lenders), and then allowing a fraction $(1-\eta)$ of consumers to have zero home-bank premium. The last column combines both sources of unobserved heterogeneity.

TABLE 7. Maximum likelihood estimation results

	(1)	(2)	(3)	(4)
Common shock: σ_ϵ	0.311	0.312	0.311	0.309
	(0.0015)	(0.0016)	(0.0018)	(0.0018)
Idiosyncratic shock: σ_u	0.0795	0.0819	0.0671	0.0624
	(0.0012)	(0.0017)	(0.0011)	(0.0011)
Idiosyncratic search cost: $\sigma_\kappa - \bar{\kappa}$	0.0708	0.0669	0.145	0.127
	(0.0026)	(0.0043)	(0.0037)	(0.0037)
Baseline search cost: $\bar{\kappa}$	0.0602	0.0574	0.0733	0.0633
	(0.0023)	(0.0048)	(0.0018)	(0.0018)
Home premium: λ	0.138	0.123	0.294	0.278
	(0.0034)	(0.0099)	(0.01)	(0.01)
Fraction with small CS: μ	0	0.373	0	0.284
		(0.0032)		(0.0032)
Fraction with $\lambda > 0$: η	1	1	0.728	0.72
			(0.11)	(0.11)
N	3000	3000	3000	3000
LLF/N	-1.9238	-1.9095	-1.8848	-1.8832

The two parameters entering the search cost distribution suggest that search frictions are economically important, and fairly stable across specifications. The baseline cost is equal to $\$6.02$, and the average is $\$19.1$ per month (i.e. $E(\kappa_i) = \bar{\kappa} + \sigma_\kappa = 0.191$). The search cost estimate, especially the private value component, tends to be larger when we introduce additional sources unobserved heterogeneity. The largest estimate, in specification (3), corresponds to an average search cost of $29 per month. The estimates also imply an important amount of dispersion. In the baseline specification, the median search cost is equal to $15, and the inter-quantile range is $14.38.

According to the model, the marginal consumer accepting the initial quote is indifferent between searching and reducing his monthly payment by $\$\kappa_i$, or accepting p^0. Over a five year period, with a discount factor equal to 0.96, these estimates correspond to an average upfront sunk cost of between $\$1,047$ and $\$1,590$.[10]

Are these number realistic? In Allen, Clark, and Houde (2011a) we document that the average discount that a mortgage broker can derive for a borrower is about 20 basis points, or approximately $16 per month on a $140,000 loan. These savings correspond to the median search cost in specifications (1) and (2), and the 25th-percentile in specifications (3) and (4). The magnitude of our search-cost estimate therefore matches fairly well with an interpretation of the role of mortgage brokers as an intermediary whose objective is to reduce consumer search cost.

In addition, Hall and Woodward (2010) calculate that a U.S. home buyer could save an average of $900 on origination fees by requesting quotes from two brokers rather than one (unlike in Canada, in the U.S. brokers act like financial institutions in that they can originate loans and do not have fiduciary duties). Our estimate of the search cost is also consistent with this estimate.

Depending on the specification, the home bank premium estimates, λ, range between $\$13.80$ and $\$29.40$. Therefore, consumers are willing to pay between $\$13.80$ and $\$29.40$ every month to combine their mortgage with their day-to-day banking service provider. Over five years, this corresponds to an upfront cost of between $759 and $1, 617. Assuming that this utility gain originates from avoiding the cost of switching bank affiliation, our result suggests that switching costs are large, and of similar order of magnitude to the cost of gathering multiple quotes. Adding heterogeneity in the value of the home bank nearly doubles the switching cost estimate. We estimate that 72% of consumers have positive home bias.

[10]The search cost is measured in terms of monthly payment units. Since the contract is written over a 60 month period, the discounted value of the search cost is equal to $\sum_{t=0}^{60} \frac{\kappa_i}{(1+r)^{60}}$. With an annual discount factor of 0.96 the monthly interest rate is 0.3%.

