Competition, Geographic Proximity and Pricing in the Retail Banking Industry*

Santiago Carbo-Valverde † Hector Perez-Saiz ‡

July 19, 2018

Abstract

We quantify the effect of competition and geographic proximity of bank branches on the adoption of bank accounts, credit cards and lines of credit by households. We propose a methodology that models the strategic decision of market presence of each financial institution and the adoption of financial products by the households. This allows us, in comparison with previous studies, to solve for potential endogeneity issues when estimating the effect of competition in the market. Using a very detailed household-level database for Canada, we find that competition significantly reduces the monthly fees paid for bank accounts and credit cards, and that geographic proximity also has a relevant effect. We also perform a counterfactual policy experiment where we estimate the effect of bank branch closures on the adoption of financial products by segment of population. In an era of the emergence of sophisticated financial technologies, we show that physical branches and proximity still matter in order to understand financial inclusion and competitiveness in the industry.

JEL classification: G2, G21, L13, O16.

Keywords: Bank accounts, credit cards, competition, market presence, entry barriers, proximity.

*For all their comments and suggestions, we wish to thank Victor Aguirregabiria, Jason Allen, Christoph Bastenm Wilko Bolt, Evren Damar, Hans Degryse, Hans Gersbarch, Kari Kemppainen, Mark Manuszak, Ouarda Merrouche, Steven Ongena, Francisco Rodriguez, John Wilson and participants at presentations at the Bank of Canada, Central Bank of Poland, DNB Retail Payments Conference, Federal Reserve Board of Governors, Midwest Finance Association, Universita Cattolica del Sacro Cuore, University of Zurich and Wolpertinger Conference. Santiago Carbo gratefully acknowledges financial support from MICINN-FEDER ECO2015-67656-P. Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the Bank of Canada or the International Monetary Fund.

†CUNEF and Bangor University, Leonardo Prieto Castro 2, 28040 Madrid, Spain. scarbo@cunef.edu

‡International Monetary Fund (formerly at the Bank of Canada), 700 19th Street, N.W., Washington, D.C. 20431, USA. hperez-saiz@imf.org

1 Introduction

In most advanced economies the retail banking industry has experienced continuous consolidation over recent decades, which has raised the concerns of regulators regarding the level of competition in the industry. At the same time, the emergence of internet and mobile banking has led to bank branch closures in many countries. Both phenomena not only have had effects on competition, but also on the degree of geographic proximity between banks and customers, affecting the way they both connect and interact. This is an important issue, which is also related to the emergence of more sophisticated financial technologies, or *FinTech*, which are transforming the traditional relationship between banks and their customers, and is attracting considerable policy debate (Draghi, 2017).

This article studies the effect of competition and the geographic proximity of financial institutions on the prices and credit limits of retail banking services. We propose a methodology that takes into account the endogeneity of the market-presence decision by financial institutions in order to estimate the effect of competition and geographic proximity on the fees, rates and credit limits of bank accounts, credit cards, and lines of credit. Without correcting this endogeneity problem, simple OLS estimates may be biased, i.e. market presence may affect rates not due to competitive reasons but because markets are attractive and financial institutions are able to set higher rates.

In our methodology, we jointly estimate a latent profit equation that determines the presence of financial institutions in geographic markets, and a set of household-level outcome equations for the fees, rates and credit limits of the financial products considered. We correct for the endogeneity of the market-presence decision by considering the optimum equilibrium strategy of each financial institution that is a potential entrant in each market and decides to be present (or not) in equilibrium. We use a detailed database that includes the transaction prices and observed characteristics of the products acquired by every household, as well as their demographic attributes and length of relationship with the financial institution. This way we take into account household-level, product-level and other characteristics that may affect the fees, rates and credit limits offered. Controlling for these variables is crucial as the types of bank accounts, credit cards and other financial products have expanded over the years and financial institutions tend to 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 exploit two dimensions that are rarely considered together: market complexity (distance, sources of market power, and market-presence strategies) and customer/product complexity (price discrimination and product differentiation). Other methodologies, such as Panzar and Rosse (1987), often use aggregate revenue information and are not well suited to studying complex multiproduct firms such as large banks.

In the first set of results, we find that, after controlling for financial product characteristics and demographic household attributes, an additional competitor in the market decreases the monthly fees of a bank account by -6.7%. In contrast, when using standard regression analysis (without considering optimum market presence strategies) we find a smaller effect (-4.5%). A similar result–and the smaller effect of OLS estimations–is found when considering credit card fees (-20.9% vs -2.1%), and lines of credit (-2.2% vs +0.94%). The differences found across products suggest that some retail products are less sensitive to competition than others. We also find that competition tends to increase the credit limits offered for lines of credit, but to reduce it for credit cards.

Second, we find that the geographic proximity to the branch of a financial institution plays an important role in explaining the purchase of financial products. The literature has mainly focused on corporate loans (Petersen and Rajan, 2002; Degryse and Ongena, 2005; DeYoung *et al.*, 2008; Agarwal and Hauswald, 2010; Ergungor, 2010) and often finds that proximity facilitates credit but also makes it more expensive due to price discrimination. The empirical results suggest that banks increase the loan rates offered to geographically close corporations because of the lower transportation costs (Hotelling, 1929; Salop, 1979), rather than decreasing them due to the lower costs of monitoring (Sussman and Zeira, 1995).

We also perform a counterfactual policy analysis where we study the effect of bank branch closures on the adoption of financial products by population segments. Our empirical methodology, which uses detailed demographic household-level data, is convenient for analysing how adoption varies across various population segments. If banks close branches and replace brick and mortar branches with mobile banking or other technology-based financial services, individuals need to have the knowledge and skills required to use these new digital financial services. The literature on the digital divide (Norris, 2001) shows that factors such as age, income and education greatly affect the ability of individuals to use digital services. Therefore, in the counterfactual exercise we consider the case of individuals with low income or low education levels who are less likely to substitute the physical branch with online banking services.

The results of the counterfactual exercise show that credit cards are the products that are the least affected by the closure of branches (followed by accounts and lines of credit). In the extreme case where most branches are closed, adoption by more elderly individuals, low-income individuals and individuals without a university degree would be reduced by up to 50%. In addition, we study the effect of bank branch closures on the fees and rates paid for the financial products considered. This is an interesting case because branch closures simultaneously affect geographic proximity and also the level of competition in the markets, and both variables have an effect on the fees and rates paid. We find that branch closures have a larger relative impact on credit card fees compared to bank accounts and lines of credit, and this is mainly due to the large effect that we find of competition on credit cards.

Consistent with the previous literature, our results show that proximity facilitates access to accounts, credit cards and lines of credit. Also, we find that proximity increases the fees paid for accounts and credit cards, which is consistent with the fact that these products do not create large risks and they require low monitoring, so transportation costs are the main source of price discrimination. However, we find that lines of credit have lower rates for geographically close households, which is consistent with the fact that this is a more sophisticated product that may be more sensitive to monitoring costs. Interestingly, lending terms (credit limits) for closer households are worse for credit cards and lines of credit.

Our paper adds to the sizeable literature that measures competition by studying the effect of concentration on the prices of financial products, and it also considers other questions related to the nature of the lender-borrower relationship.[1] Although the literature has not paid as much attention to financial product fees as to interest rates (with some exceptions, see for instance Hannan, 2006; Dvořák and Hanousek, 2009; Tennant and Sutherland, 2014; Berg *et al.*, 2016), the compression of spreads in an era of low interest rates and financial stagnation has raised the importance of non-interest income to increase profitability in the banking industry.

Recent articles have studied the effect of market concentration by analysing bank mergers (Focarelli and Panetta, 2003; Garmaise and Moskowitz, 2006; Park and Pennacchi, 2009; Erel, 2011). A common difficulty in these studies is the fact that the merger decision is endogenous, i.e. banks decide to merge to expand into markets that are economically attractive, which affects the post-merger outcomes of competition. With our methodology, we are able to take into account the endogeneity of the market-presence decision by jointly estimating the equilibrium market-presence decision of every financial institution and the product outcome equations.

While some of the articles that look at the effect of market concentration on lending rates use data at the household 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 be very relevant for pricing purposes. For instance, account fees are usually waived by financial institutions when the balance is high enough, and premium credit cards tend to have larger fees when credit limits are high. Ignoring these characteristics may bias the estimation of the effect of market concentration on the fees and limits offered, as financial institutions in markets that face more competition may offer better product characteristics or discriminate differently on household observable characteristics.

Jointly estimating the equilibrium market-presence decision and the outcomes equations provides clear advantages. A simple regression analysis would not take into account possible endogeneity effects that may bias the estimates. For instance, a market with attractive demographic characteristics attracts new competitors, which may also set high fees for the financial products provided. In order to correct for this endogeneity problem, we propose a methodology that allows us to correct for this bias by endogenizing the market presence decision and taking into account

[1] Degryse *et al.* (2009) provide an extensive review of the existing literature. Some theories, such as the contestable market theory, argue that the mere threat of entry is enough to affect market power. By contrast, entry barrier models claim that it is actual market presence and not potential entry that affects competition.

the possible correlation between the unobserved factors that affect market presence and the factors that affect the competitive outcomes.

Mazzeo (2002b) and Manuszak and Moul (2008) use a two-step methodology and a simpler game structure to estimate the effect of market presence on market outcomes (see also Ellickson and Misra, 2012). In our paper, we have a richer structural framework that allows us to fully capture the effect of firm heterogeneity on market presence and on various competitive outcomes, but we require a more complex estimation methodology that uses simulation methods, as in Berry (1992), Bajari *et al.* (2010), Perez-Saiz (2015) or Perez-Saiz and Xiao (2017). We jointly estimate the equilibrium presence and outcome equations using a simulated maximum likelihood estimator from Gourieroux and Monfort (1990). Our paper follows the recent literature that uses structural empirical methods applied to the banking industry, as in Ferrari *et al.* (2010), Aguirregabiria *et al.* (2015), Egan *et al.* (2017) or Allen *et al.* (2017).

