Staff Working Paper/Document de travail du personnel—2023-2

Last updated: January 4, 2023

Geographical and Cultural Proximity in Retail Banking

by Santiago Carbo-Valverde,[1] Hector Perez-Saiz[2] and Hongyu Xiao[3]

[1] Universitat de Valencia
Valencia, Spain
santiago.carbo@uv.es

[2] International Monetary Fund
Washington, DC
hperez-saiz@imf.org

[3] Currency Department
Bank of Canada
HXiao@bank-banque-canada.ca

DOI: https://doi.org/10.34989/swp-2023-2 | **ISSN** 1701-9397

Acknowledgements

This is a revised version of a previously circulated manuscript titled "Competition, Geographic Proximity and Pricing in the Retail Banking Industry." For all their valuable comments and suggestions, we wish to thank Victor Aguirregabiria, Jason Allen, Christoph Bastenm, Wilko Bolt, Evren Damar, Hans Degryse, Hans Gersbarch, Kim Huynh, Kari Kemppainen, Mark Manuszak, Ouarda Merrouche, Steven Ongena, Francisco Rodriguez, Alex Shcherbakov, Larry Schembri, Andre Stenzel, John Wilson and participants at presentations at the Bank of Canada, Canadian Economics Association, 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, its Executive Board, or IMF management.

Abstract

This paper measures how both geographical and cultural proximity of bank branches affect household credit choice and pricing. We examine both types of proximity jointly to separately identify the importance of soft information versus alternative mechanisms. Using a detailed household-level database for Canada, we find that both geographical and cultural proximity increase consumer credit by reducing the cost of obtaining soft information. Furthermore, soft information obtained via the two types of proximity can be either substitutes for or complements to each other, with complementarity being more likely for products that require high levels of ex-ante screening. Overall, our results suggest that ongoing branch consolidation, happening in many countries, may lead to lower financial inclusion, especially in culturally diverse neighbourhoods.

Topics: Credit and credit aggregates, Financial institutions, Financial services
JEL codes: D82, D83, G20, G21, R22, Z10, Z13

Résumé

Ce document mesure l'incidence de la proximité géographique et culturelle des succursales bancaires sur les choix des ménages en matière de crédit et sur les prix qui leur sont offerts. Nous étudions les deux types de proximité simultanément pour cerner séparément l'importance des informations difficilement quantifiables par rapport à d'autres mécanismes. En nous basant sur des données détaillées sur les ménages canadiens, nous montrons que la proximité tant géographique que culturelle accroît le crédit octroyé aux consommateurs parce qu'elle réduit les coûts liés à l'obtention d'informations difficilement quantifiables. De plus, ces informations obtenues grâce aux deux types de proximité peuvent se substituer les unes aux autres ou se compléter. La complémentarité des informations est plus probable pour les produits qui exigent une vérification préalable du crédit plus complète. Dans l'ensemble, nos résultats donnent à penser que le regroupement des succursales qui se produit en ce moment dans de nombreux pays pourrait diminuer l'inclusion financière, surtout dans les quartiers où il y a une grande diversité culturelle.

Sujets : Crédit et agrégats du crédit, Institutions financières, Services financiers
Codes JEL : D82, D83, G20, G21, R22, Z10, Z13

1 Introduction

Geographical and cultural proximity between households and bank[1] branches are important drivers of choice and pricing of retail banking products. Greater geographic proximity reduces transportation costs (Hotelling, 1929) and facilitates the flow of soft information (Liberti and Petersen, 2019), a crucial component of lending relationships between commercial banks and potential customers (Petersen and Rajan, 1994). Greater cultural proximity can also lower the cost of obtaining soft information about potential borrowers (Fisman *et al.*, 2017), as well as potential taste-based discrimination by loan officers (Becker, 1957). Since soft information acquisition can be facilitated by both geographical and cultural proximity, interaction effects could occur and be significant, as information obtained via one type of proximity may substitute or complement information obtained from the other type.

This paper quantifies the joint effect of both geographical and cultural proximity between bank branches and households on consumer financial product choice and pricing, focusing on the effects of the proximity interaction term. Examining both types of proximity jointly allows us to separately identify soft information-related versus alternative mechanisms because it is less plausible that transportation costs and taste-based discrimination have interaction effects since they are limited to one type of proximity only. Our laboratory is Canada, a culturally segmented country with two official languages (English and French), which provides significant cultural diversity across local markets. We jointly estimate a set of household-branch pair-level choice and pricing equations using data from a detailed survey on household financial choices, as well as a comprehensive branch location database. To correct for the endogeneity of branch entry decisions, we compute firms' equilibrium entry strategies, where latent profit depends on market- and firm-level variables, as well as competition from other potential entrants.

