The Pricing of Financial Products in Retail Banking: Competition, Geographic Proximity and Credit Limits*

Santiago Carbo-Valverde † Hector Perez-Saiz ‡

[PRELIMINARY DRAFT, DO NO CITE WITHOUT PERMISSION]

August 24, 2016

Abstract

We quantify the effect of market competition and geographic proximity on the provision of bank accounts, credit cards and lines of credit in the retail banking industry. We propose a methodology that models the strategic entry decision of the financial institutions in every market, which allows us to, compared to previous studies, solve for potential endogeneity issues when estimating the effect of competition on the fees, rates and credit limits offered. Using a very detailed household-level database for rural markets in Canada, we find that competition reduces significantly the monthly fees paid for bank accounts and credit cards, and that geographic proximity also plays a role in reducing these fees. Our results show that a standard regression approach that ignores strategic entry decisions typically underestimates the effect of competition on the observed outcomes, and that physical branches and proximity still matter to understand well the competititive landscape in the retail banking industry.

JEL classification: L13, L51, G21, G28.

Keywords: Bank accounts, credit cards, competition, entry barriers, credit unions.

---

*We want to thank for comments and suggestions Victor Aguirregabiria, Wilko Bolt, Evren Damar, Jason Allen, Kari Kemppainen, Mark Manuszak, Francisco Rodriguez and participants at presentations at the Bank of Canada, the DNB 2016 Retail Payments Conference and the Federal Reserve Board of Governors. Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the Bank of Canada.

†Bangor Business School, Bangor University, Bangor, Gwynedd, LL57 2DG, UK. s.carbo-valverde@bangor.ac.uk

‡Bank of Canada, Financial Stability Department. 234 Laurier, Ottawa (ON), Canada. perh@bankofcanada.ca

1 Introduction

The banking industry in many advanced economies has experienced continuous consolidation over the last decades. At the same time, the banking sector is increasingly complex, with financial institutions offering a larger portfolio of different financial products to households such as accounts, credit cards and credit lines. This market concentration has raised the fears of authorities that a lack of a competitive banking industry may affect the observed competitive outcomes of these retail products that are increasingly sophisticated and are usually priced depending on consumer characteristics.

This article studies the effect of competition and geographic proximity on the prices and credit limits of a number of financial services offered by banks and credit unions in Canadian rural markets. We propose a methodology that takes into account the endogeneity of the market presence decision by banks in order to estimate the effect of competition and geographic proximity on the fees and rates paid for bank accounts and credit cards, and on the credit limits of credit cards and lines of credit provided by these financial institutions. These are products that are commonly offered by most retailer financial institutions in Canada and are used by consumers for payment purposes and to obtain fast credit or fulfill liquidity needs.

In our methodology, we jointly estimate a latent profit equation that determines bank presence in geographic markets, and a set of outcome equations for fees, rates and credit limits for various financial products. We take into account possible endogeneities in the market presence decision by considering the optimum equilibrium strategies of financial institutions that are potential entrants in every market. We use a very detailed database that includes the transaction prices and observed characteristics of the financial products acquired by every individual household, as well as their demographic attributes. This allows us to take into account household-level and product-level characteristics that may affect how financial institutions use these characteristics to set the fees, rates and credit limits offered. Controlling for these variables is crucial as the available varieties of bank accounts, credit cards and other financial products have expanded over the years and banks tend price discriminate based on them. For instance, Allen *et al.* (2014) use a detailed transaction-level database to provide evidence of significant price dispersion in the Canadian retail mortgage market.

This methodology and database also permits us to exploit two dimensions of competition analysis that are rarely considered together: market complexity (distance, sources of market power, and market presence strategies) and product complexity (multiple product strategies and characteristics, cross-priced and cross-selling strategies). Our estimates provide a rich set of results with interesting implications.

First, after controlling for financial products characteristics and demographic household attributes, we find that an additional competitor in the market decreases the monthly fees paid for a bank account by -7.2%. In contrast, the effect obtained when using standard regression analysis (without considering market presence strategies) is smaller (-4.9%). A similar result is found when considering credit card fees (-16.2% vs -1.7%). Interestingly, we do not observe a significant effect of competition for lines of credit interest rates. Regarding credit limits, we find that competition tends to increase the credit limits offered for credit cards and lines of credit.

Second, we find that physical distance between the branch of the financial institution that provides the product and the household plays an important role to explain the purchase of financial products. This is consistent with the existing literature that mainly focuses on personal loans and find that proximity facilitates credit but makes it also more expensive (Petersen and Rajan, 2002; Degryse and Ongena, 2005; DeYoung *et al.*, 2008; Agarwal and Hauswald, 2010). For the case of credit cards and lines of credit, we find that proximity facilitates access to credit, and although fees and rates paid are smaller, lending terms (credit limits) are worse.

Our paper follows a large literature in banking that studies the effect of competition on the prices of financial products and other topics related to the nature of the lender-borrower relationship.[1] Our results are in line with previous studies but offer a number of insights on quantitative and qualitative differences when product characteristics and households demographics are jointly considered. Despite the emergence of internet and mobile banking, our results also show that physical branches and proximity still matter to understand well the competititive landscape in the retail banking industry.

Although the relationship lending literature has not paid as much attention to bank product fees as paid to interest rates (Hannan, 2006; Dvořák and Hanousek, 2009; Tennant and Sutherland, 2014), the compression of spreads in an era of low interest rates and financial stagnation has raised the importance of non interest income from bank accounts and credit cards for profitability and survival of financial institutions. Additionally, analysing the impact of fees on market power has been more and more challenging for antitrust authorites as product complexity has increased, which adds value to our analysis.

Recent work has studied the effect of competition in the banking industry by focusing on the effect of mergers (Focarelli and Panetta, 2003; Garmaise and Moskowitz, 2006; Park and Pennacchi, 2009; Erel, 2011). A common difficulty in these studies is the endogeneity of the merger decision to expand into new markets, i.e. banks decide to merge in order to expand in markets that are economically attractive, which affects the post-merger outcomes of competition. In our approach, we are able to take into account the endogeneity of the market presence decision by jointly estimating the equilibrium market presence decision of every bank and the product outcome equations.

While some of the articles that look at the effect of competition on lending rates use data at the household or individual level (e.g. Garmaise and Moskowitz, 2006; Zarutskie, 2006; Erel, 2011), they do not usually incorporate detailed information at household and product level, which can

[1] Degryse *et al.* (2009) provide a nice review of the existing literature, which is too large to be summarized here.

be very relevant for pricing purposes. For instance, account fees are usually waived when the account balance is high enough, and premium credit cards have larger fees, specially if they have large credit limits. Ignoring these characteristics may bias the estimates of competition on the fees and limits offered, as banks in markets that are more competitive may offer better product characteristics or discriminate differently on household observables.

Estimating jointly the equilibrium market presence decision and the outcomes equations provide clear advantages. A simple regression analysis used to estimate these equations would not take into account possible endogeneity effects that may create biases in the parameter estimates. For instance, a market with attractive demographic characteristics may simultaneously cause the entry of new competitors, who may also set high fees for the financial products provided. In order to correct for this endogeneity problem, we propose an structural estimation methodology that allows us to correct for this possible effect by endogenizing the market presence decision and taking into account the possible correlation between the unobserved factors that affect market presence and the factors that affect the competititve outcomes.

Recent papers have tried to estimate the effect of competition on outcome equations such as prices, revenues and others for other industries (see Mazzeo, 2002b; Manuszak and Moul, 2008; Ellickson and Misra, 2012). Mazzeo (2002b) and Manuszak and Moul (2008) use a two-step methodology that is based on assuming a simple structure of the entry game. In our paper, we have a richer structural framework that allows to fully capture the effect of firm heterogeneity on entry and on various competitive outcomes, but requires a more complex estimation methodology that uses simulation as in Berry (1992), Bajari *et al.* (2010), Perez-Saiz (2015) or Perez-Saiz and Xiao (2016). We estimate the equilibrium presence and outcome equations jointly using a simulated maximum likelihood estimator. To calculate the likelihood, we take into account the dependencies of unobserved factors affecting entry and the competitive outcomes, and we estimate the distribution of observed outcomes conditional on the observed equilibrium entry using a non-parametric (kernel density) estimator from Fermanian and Salanie (2004) that estimates these conditional probabilities using simulation methods. Because of multiplicity of equilibria, we need to solve for all the equilibria of the entry game to calculate these conditional distributions, and we assume an equilibrium selection rule that selects the equilibrium with the highest joint profits.

The Canadian banking industry has very interesting characteristics that makes it almost ideal to study this question. First, it has a significant degree of concentration, with small variation in concentration levels over the years, with the "Big Six"[2] having a predominant role. Second, the number of firms that can be considered potential entrants in the markets considered is small. This is in part due to the existence of regulatory entry barriers that affect some of the players. We are able to include 7 financial institutions as potential entrants for each market, which represents more than 90% of the banking presence in Canadian rural markets. Also, the number of players

[2]Royal Bank of Canada (RBC), Bank of Montreal (BMO), Bank of Nova Scotia (BNS), Canadian Imperial Bank of Commerce (CIBC) National Bank of Canada (NBC), and Toronto-Dominion Bank (TD).

and their competitive interaction makes this market tractable and the results easy to interpret. This is an advantage of this industry compared to other countries such as the U.S., which has a significantly much larger number of banks, and it is still subject to dynamic effects due to the constant consolidation, bank failures and expansion of the national banks during the last decades.

This paper is divided into seven sections. Section II goes into more detail regarding the evolution of Canadian banking industry. Section III examines the data we use. Section IV describes the static entry model. Section V describes the outcome equations and the estimation methodology. Section VI discuses the empirical results. Section VII concludes.

2 The Canadian banking industry

Our research is motivated by the sustained oligopolistic nature and geographical dispersion of the Canadian retail banking industry. Indeed, the Canadian banking industry is concentrated, with the "Big Six" banks controlling 98% of total banking system assets in 2008, and over 80% of the assets from all Canadian financial firms combined.

This dominance has been enhanced over the last three decades through deregulation. Traditionally, Canadian banks' activities were strictly regulated, with product portfolio regulations differentiating banks from trust and loan companies, which specialized in mortgage lending. Deregulations since the 1980s gradually weakened and eliminated some of these restrictions[3] and there was a significant subsequent industry consolidation.

Despite this consolidation, a third type of depository institution exists in Canada to provide competition to the banks. They are the credit unions. Credit unions are financial institutions founded on the cooperative principle and owned by their members. They can provide the same type of depository and lending services as banks do, although they have constraints to provide certain types of financial services. In many areas of the country, they are strong competitors to banks in the retail market. In fact, Desjardins Group is the largest financial institution in Quebec by asset size. Canada has one of the largest credit union systems in the world in percentage terms,[4] with 11 million members covering more than 40% of the economically active population.[5]

The evolution of the Canadian retail banking industry is shown in Figure 1. The total number of retail branches in the country stabilized after 2003, reaching a long-term equilibrium state after decades of long decline following deregulation of the financial sector in the 1980s and 1990s. At

[3]The 1980 Bank Act revisions allowed banks to establish subsidiaries in other financial services markets such as mortgage lending (Allen and Engert, 2007) Foreign banks were allowed to establish subsidiaries. The 1987 revisions allowed banks to acquire securities dealers. The 1992 revisions allowed banks to acquire trust companies.

[4]The largest US credit union, Navy Federal Credit Union, only has assets of $58 billion at the end of first quarter of 2014, which is less than a third of Desjardins' total assets. The total asset size of Canadian credit unions is almost C$400 billion.

[5]World Council of Credit Unions, Raw Statistical Data 2006.

a same time, a stable industry struture emerged after the merger between Canada Trust and TD Bank in 1999, characterized by the absence of entry and merger between large players.

The Canadian retail banking is not only concentrated, but also seems to exhibit a significant level of entry barriers. Indeed, no significant new entry in the industry occurred after 2000. This entry barrier still has a regulatory component, since despite dropping foreign ownership restrictions, the 2012 Bank Act still stipulates that no single shareholder or group of shareholders can have de facto control of a large bank, defined to be any bank with an equity of $12 billion or more (the "Big Five" qualify).