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, due in part to the existence of regulatory entry barriers, the number of firms that can be considered potential entrants in the markets considered is small. This makes the analysis tractable and the results easy to interpret, which is an advantage compared to other countries such as the U.S., with a much larger number of financial institutions and dynamic effects due to bank consolidation.

This paper is divided into six 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 empirical model. Section V discusses the empirical results. Section VI presents the counterfactual experiment. 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 in the Canadian financial system. This dominance has been enhanced over the last three decades through the deregulation that started in the 1980s, which gradually weakened and eliminated some market restrictions and led to significant industry consolidation.

Credit unions are another type of depository institution that exists in Canada. They are financial institutions founded on the cooperative principle and owned by their members. They can provide the same types of depository and lending services as banks do, although they have con-

[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).

straints on providing 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 credit union and the largest financial institution in Quebec. Canada has one of the largest credit union systems in the world in percentage terms,[3] with 11 million members covering more than 40% of the active population.[4]

The evolution of the Canadian retail banking industry is shown in Figure C.1 in the Appendix. 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.[5] At the same time, a stable industry structure emerged after the merger between Canada Trust and TD Bank in 1999, characterized by the absence of entry and merger between large players. Furthermore, 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".

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 market concentration and lack of entry naturally raises questions about how competitive the Canadian retail banking industry actually is. Some prior studies of competitiveness in the Canadian banking industry focused on indicators of contestability (Baumol *et al.*, 1982) on a national scale, using bank-wide variables such as total assets, such as in Allen and Liu (2007). Other studies focused on specific products such as mortgage loans (Allen *et al.*, 2014). In our paper we jointly estimate the market presence and the pricing decision of financial institutions to quantify the effect of geographic presence on market outcomes.

3 Data

3.1 Market presence data

For our model, we use financial institution 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 the 1980s and 1990s, all depository institutions can accept deposits from individuals and businesses, and they no longer have regulatory barriers the prevent them from entering each other's businesses. We therefore consider all financial institutions to be competing in the same overall market.

[3]The largest US credit union, Navy Federal Credit Union, only has assets of $58 billion as 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.

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

[5]This stable market configuration is also supported by a very low rate of entry and exit of bank branches at market level in that period (about 2% per year).

Table 1: Summary statistics for markets considered in estimation.

Summary statistics of markets considered in estimation, for all of Canada, and by province. BC: British Columbia. MT: Manitoba. NB: New Brunswick. NFL: Newfoundland and Labrador. NS: Nova Scotia. ON: Ontario. QC: Quebec. SK: Saskatchewan. Source: Stats Canada.

Variable	Canada	BC	MT	NB	NFL	NS	ON	QC	SK
Population:									
mean	7,489	8,653	2,862	3,915	2,236	5,817	14,584	5,743	4,593
min	205	305	275	355	215	440	365	230	210
p25	745	1,438	670	825	385	3,020	2,990	718	345
p50	2,000	3,238	1,250	1,190	650	4,770	7,565	1,338	555
p75	6,700	6,532	2,735	1,590	3,010	8,650	14,630	2,795	1,000
max	199,385	122,175	40,705	66,690	13,385	13,865	155,995	144,595	199,385
Per-capita income:									
mean	23,088	24,504	22,322	20,220	16,203	20,461	26,088	21,307	20,328
min	0	16,715	15,349	13,836	0	15,858	15,431	0	0
p25	20,111	21,866	19,579	18,475	16,356	19,274	22,664	18,502	18,062
p50	23,048	23,674	21,713	19,996	17,773	20,270	25,767	21,106	21,836
p75	26,636	26,674	24,589	21,661	18,664	21,489	28,785	23,324	25,275
max	47,958	38,508	38,311	30,485	27,079	27,443	39,947	47,958	32,131
Number of business:									
mean	576	746	304	235	114	385	964	332	428
min	0	0	0	12	0	61	0	0	0
p25	82	134	99	34	14	201	186	34	71
p50	217	304	152	88	35	291	556	86	126
p75	571	608	410	114	110	538	1,145	199	274
max	14,249	9,985	2,906	3,918	663	1,214	7,918	8,516	14,249
Proportion of French population (in %):									
mean	19.1	1.7	6.1	18.4	0.8	3.9	9.7	92.7	3.3
min	0.0	0.0	0.0	0.0	0.0	0.3	0.0	12.9	0.0
p25	0.7	0.9	0.7	1.1	0.0	0.9	1.0	95.0	0.0
p50	1.7	1.5	1.4	2.9	0.0	1.1	1.6	98.1	0.9
p75	7.0	2.2	2.8	11.3	0.2	1.6	5.2	99.2	2.6
max	100.0	7.7	86.7	94.6	20.2	65.8	100.0	100.0	100.0
Unemployment rate (in %):									
mean	9.3	8.9	4.5	14.9	28.8	13.5	8.1	13.7	7.0
min	0.0	0.0	0.0	2.7	0.0	5.7	0.0	0.0	0.0
p25	4.7	5.8	2.0	9.4	18.3	9.5	5.4	8.6	0.0
p50	7.2	8.2	3.6	13.2	25.0	11.6	7.1	12.8	5.4
p75	11.6	11.0	6.2	17.1	37.1	16.7	8.9	17.4	10.3
max	60.0	34.2	19.2	43.8	60.0	37.2	36.6	40.0	38.1

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 instance, Toronto, with a population of more than 2.6 million people, constitutes one census subdivision, just like Martensville, SK, a small city with fewer than 8,000 inhabitants. Apart from cities and towns, census subdivisions also include rural areas grouped together into counties, Aboriginal 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 rural and urban isolated census subdivisions to be included in our model, based on well-defined criteria. In particular, we only include census subdivisions that have between 200 and 200,000 individuals and that are separated by at least 10 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 very large cities, which are composed of multiple neighbourhoods and have an internal structure that makes it harder to get a well-defined market. Very large cities are also excluded given that our model does not take into account the number of branches a financial institution has in a market, only whether it is present in a market at all. Eliminating very large population centres minimizes this confounding factor and most of the financial institutions in our sample are present in a market with a single branch.

We then eliminate markets that are located less than 50 km away from any major urban centres. Excluding markets located close to large urban centres helps 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. 50 km 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] We then limit the area of each market to 300 square kilometres so it can be reasonably thought of as a single market rather than composed of multiple separated markets. We also exclude Aboriginal reserves from consideration, given their special administrative status, and thus avoid potential regulatory confounders. After considering all these exclusions and constraints, we have 693 markets that we use in our estimations. These isolated markets show a relatively clean relationship between population and market presence (see Figure C.2 in the Appendix). A map of one of the markets we have selected, Moose Jaw (Saskatchewan), is shown in Figures C.3 and C.4 in the Appendix.

We construct the dependent variable of financial institution 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 present in the market, we set the dependent variable to 1. Otherwise, we set the dependent variable to 0. We focus on the market-presence decision of the Big Six banks and the credit unions (which includes Desjardins and other smaller credit unions). In total we

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

consider 7 potential entrants in every market selected. Therefore, the industry structure is determined by the market presence decision of all 7 potential entrants, and can have 2^7 configurations per market. This methodology allows us to study with detail the effect of competition on market outcomes despite the lack of available data of concentration levels by financial institution and market in Canada (e.g. HHI or C4 indicators).

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 household preference for financial institutions 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 693 markets considered are shown in Table 1. Nationally, the average population is about 7,500. 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 large cities with populations of up to 200,000. Statistics per province show significant differences between them. Newfoundland and Labrador have 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 the largest retail presence in rural markets, and they have a particularly large presence in Quebec.

In addition, Table 3 shows some statistics for markets where only the Big Six are present, and for markets where only CUs are present. As we can observe in the statistics shown, the Big Six 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.

Table 2: Market entry by province and institution.

In this table we show the number of markets where every financial institution is present by province. Our sample consists of 863 markets. RBC: Royal Bank of Canada. BMO: Bank of Montreal. BNS: Bank of Nova Scotia. CIBC: Canadian Imperial Bank of Commerce. NBC: National Bank of Canada. TD: Toronto-Dominion Bank.

Province	BMO	BNS	CIBC	CU	NAT	RBC	TD	Markets with entry	Markets with no entry
British Columbia	20	24	44	37	0	34	15	64	32
Manitoba	2	6	12	31	0	15	5	39	16
New Brunswick	2	4	1	8	2	3	2	10	19
Newfoundland	4	6	4	2	0	3	1	7	22
Nova Scotia	4	11	9	10	0	12	4	24	5
Ontario	90	80	96	97	17	99	83	175	39
Quebec	6	0	12	92	27	10	6	92	34
Saskatchewan	11	15	28	57	2	30	14	81	34
Total	139	146	206	334	48	206	130	492	201

3.2 Household-level data

Financial product data at household level are obtained from Ipsos Reid's Canadian Financial Monitor (CFM) database for years 2007-2010.[7] This database includes a very complete overview of 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, checking 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 showing general attitudes 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 the relevant outcome variables for our analysis are not available (e.g. mortgages), or because the products are too specialized (e.g. mutual funds).

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

Summary statistics of markets considered in estimation, for markets where only Big6 or CU are present. 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, respectively. Source: Stats Canada.