We find that both geographical and cultural proximity positively affect credit card and line of credit choices by households. In fact, a bank having a geographically close or culturally close branch increases the likelihood of households possessing credit cards and lines of credit from that bank by more than 70% and 100%, respectively, at the sample mean, with a slightly larger effect for geographical proximity. The proximity interaction effect is also highly statistically significant, suggesting that both types of proximity affect household choice via a reduction in the cost of soft information provision. Furthermore, we find that for credit card choice, the two types of proximity are substitutes for each other in obtaining soft information, with the proximity interaction term being significantly negative. However, for line of credit choice, the two types complement each other, with the interaction effect being significantly positive.Furthermore, we find that soft information obtained via the two types of proximity are substitutes for credit card choice, with the proximity interaction term being significantly negative, and complements for line of credit choice,

[1]For the purposes of this paper, we use the term "bank" interchangeably with "financial institution" (FI), meaning all FIs taking deposits.

with the interaction effect being significantly positive. These opposing effects are consistent with banks needing less ex-ante information on the creditworthiness of the potential borrower for credit cards than they need for lines of credit. Finally, these effects are not likely to be explained by pure pricing effects since the proximity effect on prices is largely insignificant.

Our results show that, despite the emergence of Internet and mobile banking, physical branches and proximity still matter for household choice of financial services. In a counterfactual where we increase the cost of branch entry to reduce the presence of branches, we show that the likelihood of households possessing credit cards or lines of credit decreases, with the latter falling by over 30% when most branches are closed.

We make several contributions in this research. To our knowledge, this is the first article jointly estimating the effects from both measures of proximity in retail banking. By examining the proximity interaction effects, we find new evidence that lowering the cost of obtaining soft information is a significant mechanism driving the effects of both geographical and cultural proximity, while previous literature has provided mixed evidence on this score (Degryse and Ongena, 2005; Agarwal and Hauswald, 2010; Fisman *et al.*, 2017; D'Acunto *et al.*, 2021). Furthermore, we advance the state of knowledge by finding that the effects of proximity interaction differ across banking product types, likely driven by differences in soft information acquisition needs prior to loan approval. In addition, we contribute to the banking literature by focusing on the effects of proximities on consumer financial products, complementing the previous literature where the focus has been on small business lending (Rehbein *et al.*, 2020). Methodologically, we propose a rich structural empirical model that includes bank entry decisions; therefore, this is also the first study that estimates product choice, prices, and limits together, allowing the bank to flexibly adjust its offerings (Manuszak and Moul, 2008).

This paper is divided into six sections. Section 2 discusses the literature and, specifically, its theoretical predictions. Section 3 examines the data that we use. Section 4 describes the empirical model. Section 5 discusses the empirical results and presents the counterfactual. Finally, Section 6 concludes.

2 Theoretical Predictions

2.1 Geographical Proximity

There is a rich literature exploring the effects of geographical proximity on financial product choice and pricing. According to Degryse and Ongena (2005), there are two main mechanisms linking the two: transportation costs and costs of soft information acquisition.

With lower geographical proximity, transportation costs increase due to higher monetary and time costs traveling between the borrower location and the branch location. In the case of consumer lending, this cost is likely to be mostly incurred by the borrower visiting the local branch.[2] In a simple Hotelling model with uniform pricing, this would increase the borrower's likelihood of adopting a financial product from a closer branch. This implies that banks' market power over the borrower is greater with increased geographical proximity, as the distance between the borrower and competing banks gets correspondingly greater. Hence, the bank may adopt spatial price discrimination tactics to maximize its profit.