Despite the evolution of online banking, there is strong evidence that consumers still prefer to do business with banks that offer physical branch locations near them. (Amel and Starr-McCluer, 2002) The number of branches not only has been stable since the mid 2000s, but it has also slightly increased. Also, the largest online-only bank in Canada, ING Bank of Canada, only had an asset base of $23 billion by the end of 2006, which is in the same range as regional players such as ATB Financial, and much smaller than the "Big Six" or Desjardins, which have a large branch network.

This market concentration and lack of entry naturally raises questions about how competitive the Canadian retail banking industry actually is. Historically, studies in contestability have shown that concentrated markets can still be competitive (Baumol *et al.*, 1982). Indeed, some prior studies of competitiveness in the Canadian banking industry focused on indicators of contestability on a national scale, using bank-wide variables such as total assets, like in Allen and Liu (2007). They found that the industry structure was monopolistic competition. Other studies focused on specific products such as competition in mortgage loans Allen *et al.* (2014). This paper differs from the above because it studies in detail the competition between canadian banks in geographically separated local markets, by looking at the entry decision of each financial institution with respect to each market.

3 Data

3.1 Market presence data

For our model, we use bank branch location data from the 2006/2007 edition of Canadian Financial Services, a comprehensive directory of all Canadian Financial Institutions and their branches. The directory is updated annually and contains the exact address of each branch, including the 6-digit postal code. After the deregulation in Canadian banking in the 1980s and 1990s, the depository institutions can all accept deposits from individuals and businesses, and they no longer have regulatory barriers the prevent them from entering each other's businesses. So we consider all financial institutions to be competing in the same overall market.

We define markets using census subdivisions, which is a general term for municipalities in

Canada. They vary widely in area, population and other observed characteristics. For example, Toronto, with a population of more than 2.6 million people, constitutes one census division, just like Martensville, SK, a small city with less than 8000 inhabitants. Apart from cities and towns, census subdivisions also include rural areas grouped together into counties, indian reserves and other unorganized territory. We obtain market-level data such as population, unemployment and per-capita income from the 2006 census.

Because census subdivisions do not necessarily reflect the boundaries of a market, we manually select small rural isolated markets to include in our model, based on well-defined criteria. In particular, we only include census subdivisions that have between 200 and 50,000 individuals that are separated by at least 15 km. The population lower limit eliminates regions too uninhabited to support bank branches, while the upper limit exists to ensure that we do not include large cities, which are composed of multiple neighbourhoods and have an internal structure that makes harder to get a well-defined market. Large cities are also excluded given that our model does not take into account the number of branches a bank has in a market, only whether it enters into a market at all. In fact, we do not differentiate between a bank that has 1 branch in a market versus another that has 3 in the same market, despite the fact that a bank clearly has to consider different factors when considering a decision to open a first branch in a market and thus establish a presence, versus a decision to add a branch in a market where it is already present. Eliminating large population centers minimizes this confounding factor and most of the financial institutions in our sample enter in a market with a single branch.

We then eliminate markets that are located less than 50km away from any major urban centers. Excluding markets located close to large urban centers will help avoid the confounding factor of commuters. Indeed, if a worker lives in a suburb and commutes downtown for work, he or she might satisfy his banking needs at a branch closer to work than at a branch close to home. 50km can be an hour's drive to work, and according to the Canadian Census the vast majority of people do not commute that far.[6] Then, we limit the area of each market to 300 square kilometers so it can be reasonably thought of as a single market rather than composed of multiple separated markets, which occurs frequently in large rural counties. We also exclude indian reserves from consideration, given their special administrative status and thus avoid potential regulatory confounders. After considering all these exclusions and constraints, we have 448 markets that we use in our estimations. These isolated markets show a relatively clean relationship between population and entry (see Figure 2). A map of one of the markets we have selected, Moose Jaw (Saskatchewan), is shown in Figures 3 and 4.

We construct the dependent variable of bank presence/no presence into a market by looking for its branches within a 10 km radius of the centroid of a given census subdivision. If one branch of the bank is found, we state that the bank is present in the market and set the dependent variable

[6] According to the 2006 Canadian Census, the median commuting distance of workers in Canada is 7.6 kilometers. Across provinces, the median commuting distance ranges between 4.5 km (in Saskatchewan) and 8.7 km (in Ontario).

to 1. Otherwise, we set the dependent variable to 0. We focus on the market presence decision of the Big6 banks and the credit unions (that includes Desjardins and other smaller credit unions). In total we consider 7 potential entrants in every market selected. Therefore, the industry structure is determined by the entry decisions of all 7 potential entrants, and can have 2^7 configurations per market.

We take population, per-capita income, unemployment, number of businesses, and proportion of French-speaking population as market-level exogenous variables for all potential entrants. We also look at two financial institution-level exogenous variables, the asset of a bank/credit union within a province's borders and the amount held outside. We chose asset size partly because it can be a significant variable on the banks' cost function (McAllister and McManus, 1993). It can also correlate with potential consumer preference for banks that have a larger local presence, or the ones that are larger and therefore could be perceived to be safer. The latter can also be attractive due to their larger national and international presence. In addition, we use the minimum distance from the market to the historical bank headquarters, a variable that varies at market-bank level.

National and provincial market descriptive statistics for the 448 markets considered are shown in Table 1. Nationally, the average population is about 5,200. Other statistics reflect the fact that the vast majority of our markets are small rural towns with population in the few thousands, but that we also include some small cities with population up to 50,000. The average unemployment rate of 11.7% in our markets is almost double the national average of little more than 6%. Statistics per province show significant differences between provinces. Newfoundland and Labrador has the lowest per-capita income of them all, while Quebec has the highest population per census subdivision.

Table 2 shows the distribution of branches across provinces. We observe significant variation of geographic presence across provinces and financial institutions. Desjardins and other credit unions are the financial institutions with largest retail presence in rural markets, and they have a specially large presence in Quebec.

In addition, Table 3 shows some statistics for markets where only Big6 enter, and for markets where only CU enter. As we can observe in the statistics shown, Big6 tend to enter in larger markets, more attractive economically (larger income levels, more businesses and less unemployment), but with a lower proportion of French-speaking population.

3.2 Household-level data

Financial product data at household level is obtained from Ipsos Reid's Canadian Financial Monitor (CFM) database for years 2007-2010.[7] This database includes a very complete overview of

[7]We consider household data for three years (2007-2010) after the year considered for the entry game (2006). Some crucial household-level variables used in the model are only available from 2007. In addition, having a large sample size helps for the identification of the model equations. The relatively stable market structure for those years also

all the financial products and services of about 12,000 Canadian households annually. The CFM database covers most financial products offered to Canadian households, such as credit cards, chequing and savings accounts, insurance products, mortgages, personal loans, lines of credit, bonds, stocks, mutual funds, etc. The database includes some of the most relevant characteristics of these products, such as current balance, fees, interest rates, credit limits, payments usage and other product characteristics. The database also includes some detailed demographic characteristics of the households, such as income, location, education, age, marital status, employment etc. We also have a complete overview of the total assets available (real estate, cars, stock, mutual funds, precious metals, etc.) and some variables with some general attitutes about the household finances (difficulty paying its debt, use of a financial advisor, etc). For our empirical model we consider bank accounts, credit cards and lines of credit. Other financial products are not considered either because data on interesting outcome variables is not available (e.g. mortgages), or because they are too specialized products (e.g. mutual funds).

We do a careful selection of the households to be included in our empirical model that is consistent with our entry model. Figure 5 shows how these households are selected. For every census subdivision, we select CFM households that are located in a circle around the centroid of the census subdivision considered. Then, we identify the financial institutions with geographic presence in a circle around every CFM household selected. In order to estimate this geographic location with the highest detail we use 6-digit postal codes location information which we convert to latitude-longitude information for branches, CFM households, and census subdivisions.[8] In total we have 5,893 unique household-year observations in our sample.

Table 4 provides some useful descriptive statistics for the demographic characteristics of the households included in the database. There is a relatively large variation of demographic characteristics. We also observe a large variation in the characteristics of the financial products that they hold. For instance, households pay on average 6,85 dollars per month for every bank account and many accounts do not have any fees. Also, households pay on average 17.69 dollars per year for every credit card. On average, the annual interest rate of a line of credit is 5.12%. Credit limits for lines of credit are significantly larger than for credit cards.

Also, Table 4 includes information about geographic proximity of branches in a radius of 10km around the household. On average, every household considered in our sample has 4.06 financial institutions in a circle of radious 10 km around the household. Also, the probability that a financial institution that provides a financial product to the household is in a circle of radious 10 km around the household is about 0.8 for accounts, credit cards and lines of credit. This result shows that geographic proximity plays an important role to explain the adoption of a financial product of a certain bank by a household.

motivates the use of a three-year period.

[8]A 6-digit postal code covers a relatively very small geographic area. There are more than 900,000 6-digit postal codes in Canada, and in many cases they uniquely identify an area as small as a condominium building or group of houses in a suburb.

In our empirical model that considers product level data, a unit of observation is a financial product acquired by a household-year for each of the 7 financial institutions considered in the entry model. Since there are 5,893 unique household-year observations in our sample, and there are 7 financial institutions, in total we have a sample of size $N = 5,893 \times 7 = 41,251$ observations where we observe financial product characteristics (when the household has the product with the financial institution), or we do not observe them in case the household has not acquired the product with that financial institution. We do not observe in our data base interest rates for bank accounts or credit cards, but we observe monthly fees. We also observe interest rates for lines of credit, and credit limits for credit cards and lines of credit. These are the competitive outcomes that we use in our empirical model.

4 Empirical model

4.1 Geographic presence of bank branches

We propose and estimate a perfect information static game where every potential entrant decides to be present in every market considered (see Bresnahan and Reiss, 1991; Berry, 1992; Cohen and Mazzeo, 2007).[9] We assume that, each potential entrant decides independently whether to enter into every market, observing all the factors that enter into each other's profit function. Therefore, the decision to enter is treated independently in every market. Network effects could exist to some extent, for instance the size of the branch network could provide an advantage to banks (see Ishii, 2005; Dick, 2007). In our empirical model, we consider firm-level controls such as total size of the financial institution that could, at least partially, include this effect.

Market presence of potential entrant i in market m depends on expected profits given by latent variable $\pi_{i,m}$. Let denote $a_{i,m}$ an observed indicator variable which is equal to 1 if potential entrant i enters in market m, and 0 otherwise. There is entry in market m only if it is profitable, therefore

$$a_{i,m} = \begin{cases} 1 & \text{if} \quad \pi_{i,m} \geq 0 \\ 0 & \text{otherwise} \end{cases} . \quad (1)$$

The assumption of profitable entry is clearly reasonable for the case of banks, which are private companies that maximize profits, but it also applies to credit unions, who follow typically a different objective function, but cannot afford to lose money if they want to stay in business for the long-run.

Similar to Berry (1992) and Ciliberto and Tamer (2009a), we assume a reduced-form linear latent profit equation that includes fixed and variable parameters. We do not distinguish between

[9]This literature typically denotes this type of static games as "entry games", see Berry and Reiss (2007) for an extensive survey.

costs and revenues, with both their effects netted out and the net effect on profit included in the equation. If potential entrant i enters in market m, profits from entry are equal to:

$$\pi_{i,m} = \theta_0 + \theta_1 X_m + \theta_2 Z_i + \sum_{j,j\neq i} \theta_{ij} a_{j,m} + \varepsilon^{\pi}_{i,m}, \tag{2}$$

where X_m and Z_i respectively are vectors of market-level and firm-level exogenous variables that affect the firm's profit function and are observed by the firms and the econometrician. We include provincial fixed-effects, and firm fixed-effects respectively in these variables. α_0 is a constant term that represents a fixed entry cost while $\varepsilon^{\pi}_{i,m}$ is a market and firm-specific error term that is iid with variance normalized to one. $\varepsilon^{\pi}_{i,m}$ is observed by all potential entrants, but not by the econometrician.

We also model competitive effects between banks. We estimate separate competitive effects between every pair of firms or group of firms if they are both potential entrants in the market. In the profit equation, θ_{ij} is the competitive effect of bank j's entry on bank i's profit if bank j enters into the market. This is a flexible way to take into account firm-level unobserved effects that affect each bank's competitiveness against other banks. This flexible approach also allows us to differentiate between different competitive models for every financial institution. For example, credit unions could compete more aggressively because they don't always seek to maximize profits. Given the reduced-form latent profit equation, the competitive effects can also encompass other causes of differentiated competition, such as competition on service quality, differentiated funding costs, differentiated portfolios of products, etc.