	Big6 present only		CU present only		t-test: Big6 - CU
Variable	mean	se	mean	se	t
Population	6,180	545	1,632	212	7.78***
Total income	25,320	309	21,406	642	5.49***
Proportion French (in %)	5.17	1.01	29.87	3.81	-6.26***
Number of business	539	47	155	13	7.82***
Unemployment rate (in %)	7.33	0.37	9.46	0.77	-2.49**

We carefully select the households to be included in our empirical model that are consistent with our market-presence model. Figure C.5 in the Appendix 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

[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. Having a large sample size helps for the identification of the model equations. In addition, the market structure is relatively stable in this period.

geographic presence in a circle around every CFM household selected. In order to estimate this geographic location with the highest possible detail we use 6-digit postal-code location information which we convert to latitude-longitude information for branches, CFM households, and census subdivisions.[8] In total we have 11,599 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 6.82 dollars per month for every bank account and many accounts do not have any fees. Households also pay on average 16.47 dollars per year for every credit card. On average, the annual interest rate of a line of credit is 4.88%. Credit limits for lines of credit are significantly larger than for credit cards.

In addition, Table 4 includes information about the geographic proximity of branches in a 10-km radius around the household. On average, every household considered in our sample has 4.5 financial institutions in a 10-km circle of radius around it. Moreover, the probability that a financial institution that provides a financial product to the household is within a 10-km radius around it is about 80% for accounts, credit cards and lines of credit. This result shows that geographic proximity plays an important role in explaining the adoption of a financial product of a certain financial institution by a household.

In our empirical model, a unit of observation is a financial product acquired by a household-year for each of the 7 financial institutions considered in the market-presence model. Since there are 11,599 unique household-year observations in our sample, and there are 7 financial institutions, in total we have a sample of size $N = 11,599 \times 7 = 81,193$ observations where we observe financial product characteristics (when the household has the product with the financial institution), or we do not observe them in the event that the household has not acquired the product with that financial institution. We do not observe interest rates for bank accounts or credit cards in our database, but we do observe monthly fees. We also observe interest rates for lines of credit, and credit limits for credit cards and lines of credit that we use in our empirical model.

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

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)	6.82	15.26	0.00	0.00	0.00	0.00	8.00	81.00	11,961
Accounts: Balance (in $10,000)	0.50	1.88	0.00	0.00	0.03	0.15	0.45	5.50	11,961
Accounts: Checking account	0.63	0.48	0.00	0.00	0.00	1.00	1.00	1.00	11,961
Accounts: Length relationship	5.70	1.71	1.00	1.00	5.00	7.00	7.00	7.00	11,961
Cards: Fees (dollars)	16.47	39.22	0.00	0.00	0.00	0.00	0.00	165.00	11,295
Cards: Limits (1000s dollars)	9.20	8.38	0.00	0.00	3.20	7.50	13.00	35.00	11,295
Cards: Credit protection	0.23	0.42	0.00	0.00	0.00	0.00	0.00	1.00	11,295
Cards: Rewards	0.25	0.43	0.00	0.00	0.00	0.00	0.00	1.00	11,295
LOC: Rates (in %)	4.88	3.18	0.00	0.00	3.00	5.25	6.50	14.50	4,390
LOC: Limits (1000s dollars)	41.18	54.65	0.00	0.30	12.50	22.50	45.00	262.50	4,390
LOC: Fixed rate	0.32	0.46	0.00	0.00	0.00	0.00	1.00	1.00	4,390
LOC: Secured line of credit	0.57	0.50	0.00	0.00	0.00	1.00	1.00	1.00	4,390
LOC: Length relationship	5.98	1.47	1.00	2.00	5.00	7.00	7.00	7.00	4,390
Demographic variables:									
Assets (in 100,000s)	0.94	2.22	0.00	0.00	0.02	0.11	0.74	11.27	11,599
Age	54.62	15.60	18.00	22.00	43.00	56.00	66.00	86.00	11,599
Size household	2.22	1.15	1.00	1.00	1.00	2.00	3.00	6.00	11,599
Own house	0.79	0.41	0.00	0.00	1.00	1.00	1.00	1.00	11,599
Difficulty paying debt	2.97	2.53	0.00	0.00	1.00	2.00	5.00	9.00	11,599
Employed	0.91	0.29	0.00	0.00	1.00	1.00	1.00	1.00	11,599
Uses financial advisor	0.36	0.48	0.00	0.00	0.00	0.00	1.00	1.00	11,599
University degree	0.85	0.36	0.00	0.00	1.00	1.00	1.00	1.00	11,599
Married	0.74	0.44	0.00	0.00	0.00	1.00	1.00	1.00	11,599
Income (in 100,000)	0.65	0.53	0.08	0.08	0.32	0.50	0.80	2.50	11,599
Sophisticated investor	0.39	0.49	0.00	0.00	0.00	0.00	1.00	1.00	11,599
Proximity of branches:									
Number FIs in circle radious 10km	4.50	2.24	0.00	0.00	3.00	6.00	6.00	7.00	11,599
FI for ACC is in circle radious 10km	0.84	0.37	0.00	0.00	1.00	1.00	1.00	1.00	11,961
FI for CARD is in circle radious 10km	0.81	0.39	0.00	0.00	1.00	1.00	1.00	1.00	11,295
FI for LOC is in circle radious 10km	0.85	0.36	0.00	0.00	1.00	1.00	1.00	1.00	4,390

4 Empirical model

4.1 Geographic presence of financial institution branches

We estimate a perfect information static game where every potential entrant decides to be present in every market (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, there is perfect information and 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 financial institutions (see Ishii, 2005; Dick, 2007). In our empirical model, we consider firm-level controls such as total size of the financial institution, which 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 presence in market m only if it is profitable, therefore

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

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

As in 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 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 presence in the market are equal to:

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

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 $\varepsilon_m^{\pi,i}$ is a market and firm-specific, independent and identically distributed error term with variance normalized to one. $\varepsilon_m^{\pi,i}$ is observed by all potential entrants, but not by the econometrician.

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

We also model competitive effects between financial institutions that enter in the market, represented by the term $a_{j,m}$. 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 financial institution j's on financial institution i's profit if financial institution j is present in the market. This is a flexible way to take into account firm-level unobserved effects that affect each financial institution's competitiveness against other financial institutions. This approach also allows us to differentiate between different competitive models for every financial institution. For instance, credit unions could compete more aggressively because they may not 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 $a^*_m = (a^*_{1,m}, ..., a^*_{E,m})$ 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 need to solve for all Nash equilibria, and in the case of multiple equilibria, we assume that the most efficient equilibrium is selected (see identification section in the Appendix). We 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 financial institutions and credit unions are competing in the same market, for the same customers.

4.2 Accounts

The first financial product that we consider is bank accounts. Financial institutions may charge monthly fees to bank accounts, and these fees may depend on the number of transactions per month, minimum balance, or the type of accounts (checking vs saving). Financial institutions may also set different fees to clients depending on observed demographic variables. For bank accounts (and for the rest of financial products considered), we use the following vector of demographic variables:

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

We characterize the monthly fees paid by a household i for an account with financial institution b in year t as follows:

$$\begin{aligned} Fees^{ac}_{i,b,t} = \alpha^{ac}_1 + \alpha^{ac}_2 \cdot balance_{i,b,t} + \alpha^{ac}_3 \cdot checking_{i,b,t} + \alpha^{ac}_4 \cdot X_{i,t} \\ + \alpha^{ac}_5 \cdot length_{i,b,t} + \alpha^{ac}_6 \cdot Close_{i,b,t} + \alpha^{ac}_7 \cdot Number_{i,t} + B_b + \varepsilon^{ac,F}_{i,b,t}. \end{aligned} \tag{5}$$

Variable *Fees* is the monthly fee paid for the bank account, expressed in logs. Variable *balance* is the current 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 financial institution that provides the service and the household. $Close_{i,b}$ is an indicator variable equal to 1 if financial institution b has geographic presence in a 10-km radius around the location of household i. *Number* is the number of financial institutions present in a 10-km radius around the location of the household. This variable represents the competitive effect of the market presence of financial institutions on the account fees. We also include financial institution-fixed effects (B_b). These fixed effects are considered in Eq. (5) to take into account any pricing policy set by financial institutions 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 financial institution b in year t. Although bank accounts are very common financial products, households rarely have accounts with all 7 financial institutions considered in our sample. We consider that the decision of a household i to have a bank account with financial institution b is given by the following latent equation:

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

and we observe that household i has an account with financial institution b, i.e. $D^{ac}_{i,b,t} = 1$, when the latent variable $D^{*\,ac}_{i,b,t} \geq 0$:

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

Demographic characteristics should influence the decision of having an account given by Eq. (6). For instance, 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 of using 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 payment channels (online, mobile, branches, ABM, etc) over one month. We would expect that households with high usage of various payment channels would be more likely to use bank accounts.

The selection of the variables to be included in every equation for accounts and for other products is based on broad perceptions about how households and banks make their economic decisions. For instance, $heavy\_usage_{i,t}$ should affect the demand for bank accounts by households, but not the fees paid, since financial institutions 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 (especially for new clients).

Accounts are products that are broadly used by most households but there are other products that are not so common such as credit cards and/or lines of credit. For these products, we use indicators for financial sophistication and financial 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. This implies that these variables are relatively opaque for the financial institutions. We also use risk variables that affect the limits granted for financial products by financial institutions. Credit limits granted by financial institutions should depend highly on the 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. These exclusion restrictions of variables used in our empirical model are explained in the Appendix in greater detail.