The remaining parameters associated with the firms' profit function suggest that firms are more or less symmetric when it comes to the cost of lending. Most of the unobserved heterogeneity between consumers is common across firms, since the variance of the unobserved match value (i.e. σ_u) is four times smaller than the common shock (i.e. σ_ϵ). Using our lending cost estimates, we can decompose the total variance into four components: (i) idiosyncratic shock (i.e. $L_i \times u_{ij}$), (ii) fixed differences across lenders (i.e. $L_i \times \bar{c}_j$ or fixed-effect), (iii) unobserved consumer heterogeneity (i.e. $L_i \times \epsilon_i$). The standard-deviation in lending cost across consumers, including the contribution of financial characteristics, is equal to $\$390$. Fixed differences across lenders account for a small fraction of this dispersion: $\$8.37$. The standard-deviation of the idiosyncratic match values is equal $\$14.10$, and the standard-deviation of the common component is equal to $\$77.03$. Cost differences across banks, observed and unobserved, therefore account for a small fraction of the total variance in lending cost.

This decomposition has important implications for competition. Abstracting from loyalty issues, the average difference between the first and second highest surplus is close to zero in the average market with six lenders. Moreover, our bank fixed-effect estimates imply very little systematic differences across lenders. As a result, the market for "non-loyal" consumers is very competitive: banks are nearly homogeneous and have similar cost structures, which leads to a Bertrand-type equilibrium.

This is not to say that every consumer is benefiting from this favorable structure. The number of available lenders varies greatly across consumers, and we estimate that about 30% of consumers only consider dominant lenders in their choice-set. For these consumers the average number of lender drops to three, which can increase significantly the profit margin of banks.

6.3. **Market power and efficiency.** Overall, our results suggest that search costs and consumer loyalty are the two main sources of market power in the Canadian mortgage market. Absent these factors, the model would predict a nearly competitive market, especially for consumers in neighborhoods with a larger number potential lenders. Lenders with a large consumer base have substantial control over prices because they receive a larger fraction of "first-visits". This allows them to exploit the search costs consumers face and serve a larger proportion of non-searchers. The second source of market power originates from brand loyalty or switching costs, and implies that, even conditional on facing competition, a home bank is more likely to retain consumers.

In this section, we use the parameter estimates from specification (4) in Table 7 to measure the implied profit margins of lenders, and infer the importance of the search cost frictions in distorting the allocation of contracts. Recall that this specification includes

TABLE 8. Summary statistics on the distribution of profit margins and transaction characteristics

Statistics	Payment $/month	Profit $/month	Markup $(p-c)/p$	Loan $\times 1000	Switch	Highest surplus
Non-Searchers						
Mean	966	32	.041	135	0	.67
Std-dev	378	14	.027	55	0	.47
p_{10}	517	19	.015	69	0	0
p_{90}	1504	51	.08	213	0	1
Searchers						
Mean	975	16	.019	140	.45	1
Std-dev	388	14	.018	56	.5	0
p_{10}	512	2	.0021	72	0	1
p_{90}	1541	33	.044	222	1	1

two sources of consumer heterogeneity not in the baseline specification: (i) a fraction of consumers ($\hat{\mu} = 28.4\%$) face a restricted consideration set, and (ii) a fraction of consumers ($1 - \hat{\eta} = 28\%$) have zero switching costs. This allows us to contrast profits among different groups of consumers facing different levels of competition. This is because, on average, consumers with zero switching costs and a full choice-set benefit the most from competition, while consumers with positive switching costs and restricted choice-sets benefit the least.

Recall that a lender's profits from a transaction are given by the realized transaction price (i.e. monthly payment), and its cost:

$$\pi_{ij} = P_i - L_i \times (Z_{ij}\gamma + \epsilon_i - u_{ij}). \tag{22}$$

We calculate this profit level for transaction (i, j) by simulating the random shocks that determine the outcome: (i) the common shock ϵ_i, (ii) the idiosyncratic match value u_{ij} for each $j \in \mathcal{N}_i$, (iii) the identity of the home bank E_{ij}, and (iv) the search cost value κ_i. The first three shocks determine the transaction surpluses for each lender in $\mathcal{N}_i$, which in turn identify the first and second best options and the value of the competitive price (i.e. outside option W_i). The search-cost realization determines whether or not the consumer pays the competitive price, or the initial quote. We also sample a consumer type from two binomial distributions: (i) full or restricted choice-set, (ii) zero or positive switching cost.