A Nash equilibrium in pure strategies in a market m is given by the vector a^*_m for all potential entrants in the market, and is obtained by the following set of inequalities:

$$\pi_{i,m}(a^*_{1,m}, ..., a^*_{i,m}, ..., a^*_{E,m}) \geq \pi_{i,m}(a^*_{1,m}, ..., a_{i,m}, ...a^*_{E,m}) \quad \text{for any } i \in E \text{ and any } a_{i,m}\ , \tag{3}$$

where $E = 7$ is the set of potential entrants. We do assume that each competitor affects other potential entrants' profit only through the competitive term, given that our markets are isolated and therefore, there are no network effects. Our model assumes that all banks and credit unions are competing in the same market, for the same customers.

4.2 Accounts

The first financial product that we consider is a bank account. Banks may charge monthly fees to bank accounts, and these fees may depend on aspects such as number of transactions per month, minimum balance, and the type of accounts (checking vs saving). Banks may also price discriminate clients using observed demographic data. For bank accounts (and for the rest of financial products considered), we use as demographic controls the following vector of variables:

$$X_{i,t} = [agehead_{i,t}, assets_{i,t}, hldsize_{i,t}, income_{i,t}, married_{i,t}, ownrent_{i,t}, university_{i,t}, province_{i,t}, year_{i,t}]. \quad (4)$$

We propose the following equation to characterize the monthly fees paid by a household i for an account with bank b in year t:

$$\begin{aligned} Fees^{ac}_{i,b,t} &= \alpha^{ac}_1 + \alpha^{ac}_1 \cdot balance_{i,b,t} + \alpha^{ac}_2 \cdot checking_{i,b,t} + \alpha^{ac}_3 \cdot X_{i,t} \\ &+ \alpha^{ac}_4 \cdot length_{i,b,t} + \alpha^{ac}_5 \cdot Close_{i,b,t} + \alpha^{ac}_6 \cdot Number_{i,t} + B_b + \varepsilon^{ac,P}_{i,b,t}. \end{aligned} \quad (5)$$

Variable *balance* is the balance of the account and *checking* is an indicator equal to 1 if it is a checking account. *length* is a categorical variable that shows the length of the relationship in years between the bank that provides the service and the household. $Close_{i,b}$ is an indicator variable equal to 1 if bank b has geographic presence in a circle of radious10 km around the location of the household. $Number_i$ is the number of financial institutions present in a circle of radious 10 km around the location of the household. This variable represents the competitive effect on the fees that we want to estimate. We also include financial institution-fixed effects (B_b). These fixed effects are added to consider any pricing policy set by banks that is not dependent on household demographic variables, product characteristics or other variables. A more detailed definition of these and other variables can be found in the appendix.

A unit of observation is an account of household-year i with bank b in year t. Although bank accounts are very common financial products, households rarely have accounts with all 7 banks considered in our sample. We consider that the decision of a household i to have a bank account with bank b is given by the following latent equation:

$$D^{*\,ac}_{i,b,t} = \gamma^{ac}_1 + \gamma^{ac}_2 \cdot X_{i,t} + \gamma^{ac}_3 \cdot heavy\_usage_{i,t} + \gamma^{ac}_4 \cdot Close_{i,b,t} + \varepsilon^{ac,D}_{i,b,t}, \quad (6)$$

and we observe that household i has an account with bank b when the latent variable $D^{*\,ac}_{i,b,t} > 0$:

$$D^{ac}_{i,b,t} = \begin{cases} 1 & if \quad D^{*\,ac}_{i,b,t} \geq 0 \\ 0 & if \quad D^{*\,ac}_{i,b,t} < 0 \end{cases} . \quad (7)$$

Demographic characteristics should influence the decision of having an account. For example, individuals that are relatively old may be more reluctant to use many bank accounts. Variable $heavy\_usage_{i,t}$ represents the importance for the household to use different electronic payment channels. This variable is constructed using information on the payment channel usage habits section from the CFM and is equal to the total number of transactions made over a variety of channels (online, mobile, branches, ABM, etc) over one month. We would expect that consumers with high use of various payment channels may be more likely to use bank accounts of various banks.

4.3 Credit cards

The second financial product that we consider is a credit card. Banks obtain revenues from selling credit cards to customers from the annual fees they charge to them, the interests charged on revolving accounts, and also other fees such as fees charged to merchants that accept these cards. We do not observe interest rates on credit cards in the CFM database, but we observe annual fees. As in the case of bank accounts, we assume that the annual fees paid for a credit card with bank b by household i are determined by the following equation:

$$\begin{aligned} Fees^{cc}_{i,b,t} &= \alpha^{cc}_1 + \alpha^{cc}_2 \cdot X_{i,t} + \alpha^{cc}_3 \cdot protection_{i,b,t} + \alpha^{cc}_4 \cdot rewards_{i,b,t} + \alpha^{cc}_5 \cdot limit_{i,b,t} \\ &+ \alpha^{cc}_6 \cdot Close_{i,b,t} + \alpha^{cc}_7 \cdot Number_{i,t} + B_b + \varepsilon^{cc,P}_{i,b,t}. \end{aligned} \tag{8}$$

We use similar variables as in the case of bank accounts. Additionally, we consider *protection* which is an indicator variable for credit protection of the credit card, and also another indicator variable for credit card rewards. Both variables should positively affect the annual fees paid as they provide additional benefits to the card holders. The credit limit of the credit card (*limit*) is also added, as premium cards with very large limits usually have large annual fees.

Contrary to bank accounts, credit cards are risky products for banks because banks extend credit to card holders. A crucial variable used by banks to control risk is the credit limit of the credit card. Typically, consumers can easily get approved applications for credit cards from most banks, but banks use the amount of credit limit granted to control the risk. We propose the following equation to explain the credit limit given by bank b to a household i in year t:

$$\begin{aligned} Limit^{cc}_{i,b,t} = \beta^{cc}_1 + \beta^{cc}_2 \cdot X_{i,t} + \beta^{cc}_3 \cdot heavy\_usage_{i,t} + \beta^{cc}_4 \cdot unemployment_{i,t} + \\ + \beta^{cc}_5 \cdot DifficultyPayDebt_{i,t} + \beta^{cc}_6 \cdot Close_{i,b} + \beta^{cc}_7 \cdot Number_{i,t} + B_b + \varepsilon^{cc,L}_{i,b,t}. \end{aligned} \tag{9}$$

In addition to demographics, we consider two variables that should affect the perceived riskiness of households by the bank, and therefore should determine the total credit provided. These are an indicator for unemployment, and a variable (with values between 0 and 9) that the household uses to rank its perceived difficulties for paying its debt.

As for the case of accounts, we consider that the decision of having a credit card with bank b is given by the following equation:

$$\begin{aligned} D^{*\,cc}_{i,t} &= \gamma^{cc}_1 + \gamma^{cc}_2 \cdot X_i + \gamma^{cc}_3 \cdot heavy\_usage_{i,t} \\ &+ \gamma^{cc}_4 \cdot sophisticated_{i,t} + \gamma^{cc}_5 \cdot advice_{i,t} + \gamma^{cc}_6 \cdot Close_{i,b,t} + \varepsilon^{cc,D}_{i,b,t}. \end{aligned} \tag{10}$$

As for the case of the accounts, we also consider demographic characteristics to influence the decision of having an account. Variable $sophisticated_i$ is an indicator variables for households that have more than 20% of their total assets either in stock exchange assets or mutual funds. Also, variable *advice* is an indicator variable equal to 1 if the household uses regularly a financial advisor.

4.4 Lines of credit

The third and last financial product that we consider is a line of credit. This is defined as a pre-approved loan that the household can draw on at anytime using a cheque, credit card or ABM. We do not observe the annual fees paid to use this line of credit, but we observe the annual interest rate paid. We assume that these rates are determined by the following equation:

$$\begin{aligned} Rates^{loc}_{i,b,t} = \alpha^{loc}_1 + \alpha^{loc}_2 \cdot X_{i,t} + \alpha^{loc}_3 \cdot fixed\_rate_{i,b,t} + \alpha^{loc}_4 \cdot secured_{i,b,t} + \alpha^{loc}_5 \cdot limit_{i,b,t} \\ + \alpha^{loc}_6 \cdot length_{i,b,t} + \alpha^{loc}_7 \cdot Close_{i,b,t} + \alpha^{loc}_8 \cdot Number_{i,t} + B_b + \varepsilon^{loc,P}_{i,b,t}. \end{aligned} \quad (11)$$

We use two indicator variables to characterize the line of credit: An indicator variable for lines of credit with a fixed rate, and an indicator variable for lines of credit that are secured with some form of collateral (such as a house). These variables are also used in the credit limit equation for lines of credit:

$$\begin{aligned} Limit^{loc}_{i,b,t} = &\beta^{loc}_1 + \beta^{loc}_2 \cdot X_{i,t} + \beta^{loc}_3 \cdot fixed\_rate_{i,b,t} + \beta^{loc}_4 \cdot secured_{i,b,t} \\ &+ \beta^{loc}_5 \cdot unemployment_{i,t} + \beta^{loc}_6 \cdot DifficultyPayDebt_{i,t} \\ &+ \beta^{loc}_7 \cdot length_{i,b,t} + \beta^{loc}_8 \cdot Close_{i,b,t} + \beta^{loc}_9 \cdot Number_{i,t} + B_b + \varepsilon^{loc,L}_{i,b,t}. \end{aligned} \quad (12)$$

We also use in this equation similar variables to the ones used in the credit limit equation for credit cards. As for the other financial products, the decision of having a line of credit with a bank b depends on the following latent variable:

$$D^{*\,loc}_{i,b,t} = \gamma^{loc}_1 + \gamma^{loc}_2 \cdot X_{i,t} + \gamma^{loc}_3 \cdot sophisticated_{i,t} + \gamma^{loc}_4 \cdot advice_{i,t} + \gamma^{loc}_5 \cdot Close_{i,b,t} + T_t + \varepsilon^{loc,D}_{i,b,t} \quad (13)$$

4.5 Covariance matrices

The elements of the covariance matrices of the different error terms of the equations shown previously affect the potential endogeneity issues. We could potentially allow for arbitrary correlation among the elements of the covariance matrices, but in practice, we restrict their elements for tractability and focus on several key issues. In particular, we consider that the unobserved variables that affect entry in the markets are correlated with the unobserved characteristics that affect the fees, rates and limits set by these financial institutions for the products they sell to clients in those markets. In addition, we assume that there exists a correlation between the error term that affects the decision to buy a product p from a bank (latent variable $D^{*p}_{i,b,t}$), and the observed fees/rates and limits of the product.

In addition, we assume that error terms between the three products are uncorrelated. This is an important assumption that reduces the set of parameters to estimate, and also greatly simplifies the calculation of the likelihood used in the estimation.