4.3 Credit cards

The second financial product that we consider is credit cards. Financial institutions 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 the interest rate charged on the outstanding balance in credit cards in the CFM database, but we observe the annual fees paid. As in the case of bank accounts, we assume that the annual fees paid for a credit card with financial institution b by household i are determined by the following equation:

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

Variable *Fees* is the annual fee paid for the credit card, expressed in logs. We use similar control 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 financial institutions because of the risk of default on the card balance. A crucial variable used by financial institutions to control risk is the credit limit of the credit card. Typically, households can easily get applications approved for credit cards from most financial institutions, but financial institutions use the amount of credit limit granted to control the risk of default. We propose the following equation to explain the credit limit given by financial institution 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 financial institution, and therefore should determine the total credit provided. We use an indicator for unemployment, and a variable (with values between 0 and 9) that the household uses to rank the perceived difficulty for paying the debt.

We assume that the decision of having a credit card with financial institution b is based on the following latent variable:

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

As in the case of accounts, we also consider demographic characteristics to influence the decision of having an account. Variable *sophisticated* is an indicator variable of financial sophistication, which is assumed equal to one when a household has more than 20% of the value of its total assets either in stock exchange assets or mutual funds. Also, variable *advice* is an indicator variable equal to 1 if the household regularly uses a financial advisor.

4.4 Lines of credit

The third and last financial product that we consider is lines of credit. This is defined as a pre-approved loan that the household can draw on at any time 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:

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

Variable *Rates* is the annual interest rate charged on the outstanding balances of the line of credit, expressed in logs. In addition to other variables used in previous equations, 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). Variables used in Eq. (11) are also used in the credit limit equation for lines of credit:

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

We also use in Eq. (12) similar variables as those proposed to explain the credit limit equation for credit cards. As for the other financial products considered, the decision of having a line of credit with a financial institution 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

We could potentially allow for arbitrary correlation between the elements of the covariance matrices of the error terms of the equations of the empirical model. In practice, we restrict the number of correlations for tractability reasons and we focus on several key issues. In particular, we consider that the unobserved variables that affect market presence 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 financial institution (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 a relevant 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:

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

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

$$\sum^{cc} = \begin{array}{c} \\ \varepsilon^{\pi,1} \\ \dots \\ \varepsilon^{\pi,7} \\ \varepsilon^{cc,D} \\ \varepsilon^{cc,F} \\ \varepsilon^{cc,L} \end{array} \begin{array}{c} \begin{array}{cccccc} \varepsilon^{\pi,1} & \dots & \varepsilon^{\pi,7} & \varepsilon^{cc,D} & \varepsilon^{cc,F} & \varepsilon^{cc,L} \end{array} \\ \begin{bmatrix} 1 & & & & & \\ \dots & \dots & & & & \\ \rho_\pi & \dots & 1 & & & \\ \rho^{cc}_{\pi,D} & \dots & \rho^{cc}_{\pi,D} & 1 & & \\ \rho^{cc}_{\pi,F} & \dots & \rho^{cc}_{\pi,F} & \rho^{cc}_{D,F} & \sigma^{cc}_F & \\ \rho^{cc}_{\pi,L} & \dots & \rho^{cc}_{\pi,L} & \rho^{cc}_{D,L} & 0 & \sigma^{cc}_L \end{bmatrix} \end{array} \quad (15)$$

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

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

4.6 Estimation

This section explains in detail the simulated maximum likelihood methodology used to estimate the empirical model. An observation in our empirical model is a financial product that a household i has acquired from financial institution b in a given year t. If the household has the product with financial institution b, then we will observe fees/rates or credit limits. Therefore, for the three products $p \in \{ac, cc, loc\}$, 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 the empirical model.

In addition, market presence by financial institution b is endogenous in the model. Therefore, variables $Close_{i,b,t}$ and $N_{i,t}$ are endogenous and calculated after solving the market equilibrium condition in Eq. (3) for every market m where households are located.

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

$$
\begin{aligned}
\Pr(Fees^{p}_{i,b,t}, Limit^{p}_{i,b,t}, D^{p}_{i,b,t}, Close_{i,b,t}, N_{i,t}) = &\left[\Pr(D^{p}_{i,b,t} = 0, Close_{i,b,t}, N_{i,t})\right]^{(1-D^{p}_{i,b,t})} \cdot \\
&\left[f(Fees^{p}_{i,b,t}, Limit^{p}_{i,b,t} / D^{p}_{i,b,t} = 1, Close_{i,b,t}, N_{i,t}) \cdot \Pr(D^{p}_{i,b,t} = 1, Close_{i,b,t}, N_{i,t})\right]^{D^{p}_{i,b,t}} .
\end{aligned} \tag{17}
$$

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:

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

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

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

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 in 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 that requires significant computation is

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

which is calculated by solving the market presence equilibrium using Eq. (3). This term represents the predicted probability of observing the presence of a financial institution b and a number of financial institutions N in a circle around the household i. Since there is no closed form solution for this predicted probability of market presence, we need to numerically estimate it, i.e. for each draw, we have to numerically solve for all Nash-equilibria in every market using Eq. (3) and choose the most profitable one (equilibrium selection rule). This is the approach used in Bajari *et al.* (2010) to estimate static games of perfect information, which has also been recently used by Perez-Saiz (2015). Note that the financial institutions decide to be present in a given market m (a census subdivision) taking into account the demographic characteristics of the market, and the market-presence decision of other financial institutions. Using the market-presence equilibrium, the variables $Close_{i,b,t}$ and $N_{i,t}$ in Eq. (20) can be calculated for every household i and financial institution b.

Note that because error terms across products are uncorrelated, the likelihood function 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. Therefore, these assumptions significantly simplify 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 in 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 81,193 household-bank-year product observations. A detailed discussion about the identification of the empirical model is in the Appendix.

5 Empirical results

5.1 Estimates of market presence model

We first discuss the estimates of the market-presence 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 in Eq. 2 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 close to zero coefficient on business activity and a negative one on the proportion of Francophone population, whereas Big Six banks have a larger coefficient on business activity and positive on Francophone population. The coefficient for unemployment has opposite signs for both types of institutions. The coefficient for per-capita income is positive for both, but larger for the Big Six. These results show that the Big Six are particularly focused on markets with large business activity, and high per-capita income, whereas credit unions are much more focused on populated markets with few businesses.[10]

Table 5 also shows the competitive effects between types of financial institutions. Interest-

[10]There is a relatively large variation of French-speaking populations 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).

ingly, only the effect of the presence of CU on the profits of the Big Six 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 are present, credit unions banks are tough competitors for Big Six banks.

Table 6 shows results for provincial and individual financial institution 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 costs than the Big Six in general, so they are present in markets that are less attractive. Table 7 provides estimates of entry costs for all provinces and for all financial institutions. The constant term (intercept) represents the entry cost for RBC in Ontario. In most cases, the entry costs are economically relevant. These estimates show that barriers to entry into new markets are significant, which provides some evidence against the hypothesis of a perfectly contestable market that requires zero sunk entry costs.

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

All these results presented suggest that CU and the Big Six focus their market-presence 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 also be 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 also be advantageous for credit unions regarding this relationship, which may affect the quality of service in general.[11]

Moreover, it could be the case that credit unions face provincial prudential regulations that are different from their federal counterparts. The existence of different regulatory authorities in Canada, at provincial and federal level, could affect the effective implementation of the regulation and supervision of the industry (see for instance Agarwal *et al.*, 2014, who show that state regulators tend to be more lenient than federal regulators in the US).

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

Table 5: Estimates of market presence model (I).

We show estimates of the elements of the profit function (Eq. 2) in the market presence model. Demographic variables have been normalized by their mean. We have estimated the separate effects of the demographic variables for all Big6 banks, and for credit unions. "Competitive effect of X on Y" is the effect on the profit of financial institution Y if financial institution X is present in the market. Standard errors in parentheses obtained using the bootstrap method.

Variable	Market presence model
Panel A: Competitive effects:	
Competitive effect of BIG6 on BIG6	0.26326
	(0.02863)
Competitive effect of CU on BIG6	-0.06348
	(0.03233)
Competitive effect of BIG6 on CU	0.24106
	(0.03497)
Panel B: Demographic variables:	
Intercept	-1.95510
	(0.17070)
Population BIG6	-0.06412
	(0.02691)
Population CU	0.00785
	(0.02125)
Income per capita BIG6	0.49940
	(0.05857)
Income per capita CU	0.18517
	(0.03357)
Unemployment BIG6	0.04976
	(0.03146)
Unemployment CU	-0.00989
	(0.02859)
Business activity BIG6	0.29849
	(0.03272)
Business activity CU	0.03344
	(0.02241)
Proportion French BIG6	0.00696
	(0.01997)
Proportion French CU	-0.29000
	(0.05638)

Table 6: Estimates of market presence model (II).

We show estimates of the elements of the profit function (Eq. 2) in the market presence model. Some indicator variables are omitted due to perfect multicollinearity. Size and distance to headquarter are normalized by their mean. Standard errors in parentheses obtained using the bootstrap method.

Variable	Market presence model
Panel C: Provincial effects:	
British Columbia	0.25702
	(0.03366)
Manitoba	0.26106
	(0.03141)
New Brunswick	0.42259
	(0.04010)
Newfoundland and Labrador	0.36826
	(0.03531)
Nova Scotia	0.69900
	(0.06863)
Quebec	0.10300
	(0.03063)
Saskatchewan	0.52476
	(0.04949)
Panel B: Firm-level effects:	
National Size	-0.06038
	(0.04953)
Regional Size	0.17012
	(0.02999)
BMO	-0.04420
	(0.02634)
BNS	0.08511
	(0.02764)
CIBC	0.50230
	(0.05056)
CU	1.12893
	(0.11324)
NBC	-2.17158
	(0.18631)
distance to historical HQ	0.34095
	(0.04322)
distance to historical HQ (square)	-0.46559
	(0.03582)

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 we compare it with the case of 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 geographic proximity of a financial institution on the adoption of 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 often requires physical presence at the branch, and it could also be related to the usage of services related with the account (e.g. cash withdrawals). On the other hand, credit cards are typically used at the point of sale, and do not often involve interactions with the bank branch. Also, lines of credit are products that are not particularly related to payment activity by the households and their proximity coefficient is similar to credit cards.