Conditional on a choice-set, the competitive profit level is determined by the switching cost parameter, λ, and the cost difference between first and second lenders. The level of switching cost raises the profit of the home bank, and decreases the profit level of competing lenders. This is because loyal consumers who gather multiple quotes pay a premium

proportional to their switching cost, while "switching" consumers are compensated for not remaining loyal:

$$\pi_i^{\mathbf{c}} = \begin{cases} \lambda + C_{(2)} - C_{i,h_i} & \text{If } V_{i,h_i} = V_{(1)}, \\ -\lambda + C_{i,h_i} - C_{(1)} & \text{If } V_{i,h_i} = V_{(2)}, \\ C_{(2)} - C_{(1)} & \text{Otherwise.} \end{cases} \tag{23}$$

Expression (23) highlights the importance of the cost differences between lenders: the deterministic component (i.e. bank fixed-effects), and the idiosyncratic match values (u_{ij}). The latter is driven by the size of consumers' choice-sets, and the variance of match values (i.e. σ_u).

Non-competitive transactions are generated by consumers accepting the initial quote (see equation 8). If the home bank offers the "best-match", the transaction profit is increasing in the average private-value search cost of consumers (i.e. σ_κ), and in the cost difference relative to the next-best alternative. Otherwise, when consumers are initially matched with a low-ranked home bank, the transaction profit margin is constant and equal to σ_κ.

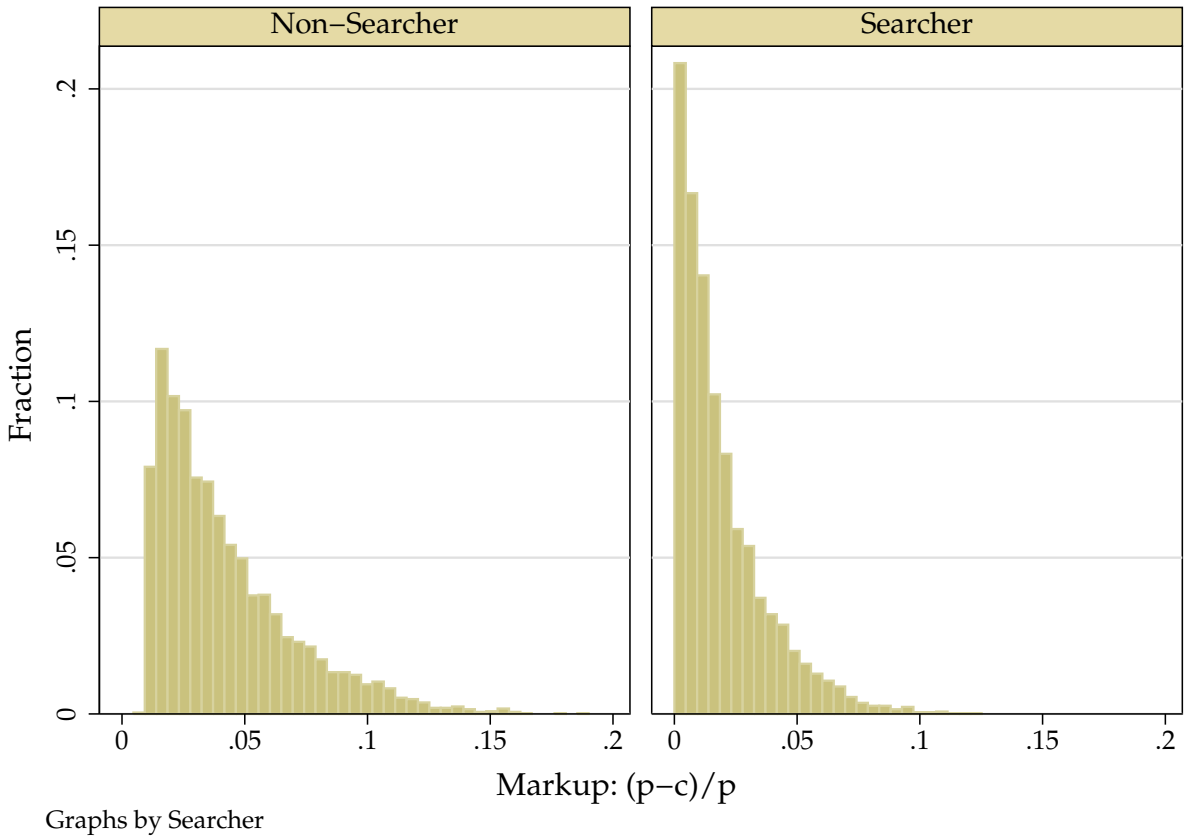
$$\pi_i^{\mathbf{n}} = \begin{cases} \sigma_\kappa + \lambda + C_{(2)} - C_{i,h_i} & \text{If } V_{i,h_i} = V_{(1)}, \\ \sigma_\kappa & \text{Otherwise.} \end{cases} \tag{24}$$