A fundamentally different mechanism relating geographical distance to financial product choice and pricing is the cost of soft information acquisition. Soft information is information that is difficult to summarize in a numeric score, requiring knowledge of its context to be fully understood (Liberti and Petersen, 2019). Soft information is especially valuable for retail banks because it is both costly to acquire and not easily transmittable once acquired.[3]

From established models in finance (Hauswald and Marquez, 2006), signal precision about borrower quality decreases with increasing distance. In other words, the closer a bank branch is to a potential borrower, the less it needs to spend to acquire the same amount of soft information about said borrower. Given adverse selection in retail banking where riskier borrowers are willing to accept higher loan prices (Stiglitz and Weiss, 1981; Agarwal *et al.*, 2010), financial institutions (FIs) are willing to extend loan offers to ex-ante lower quality borrowers at closer distances, increasing the likelihood of households choosing their product. Since closer geographical proximity decreases a bank's marginal screening cost, it could use the cost savings to attract closer borrowers by lowering prices. On the other hand, competing banks also face increasing adverse selection problems when approaching borrowers closer to the focal bank, which increases the latter's market power over closer borrowers and may lead to spatial price discrimination.[4]

Both lower transportation costs and soft information acquisition costs are mechanisms that may affect pricing differently depending on banks' pricing strategies. Market power, whether due to transportation costs or adverse selection, is greater near the location of the branch. Hence, spatial price discrimination would mean that financial product prices are higher near the branch and lower farther away. On the other hand, pricing could also be based on marginal costs. Marginal costs are higher when consumers are further away from the branch, and this would imply that prices are higher for those further away from the branch.

The existing literature focused on small business lending has found support for both types of pricing strategies. On the one hand, Degryse and Ongena (2005) and Agarwal and Hauswald (2010) find empirical support for spatial price discrimination due to both transportation costs and

[2]Unlike small business lending, which the literature has traditionally focused on, it is unlikely that bank employees will visit consumers to monitor them, since that would not be a cost-effective use of their time.

[3]While lower monitoring costs is another mechanism that can link geographical proximity to financial product choice and pricing (Degryse and Ongena, 2005), continuous monitoring is less likely for consumer lending products than for business loans, so we ignore them here.

[4]While common prices and rates are posted nationally for many products such as mortgages, there is evidence that actual prices and contract terms are determined through a search and negotiation process (Allen *et al.*, 2014). Similar to mortgages, we assume that prices for credit cards and lines of credit are negotiable.

adverse selection. On the other hand, Knyazeva and Knyazeva (2012) and Bellucci *et al.* (2013) find support for marginal cost pricing. There is little evidence on the impact of geographical proximity on financial product choice and pricing in consumer lending.

Table 1: Theoretical Models Linking Proximity and Product Choice/Pricing

This table lists models relating the effects of geographical and cultural proximity on financial product choice and pricing.

Models/Channels	**Impact on Likelihood of Choosing Product**		
	Greater Geographical Proximity	**Greater Cultural Proximity**	**Proximity Interaction**
Transportation Costs	positive	none	none
Taste-Based Discrimination	none	positive	none
Soft Information Costs	positive	positive	positive/negative
	Impact on Pricing with Price Discrimination		
Transportation Costs	positive	none	none
Taste-Based Discrimination	none	none	none
Soft Information Costs	positive	positive	positive/negative
	Impact on Pricing with Marginal Cost Pricing		
Transportation Costs	negative	none	none
Taste-Based Discrimination	none	negative	none
Soft Information Costs	negative	negative	positive/negative

2.2 Cultural Proximity

There is also established literature on the relationship between cultural proximity and financial decisions (Grinblatt and Keloharju, 2001), though the literature linking culture and retail banking product choice and pricing decisions is more recent, but growing rapidly. There are two main channels discussed: taste-based discrimination and soft information acquisition costs.

First, cultural proximity can lead to preference-based (taste-based) discrimination. In this channel, an individual can prefer to work with individuals from the same culture, even though it might be costly for them to do so (Fisman *et al.*, 2017). For example, a loan officer may reduce their performance (and hence career progression) in order to indulge their discriminatory preferences and lend to marginal borrowers from their own culture. At the same time, this expectation would be known by potential borrowers, who would expect more favorable treatment with culturally close lenders. Preference-based discrimination would lead to higher choice likelihood and lower pricing for culturally aligned borrowers, *ceteris paribus*.

Second, cultural proximity can mitigate information frictions in lending by reducing the cost of obtaining soft information, thus improving the precision of the signal that the officer obtains about a potential borrower's creditworthiness (Cornell and Welch, 1996). Fisman *et al.* (2017) showed that benefits of cultural proximity are distinct from and additive to those that are derived from a borrower's observable track record with the lending institution or those that come from repeated interaction between officer and borrower. In other words, cultural proximity provides additional soft information that cannot be obtained through repeated transactions alone.