We also find that certain demographic characteristics are key to explaining adoption. Households 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 market-presence 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 a geographic area close to the location of the household decreases the monthly rate for an account by -6.7%. In contrast, the effect in the OLS estimates is smaller (-4.5%). A similar effect is found when considering credit cards (-20.9% vs -2.1%) and lines of credit (-2.2% vs +0.9%). These differences between structural and OLS estimates can be understood from the estimates of the covariance matrix shown in Table 11. The correlation term $\rho_{\pi,F}$ and $\rho_{\pi,R}$ is positive and relatively large for credit cards and lines of credit. This implies that unobservable variables that increase the fees/rates paid by households for credit cards and lines of credit are positively correlated with unobservable variables that positively affect entry in those markets. Therefore, if we do not take into account this positive correlation, we would be significantly underestimating the effect of market presence in the fees/rates paid for these products.

Another interesting result is the effect of the geographic proximity on the fees and rates paid for the financial products. Households with an account provided by a financial institution that is close (in a circle of radius 10 km) tend to pay higher fees (10.3%). In contrast OLS gives a much smaller estimate (2.1%). For the case of credit cards, we find a positive effect of proximity on the fees (9.4%), whereas OLS gives a positive and larger value (16.2%). This result obtained for accounts and credit cards is consistent with the fact that these products do not create large risks and they require low monitoring from the financial institution, so transportation costs should be

Table 7: Entry barriers by potential entrant and province

In this table, we show total fixed effects by firm and province. They are calculated by adding the constant term, the financial institution fixed effects, and the provincial fixed effects.

Province	**BMO**	**BNS**	**CIB**	**DES**	**LCU**	**NAT**	**RBC**	**SCU**	**TD**
British Columbia	-2.4	-2.6	-2.2	-1.8	-0.7	-3.3	-2.8	-1.4	-2.8
Manitoba	-2.5	-2.6	-2.3	-1.9	-0.8	-3.4	-2.9	-1.5	-2.9
New Brunswick	-2.1	-2.2	-1.9	-1.4	-0.3	-2.9	-2.4	-1.0	-2.4
Newfoundland and Labrador	-2.1	-2.2	-1.9	-1.5	-0.4	-3.0	-2.5	-1.1	-2.5
Nova Scotia	-2.2	-2.4	-2.0	-1.6	-0.5	-3.1	-2.6	-1.2	-2.6
Ontario	-2.5	-2.7	-2.3	-1.9	-0.8	-3.4	-2.9	-1.5	-2.9
PEI	-2.2	-2.3	-2.0	-1.5	-0.4	-3.0	-2.5	-1.1	-2.5
Quebec	-2.6	-2.7	-2.4	-1.9	-0.8	-3.4	-2.9	-1.5	-2.9
Saskatchewan	-2.5	-2.6	-2.3	-1.9	-0.8	-3.4	-2.8	-1.5	-2.8

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 the bootstrap method (for structural model).

	Accounts		**Credit cards**		**Lines of credit**	
Variable	Structural	Probit	Structural	Probit	Structural	Probit
Financial advisor			0.05609	0.03725	0.08457	0.06932
			(0.00898)	(0.01229)	(0.01326)	(0.01617)
Heavy usage	0.06863	0.03703	0.07292	0.01878		
	(0.01388)	(0.00267)	(0.02075)	(0.00269)		
Sophisticated			0.01666	-0.00484	0.03255	0.01616
			(0.01000)	(0.01361)	(0.01199)	(0.01842)
Close provider	0.66654	0.64877	0.56575	0.55050	0.56214	0.54305
	(0.06428)	(0.01351)	(0.05028)	(0.01333)	(0.05647)	(0.01899)
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	81193	81193	81193	81193	81193	81193

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 the bootstrap method (for structural model).

	Account fees		**Credit card fees**		**Line of credit rates**	
Variable	Structural	OLS	Structural	OLS	Structural	OLS
Account balance	-0.01746	-0.01872				
	(0.00217)	(0.01351)				
Checking account	0.10938	0.10784				
	(0.01123)	(0.03892)				
Card protection			0.35268	0.35261		
			(0.03570)	(0.06037		
Rewards			1.63739	1.63721		
			(0.16497)	(0.07452		
Limit (in logs)			0.22268	0.22233	0.10165	0.10175
			(0.02119)	(0.01212)	(0.01011)	(0.03952)
Fixed rate					0.24952	0.24952
					(0.02612)	(0.07952)
Secured					0.08798	0.08801
					(0.00864)	(0.08157)
Length relationship	0.08370	0.07063			0.00413	0.00446
	(0.01384)	(0.01249)			(0.00175)	(0.02585)
Close provider	0.10383	0.02192	0.09425	0.16208	-0.24639	0.02825
	(0.00998)	(0.06665)	(0.00916)	(0.08453)	(0.02285)	(0.13119)
Number of competitors	-0.06765	-0.04515	-0.20952	-0.02106	-0.02275	0.00942
	(0.01419)	(0.01146)	(0.02012)	(0.01557)	(0.00280)	(0.02234)
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	11961	11961	11295	11295	4390	4390

the main source of price discrimination. On the contrary, for the case of lines of credit, we find a negative effect (-24.6%), much larger than OLS (+2.8%). This is consistent with the fact that lines of credit are more sophisticated products and may be more sensitive to monitoring costs.

The estimate of the variable that provides the effect of the length of relationship with the financial institution on the fees/rates paid is positive for all products, but economically much larger for accounts. One interpretation of this result is that bank accounts create large switching costs on households, whereas a line of credit is a more sophisticated and less common 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 10.9% more expensive than a saving account, and if the account balance increases, the annual fees tend to decrease. A household with a rewards credit card pays 163.7% more fees, whereas a credit card with credit protection has fees that are 35.2% higher. Also, the higher the credit 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 8.7% and 24.9% more expensive (in terms of interest rates paid), and lines of credit with a higher limit also pay a higher interest rate. The size of these coefficients show that product characteristics are very important in explaining the fees and rates paid for these products.

Table 10 shows the estimation of the credit limit equation. We find that proximity reduces credit to households, but that an increase in the number of competitors increases it for lines of credit, and reduces it for credit cards. One interpretation of this result is that lines of credit are products that are more sensitive to competition, and therefore financial institutions compete to offer better lending terms. We also find that cards with credit protection and rewards typically have larger limits. Additionally, after controlling for aspects such as total assets and other demographic characteristics, we find that households that are employed and have difficulties in paying their debt (although the effect is very small) have larger limits.

For lines of credit, we find that a longer relationship with the financial institution increases the credit provided, which is intuitive. Households that have difficulties paying their debt have lower access to credit in the future. Households with employed members also have a positive effect on the credit limit, with a similar effect to credit cards. By comparing this result with the previous result of the effect on the fees/rates paid, we find that the length of the relationship increases the amount of credit offered, but it also makes it more expensive.

In summary, consistent with some of the main results found in the existing literature for the case of corporate clients and personal loans, our results show that proximity facilitates credit as adoption rates are higher, but also makes fees more expensive for accounts and credit cards, and reduces the loan rates for lines of credit, and credit limits tend to be smaller. On the other hand, a higher number of competitors tends to reduce fees and rates, and increases the provision of credit for lines of credit. These results provide evidence on the importance of the different channels

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 the bootstrap method (for structural model).

	Credit card limits		**Line of credit limits**	
Variable	Structural	OLS	Structural	OLS
Card protection	0.13278	0.13351		
	(0.01285)	(0.03681)		
Rewards	0.39125	0.39202		
	(0.04103)	(0.04162)		
Fixed rate			-0.38062	-0.38156
			(0.03883)	(0.03492)
Secured			0.73580	0.73394
			(0.07046)	(0.02998)
Difficulty debt	0.00269	0.01230	-0.00711	-0.01647
	(0.00473)	(0.00672)	(0.00241)	(0.00676)
Employed	0.20333	0.20618	0.18374	0.18374
	(0.01867)	(0.07316)	(0.01815)	(0.07660)
Length relationship			0.02946	0.01289
			(0.00501)	(0.01190)
Close provider	-0.08274	-0.07993	-0.10402	-0.10713
	(0.00814)	(0.05602)	(0.00936)	(0.05912)
Number of competitors	-0.00951	0.00783	0.04212	0.02513
	(0.00929)	(0.01057)	(0.00639)	(0.00981)
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	11295	11295	4390	4390

of price discrimination (i.e. transportation costs or monitoring costs). Controlling for all these aspects, households 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 in explaining the fees/rates and credit limits of these products.

Our results suggest that some products may be more contestable than others, as they are less sensitive to the market presence of additional competitors. In the case of these products, the actual presence of financial institutions in a given market is not enough to affect competitive outcomes. In other words, these products are more contestable.

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 the bootstrap method.