Therefore, unlike in the competitive case, the profits from non-search transactions are bounded below by the search-cost parameter.

Table 8 presents summary statistics on profit margins and other transaction characteristics, for a simulated sample of consumers. We measure market power using monthly profit margins, and the implied markups. Notice that, on average, searchers incur larger monthly payments than non-searchers despite the fact that profit margins on searchers are smaller. These differences are due to the fact that consumers financing larger loans are more likely to search (see column 5). Not surprisingly, we estimate that the market is very competitive, since lenders exhibit little heterogeneity in costs. This is especially true for consumers choosing to gather multiple quotes. The unconditional average markup is equal to 3%, the average among searchers is 2%, and the average among non-searchers is 4%.

Although markups are on average small, there is substantial dispersion, especially among competitive transactions. Figure 5 illustrates the distribution of markups for searchers and non-searchers. Since σ_u is small, we estimate that nearly 25% of searchers get a quote that is very close to the perfectly competitive level. The distribution is highly skewed, with nearly 10% of competitive transactions generating more than 5% markups (up to 14%). The markup distribution among non-competitive transactions is also skewed, but is less concentrated around the lowest profit level.

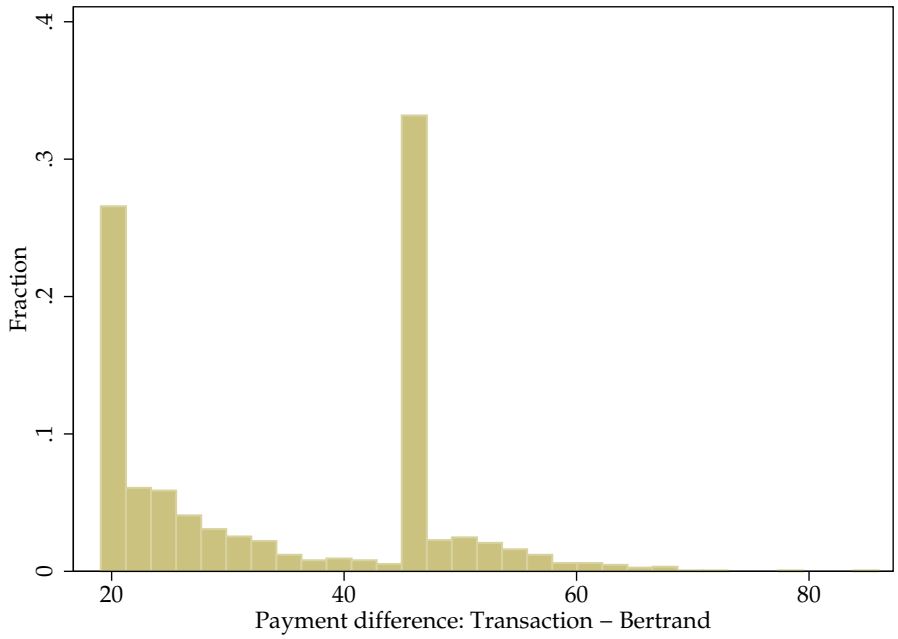
FIGURE 5. Distribution of markups for searchers and non-searchers