Given these assumptions, the structure of the covariance matrix for bank accounts is the following:

$$\sum^{ac} = \begin{array}{c} \\ \varepsilon^{\pi,1} \\ ... \\ \varepsilon^{\pi,7} \\ \varepsilon^{ac,D} \\ \varepsilon^{ac,F} \end{array} \begin{array}{c} \begin{array}{ccccc} \varepsilon^{\pi,1} & ... & \varepsilon^{\pi,7} & \varepsilon^{ac,D} & \varepsilon^{ac,F} \end{array} \\ \begin{bmatrix} 1 & & & & \\ ... & .. & & & \\ \rho_{\pi} & ... & 1 & & \\ \rho^{ac}_{\pi,D} & ... & \rho^{ac}_{\pi,D} & 1 & \\ \rho^{ac}_{\pi,F} & ... & \rho^{ac}_{\pi,F} & \rho^{ac}_{D,F} & \sigma^{ac}_{F} \end{bmatrix} \end{array} \tag{14}$$

The structure of the covariance matrix for credit cards is the following:

$$\sum^{cc} = \begin{array}{c} \\ \varepsilon^{\pi,1} \\ ... \\ \varepsilon^{\pi,7} \\ \varepsilon^{cc,D} \\ \varepsilon^{cc,F} \\ \varepsilon^{cc,L} \end{array} \begin{array}{c} \begin{array}{cccccc} \varepsilon^{\pi,1} & ... & \varepsilon^{\pi,7} & \varepsilon^{cc,D} & \varepsilon^{cc,F} & \varepsilon^{cc,L} \end{array} \\ \begin{bmatrix} 1 & & & & & \\ ... & ... & & & & \\ \rho_{\pi} & ... & 1 & & & \\ \rho^{cc}_{\pi,D} & ... & \rho^{cc}_{\pi,D} & 1 & & \\ \rho^{cc}_{\pi,F} & ... & \rho^{cc}_{\pi,F} & \rho^{cc}_{D,F} & \sigma^{cc}_{F} & \\ \rho^{cc}_{\pi,L} & ... & \rho^{cc}_{\pi,L} & \rho^{cc}_{D,L} & 0 & \sigma^{cc}_{L} \end{bmatrix} \end{array} \tag{15}$$

And the structure of the covariance matrix for lines of credit is the following:

$$\sum^{loc} = \begin{array}{c} \\ \varepsilon^{\pi,1} \\ ... \\ \varepsilon^{\pi,7} \\ \varepsilon^{loc,D} \\ \varepsilon^{loc,L} \\ \varepsilon^{loc,R} \end{array} \begin{array}{c} \begin{array}{cccccc} \varepsilon^{\pi,1} & ... & \varepsilon^{\pi,7} & \varepsilon^{loc,D} & \varepsilon^{loc,L} & \varepsilon^{loc,R} \end{array} \\ \begin{bmatrix} 1 & & & & & \\ ... & ... & & & & \\ \rho_{\pi} & ... & 1 & & & \\ \rho^{loc}_{\pi,D} & ... & \rho^{loc}_{\pi,D} & 1 & & \\ \rho^{loc}_{\pi,L} & ... & \rho^{loc}_{\pi,L} & \rho^{loc}_{D,L} & \sigma^{loc}_{L} & \\ \rho^{loc}_{\pi,R} & ... & \rho^{loc}_{D,R} & \rho^{loc}_{D,R} & 0 & \sigma^{loc}_{R} \end{bmatrix} \end{array} \tag{16}$$

4.6 Estimation

This section explain with detail the estimation strategy that we use. An observation in our empirical model is a possible financial product that a household-year i has bought to bank b. If the household has the product with bank b, then we will observe fees/rates or credit limits. Therefore, for the three products $p \in \{acc, cc, loc\}$, observed variables $Fees^p_{i,b,t}, Rates^p_{i,b,t}, Limit^p_{i,b,t}$ and $D^p_{i,b,t}$are endogenous.

In addition, entry in the market by bank b is endogenous in the model. Therefore, variables $Close_{i,b,t}$ and $N_{i,t}$ are endogenous and obtained by solving Eq. (3) in the entry model for every market.

The goal of our estimation strategy is to maximize the probability of observing the endogenous variables for a given bank b and a household i. For the case of a given product p, this probability that is used to calculate the likelihood can be expressed as follows:

$$\Pr(Fees^p_{i,b,t}, Limit^p_{i,b,t}, D^p_{i,b,t}, Close_{i,b,t}, N_{i,t}) = \left[\Pr(D^p_{i,b,t} = 0, Close_{i,b,t}, N_{i,t})\right]^{(1-D^p_{i,b,t})} \cdot \left[f(Fees^p_{i,b,t}, Limit^p_{i,b,t}/D^p_{i,b,t} = 1, Close_{i,b,t}, N_{i,t}) \cdot \Pr(D^p_{i,b,t} = 1, Close_{i,b,t}, N_{i,t})\right]^{D^p_{i,b,t}}. \quad (17)$$

In this equation we denote f the probability density function of the continuous variables $Fees^p_{i,b,t}$ and $Limit^p_{i,b,t}$. The rest of the endogenous variables $D^p_{i,b,t}, Close_{i,b,t}$ and $N_{i,t}$ are discrete, so we use probabilities rather than probability density functions in the likelihood.

For the case of accounts, we have a similar equation but we do not consider credit limits. For the case of lines of credit, we consider rates rather than fees. This equation (17) can be rewritten using conditional probabilities as follows:

$$\Pr(Fees^p_{i,b,t}, Limit^p_{i,b,t}, D^p_{i,b,t}, Close_{i,b,t}, N_{i,t}) = \left[\Pr(D^p_{i,b,t} = 0/Close_{i,b,t}, N_{i,t}) \cdot \Pr(Close_{i,b,t}, N_{i,t})\right]^{(1-D^p_{i,b,t})} \cdot \left[f(Fees^p_{i,b,t}, Limit^p_{i,b,t}/D^p_{i,b,t} = 1, Close_{i,b,t}, N_{i,t}) \cdot \Pr(D^p_{i,b,t} = 1/Close_{i,b,t}, N_{i,t}) \cdot \Pr(Close_{i,b,t}, N_{i,t})\right]^{D^p_{i,b,t}}. \quad (18)$$

Following the assumed covariance matrices, we need to estimate the conditional probabilities or conditional density functions such as $\Pr(D^p_{i,b,t} = 1/Close_{i,b,t}, N_{i,t})$ or $f(Fees^p_{i,b,t}, Limit^p_{i,b,t}/D^p_{i,b,t} = 1, Close_{i,b,t}, N_{i,t})$. These conditional probabilities do not have a closed-form solution, and we estimate them using simulation methods. Fermanian and Salanie (2004) show that we can estimate the conditional density function

$$f(Fees^p_{i,b,t}, Limit^p_{i,b,t}/D^p_{i,b,t} = 1, Close_{i,b,t}, N_{i,t}), \quad (19)$$

using a simple non-parametric (kernel density) estimator. These estimators are relatively standard and are usually available in most statistical packages such as Stata or Matlab. In order to estimate (19), we need to generate a large number of simulation draws. In the Appendix we explain with detail the steps necessary to calculate Eq. (19) and other conditional probabilities in Eq. (18) for a given product p.

The calculation of $\Pr(D^p_{i,b,t} = 1/Close_{i,b,t}, N_{i,t})$ and $\Pr(D^p_{i,b,t} = 0/Close_{i,b,t}, N_{i,t})$ follows a similar procedure. Since these are discrete variables, we use a simple frequency estimator to calculate these probabilities instead of a kernel density.

A probability term that requires significant computation is

$$\Pr(Close_{i,b,t}, N_{i,t}), \tag{20}$$

which is calculated by solving the entry equilibrium using Eq. (3). This term represents the predicted probability of observing entry of a bank b in the market considered. Since there is not closed form solution for this predicted probability of entry, we need to use simulations to numerically estimate it, i.e. for each draw, we have to numerically solve for all Nash-equilibria using Eq. (3) and choose the most profitable one. This is the approach from Bajari *et al.* (2010) to estimate static games of perfect information, which has also been recently used by Perez-Saiz (2015).

Note that because error terms across products are uncorrelated, the likelihood is separable for each of the three products considered. Also, since error terms for fees/rates and limits are uncorrelated, the conditional probability density term $f(Fees^p_{i,b,t}, Limit^p_{i,b,t}/D^p_{i,b,t} = 1, Close_{i,b,t}, N_{i,t})$ is separable in fees and limits. This assumption simplifies significantly the estimation procedure.

Using the simulated probability in Eq. (18) for every observation in our sample of size M and product p, we can estimate the full model by maximizing the simulated log likelihood with respect to the parameters of all the equations of our model:

$$\max_{\alpha,\beta,\gamma,\theta} \sum_{p\in\{acc,cc,loc\}} \sum_{i=1}^{M} \sum_{b=1}^{7} \log \widehat{\Pr}(Fees^p_{i,b,t}, Limit^p_{i,b,t}, D^p_{i,b,t}, Close_{i,b,t}, N_{i,t}), \tag{21}$$

where $\widehat{\Pr}(Fees^p_{i,b,t}, Limit^p_{i,b,t}, D^p_{i,b,t}, Close_{i,b,t}, N_{i,t})$ is calculated using simulation techniques as explained with detail in Box 1 in the Appendix. The asymptotic distribution of this maximum likelihood estimator has been studied by Gourieroux and Monfort (1990). The total number of observations that we have in our model are 41,251 household-bank-year product observations.

4.7 Identification

Identification of the parameters in our entry model is achieved in two ways. Firstly, we use exclusion restrictions in the profit function (variables that affect the profit of one financial institution, but not the profit of the rest of the financial institutions). This is a well known approach used in the literature to identify static entry games, as in Berry (1992) or Bajari *et al.* (2010). There are several variables that we use for this purpose. First, we use distance to the main historical headquarters of the banks. This is a variable that is constructed by obtaining the geographical presence of the banks in 1972. We use the location of the headquarter (which we assume it is the largest city in the province) and generate the minimum distance from any market to a headquarter.[10] We expect significant inertia in the subsequent expansion of banks over the decades, therefore this variable should be correlated with the geographic presence in 2006. Also, this variable varies across markets for most banks considered.

Other variables we use are the total asset size of every bank (which does not vary across markets), and regional (provincial) size of every bank (which varies across provinces but not across markets within a province). Total asset size includes all geographic markets, including international markets and any business line (such as investment or wholesale banking). "Big 6" banks are global banks with significant presence in other countries, and have a considerable non-retail activity. Therefore, total asset size can be considered to be, to a large extent, an exogenous variable. Also, regional size includes urban markets, which are markets that are not included in our data base. Urban markets, which are larger and more profitable than the rural markets, have been covered by financial institutions probably much earlier than rural markets. Therefore, regional size can be also considered, to a certain extent, an exogenous variable.

The second strategy we use to identify the entry model is related with the existence of multiple equilibria. Given the assumptions of our model, multiple equilibria are possible, contrary to other papers (Mazzeo, 2002a; Cohen and Mazzeo, 2007) where additional assumptions guarantee a unique equilibrium. This poses a problem for identification of the model. In particular, Eq. (20) would not be defined in presence of multiple equilibria. There have been a number of solution proposed in the literature to solve this problem. We use the recent approach from Bajari *et al.* (2010) that includes an equilibrium selection rule and allows to point-identify the parameters of the model.[11] To identify the equilibrium we use, we compute all possible equilibria using Eq. (3) and then select the most efficient with probability one.[12]

Regarding the rest of the equations of the model for the outcome equations, our model is very similar the well-known Heckman selection mode. In practice, having variables that are present in one equation but not in others is useful. In our case, we consider several variables that are unique

[10] More precisely, using the geographic presence of every bank in 1972 (see Canadian Bankers Association, 1972), we determine the market share of every bank in every province and we use this market share to generate a weighted measure of distance to headquarter. Since there has been a significant number of mergers in Canada since 1972, the geographic presence of a bank is generated using the geographic presence of other banks that will be acquired by the bank between 1972 and 2006.

[11] An interesting recent literature has developed a partial identification approach to solve these issues. See Ciliberto and Tamer (2009b), among others.

[12] Bajari *et al.* (2010) consider a richer framework to identify the probability that a Nash equilibrium with different characteristics (efficient equilibrium, mixed strategies equilibrium, Pareto dominated, etc) is selected.

to every equation. Usage of payments in different payment channels is a variable that affects the demand for a bank account, but not the fees paid. Intuitively, this assumes that banks may be able to discriminate on fees using observed demographics (age, income, etc) of the household, but not using the potential channel usage, which should be a variable that is private information for households, specially for new clients. We use also indicators for financial sophistication and advice as variables that affect the demand for credit cards and lines of credit, but they do not affect the rates, fees or limits on these products. Again, this implies that these are variables that are relatively opaque for banks. We also use risk variables that affect the limits granted for financial products by financial institutions. Credit limits granted by banks should highly depend on riskiness of the clients, therefore unemployment and difficulty to pay the debt should be especially related with these limits, but not with the demand for these products.

5 Empirical results

5.1 Estimates of entry model

We first discuss the estimates of the entry model. Tables 5 and 6 present the baseline model estimation results with competitive effects between financial institutions, along with the standard deviations. Most variables have been divided by their mean in order to facilitate the comparison and the variance of the error term is normalized to 1. The coefficients of the demographic variables in Table 5 mostly follow expectations on the sign but present important differences by type of institution. Credit unions have a negative coefficient on business activity and positive one on the proportion of Francophone population, whereas Big6 have the opposite sign in both variables. The coefficient for unemployment is negative for both types of institutions. Interestingly, the coefficient for population is negative for Big6 (but economically very small), whereas it is positive for CU and large. The coefficient for per capita income is positive for both, but very small for Big6. These results show that Big6 banks are specially focused on markets with large business activity, and low unemployment, whereas credit unions are much more focused on populated markets with few businesses, and credit unions are specially focused in markets with high French-speaking population.[13]

Table 5 also shows the competitive effects between types of financial institutions. Interestingly, only the effect of the entry of Big6 on the profits of CU is negative. This result suggests that after considering potential advantages for every financial institutions in terms of the demographic characteristics of the market where they enter, Big6 banks are tough competitors for credit unions.