Variable	Accounts	Credit cards	Lines of credit
σ_F or σ_R	2.00799	2.50491	2.37805
	(0.20148)	(0.23389)	(0.24863)
σ_L		1.71613	0.99006
		(0.16990)	(0.09941)
$\rho_{\pi,D}$	-0.70440	0.06271	0.11744
	(0.07881)	(0.02832)	(0.02910)
$\rho_{\pi,F}$ or $\rho_{\pi,R}$	-0.09483	0.78752	0.13708
	(0.01888)	(0.07460)	(0.01791)
$\rho_{\pi,L}$		0.37774	-0.39337
		(0.04161)	(0.04565)
$\rho_{D,F}$ or $\rho_{D,R}$	-0.09989	0.52975	0.39453
	(0.01137)	(0.04921)	(0.03900)
$\rho_{D,L}$		0.44508	-0.23311
		(0.04680)	(0.02479)

6 Effect of bank branch closures

We study the effect of closure of bank branches by financial institutions by considering a counterfactual experiment where we reduce the number of bank branches in Canada and evaluate the probability of adoption of each financial product by each individual in our database. The closure of bank branches affects the adoption of financial products because proximity indicators of branches to each individual are reduced, and this reduces the probability of adoption. In order to understand the effect of branch closures, we randomly eliminate bank branches in all the markets in Canada, recalculate the variable $Close_{i,b,t}$ and estimate the probability of adoption in equations (6), (10) and (13).

In a market environment where banks close branches and develop mobile banking services or other innovative technology-based financial services (i.e. FinTech), individuals need to have the knowledge and skills required to use these digital financial services. The literature on digital divide (Norris, 2001; Rice and Katz, 2003; Billon *et al.*, 2009) shows that factors such as age, income or education highly affect the ability of individuals to use these services. Therefore, in our counterfactual experiment we only consider the case of individuals that are likely to be affected by branch closures because they are more dependent on physical branches and are less likely to be able to substitute the physical branch by online banking services.

Figure 1: Adoption of financial products for old individuals and change in proximity indicators due to branch closures

This diagram shows the relative change of adoption of financial products for individuals older than 45 vs changes in branch proximity indicators due to bank branch closures. The graph is generated by randomly eliminating bank branches in all provinces in our database and simulating adoption indicators by product in individuals who are older than 45. The observed case for 2006 is at the extreme right of the graph (proximity indicator equal to 0.7). For comparison reasons, we normalize to one the indicators of adoption for every product.

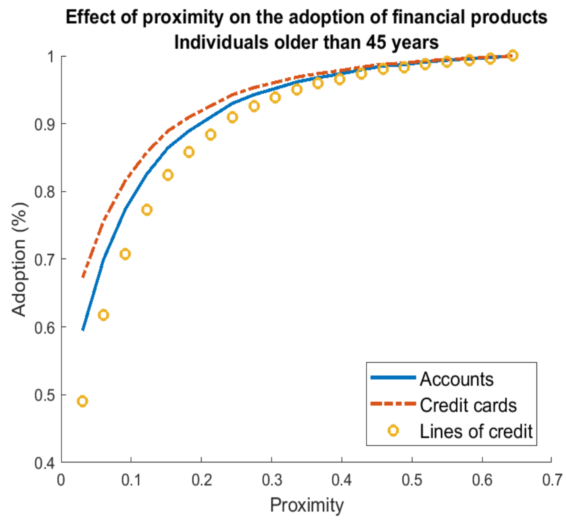


Figure 2: Adoption of financial products for low-income individuals and change in proximity indicators due to branch closures

This diagram shows the relative change of adoption of financial products for individuals with low income vs changes in branch proximity indicators due to bank branch closures. The graph is generated by randomly eliminating bank branches in all provinces in our database and simulating adoption indicators by product in individuals who have an annual income of less than 30,000 Canadian dollars. The observed case for 2006 is at the extreme right of the graph (proximity indicator equal to 0.7). For comparison reasons, we normalize to one the indicators of adoption for every product.

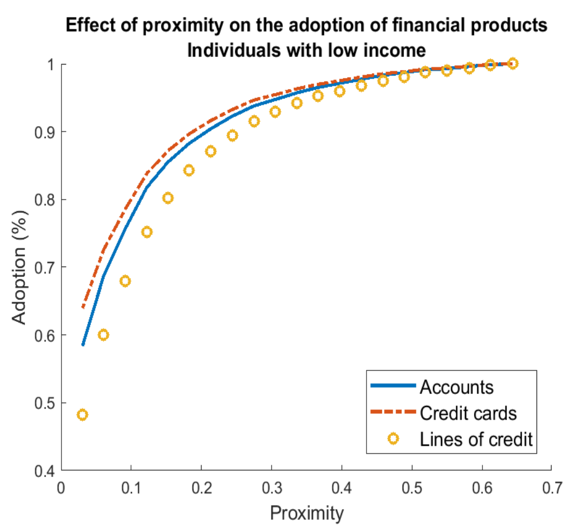


Figure 3: Adoption of financial products for uneducated individuals and change in proximity indicators due to branch closures

This diagram shows the relative change of adoption of financial products for individuals without a university degree vs changes in branch proximity indicators due to bank branch closures. The graph is generated by randomly eliminating bank branches in all provinces in our database and simulating adoption indicators by product in individuals who do not have a university degree. The observed case for 2006 is at the extreme right of the graph (proximity indicator equal to 0.7). For comparison reasons, we normalize to one the indicators of adoption for every product.

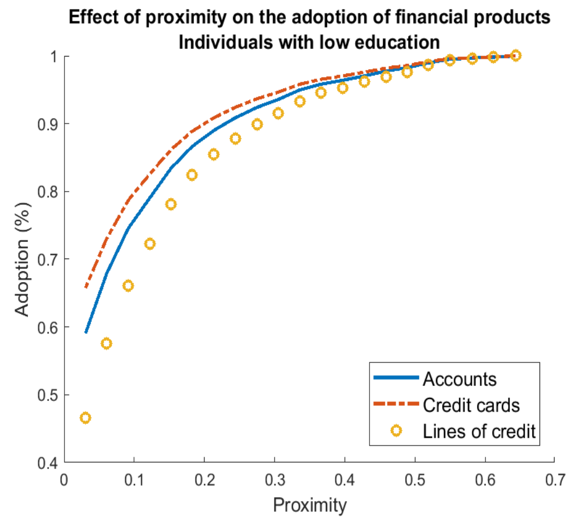
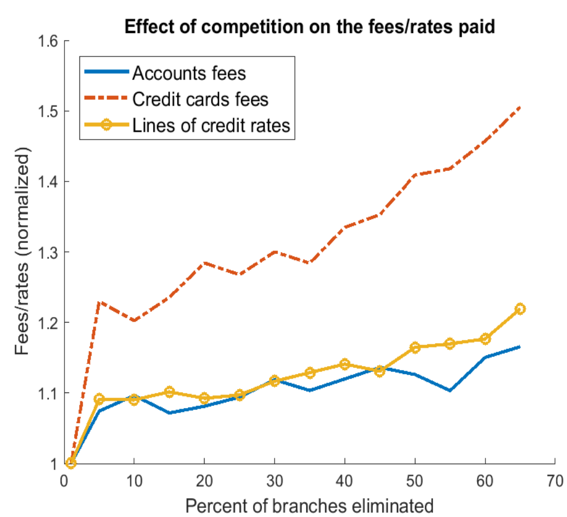


Figure 4: Fees and rates paid for financial products and change in proximity indicators due to branch closures

This diagram shows the relative change in the fees and rates paid for financial products and changes in branch proximity indicators due to bank branch closures. The graph is generated by randomly eliminating bank branches in all provinces in our database and simulating adoption indicators by product. The observed case for 2006 is at the extreme left of the graph (0 branches eliminated). For comparison reasons, we normalize to one the observed fees and rates.



We present the results for the case of elderly individuals, low-income individuals and uneducated individuals in Figures 1, 2 and 3. The extreme right of the figures show the current observed case, where we normalize adoption to 1 for all three products considered. The figures show that as banks close branches, credit cards are the products the least affected by the elimination of physical branches (followed by accounts and lines of credit). In the extreme case where most branches are closed, adoption of lines of credit would be reduced by more than 50% (compared to the observed case), whereas adoption of credit cards from financial institutions would be reduced by about 35%. The effect of proximity is non-linear, and we only start observing a relevant effect in the adoption levels when proximity is roughly reduced by 50% (from 0.7 to 0.35).[12]

In Figure 4, we show how the fees and rates for each financial product change with the closures of bank branches. As previously discussed, fees and rates depend on proximity indicators (variable $Close_{i,b,t}$), and on competition indicators ($Number_{i,t}$). As we reduce the number of branches in the markets, both competition and proximity indicators are simultaneously reduced. Figure 4 shows the joint effect of reducing proximity and competition on the fees and rates paid as branches are closed. For the case of accounts, the closure of branches reduces the level of competition, which increases fees, but geographic proximity is also reduced, which has an opposite effect on fees. Although both variables have opposite effects, the simulations show that the competitive effect is more important and account fees tend to increase as branches are closed.

In the case of lines of credit, the estimated effect of proximity is negative. The effect of competition is negative, but smaller than the case of bank accounts. Therefore, the effect of proximity and competition reinforce each other for lines of credit, which gives a total effect that is similar to the case of bank accounts. Finally, we find a much larger effect for credit cards than for accounts and lines of credit. This is not surprising because the estimated coefficients in the model suggest that the effect of competition on credit card fees is much larger than in bank accounts (and lines of credit), and the effect of geographic proximity in credit card fees is similar to the case of bank accounts. Therefore, the total effect of branch closures is significantly larger in credit cards than in the other two financial products.