Within the non-competitive markup distribution, transactions associated with constant profits are associated with "inefficient" matches, since the transaction surplus is lower than the higher surplus option (i.e. $V_{(1)} > V_{i,h}$). In Table 8 we estimate that 67% of non-searcher transactions are efficient. Therefore, the search-cost friction implies that 33% of non-searcher transactions are miss-allocated, or 17% of all transactions. This proportion is relatively small, compared to the number of lenders in consumers' choice-sets (i.e. six on average). The presence of a large switching cost raises significantly the probability that the home bank is ranked first in the surplus distribution, which greatly reduces the number of inefficient transactions. In this sense, the timing of the model is optimal for consumers with positive home-bank bias.

Figure 6 illustrates the distribution of payment differences between each simulated transaction price, and the counter-factual competitive quote for inefficient matches. The difference is bounded below by the average search cost, and exhibits two modes. The first mode corresponds to consumers who do incur switching costs (i.e. either consumers

FIGURE 6. Distribution of payment differences relative to competitive prices for inefficient matches



with no home-bank in their neighborhood, or more likely consumers with zero home-bank premium), and the second to f consumers who have positive switching costs and therefore must incur a much higher payment differential. The low density beyond these two points is a reflection of the low dispersion in u's across lenders, and the fact that among inefficient matches, the home bank is most likely to be ranked second (i.e. the winning bank would have to compensate consumers for switching).

6.4. **Merger simulation.** In this section, we simulate the effect of a hypothetical merger between two of Canada's largest banks. Our objective is to provide an additional measure of market power, and decompose the effect of losing a lender across different types of consumers.

The effect of the merger is to combine the set of consumers with prior experience with either bank, and reduce the number of lenders by one for consumers with both lenders in their choice-set. Our analysis focuses exclusively on this group of "treated" consumers who lose an option. We also abstract from cost-efficiency effects of the merger. To do so, we alter the model by eliminating systematic differences across lenders (i.e. dropping bank fixed-effects). We also arbitrarily drop the match value of one of the two merging parties, instead of using the highest match value of the two.

TABLE 9. Effect of a hypothetical merger between two banks on monthly payments

VARIABLES	(1)	(2)	(3)	(4)	(5)
Merger X Restricted choice-set		0.331**	3.351***	4.562***	3.271***
		(0.131)	(0.0891)	(0.148)	(0.106)
Constant	5.690***	3.069***	3.881***	5.217***	4.374***
	(0.0492)	(0.0553)	(0.0399)	(0.0641)	(0.0481)
Observations	264,692	25,956	72,186	93,827	43,950
Searcher	both	no	no	yes	yes
Best match	both	no	yes	yes	yes
Switching cost	both	both	both	yes	no