[13] There is a relatively large variation of French-speaking population across Canadian provinces. Quebec is a province with a large majority of French-speaking population, but other provinces such as New Brunswick and Ontario have a larger variation in French-speaking population across markets, which provides a nice source of variation to identify this effect (see Table 1).

Table 6 shows results for provincial and individual bank effects. The coefficient for CU is positive and relatively large. The rest of the banks (except CIBC) have a negative or close to zero effect. This shows that CU face lower entry barriers than Big6 in general, so they are able to enter in markets that are less attractive. Table 7 provides estimates of entry costs for all provinces and for all banks. The constant term (intercept) represents the entry cost for RBC in Ontario. In most cases, the entry barriers are negative.

Finally, as we would expect, the coefficient on distance affects negatively profits. This gives an advantage to regional players that expand to areas close to large population centers where they have their headquarters or main centers of activity.

All these results presented suggest that CU and Big6 focus their entry strategies in markets that are relatively different in terms of size, economic attractiveness, and cultural background. There are several alternative explanations to explain these differences. One potential explanation is that credit unions do not need to focus solely on the goal of maximizing profit, meaning that they can afford to lower prices more than commercial banks. They could also face lower entry barriers in local towns, given that some people might be intrinsically attracted to do business with a locally-owned financial institution, similar to how local farmers' markets are able to thrive. Furthermore, they may be more nimble than larger national banks and tailor their product offerings to the specific town they serve. This could be also related to a superior use of soft information by credit unions which improves the lender-borrower relationship (see Allen *et al.*, 2016, for a recent example for Canada). A closer proximity or superior knowledge of their members could be also advantageous for credit unions regarding this relationship, which may affect the quality of service in general.[14]

A fourth alternative is that in Canada, credit unions face provincial prudential regulations that are different than their federal counterparts with the notable exception of Quebec (Pigeon and Kellenberger, 2012). These different regulations are related to different key indicators such as capital requirements[15] or deposit insurance limits.[16] These provincial regulations could help the credit unions to be more aggressive competitors agains the larger banks. Also, the existence of different regulatory authorities in Canada, at provincial and federal level, could affect the implementation of the regulation and supervision see for example Agarwal *et al.* (2014), who show that state regulators tend to be more lenient than federal regulators in the US).

[14]There is some evidence that credit unions are highly ranked consistently in terms of customer satisfaction in Canada (Brizland and Pigeon, 2013).

[15]For example, Ontario credit unions face a leverage cap of 25 while federal banks face a cap of 20. (Ontario regulations 237/09) BC credit unions do not face leverage requirements. (BC Internal Capital Target for Credit Unions, March 2013)

[16]At least for the early 2000s, most Canadian provinces have set higher deposit insurance limits for their credit unions than the federal limits set by the Canadian Deposit Insurance Corporation (CDIC) for federally regulated banks.

5.2 Estimates of the outcome equations

We now discuss the results of the estimation of the adoption equations and the fees/rates and credit limit equations. We show in all cases the estimates of the structural model and the case of simple OLS or probit estimates.

Table 8 shows the estimation of the adoption equations for the three financial products. In all cases we observe a positive effect of the bank being relatively close to the customer to adopt the financial product. The effect is relatively larger for accounts than for credit cards or lines of credit. This is perhaps due to the fact that opening an account implies going often to the branch for cash withdrawls, etc. On the other hand, credit cards are typically used at the point of sale, and do not often involve operations at the bank branch. Also, lines of credit are products that are not very related to payments usage and their proximity coefficient is similar to credit cards.

We also find that certain demographic characteristics are key to explain adoption. Consumers with a heavy payment usage are more likely to adopt accounts and credit cards. Also, sophisticated households are more likely to adopt credit cards or lines of credit. Structural results tend to be a bit higher than probit results.

Table 9 shows estimates for the fees/rates equations. Fees and rates are expressed in logs. The endogeneity of the entry variable significantly affects the OLS estimates when compared to the structural estimates. The estimated effect is relatively large. The presence of an extra competitor in an area close to the geographic location of the household decreases the monthly rate for an account by -7.2%. In contrast, the effect in the OLS estimates is smaller (-4.9%). A similar effect is found when considering credit cards (-16.2% vs -1.7%). On the other hand, we find a positive effect for the case of lines of credit, about 6% for structural estimates and OLS.

Another interesting result is the effect of the geographic proximity of the financial institution. Households with an account provided by a bank that is close (in a circle of radious 10 km) tend to pay lower fees (-37.8%). In contrast OLS gives a positive estimate (13.3%). For the case of lines of credit, we find a negative effect (-26.1%), slightly larger than OLS (-25%). For the case of credit cards, we find a negative effect of proximity on the fees (-41.6%), whereas OLS gives a positive value (15.2%).

The estimate of the variable that provides the effect of the length of relationship with the bank on the fees/rates paid is positive for accounts, but negative for lines of credit (but insignificant). Perhaps one interpretation of this result is that bank accounts create large switching costs on consumers, whereas a line of credit is a more sophisticated product that creates smaller switching costs.

For the rest of the estimates we find intuitive results, and comparable when considering OLS and structural estimates. A checking account tends to be 16.9% more expensive than a saving account and increasing the account balance tends to decrease the annual fees. A household with a rewards credit card pays 195.5% more fees, whereas a credit card with credit protection has fees that are 34.6% higher. Also, the higher the limit, the higher the fees to be paid. For lines of credit, we find that secured lines of credit and fixed-rate lines of credit are respectively 18.2% and 13.9% more expensive (in terms of interest rates paid), and lines of credit with a higher limit pay also a higher interest rate. The size of these coefficients show that product characteristics are very important to explain the fees and rates paid for these products.

Table 10 shows results for the credit limit equation. We find that proximity reduces credit to households, but that competition increases it. Interestingly, structural estimates give a much larger effect of competition for the case of credit cards. We also find that cards with credit protection and rewards typically have larger limits. Also, after controlling for aspects such as total assets and other demographic characteristics, we find that consumers that have difficulties to pay their debt and are employed have larger limits.

For lines of credit, we find that a longer relationship with the bank increases the credit provided, which is intuitive. Households that have difficulties paying their debt have lower credit but the effect is economically much smaller than for credit cards. Consumers with employment have also a positive effect in the credit limit, with a much larger effect than for credit cards.

In summary, consistent with the existing literature, our results show that proximity facilitates credit as adoption rates are higher, but makes it also more expensive (rates and fees tend to be larger), and credit limits tend to be smaller. On the other hand, competition tends to reduce fees and rates, and increase the provision of credit. Controlling for all these aspects, houlseholds with a longer relationship with the financial institution pay higher fees or rates, but have more credit. We also find that product characteristics and household demographics play a very important role to explain the fees/rates and credit limits of these products.

6 Conclusion

In this paper, we study the effect of competition and geographic proximity on the prices and credit limits of a number of financial services offered by banks and credit unions. We use a database and propose a methodology that permits us to exploit two dimensions of competition analysis that are rarely considered together: market complexity (distance, sources of market power, and market presence strategies) and product complexity (multiple product strategies and characteristics, cross-priced and cross-selling strategies).

Our results show that competition tends to reduce fees and rates, and increase the credit supplied, whereas proximity with the financial institution that provides the product increases the likelihood of adopting the financial product, reduces the fees paid, and restricts the credit supplied. Despite the emergence of internet and mobile banking, our results show that physical branches and proximity still matter to understand well the compettitive landscape in the retail banking industry.

Policymakers keep professing their desire to increase competition in the banking industry. Our results show evidence about how different types of financial institutions compete, and how competition can positively affect the outcomes of different retail financial products in an era of financial stagnation and low interest rates.

References

AGARWAL, S. and HAUSWALD, R. (2010). Distance and private information in lending. *Review of Financial Studies*, **23** (7), 2757–2788.

—, LUCCA, D., SERU, A. and TREBBI, F. (2014). Inconsistent regulators: Evidence from banking. *The Quarterly Journal of Economics*, **129** (2), 889–938.

ALLEN, J., CLARK, R. and HOUDE, J.-F. (2014). The effect of mergers in search markets: Evidence from the Canadian mortgage industry. *American Economic Review*, **104** (10), 3365–96.

—, DAMAR, H. E. and MARTINEZ-MIERA, D. (2016). Consumer bankruptcy, bank mergers, and information. *Review of Finance*, **20(4)**, 1289–1320.

— and ENGERT, W. (2007). Efficiency and competition in Canadian banking. *Bank of Canada Review*, **2007** (Summer), 33–45.

— and LIU, Y. (2007). Efficiency and economies of scale of large Canadian banks. *Canadian Journal of Economics*, **40** (1), 225–244.

AMEL, D. F. and STARR-MCCLUER, M. (2002). Market definition in banking: Recent evidence. *Antitrust Bulletin*, **47**, 63.

BAJARI, P., HONG, H. and RYAN, S. (2010). Identification and estimation of a discrete game of complete information. *Econometrica*, **78** (5), 1529–1568.

BAUMOL, W., PANZAR, J. and WILLIG, R. (1982). *Contestable Markets and the Theory of Industrial Structure*. Harcout Brace Javanovich Ltd., New York.

BERRY, S. (1992). Estimation of a model of entry in the airline industry. *Econometrica*, **60** (4), 889–917.

— and REISS, P. (2007). Empirical models of entry and market structure. *Handbook of Industrial Organization*, **3**, 1845–1886.

BRESNAHAN, T. and REISS, P. (1991). Empirical models of discrete games. *Journal of Econometrics*, **48** (1-2), 57–81.

BRIZLAND, S. and PIGEON, M.-A. (2013). *Canada's Credit Unions and the State of the System*. Credit Union Central of Canada.

CANADIAN BANKERS ASSOCIATION (1972). *Bank Directory of Canada*. Toronto: Houston's Standard Publications.

CILIBERTO, F. and TAMER, E. (2009a). Market structure and multiple equilibria in airline markets. *Econometrica*, **77** (6), 1791–1828.

— and — (2009b). Market structure and multiple equilibria in airline markets. *Econometrica*, **77** (6), 1791–1828.

Cohen, A. M. and Mazzeo, M. J. (2007). Market structure and competition among retail depository institutions. *The Review of Economics and Statistics*, **89** (1), 60–74.

Degryse, H., Kim, M. and Ongena, S. (2009). *Microeconometrics of banking: Methods, applications, and results*. Oxford University Press, USA.

— and Ongena, S. (2005). Distance, lending relationships, and competition. *The Journal of Finance*, **60** (1), 231–266.

DeYoung, R., Glennon, D. and Nigro, P. (2008). Borrower–lender distance, credit scoring, and loan performance: Evidence from informational-opaque small business borrowers. *Journal of Financial Intermediation*, **17** (1), 113–143.

Dick, A. A. (2007). Market size, service quality, and competition in banking. *Journal of Money, Credit and Banking*, **39** (1), 49–81.

Dvořák, P. and Hanousek, J. (2009). *Paying for banking services: What determines the fees?* CERGE-EI.

Ellickson, P. B. and Misra, S. (2012). Enriching interactions: Incorporating outcome data into static discrete games. *Quantitative Marketing and Economics*, **10** (1), 1–26.

Erel, I. (2011). The effect of bank mergers on loan prices: Evidence from the united states. *Review of Financial Studies*, **24** (4), 1068–1101.

Fermanian, J.-D. and Salanie, B. (2004). A nonparametric simulated maximum likelihood estimation method. *Econometric Theory*, **20** (04), 701–734.

Focarelli, D. and Panetta, F. (2003). Are mergers beneficial to consumers? evidence from the market for bank deposits. *The American Economic Review*, **93** (4), 1152–1172.