7 Conclusion

In this paper, we quantify the effect of competition and geographic proximity of financial institutions on the adoption of bank accounts, credit cards and lines of credit by households. We propose a methodology that models the strategic decision of market presence of each financial institution and considers the household-level decision to adopt financial products. Our methodology permits us to exploit two dimensions of competition analysis that are rarely considered together:

[12]Note that in our analysis we only consider credit cards provided by financial institutions. In practice, this product is probably the financial product that has the strongest presence of non-bank competitors, such as retailers or airlines. In practice, households could have access to credit cards from other non-bank providers as bank branches are closed.

market complexity (distance, sources of market power, and market presence strategies) and customer/product complexity (price discrimination and product differentiation).

Our results show that market presence tends to reduce fees and rates, and increase the credit supplied. We also find that proximity with the financial institution that provides these products increases the likelihood of adopting the financial product, increases the fees paid, and restricts the credit supplied. A counterfactual policy experiment helps us understand in greater detail the effect of branch closures on the adoption and fees paid for these products for several segments of population that are more likely to be affected. Despite the emergence of internet and mobile banking, our results show that physical branches and proximity still matter in understanding the competitive landscape in the retail banking industry. Barriers to entry are significant, and therefore some of the products analysed are not contestable, as they are affected by the effective presence of financial institutions in the market.

Policymakers keep professing their desire to increase competition in the banking industry. Our results show evidence of the importance of market presence in affecting 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.

AGUIRREGABIRIA, V., CLARK, C. R. and WANG, H. (2015). Diversification of geographic risk in retail bank networks: Evidence from bank expansion after the Riegle-Neal Act. *forthcoming, The Rand Journal of Economics*.

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.

—, — and — (2017). Search frictions and market power in negotiated price markets. *forthcoming, Journal of Political Economy*.

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

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

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.

BERG, T., SAUNDERS, A. and STEFFEN, S. (2016). The total cost of corporate borrowing in the loan market: Don't ignore the fees. *The Journal of Finance*, **71** (3), 1357–1392.

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.

BILLON, M., MARCO, R. and LERA-LOPEZ, F. (2009). Disparities in ict adoption: A multidimensional approach to study the cross-country digital divide. *Telecommunications Policy*, **33** (10), 596–610.

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.

DRAGHI, M. (2017). Hearing of the Committee on Economic and Monetary Affairs of the European Parliament. *Introductory statement at the ECON committee of the European Parliament, Brussels, 29 Ma*.

DVOŘÁK, P. and HANOUSEK, J. (2009). Paying for banking services: What determines the fees? *Working Paper, CERGE-EI*.

EGAN, M., HORTAÇSU, A. and MATVOS, G. (2017). Deposit competition and financial fragility: Evidence from the US banking sector. *American Economic Review*, **107** (1), 169–216.

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.

ERGUNGOR, O. E. (2010). Bank branch presence and access to credit in low-to moderate-income neighborhoods. *Journal of Money, Credit and Banking*, **42** (7), 1321–1349.

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

FERRARI, S., VERBOVEN, F. and DEGRYSE, H. (2010). Investment and usage of new technologies: Evidence from a shared ATM network. *American Economic Review*, **100** (3), 1046–79.

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.

GOETZ, M. R., LAEVEN, L. and LEVINE, R. (2013). Identifying the valuation effects and agency costs of corporate diversification: Evidence from the geographic diversification of US banks. *The Review of Financial Studies*, **26** (7), 1787–1823.

—, — and — (2016). Does the geographic expansion of banks reduce risk? *Journal of Financial Economics*, **120** (2), 346–362.

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.

HOTELLING, H. (1929). Stability in competition. *The Economic Journal*, **39** (153), 41–57.

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.

NORRIS, P. (2001). *Digital divide: Civic engagement, information poverty, and the Internet worldwide*. Cambridge University Press.

PANZAR, J. C. and ROSSE, J. N. (1987). Testing for" monopoly" equilibrium. *The Journal of Industrial Economics*, pp. 443–456.

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. (2017). Cultural affinity, regulation and market structure: Evidence from the canadian retail 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.

RICE, R. E. and KATZ, J. E. (2003). Comparing internet and mobile phone usage: Digital divides of usage, adoption, and dropouts. *Telecommunications Policy*, **27** (8), 597–623.

SALOP, S. C. (1979). Monopolistic competition with outside goods. *The Bell Journal of Economics*, pp. 141–156.

SUSSMAN, O. and ZEIRA, J. (1995). Banking and development. *CEPR Discussion Paper 1127*.

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 account balance 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.
- Financial institution geographic presence:
  - Number of financial institutions with presence in a 10-km radius around the household.
  - Close: Indicator variable equal to 1 if the financial institution has presence in a 10-km radius 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 household.
- 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 regularly uses 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 details

B.1 Algorithm

We explain in 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 $\Gamma = \{\varepsilon_s^{ac}, \varepsilon_s^{cc}, \varepsilon_s^{loc}, \varepsilon_s^{\pi}\}_{s=1}^{s=S}$
3. Transform the set Γ in another set $\widetilde{\Gamma}$ that is distributed following the variance-covariance matrix Σ^p
4. Calculate $D^{*\,p}_{i,b,t}$ for all $\widetilde{\Gamma}$. Find the set of draws $\widetilde{\Gamma}_D$ such that $D^{*\,p}_{i,b,t} > 0$
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 $Fees^p_{i,b,t}$ and $Limit^p_{i,b,t}$ for the set of errors $\widetilde{\Gamma}_{\cap}$
8. Estimate $f(Fees^p_{i,b,t}, Limit^p_{i,b,t} / D^p_{i,b,t} = 1, Close_{i,b,t}, N_i)$ with a kernel density estimator using values from previous stage.

Using this algorithm, we construct the simulated likelihood from Eq. (21), which we maximize using a state-of-the-art optimizer (MATLAB/KNITRO) and the computing cluster of the Bank of Canada.

B.2 Identification

Identification of the parameters in our market presence 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 financial institutions. This distance variable is an appropriate measure that accounts for the existing branch-network economies of density, and similar distance measures have been used in the literature. As shown in Goetz *et al.* (2013), Aguirregabiria *et al.* (2015) and Goetz *et al.* (2016), banking services exhibit economies of density because banks usually have greater familiarity with the economic conditions of closer markets and face lower costs for establishing and maintaining branches than in farther markets. We use the geographical presence of the banks in 1972 to construct this variable, since it occurs before the financial deregulations that permitted the formation of universal banks. We use the location of the headquarters (which we assume is the largest city in the province) and generate the minimum distance from any market to a headquarters.[13] We expect significant inertia in the subsequent expansion of financial institutions over the decades, therefore this variable should be correlated with the geographic presence in 2006. Also, this variable varies across markets for most financial institutions considered.

Other variables we use are the total asset size of every financial institution (which does not vary across markets), and regional (provincial) size of every financial institution (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 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 fully included in our database. Urban markets, which are larger and more profitable than the rural markets, were probably covered by financial institutions much earlier than rural markets. Therefore, regional size can also be considered, to a certain extent, an exogenous variable.

The second strategy we use to identify the market-presence 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 the presence of multiple equilibria. There have been a number of solutions 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.[14] To identify the equilibrium we use, we compute all possible equilibria using Eq. (3) and then select the most efficient with probability one.[15]

Regarding the rest of the equations of the model for the outcome equations, our model is very similar to 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 to every equation. Usage of payments in different payment channels is a variable that

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

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

[15]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.

affects the demand for a bank account, but not the fees paid. Intuitively, this assumes that financial institutions 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, especially for new clients. We also use 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 financial institutions. We also use risk variables that affect the limits granted for financial products by financial institutions. Credit limits granted by financial institutions should highly depend on the riskiness of the clients, therefore unemployment and difficulty to pay the debt should be particularly related with these limits, but not with the demand for these products.

C Figures

Figure C.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.