By lowering costs of acquiring soft information, cultural proximity increases the likelihood of households choosing financial products from a culturally close FI, *ceteris paribus*. However, it may lead to one of two pricing outcomes. If financial product pricing is based on marginal costs, then cultural proximity should lead to lower pricing for culturally aligned borrowers. Alternatively, if pricing is based on willingness-to-pay, then cultural proximity could paradoxically lead to higher pricing for culturally aligned borrowers. This is because the culturally aligned bank would obtain more precise signals on the creditworthiness of culturally aligned borrowers. Competing banks would face an adverse selection problem and be less likely to lend to these borrowers. The culturally aligned institution would exercise this greater market power and increase pricing for the latter.

D'Acunto *et al.* (2021) show empirically that culture-based discrimination in lending exists and could increase lending to same-culture borrowers despite them having higher risk, while Fisman *et al.* (2017) present evidence that cultural proximity increases personal loan lending by facilitating soft information acquisition, which lowers lending risk. The authors also find that loan prices decrease with cultural proximity, suggesting that banks employ a marginal cost pricing strategy.

2.3 Proximity Interaction

Is there an interaction effect between cultural and geographical proximity on financial product choice and pricing? We hypothesize that the interactive effect between cultural proximity and transportation costs on both product choice and pricing is expected to be zero, since they come from completely different mechanisms. Similarly, we expect that taste-based discrimination would have a zero interaction effect with both mechanisms driving the effects of geographical proximity.

On the other hand, the soft information obtained via closer geographical proximity and that obtained through closer cultural proximity may complement or substitute each other, which could lead to a non-zero interactive effect between geographical and cultural proximity on financial product choice and pricing. We conjecture that if the soft information obtained via closer distance is a substitute for that obtained from cultural alignment, then the choice/pricing effect for more distant borrowers would be mitigated by cultural alignment (interaction effect in the same direction as the cultural effect, opposite direction as the more distant geographical effect). On the other hand, if the soft information obtained via closer distance is complementary to that obtained from cultural alignment, then the pricing effect for more distant borrowers would be exacerbated by cultural alignment (interaction effect in the same direction as the more distant geographical effect, opposite direction to the cultural effect). This would be the case no matter whether marginal cost pricing or willingness-to-pay pricing is practiced.

The only article that we are aware of that studies this subject is Rehbein *et al.* (2020). The authors use US county-to-county Facebook linkages to show that social connections can alleviate the negative effects of between-county physical and cultural distance on the volume of small-business lending and mortgages. Our paper differs by explicitly evaluating the interaction effects of cultural and physical proximity on both product choice and pricing and doing so at the individual household level.

To conclude, Table 1 summarizes the theoretical predictions from the literature. Empirical evidence is mixed with support found for all four mechanisms, as well as both marginal cost and price discrimination pricing strategies.

3 Data

3.1 Constructing Proximity Indicators in Canada

We obtain the cultural origin of FIs by looking at their past. The Canadian retail banking industry is composed mainly of a small number of large banks (Big Six banks)[5] and other institutions (credit unions) that provide a broad range of retail financial products to their customers. Canada's oldest bank, Bank of Montreal, was founded in 1817 by English-speaking merchants. Five of the six largest banks today (BMO, CIBC, BNS, RBC and TD) can trace their history to English-speaking founders, while National Bank was founded and controlled by French speakers. Despite this varied history, all six banks provide universal services across almost all of the provinces, in both English- and French-speaking areas. Credit unions (*caisses populaires* in French), another type of depository institution in Canada, compete with the banks. These FIs are founded on cooperative principles and owned by their members. The largest credit union is Desjardins, created in 1900 to serve French-speaking Canadians (Calomiris and Haber, 2014). Table 2 summarizes the classification of Canadian FIs by cultural origin.

We obtain precise FI branch locations from the 2006/2007 edition of Canadian Financial Ser-

[5] After decades of consolidation, the Canadian retail banking industry is highly concentrated. For example, in 2008, the Big Six banks controlled 98% of total banking system assets and over 80% of the assets in the Canadian financial system.

Table 2: Financial Institution Classification

This table classifies FIs according to type and cultural origin.