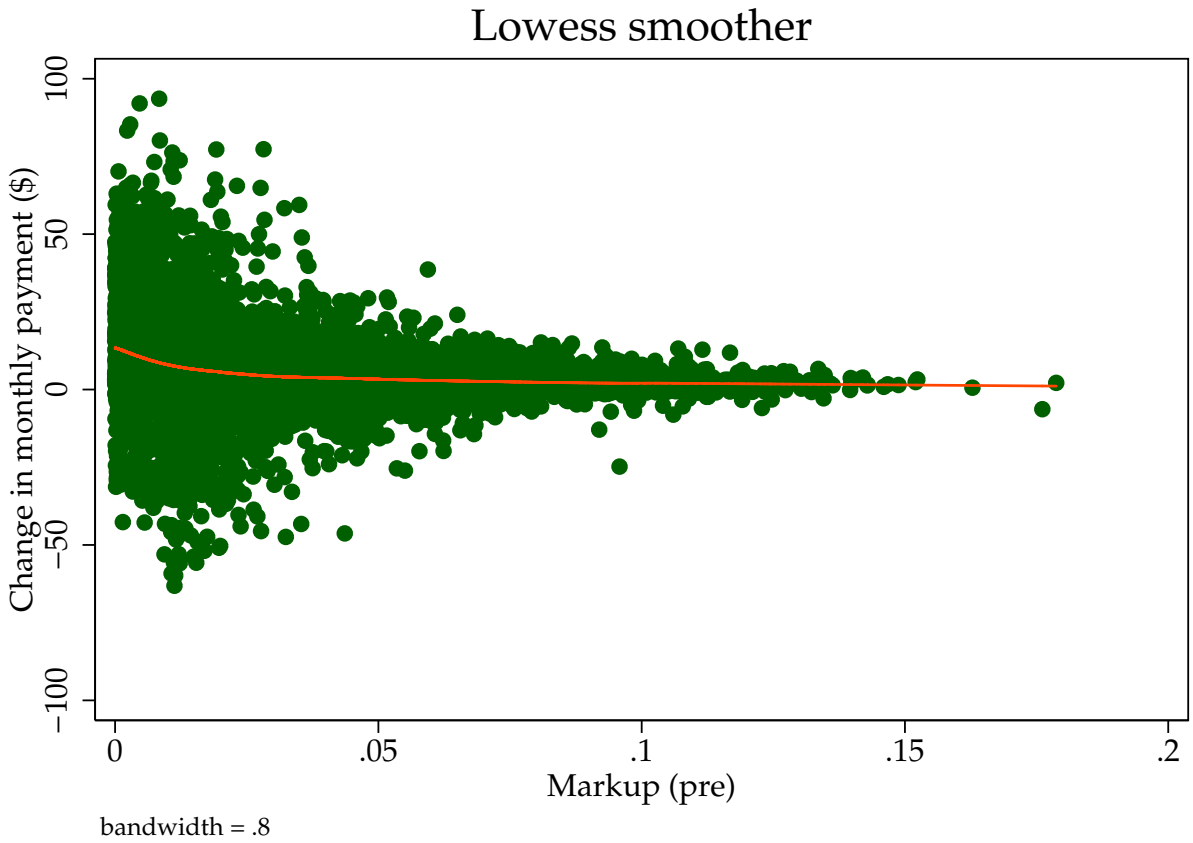
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

In order to compute simulated price changes, we simulate the model under the status-quo and counter-factual market-structures by holding fixed the realization of all random variables. We estimate the average effect of the merger on monthly payments in Table 9. Each observation is a payment change for a simulated consumer. In the first column, we estimate that the simulated merger raised the monthly payment of the average consumer by $ 5.69. While this number is small relative to the average monthly payments, it represents a sizeable increase relative to the average monthly pre-merger profits ($23.36).

In the remaining columns, we decompose the merger effect across competitive and non-competitive transactions, and across groups of consumers (positive or zero switching cost). The top-line coefficients are estimates of the difference between the merger effect on consumers with restricted choice sets (i.e. only dominant banks), and on those with full choice-sets (i.e. constant). Comparing the estimates across columns, the results show that consumers gathering more quotes are more affected than consumers accepting the initial quote. As we discussed above, this result stems from the fact that the non-competitive quote is a function of the outside option only if the home-bank offers the highest surplus (i.e. efficient match). The remaining consumers are affected by the merger only to the extent that it changes the match value of the merged entity. The comparison between columns (2) and (3) illustrates this point by displaying estimates of the effect of the merger on non-searchers for efficient and inefficient matches (i.e. Best match).

Moreover, consumers facing restricted choice-sets are more affected by the merger. The market power increase in the model translates into a change in the distribution of the second-order statistic of the value distribution. Since there is little dispersion, the effect

FIGURE 7. Distribution of merger effects across different levels of pre-merger markups



of reducing the number of options is small for consumers who have more than three or four options. On the other hand, we estimate that the merger effect is nearly doubled for consumers who consider only dominant banks in their neighborhood.