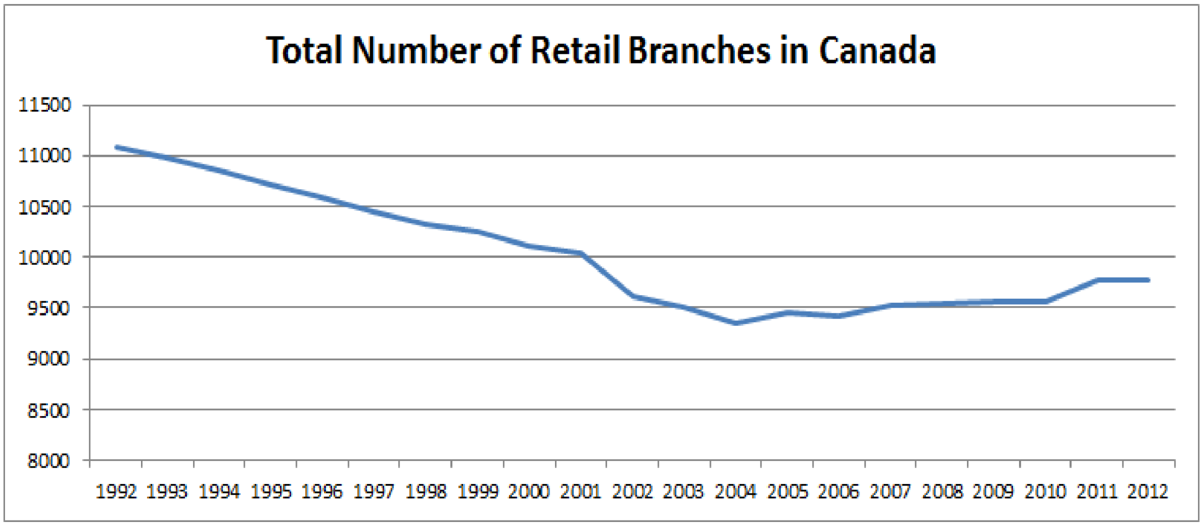


Figure C.2: Market presence and population

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

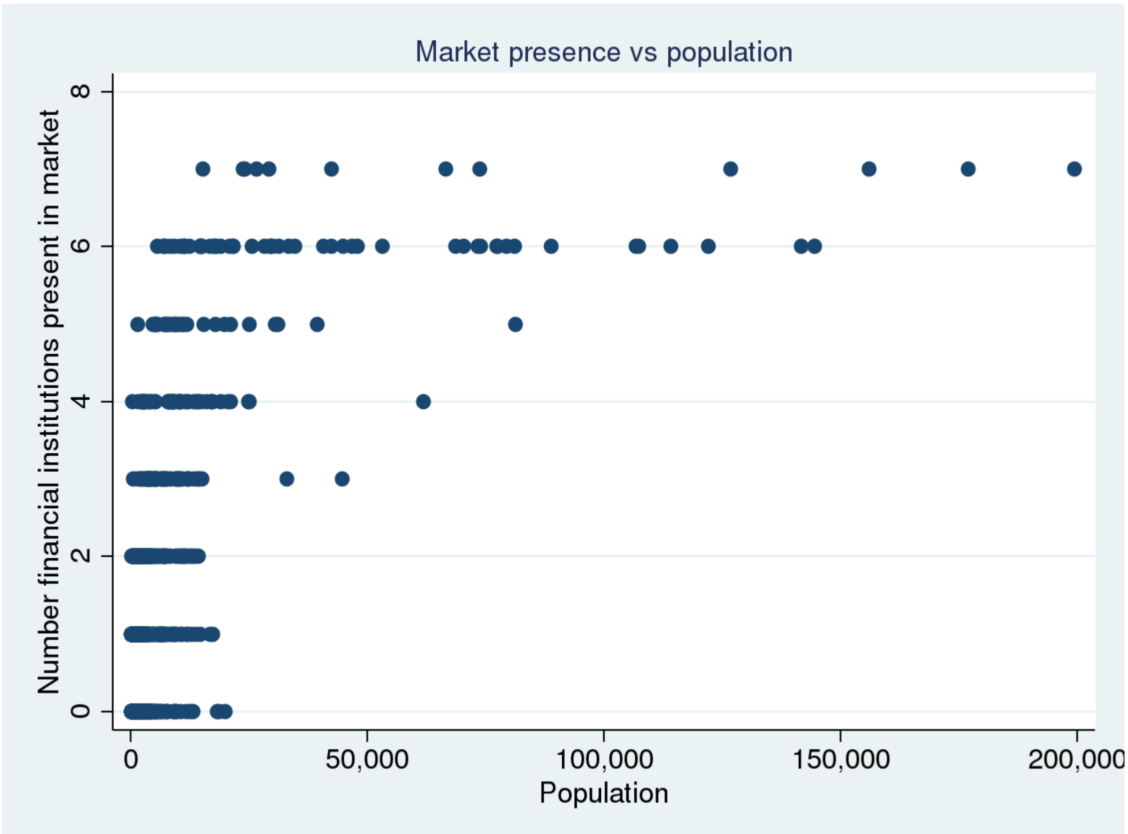


Figure C.3: Example of market: Moose Jaw in Saskatchewan (I)

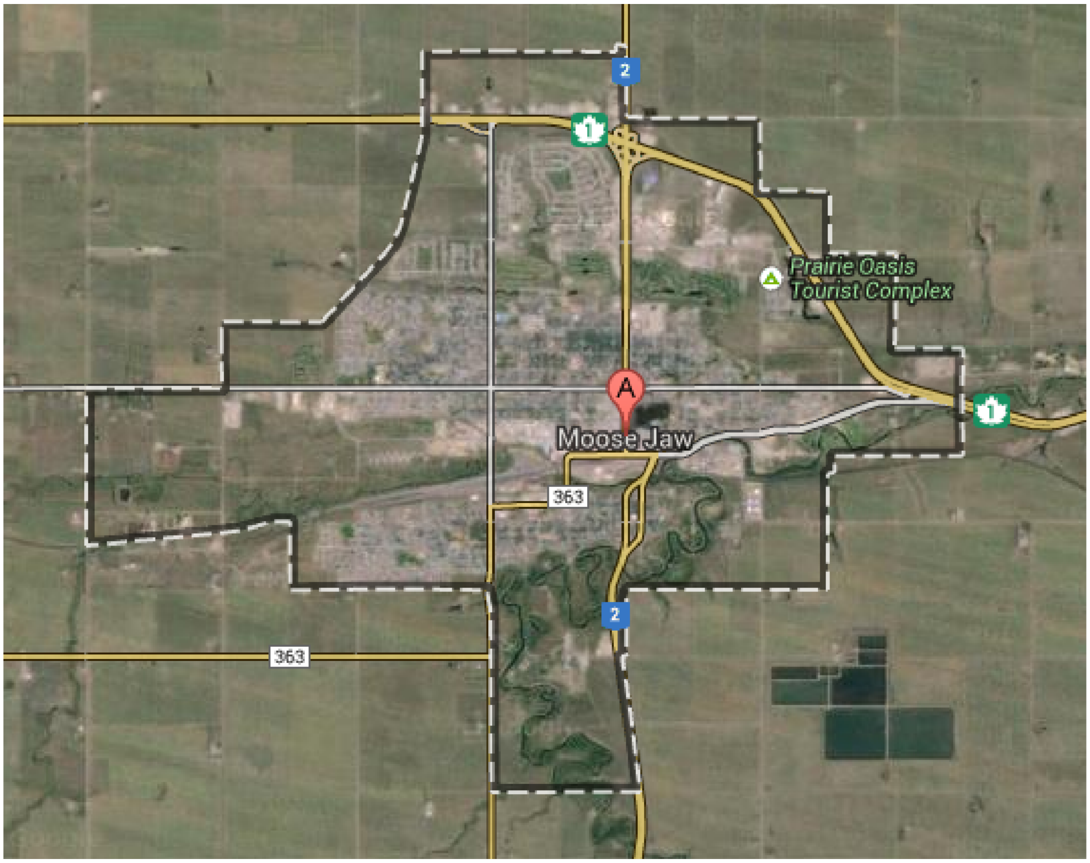


Figure C.4: Example of market: Moose Jaw in Saskatchewan (II)

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

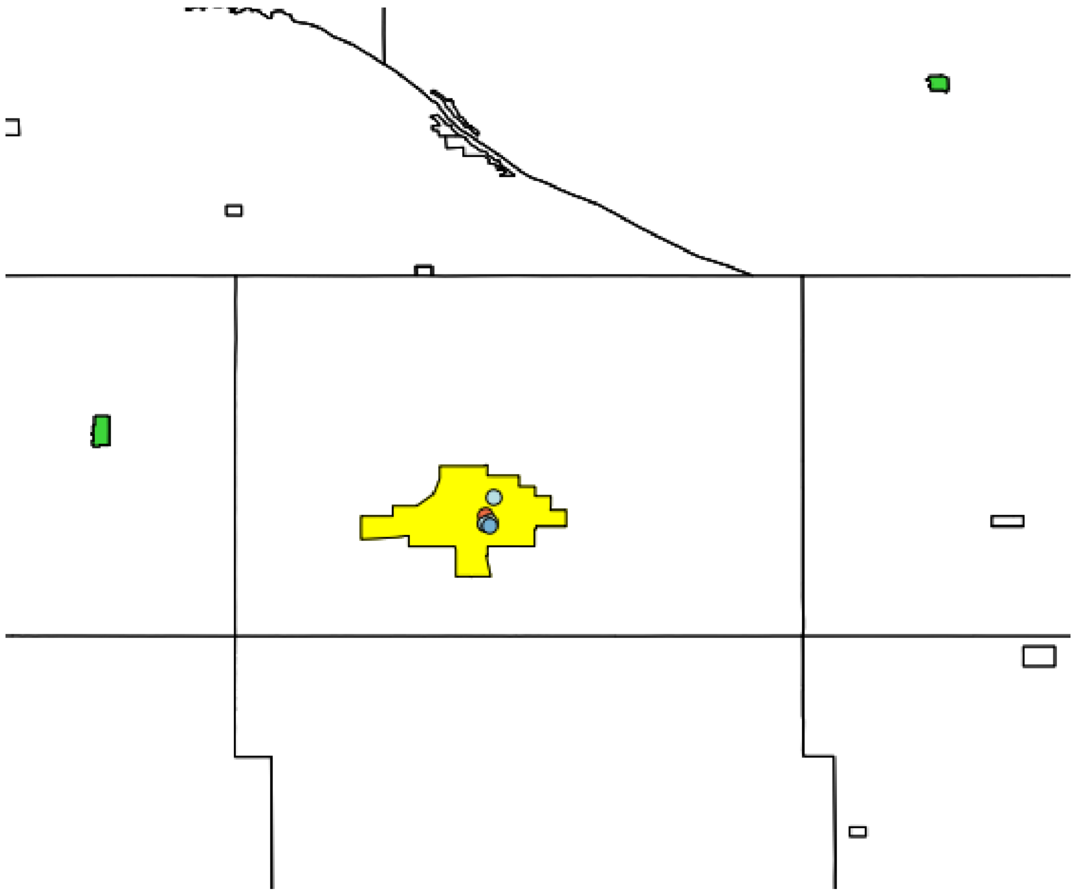


Figure C.5: Markets considered and CFM households

This diagram shows the selection of markets and CFM households included in our estimation, and the calculation of proximity indicators.