Type	Cultural Origin	
	English	**French**
Banks	BMO	National Bank
	BNS	
	CIBC	
	RBC	
	TD	
Credit Unions	Credit Unions	Desjardins Group

vices, which is a comprehensive directory of all Canadian FIs and their branches. This directory is updated annually, and it contains the exact address of each branch, including the six-digit postal code.

All our household information comes from Ipsos Reid's Canadian Financial Monitor (CFM) database, which is a detailed survey on household-level usage of financial services. The survey identifies household location according to postal code, which provides precise location determination.[6] The survey also indicates the language spoken at home by each household, which we use as an indicator for the household's cultural origin.

We construct the binary cultural proximity indicator by comparing each bank's cultural origin with the language spoken by the household at home. We set it equal to 1 if they are the same (i.e., English/English or French/French), and set it to 0 otherwise.

We construct the binary geographical proximity indicator in several steps. First, we convert postal codes to geographical coordinates using the 2006 Postal Code Conversion File (Wilkins and Khan, 2010). We then compute the straight-line/geodesic distance for every potential CFM household-FI branch pair. For each CFM household, if there is an FI branch located within 5 km of its location, the geographical proximity indicator equals to 1 for that household-FI pair. If all branches from an FI are more than 5 km away, then the geographical proximity indicator equals to 0 for that household-FI pair.

From Table 3, we can see that for all in-sample CFM household-FI pairs, 61% are geographically close and 73% are culturally close. Combining the two proximities, just over half of all pairs are both geographically and culturally close.

[6] A six-digit postal code covers a relatively small geographic area. In 2010, Statistics Canada estimated that there were an estimated 830,000 postal codes in active use, corresponding to a population of around 34 million. This implies that on average, there are only 41 people per postal code, which can correspond to a condominium building or a group of houses.

3.2 Household-Level Data

We now turn to detailed data on household-level consumption of financial services. Financial-product data at the household level are obtained from Ipsos Reid's Canadian Financial Monitor (CFM) database for 2007–2010.[7] This database includes a complete overview of all the financial products and services used by about 12,000 Canadian households annually. The CFM database covers most financial products offered to Canadian households, such as credit cards, chequing and savings accounts, insurance products, mortgages, personal loans, lines of credit, bonds, stocks, mutual funds, and so on. The database includes some of the most relevant characteristics of these products, such as fees, interest rates, credit limits, payments usage, and other product characteristics. This database also includes some detailed demographic characteristics of the households, such as age, income, employment, and so on. We also have a complete overview of the total assets available (e.g., real estate, cars, stock, mutual funds, precious metals) and some variables showing general attitudes.

3.3 Market and Sample Selection

We cannot use the full CFM household-FI branch database for estimation because the geographical proximity variable is endogenous due to strategic entry decisions by FIs in each market.[8] To take this endogeneity into account, we estimate a static entry model to obtain FIs' equilibrium entry decisions. Therefore, we need to select markets and CFM households where this entry model can be appropriately used.

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.[9] We obtain market-level data such as population, unemployment, and per capita income from the 2006 census (Statistics Canada, 2006).

Because census subdivisions do not necessarily reflect the boundaries of a market, we manually select rural and small urban isolated census subdivisions to be included in our model, based on well-defined criteria. We obtain 550 markets in total.[10] The geographical proximity variable

[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 onward. A large sample size helps in the identification of the model equations. In addition, the market structure was relatively stable in this period.

[8] We assume that households do not make strategic location decisions based on proximity to bank branches. Indeed, the 2017 American Housing Survey has shown that households are most likely to move for reasons related to living in a better home or neighborhood, and then for job-related or family-related reasons, such as being closer to a new work location or other family members (Ford, 2019).

[9] For instance, Toronto, with a population of more than 2.6 million people, constitutes one census subdivision, and Martensville, SK, a small city with fewer than 8,000 inhabitants, also constitutes one census subdivision.

[10] In particular, we only include census subdivisions that have between 200 and 200,000 individuals to eliminate uninhabited areas and large cities. We then choose census subdivisions that are separated by at least 10 km, and are located more than 50 km away from any major urban centres to avoid commuting patterns confounding our market definition. We then drop rural subdivisions larger than 300 square kilometers in area. Finally, we exclude Indigenous reserves and

is the dependent variable in our entry model, which we identify using a 5-km radius from the centroid of a given census subdivision. If at least one branch of the bank is present within that circle, then we set the dependent variable to 1; otherwise, we set the dependent variable to 0.