Garmaise, M. J. and Moskowitz, T. J. (2006). Bank mergers and crime: The real and social effects of credit market competition. *The Journal of Finance*, **61** (2), 495–538.

Gourieroux, C. and Monfort, A. (1990). Simulation based inference in models with heterogeneity. *Annales d'Economie et de Statistique*, pp. 69–107.

Hannan, T. H. (2006). Retail deposit fees and multimarket banking. *Journal of Banking & Finance*, **30** (9), 2561–2578.

Ishii, J. (2005). Compatibility, competition, and investment in network industries: Atm networks in the banking industry. *Working paper, Stanford University*.

MANUSZAK, M. D. and MOUL, C. C. (2008). Prices and endogenous market structure in office supply superstores. *The Journal of Industrial Economics*, **56** (1), 94–112.

MAZZEO, M. (2002a). Product choice and oligopoly market structure. *RAND Journal of Economics*, **33** (2), 221–242.

MAZZEO, M. J. (2002b). Competitive outcomes in product-differentiated oligopoly. *Review of Economics and Statistics*, **84** (4), 716–728.

MCALLISTER, P. H. and MCMANUS, D. (1993). Resolving the scale efficiency puzzle in banking. *Journal of Banking & Finance*, **17** (2), 389–405.

PARK, K. and PENNACCHI, G. (2009). Harming depositors and helping borrowers: The disparate impact of bank consolidation. *Review of Financial Studies*, **22** (1), 1–40.

PEREZ-SAIZ, H. (2015). Building new plants or entering by acquisition? Firm heterogeneity and entry barriers in the U.S. cement industry. *RAND Journal of Economics*, **46** (3), 625–649.

— and XIAO, H. (2016). Being local or going global? competition and entry barriers in the canadian banking industry. *Bank of Canada Working Paper*.

PETERSEN, M. A. and RAJAN, R. G. (2002). Does distance still matter? The information revolution in small business lending. *The Journal of Finance*, **57** (6), 2533–2570.

PIGEON, M.-A. and KELLENBERGER, S. (2012). *Implications of Recent Federal Policy Changes for Canada's Credit Unions*. Credit Union Central of Canada.

TENNANT, D. and SUTHERLAND, R. (2014). What types of banks profit most from fees charged? a cross-country examination of bank-specific and country-level determinants. *Journal of Banking & Finance*, **49**, 178–190.

ZARUTSKIE, R. (2006). Evidence on the effects of bank competition on firm borrowing and investment. *Journal of Financial Economics*, **81** (3), 503–537.

Appendices

A Variables definition

- Bank accounts:
  - Fees: Service charge in Canadian dollars paid in the last month by the household.
  - Balance: Current balance of account in Canadian dollars.
  - Checking account: Indicator equal to 1 if the account is a checking account, 0 if it is a savings account or other type of account.
  - Length of relationship with institution: Categorical variable with the following values. =1 if length of relationship is less than one year, =2 if between 1 and 3 years, =3 if between 4 and 6 years, =4 if between 7 and 9 years, =5 if between 10 and 14 years, =6 if between 15 and 19 years, =7 if more than 20 years.
- Credit cards:
  - Fees: Annual fee of the credit card in Canadian dollars.
  - Limit: Total credit card spending limit in Canadian dollars.
  - Protection: Indicator equal to 1 if the card has an insurance that will pay off the debt if the borrower falls ill or passes away while the policy is in force.
  - Rewards: Indicator equal to 1 if the card includes some loyalty program that provides miles, points, etc.
- Lines of credit:
  - Rate: Annual interest rate (in %) charged on the outstanding balances of the line of credit.
  - Fixed rate: Indicator equal to 1 if the interest rate charged on outstanding balances is fixed.
  - Limit: Credit limit on the line of credit in Canadian dollars.
  - Secured: Indicator equal to 1 if the line of credit is secured against an asset (e.g. a house).
  - Length of relationship with institution (also available for accounts): Categorical variable with the following values. =1 if length of relationship is less than one year, =2 if between 1 and 3 years, =3 if between 4 and 6 years, =4 if between 7 and 9 years, =5 if between 10 and 14 years, =6 if between 15 and 19 years, =7 if more than 20 years.
- Bank geographic presence:
  - Number of financial institutions with presence in a circle of radious 10 km around the household.
  - Close: Indicator variable equal to 1 if the financial institution has presence in a circle of radious 10 km around the household.
- Demographic variables for households:
  - Age: Age in years of the head of the house.
  - Assets: Total assets of household in Canadian dollars (in logs). It includes total balance in accounts, value of bonds, mutual funds, stock, real estate, other liquid assets, illiquid assets, etc.

- Difficulty paying debt: Indicator between 0 and 9 where the household reports its perceived difficulty to pay the debt (0=Low difficulty, 9=High difficulty).
- Employed: Indicator equal to 1 if the head of the house is employed.
- Income: Total annual income of the household.
- Married: Indicator equal to 1 if the head of the household is married.
- Own house: Indicator equal to 1 if the house is owned by the houlsehold.
- Size: Number of family members living in the household.
- Sophisticated investor: Indicator equal to 1 if the more than 20% of total assets are either stock exchange assets or mutual funds.
- Unemployment: Indicator equal to 1 if the head of the household is unemployed.
- University degree: Indicator equal to 1 if the head of the household has a university degree.
- Uses financial advisor: Indicator equal to 1 if the household uses regularly a financial advisor .
- Usage payments: Total number of payment transactions per month, including transactions in ATMs, phone payments, online, and mobile payments.

- Demographic variables for markets (census subdivisions)
  - Population in the market.
  - Income: Per capita income in the market.
  - Unemployment: Unemployment rate in the market.
  - Business activity: Number of businesses in the market.
  - Proportion French: Proportion of francophone population in the market.
  - Distance historical HQ: Distance to the closest headquarter of the financial institution as in 1972.

B Computational procedure

We explain with detail in the next box the steps necessary to calculate Eq. (19) and other conditional probabilities in Eq. (18) for a given product p:

BOX 1: ALGORITHM TO SIMULATE CONDITIONAL PROBABILITIES:

1. Select a large number of simulations draws S
2. Generate a set of independent random draws $\Gamma = \{\varepsilon_s^{acc}, \varepsilon_s^{card}, \varepsilon_s^{loc}, \varepsilon_s^{\pi}\}_{s=1}^{s=S}$
3. Transform the set Γ in another set $\widetilde{\Gamma}$ that is distributed following the variance-covariance matrix Σ^p
4. Calculate $D_{i,b,t}^{*\,p}$ for all $\widetilde{\Gamma}$. Find the set of draws $\widetilde{\Gamma}_D$ such that $D_{i,b,t}^{*\,p} > 0$
5. Solve the entry equilibrium for all $\widetilde{\Gamma}$ using Eq. (3). Find the set of draws $\widetilde{\Gamma}_{E,N}$ such that $E_{i,b,t}, N_i$ is an equilibrium
6. Determine the subset of $\widetilde{\Gamma}_\cap = \widetilde{\Gamma}_D \cap \widetilde{\Gamma}_{E,N}$.
7. Calculate $Fees_{i,b,t}^p$ and $Limit_{i,b,t}^p$ for the set of errors $\widetilde{\Gamma}_\cap$
8. Estimate $f(Fees_{i,b,t}^p, Limit_{i,b,t}^p / D_{i,b,t}^p = 1, Close_{i,b,t}, N_i)$ with a kernel density estimator using values from previous stage.

C Results

Table 1: Summary statistics for markets considered in estimation.

Summary statistics of markets considered in estimation, for all Canada, and by province. Source: Stats Canada.

Variable	Canada	BC	MT	NB	NFL	NS	ON	PEI	QC	SK
Population:										
mean	5,281	6,120	3,462	3,430	4,034	3,940	10,406	16,236	4,866	2,178
min	211	341	312	433	211	444	382	299	361	225
p25	884	1,384	718	960	578	1,401	2,354	299	921	449
p50	1,880	3,474	1,428	1,623	1,808	2,986	6,467	16,236	1,868	866
p75	5,640	7,538	2,880	4,638	4,827	5,815	14,822	32,174	3,790	1,384
max	48,821	35,944	41,511	18,129	21,966	11,765	48,821	32,174	42,240	34,138
Per-capita income:										
mean	22,623	24,348	23,059	20,780	18,774	20,240	26,334	19,815	21,336	21,163
min	0	16,715	15,349	15,101	0	17,771	16,164	17,425	13,243	0
p25	19,472	21,596	19,962	18,403	16,343	18,911	22,761	17,425	18,508	18,911
p50	22,411	23,392	22,756	20,781	18,110	20,132	25,884	19,815	21,468	22,616
p75	25,465	25,894	24,956	22,835	21,057	20,413	29,979	22,205	23,256	24,949
max	39,947	38,984	34,900	34,145	36,091	27,443	39,947	22,205	38,234	33,031
Number of business:										
mean	425	570	349	224	205	348	724	1,478	304	312
min	0	0	0	26	5	0	0	0	0	0
p25	92	154	111	71	23	163	159	0	45	120
p50	218	359	218	136	87	214	511	1,478	152	200
p75	512	655	443	278	233	538	1,034	2,955	290	302
max	3,276	3,276	2,906	1,074	1,197	1,214	2,836	2,955	2,460	2,183
Proportion of French population (in %):										
mean	20.7	1.6	5.7	32.3	0.9	5.8	10.4	1.2	87.3	2.1
min	0.0	0.0	0.0	0.0	0.0	0.3	0.2	0.0	0.0	0.0
p25	0.7	0.8	0.9	2.1	0.0	1.0	1.0	0.0	91.2	0.0
p50	1.8	1.4	1.9	3.5	0.0	1.5	2.0	1.2	95.1	0.8
p75	14.5	2.1	3.7	83.1	0.3	2.8	6.0	2.3	97.0	2.0
max	100.0	7.8	59.0	92.1	20.2	64.6	82.4	2.3	100.0	40.2
Unemployment rate (in %):										
mean	11.7	9.6	7.1	14.5	26.8	12.7	8.9	18.8	13.5	10.3
min	1.1	3.1	1.1	3.1	10.3	5.7	2.9	10.3	2.6	1.9
p25	6.2	6.2	4.3	9.2	18.3	8.7	5.6	10.3	8.6	5.4
p50	9.0	8.8	5.9	13.4	26.4	12.3	7.3	18.8	11.3	7.5
p75	14.3	11.3	8.1	17.6	32.4	16.0	9.5	27.3	17.3	11.8
max	59.1	26.9	19.2	43.8	59.1	21.6	36.6	27.3	37.3	38.1

Table 2: Market entry by province and institution.

In the next table we show the number of markets where every financial institution is present by province. Our sample consists of 863 markets.

Province	BMO	BNS	CIBC	CU	NAT	RBC	TD	Markets with entry	Markets with no entry
British Columbia	19	23	40	36	0	34	15	59	10
Saskatchewan	13	16	24	54	0	29	15	72	18
Manitoba	7	9	11	32	0	26	12	40	5
Ontario	38	35	44	48	5	43	39	76	9
Quebec	6	0	16	60	26	8	2	62	14
New Brunswick	7	13	3	20	8	9	5	28	5
PEI	1	1	1	1	1	1	1	1	1
Nova Scotia	5	12	8	8	0	14	5	18	0
Newfoundland	10	12	6	4	0	5	3	15	15
Total	106	121	153	263	40	169	97	371	77

Table 3: Summary statistics for markets considered in estimation by type of entrant.

Summary statistics of markets considered in estimation, for markets where only Big6 or CU enter. A t-test for significant differences in the means of the variables appear in the last column. *, **, and *** indicate statistical significance at the 0.10, 0.05, and 0.01 levels, respectivelySource: Stats Canada.

	Big6 enter only		CU enter only		t-test: Big6 - CU
Variable	mean	se	mean	se	t
Population	5,803	728	1,502	253	5.58***
Total income	23,773	486	21,116	696	3.13***
Proportion French (in %)	4.62	1.01	32.27	5.23	-5.19***
Number of business	475	56	143	17	5.72***
Unemployment rate (in %)	10.53	0.70	14.00	1.15	-2.56**

Table 4: Summary statistics for demographic and product characteristics, and bank proximity.

Summary statistics for demographic characteristics of households, fees/rates, credit limits and product characteristics and proximity of branches. Variables are defined in Appendix. Source: CFM database and bank branches database.