Figure 7 plots the distribution of payment changes across different levels of pre-merger markups in the full sample. The orange line corresponds to a non-parametric regression line. The figure highlights two key features.

First, the effect of the merger on payments is highly dispersed, especially among low-markup consumers. Recall that consumers at the bottom of the markup distribution are paying a competitive price and likely have a large number of lender options. For these transactions, a merger can have two types of effects. First, it generates a market power increase by reducing the value of the second-best option. Second, it can increase or decrease prices by changing the identity of the highest or second-highest option. On the one hand, the transaction price will increase in proportion to the value of the switching cost if the home-bank moves up in the ranking, and becomes the highest valued option. On the

other hand, the transaction price will decrease by the same proportion if the home-bank goes from the highest to a lower value rank. As we discussed above, the home bank is very likely to be ranked first or second in the value distribution, and therefore the merger can easily induce large negative or positive changes in monthly payments.

The second feature of the payment change distribution is the fact consumers at the bottom of the markup distribution are more likely to be adversely affected by the merger, than consumers at the top. Recall that consumers at the top are mostly composed of non-searchers, and loyal consumers who are paying a premium over their next-best alternative. Within the set of non-searchers, consumers matched with a inefficient lender do not experience a market power increase, since their quote does not reflect the value of their outside option. In the latter case, loyal consumers facing a relatively weak outside option are unlikely to be affected by a merger since the merger is unlikely to change the identity of the outside option (in the initial stage), or the highest-match value (in the second stage).

In Allen, Clark, and Houde (2011b) we exploit the variation generated by a merger between a Trust and a charted bank using the same data, and estimate the treatment effect of losing a lender in consumers' choice-sets. The results are largely consistent with the simulation results obtained here. First, we estimate that the merger raised the average transaction interest rates by 5.5 basis points, which correspond to a $4.75 increase in monthly payment for the average loan in our data (assuming a base interest rate of 7%). The model prediction, equal to $5.69, is therefore well within the confidence interval of the estimate we obtain using the natural experiment.

Second, we estimate that the merger effect varies across the quantiles of the conditional price distribution, in a way that is similar to what we find here. In particular, we find that consumers at the top of the price distribution are not affected by the merger, while consumers around and below the median experienced a significant price increase (up to 8 basis points).The merger simulation results predict similar patterns: consumers at the bottom of the markup distribution face the largest price increase, while non-searchers at the top of the markup distribution are mostly unaffected.

7. Conclusion

Although mortgage markets have recently been under great scrutiny, researchers have largely ignored the degree of competition among lenders in their analysis. However, knowing the extent of market power is crucial for evaluating policies designed to regulate mortgage markets. These include antitrust policies regarding the approval of bank mergers, restrictions on the compensation of financial intermediaries, regulations constraining the scope of bank activities, and policies affecting banks' costs of funding such as those

targeting capital or securitization. Moreover, the degree of market power can also affect the transmission mechanism of monetary policy and the effectiveness of macroprudential tools such as mortgage insurance guidelines and/or down-payment requirements.

Our analysis suggests that much of the market power in the mortgage market stems from search and switching costs. Policies designed to increase competition, such as restrictions on merger activity, may therefore not be effective for those consumers unwilling to search or unable to negotiate. Instead, policies designed to increase information about the market, contracts, or the availability of different lenders would be beneficial to consumers. Similarly, policies that encourage consumers to consider lenders other than their main financial institutions would reduce overall market power.

REFERENCES

Adams, W., L. Einav, and J. Levin (2009). Liquidity constraints and imperfect information in subprime lending. *American Economic Review 99*, 49–84.

Allen, J., R. Clark, and J. F. Houde (2011a). Discounting in mortgage markets. Bank of Canada Working Paper No. 2011-3.

Allen, J., R. Clark, and J. F. Houde (2011b). Price competition and concentration in search and negotiation markets: Evidence from mortgage lending.

Allen, J. and D. McVanel (2009). Price movements in the Canadian residential mortgage market. Bank of Canada Working Paper 2009-13.