We carefully select the households that are consistent with our market-presence model to be included in our empirical model . Figure C.4 in the Appendix shows how these households are selected. For every in-sample census subdivision, we select CFM households that are located within a 5 km radius of the centroid of the census subdivision being considered.

In our empirical model, a unit of observation is a financial product acquired by a household-year for each of the seven FIs that are considered in the market-presence model. Given that there are 8,451 unique household-year observations in our sample and seven FIs, in total we have a sample of $N = 8,451 \times 7 = 59,157$ observations where we observe financial product characteristics (when the household has the product with the FI).

Table 3 provides useful descriptive statistics for the demographic characteristics of the households included in our sample. The average age is fairly high at 54 years old, with a distribution from the early 20s to the late 80s. The average annual income is $64,000, with the range being $8,000 to $250,000. Around 90% of all households have at least one member who is employed, while 77% of all households own at least one house. In summary, this shows that the households in our sample skew older and wealthier than the average household in Canada.

On financial product choice, we can see that credit cards are chosen in 14% of all household-FI pairs, which averages to around one credit card-FI relationship per household. On the other hand, line of credit choice is significantly lower at 5% of all household-FI pairs. Households pay on average $16.84 per year for a credit card. On average, the annual interest rate of a line of credit is 4.94%. The average credit limit for lines of credit is more than four times larger than the average limit for credit cards, signaling that FIs are taking a much larger risk with the former product. Understandably, lines of credit are rarely offered to customers without longer-term banking relationships, with more than 75% of all lines of credit held by households that have a relationship of more than five years with the same FI.

4 Empirical Model

4.1 Methodology

In our methodology, we jointly estimate a set of household-level outcome equations for the fees, rates and credit limits of the financial products considered and a latent profit equation that deter-

markets from Alberta from consideration to avoid potential regulatory confounders. National and provincial market descriptive statistics for the markets considered are shown in Table C.1. 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 that we have selected, Moose Jaw (Saskatchewan), is shown in Figure C.3 in the Appendix.

Table 3: Summary Statistics for Proximity Indicators, Demographic and Product Characteristics

This table shows summary statistics for household-branch proximity indicators, choice indicators, fees/rates, credit limits, and other product characteristics, as well as demographic characteristics of households. Variables are defined in the Appendix. Source: CFM database and bank branches database.

Variable	mean	sd	min	p1	p25	p50	p75	p99	N
Proximity Indicators:									
Geographical Proximity	0.61	0.49	0.00	0.00	0.00	1.00	1.00	1.00	59,157
Cultural Proximity	0.73	0.45	0.00	0.00	0.00	1.00	1.00	1.00	59,157
Proximity Interaction	0.51	0.50	0.00	0.00	0.00	1.00	1.00	1.00	59,157
Number of FIs in 5-km radius	4.28	2.23	0.00	0.00	2.00	5.00	6.00	7.00	59,157
Cards: Chosen	0.14	0.34	0.00	0.00	0.00	0.00	0.00	1.00	59,157
Cards: Fees (dollars)	16.84	39.02	0.00	0.00	0.00	0.00	0.00	170.00	8,085
Cards: Limits (1000s dollars)	9.09	8.29	0.00	0.00	3.00	7.50	12.50	35.00	8,085
Cards: Credit protection	0.24	0.43	0.00	0.00	0.00	0.00	0.00	1.00	8,085
Cards: Rewards	0.26	0.44	0.00	0.00	0.00	0.00	1.00	1.00	8,085
LOC: Chosen	0.05	0.22	0.00	0.00	0.00	0.00	0.00	1.00	59,157
LOC: Rates (in %)	4.94	3.25	0.00	0.00	3.18	5.25	6.75	14.99	3,072
LOC: Limits (1000s dollars)	38.55	51.65	0.00	0.30	12.50	22.50	45.00	237.50	3,072
LOC: Fixed rate	0.32	0.47	0.00	0.00	0.00	0.00	1.00	1.00	3,072
LOC: Secured line of credit	0.56	0.50	0.00	0.00	0.00	1.00	1.00	1.00	3,072
LOC: Length relationship	5.98	1.47	1.00	2.00	5.00	7.00	7.00	7.00	3,072
Demographic variables:									
Assets (in 100,000s)	0.92	2.16	0.00	0.00	0.01	0.10	0.73	10.40	59,157
Age	54.74	15.83	18.00	22.00	44.00	56.00	66.00	87.00	59,157
Income (in 100,000)	0.64	0.53	0.08	0.08	0.28	0.50	0.85	2.50	59,157
Own house	0.77	0.42	0.00	0.00	1.00	1.00	1.00	1.00	59,157
Difficulty paying debt	2.97	2.54	0.00	0.00	1.00	2.00	5.00	9.00	59,157
Employed	0.90	0.30	0.00	0.00	1.00	1.00	1.00	1.00	59,157
Uses financial advisor	0.35	0.48	0.00	0.00	0.00	0.00	1.00	1.00	59,157
Sophisticated investor	0.39	0.49	0.00	0.00	0.00	0.00	1.00	1.00	59,157