Variable	mean	sd	min	p1	p25	p50	p75	p99	N
Accounts: Fees (dollars)	6.85	15.32	0.00	0.00	0.00	0.00	8.00	81.00	6,135
Accounts: Balance (in 10,000 dollars)	0.49	1.52	0.00	0.00	0.03	0.15	0.45	5.50	6,135
Accounts: Checking account	0.61	0.49	0.00	0.00	0.00	1.00	1.00	1.00	6,135
Accounts: Length relationship	5.81	1.67	1.00	1.00	5.00	7.00	7.00	7.00	6,135
Cards: Fees (dollars)	17.69	39.86	0.00	0.00	0.00	0.00	15.00	160.00	5,581
Cards: Limits (1000s dollars)	9.08	8.56	0.00	0.00	3.00	7.00	13.00	36.00	5,581
Cards: Credit protection	0.25	0.43	0.00	0.00	0.00	0.00	0.00	1.00	5,581
Cards: Rewards	0.26	0.44	0.00	0.00	0.00	0.00	1.00	1.00	5,581
LOC: Rates (in %)	5.12	3.34	0.00	0.00	3.25	5.34	7.00	14.80	2,162
LOC: Limits (1000s dollars)	35.26	47.65	0.00	0.30	12.50	22.50	45.00	212.50	2,162
LOC: Fixed rate	0.33	0.47	0.00	0.00	0.00	0.00	1.00	1.00	2,162
LOC: Secured line of credit	0.55	0.50	0.00	0.00	0.00	1.00	1.00	1.00	2,162
LOC: Length relationship	6.05	1.41	1.00	2.00	5.00	7.00	7.00	7.00	2,162
Demographic variables:									
Assets (in 100,000s)	0.86	2.10	0.00	0.00	0.02	0.10	0.71	10.25	5,893
Age	55.21	15.25	18.00	22.00	45.00	56.00	66.00	86.00	5,893
Size household	2.24	1.13	1.00	1.00	2.00	2.00	3.00	6.00	5,893
Own house	0.82	0.39	0.00	0.00	1.00	1.00	1.00	1.00	5,893
Difficulty paying debt	2.96	2.52	0.00	0.00	1.00	2.00	5.00	9.00	5,893
Employed	0.90	0.29	0.00	0.00	1.00	1.00	1.00	1.00	5,893
Uses financial advisor	0.35	0.48	0.00	0.00	0.00	0.00	1.00	1.00	5,893
University degree	0.83	0.37	0.00	0.00	1.00	1.00	1.00	1.00	5,893
Married	0.76	0.43	0.00	0.00	1.00	1.00	1.00	1.00	5,893
Income (in 100,000)	0.62	0.50	0.08	0.08	0.28	0.50	0.80	2.50	5,893
Sophisticated investor	0.38	0.49	0.00	0.00	0.00	0.00	1.00	1.00	5,893
Proximity of branches:									
Number FIs in circle radious 10km	4.06	1.93	0.00	0.00	2.00	5.00	6.00	7.00	5,893
FI for ACC is in circle radious 10km	0.84	0.37	0.00	0.00	1.00	1.00	1.00	1.00	6,135
FI for CARD is in circle radious 10km	0.80	0.40	0.00	0.00	1.00	1.00	1.00	1.00	5,581
FI for LOC is in circle radious 10km	0.84	0.37	0.00	0.00	1.00	1.00	1.00	1.00	2,162

Table 5: Estimates of entry model (I).

We show estimates of the elements of the profit function (Eq. 2) in the the entry model. Demographic variables have been normalized by their mean. "Competitive effect of X on Y" is the effect on the profit of financial institution Y if financial institution X enters in the market. Standard errors in parentheses obtained using bootstrap.

Variable	Entry model
Panel A: Competitive effects:	
Competitive effect of BIG6 on BIG6	0.04499
	(0.02179)
Competitive effect of CU on BIG6	0.24508
	(0.03198)
Competitive effect of BIG6 on CU	-0.06866
	(0.02342)
Panel B: Demographic variables:	
Intercept	-0.96986
	(0.08476)
Population BIG6	-0.00096
	(0.02264)
Population CU	0.20220
	(0.02113)
Income BIG6	0.00680
	(0.02097)
Income CU	0.17917
	(0.02430)
Unemployment BIG6	-0.26600
	(0.03510)
Unemployment CU	-0.29688
	(0.02927)
Business activity BIG6	0.19471
	(0.02312)
Business activity CU	-0.11042
	(0.01934)
Proportion French BIG6	-0.16209
	(0.02738)
Proportion French CU	0.19254
	(0.02401)

Table 6: Estimates of entry model (II).

We show estimates of the elements of the profit function (Eq. 2) in the the entry model. Size and distance to headquarter are normalized by their mean. Standard errors in parentheses obtained using bootstrap.

Variable	Entry model
Panel C: Provincial effects:	
British Columbia	0.18568
	(0.02193)
Manitoba	0.26834
	(0.02705)
New Brunswick	0.46453
	(0.04539)
Newfoundland and Labrador	0.45780
	(0.04642)
Nova Scotia	0.51770
	(0.05128)
Prince Edward Island	0.45050
	(0.04790)
Quebec	0.16603
	(0.02357)
Saskatchewan	0.14108
	(0.01756)
Panel B: Firm-level effects:	
National Size	0.44322
	(0.04583)
Regional Size	0.05438
	(0.02134)
BMO	0.14693
	(0.02125)
BNS	-0.06172
	(0.01642)
CIBC	0.28283
	(0.03228)
CU	0.52861
	(0.05185)
NBC	-0.99947
	(0.10124)
distance to historical HQ	-0.26752
	(0.03353)
distance to historical HQ (square)	-0.01513
	(0.01987)

Table 7: Entry barriers by financial institution and province

We estimate the level of entry barriers by financial institution and province. They are calculated by addying the constant term, the bank fixed effects, and the provincial fixed effects from Table 6.

Province	BMO	BNS	CIB	CU	NAT	RBC	TD
British Columbia	-0.6	-0.8	-0.5	-0.3	-1.8	-0.8	-0.8
Manitoba	-0.6	-0.8	-0.4	-0.2	-1.7	-0.7	-0.7
New Brunswick	-0.4	-0.6	-0.2	0.0	-1.5	-0.5	-0.5
NFL-Labrador	-0.4	-0.6	-0.2	0.0	-1.5	-0.5	-0.5
Nova Scotia	-0.3	-0.5	-0.2	0.1	-1.5	-0.5	-0.5
Ontario	-0.8	-1.0	-0.7	-0.4	-2.0	-1.0	-1.0
PEI	-0.4	-0.6	-0.2	0.0	-1.5	-0.5	-0.5
Quebec	-0.7	-0.9	-0.5	-0.3	-1.8	-0.8	-0.8
Saskatchewan	-0.7	-0.9	-0.5	-0.3	-1.8	-0.8	-0.8

Table 8: Estimates for adoption of financial products.

We show estimates of the adoption equation for the three financial products considered. We also show probit estimates. Household variables, and provincial and year fixed effects used in all cases. Standard errors in parentheses obtained using bootstrap (for structural model).

	Accounts		Credit cards		Lines of credit	
Variable	Structural	Probit	Structural	Probit	Structural	Probit
Financial advisor			0.05698	0.03409	0.07558	0.04260
			(0.00751)	(0.01781)	(0.01111)	(0.02352)
Heavy usage	0.08473	0.04575	0.07960	0.02176		
	(0.01438)	(0.00414)	(0.01884)	(0.00425)		
Sophisticated			0.04943	0.02608	0.10536	0.07235
			(0.00826)	(0.01929)	(0.01455)	(0.02611)
Close provider	0.82572	0.80525	0.66973	0.63894	0.68709	0.64906
	(0.07048)	(0.01831)	(0.07368)	(0.01795)	(0.06635)	(0.02593)
Household variables	YES	YES	YES	YES	YES	YES
Provincial fixed effects	YES	YES	YES	YES	YES	YES
Year fixed effects	YES	YES	YES	YES	YES	YES
Number of household-year-bank product	41251	41251	41251	41251	41251	41251

Table 9: Estimates for fees/rates equations.

We show estimates of the fees/rates equations for the three financial products considered. We compare structural estimates, and estimates from OLS. Account/card fees and lines of credit rates are expressed in logs. Account balance expressed in 10,000 dollars. Length relationship is a categorical variable with values 1-7 (see Appendix for definitions). LOC annual interest rates are expressed in logs. Credit limit expressed in logs. Household variables, bank, provincial and year fixed effects used in all cases. Standard errors in parentheses obtained using bootstrap (for structural model).

	Accounts		**Credit cards**		**Lines of credit**	
Variable	Structural	OLS	Structural	OLS	Structural	OLS
Account balance	-0.03601	-0.03610				
	(0.00387)	(0.02468)				
Checking account	0.16995	0.16979				
	(0.01771)	(0.05353)				
Card protection			0.34655	0.34683		
			(0.03744)	(0.08595		
Rewards			1.95571	1.95581		
			(0.19593)	(0.10687		
Limit (in logs)			0.19425	0.19708	0.16601	0.16556
			(0.01862)	(0.01700)	(0.01644)	(0.05803)
Fixed rate					0.13933	0.13914
					(0.01413)	(0.11107)
Secured					0.18260	0.18224
					(0.01772)	(0.11343)
Length relationship	0.05990	0.05966			0.00198	-0.00240
	(0.00745)	(0.01751)			(0.00288)	(0.03849)
Close provider	-0.37887	0.13316	-0.41602	0.15251	-0.26144	-0.25080
	(0.03279)	(0.08256)	(0.04182)	(0.10951)	(0.02503)	(0.16550)
Number of competitors	-0.07236	-0.04943	-0.16207	-0.01766	0.05554	0.06022
	(0.00852)	(0.01620)	(0.01608)	(0.02398)	(0.00703)	(0.03313)
Household variables	YES	YES	YES	YES	YES	YES
Bank fixed effects	YES	YES	YES	YES	YES	YES
Provincial fixed effects	YES	YES	YES	YES	YES	YES
Year fixed effects	YES	YES	YES	YES	YES	YES
Number of household-year-bank product	6135	6135	5581	5581	2162	2162

Table 10: Estimates for limit equations.

We show estimates of the credit limit equations for credit cards and lines of credit. Credit limits are expressed in logs. We compare structural estimates, and estimates from OLS. Length relationship is a categorical variable with values 1-7 (see Appendix for definitions). Household variables, bank, provincial and year fixed effects used in all cases. Standard errors in parentheses obtained using bootstrap (for structural model).

	Credit cards		**Lines of credit**	
Variable	Structural	OLS	Structural	OLS
Card protection	0.24063	0.23801		
	(0.02223)	(0.04907)		
Rewards	0.36677	0.36375		
	(0.03794)	(0.06392)		
Fixed rate			-0.42414	-0.42417
			(0.03916)	(0.04928)
Secured			0.61390	0.61399
			(0.06319)	(0.04350)
Difficulty debt	0.03802	0.01083	-0.00578	-0.00509
	(0.00614)	(0.00981)	(0.00157)	(0.00901)
Employed	0.02202	0.01182	0.23269	0.23269
	(0.00325)	(0.09969)	(0.02342)	(0.13206)
Length relationship			0.03370	0.03491
			(0.00382)	(0.01873)
Close provider	-0.11024	-0.11930	-0.19480	-0.19478
	(0.01046)	(0.07230)	(0.01928)	(0.07288)
Number of competitors	0.04290	0.01088	0.01114	0.01325
	(0.01062)	(0.01605)	(0.00246)	(0.01399)
Household variables	YES	YES	YES	YES
Bank fixed effects	YES	YES	YES	YES
Provincial fixed effects	YES	YES	YES	YES
Year fixed effects	YES	YES	YES	YES
Number of household-year-bank product	5581	5581	2162	2162

Table 11: Estimates of elements of the covariance matrix.

We show estimates of the elements of the variance-covariance matrix used in our econometric model. Standard errors have been generated using bootstrap.