Armstrong, M. and J. Zhou (2011). Exploding offers and buy-now discounts.

Berry, S., J. Levinsohn, and A. Pakes (2004, Feb). Differentiated products demand systems from a combination of micro and macro data: The new car market. *The Journal of Political Economy 112*(1), 68.

Bester, H. (1993). Bargaining versus price competition in markets with quality uncertainty. *The American Economic Review 83*(1), 278–288.

Brannan, L. and L. M. Froeb (2000, May). Mergers, cartels, set-asides, and bidding preferences in asymmetric oral auctions. *The Review of Economic and Statistics 82*(2), 283–290.

Capps, C., D. Dranove, and M. Satterthwaite (2003). Competition and market power in option demand markets. *RAND Journal of Economics 34*(4), 737–763.

Crawford, G. and A. Yurukoglu (2011). The welfare effects of bundling in multichannel television.

Dafny, L. (2010). Are health insurance markets competitive? *American Economic Review 100*(4), 1399–1431.

De Los Santos, B., A. Hortacsu, and M. Wildenbeest (2011). Testing models of consumer search using data on web browsing and purchasing behavior. *American Economic Review (Forthcoming)*.

Dranove, D., M. Satterthwaite, and A. Sfekas (2008). Boundedly rational bargaining in option demand markets: An empirical application.

Dubin, J. A. and D. L. McFadden (1984, March). An econometric analysis of residential electric appliance holdings and consumption. *Econometrica 52*(2), 345–362.

Grennan, M. (2011). Price discrimination and bargaining: Empirical evidence from medical devices.

Hall, R. E. and S. Woodward (2010). Diagnosing consumer confusion and sub-optimal shopping effort: Theory and mortgage-market evidence. NBER working paper 16007.

Hanemann, M. (1984, May). Discrete/continuous models of consumer demand. *Econometrica 52*(3), 541–562.

Hong, H. and M. Shum (2006). Using price distributions to estimate search costs. *RAND Journal of Economics 37*(2), 257–275.

Hortaçsu, A. and C. Syverson (2004). Product differentiation, search costs, and competition in the mututal fund industry: A case study of S&P 500 index funds. *Quarterly Journal of Economics*, 403–456.

Langer, A. (2011, October). Demographic preferences and price discrimination in new vehicle sales. working paper, University of Michigan.

Postel-Vinay, F. and J.-M. Robin (2002). Equilibrium wage dispersion with worker and employer heterogeneity. *Econometrica 70*(2), 2295–2350.

Sorensen, A. (2000, August). Equilibrium price dispersion in retail markets for prescription drugs. *Journal of Political Economy 108*(4).

Town, R. and G. Vistnes (2001). Competition in networks: An analysis of hospital pricing behavior. *Journal of Health Economics 20*, 719–753.

Wolinsky, A. (1986). True monopolistic competition as a result of imperfect information. *Quarterly Journal of Economics 101*(3), 493–511.

TABLE 10. Definition of Household / Mortgage Characteristics

Name	Description
FI	Type of lender
Source	Identifies how lender generated the loan (branch, online, broker, etc)
Income	Total amount of the borrower(s) salary, wages, and income from other sources
TSD	Ratio of total debt service to income
Duration	Length of the relationship between the borrower and FI
R-status	Borrowers residential status upon insurance application
FSA	Forward sortation area of the mortgaged property
Market value	Selling price or estimated market price if refinancing
Applicant type	Quartile of the borrowers risk of default
Dwelling type	10 options that define the physical structure
Close	Closing date of purchase or date of refinance
Loan amount	Dollar amount of the loan excluding the loan insurance premium
Premium	Loan insurance premium
Purpose	Purpose of the loan (purchase, port, refinance, etc.)
LTV	Loan amount divided by lending value
Price	Interest rate of the mortgage
Term	Represents the term over which the interest rate applies to the loan
Amortization	Represents the period the loan will be paid off
Interest type	Fixed or adjustable rate
$CREDIT$	Summarized application credit score (minimum borrower credit score).

Some variables were only included by one of the mortgage insurers.