Variable	Accounts	Credit cards	Lines of credit
σ_F or σ_R	2.00197	2.56953	2.38778
	(0.20246)	(0.25707)	(0.25082)
σ_L		1.78253	-0.08745
		(0.17706)	(0.00882)
$\rho_{\pi,D}$	-0.17674	-0.78654	0.06195
	(0.01916)	(0.08470)	(0.01242)
$\rho_{\pi,F}$ or $\rho_{\pi,R}$	0.00043	0.57543	-1.25431
	(0.01260)	(0.05366)	(0.12708)
$\rho_{\pi,L}$		-0.68954	0.37487
		(0.07702)	(0.03904)
$\rho_{D,F}$ or $\rho_{D,R}$	0.46186	0.77267	-0.00097
	(0.04822)	(0.07717)	(0.00441)
$\rho_{D,L}$		-1.20804	-0.08745
		(0.13194)	(0.00882)

Figure 1: Evolution of total number of bank branches

The evolution of total number of bank branches in Canada shows a steady decrease in the 1990s and early 2000s, with a stabilization after the mid 2000s.

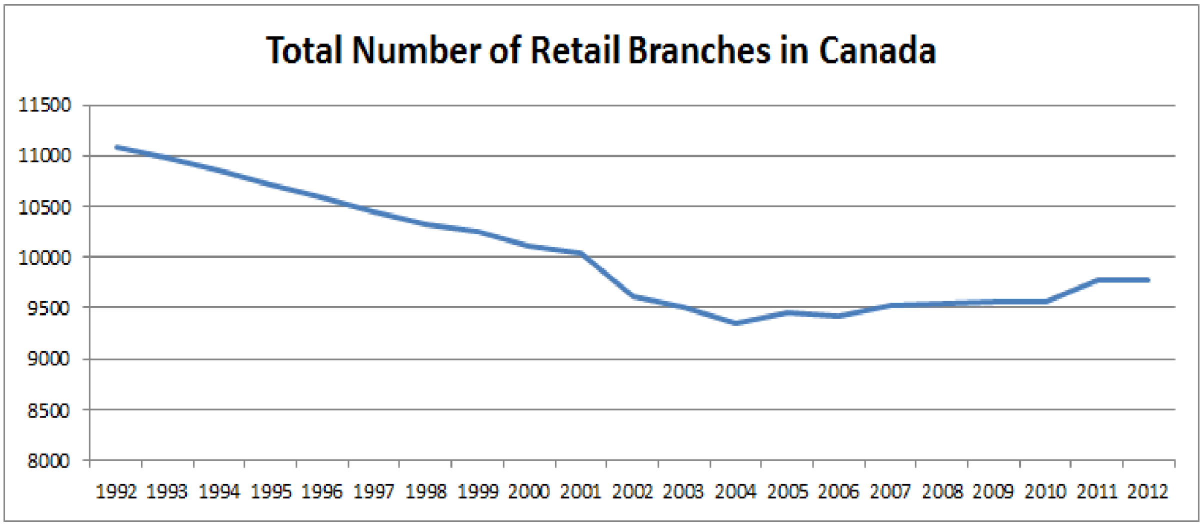

Figure 2: Entry and population

This figure of number of entrants vs. population shows a clear positive correlation between the two. We also see that most markets have 6 entrants or less.

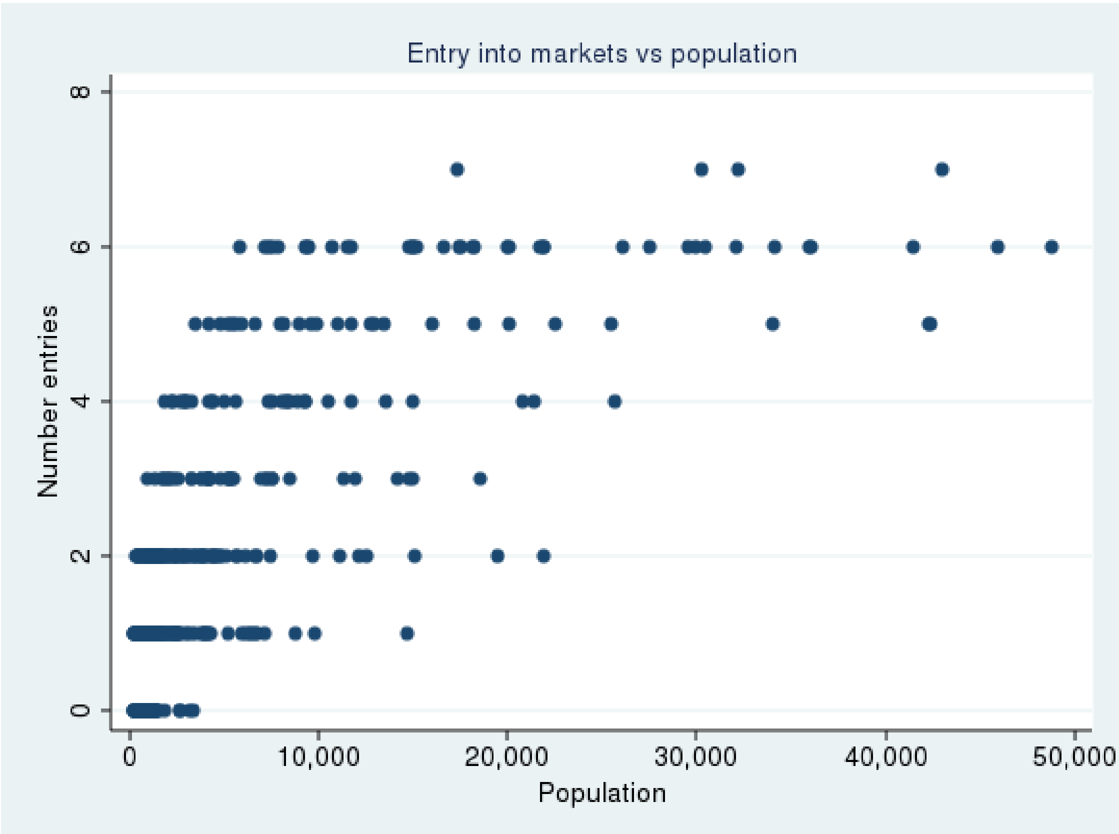

Figure 3: Example of market: Moose Jaw in Saskatchewan (I)

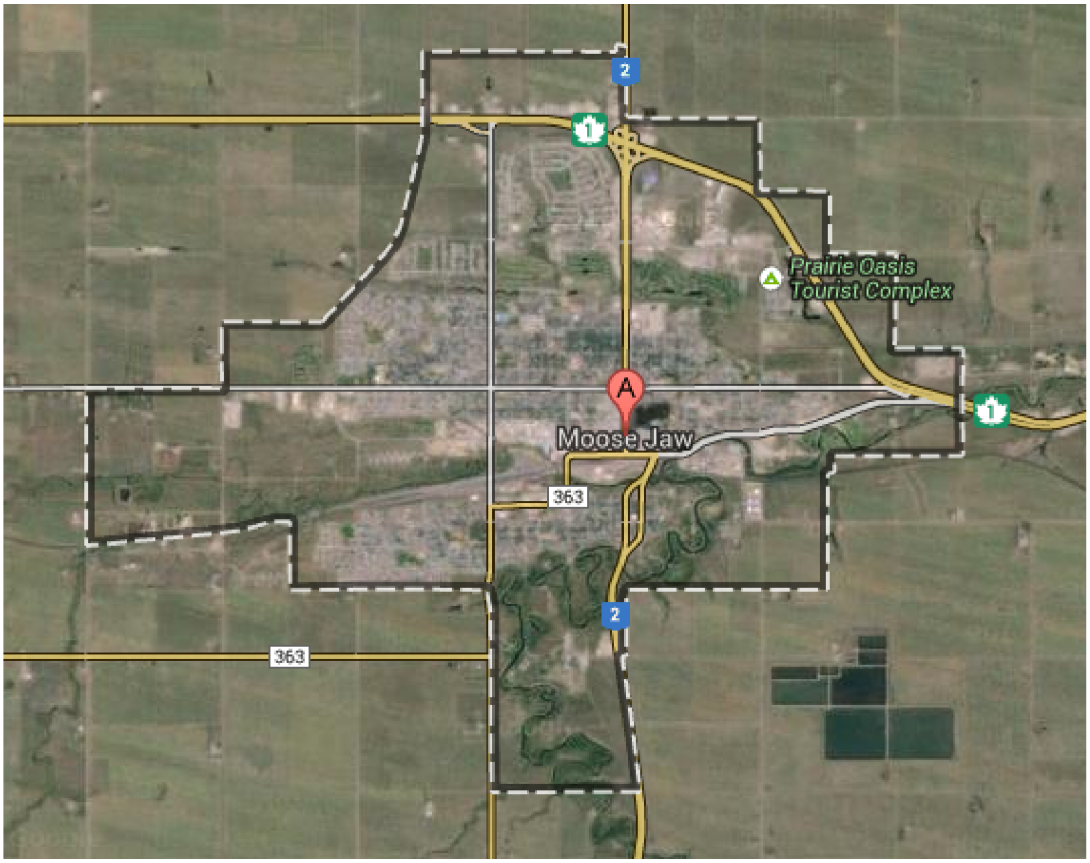

Figure 4: Example of market: Moose Jaw in Saskatchewan (II)

This map highlights the various bank branches in Moose Jaw (Saskatchewan). Each dot represents a branch, and different colors represent different institutions.

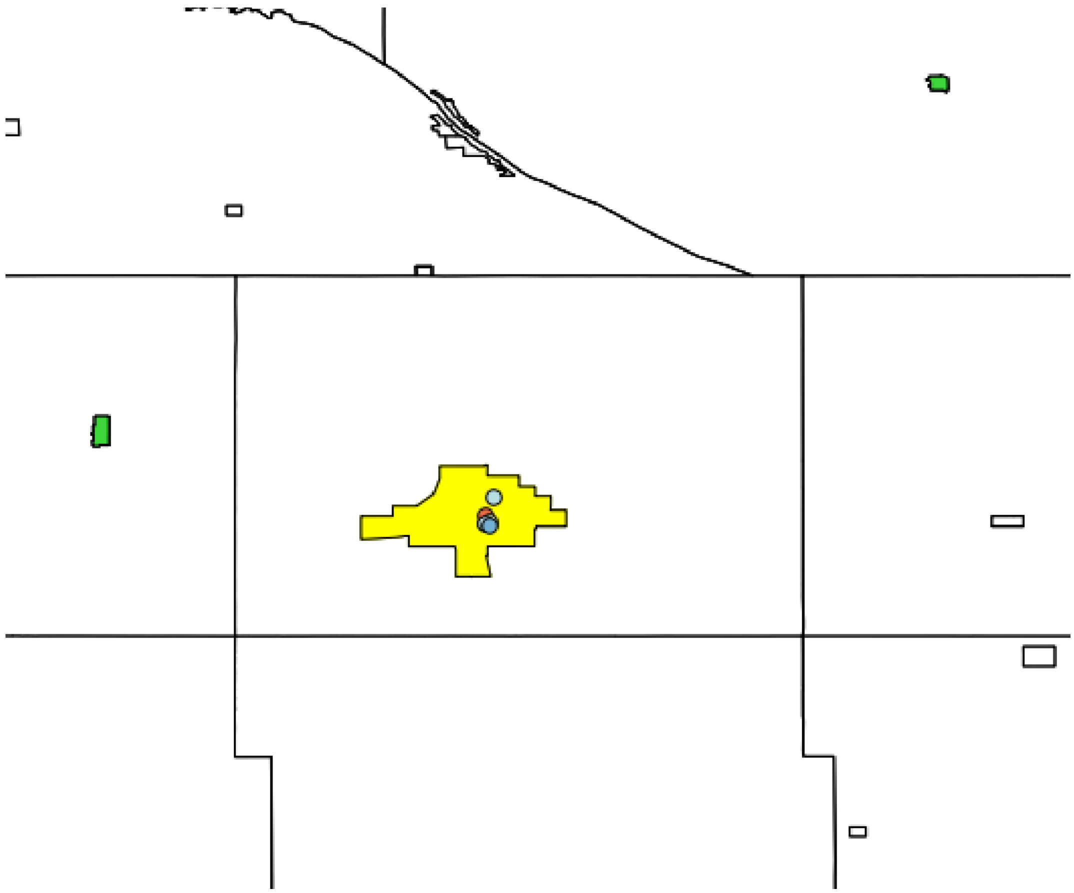

Figure 5: Markets considered and CFM households

This diagram shows the selection of markets and CFM households included in our estimation.