mines the presence of FIs in geographically isolated markets. The joint estimation of these two sets of equations is necessary to correct for the well-known selection problem that creates a bias in the estimates when we estimate the individual equations with simple regression methods. Intuitively, we correct for the endogeneity of the market-presence decision in the set of household-level outcome equations by considering the optimum equilibrium strategy of each FI that is a potential entrant in each market and decides to be present (or not) in equilibrium. [11] [12]

We also use a detailed database that includes the transaction prices and observed characteristics of the products acquired by every household, as well as its demographic attributes and length of relationship with the FI. This allows us to 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 because the types of credit cards and lines of credit have expanded over the years and FIs 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.

4.2 Credit Cards

Credit cards are the first financial product that we consider. Financial institutions obtain revenues from the annual fees of credit cards they sell to customers, the interest charged on revolving accounts, and other fees such as fees charged to merchants that accept these cards. We do not observe the interest rate charged on the outstanding balance on credit cards in the CFM database, but we observe the annual fees paid. For credit cards, we consider the following vector of demographic variables and card characteristics:

$$X_{i,t} = [agehead_{i,t}, assets_{i,t}, income_{i,t}, ownrent_{i,t}, province_{i,t}, year_{i,t}]. \tag{1}$$

$$C_{i,b,t} = [protection_{i,b,t}, rewards_{i,b,t}]. \tag{2}$$

We assume that the annual fees paid for a credit card with FI b by household i are determined by the following equation:

$$\begin{aligned} Fees^{cc}_{i,b,t} = {} & \alpha^{cc}_1 + \alpha^{cc}_2 \cdot X_{i,t} + \alpha^{cc}_3 \cdot C_{i,b,t} + \alpha^{cc}_4 \cdot Close_{i,b,t} + \alpha^{cc}_5 \cdot Cult_{i,b,t} \\ & + \alpha^{cc}_6 \cdot Close_{i,b,t} \cdot Cult_{i,b,t} + \alpha^{cc}_7 \cdot N_{i,t} + \alpha^{cc}_8 \cdot Limit_{i,b,t} + B_b + \varepsilon^{cc,F}_{i,b,t}. \end{aligned} \tag{3}$$

Variable *Fees* is the annual fee paid for the credit card and is expressed in logs. $Close_{i,b,t}$ is the binary geographical proximity indicator variable that is equal to 1 if FI b has a branch within a

[11] 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. However, we require a more complex estimation methodology that uses simulation methods, as in Berry (1992), Bajari *et al.* (2010), Perez-Saiz (2015), and Perez-Saiz and Xiao (2022). We jointly estimate the equilibrium presence and outcome equations using a simulated maximum likelihood estimator from Gourieroux and Monfort (1993). 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.* (2016), Egan *et al.* (2017), and Allen *et al.* (2019).

[12] This joint estimation methodology also allows us to obtain new insights into cross-product synergies. For instance, premium credit cards tend to have larger fees and their credit limits are high. Ignoring these characteristics may bias the estimation of the effect of market concentration on the fees and limits offered because FIs in markets that face more competition may offer better product characteristics or discriminate differently on observable household characteristics.

5-km radius of the location of household i. $Cult_{i,b,t}$ is the binary cultural proximity indicator that is equal to 1 if FI b comes from the same cultural origin as household i.

For control variables, $N_{i,t}$ is the number of FIs present in a 5-km radius of the location of the household. This variable captures the competitive effect of the market presence of FIs on the credit card fees. The credit limit of the credit card ($Limit_{i,b,t}$) is also added because premium cards with large limits usually have large annual fees. We include household demographic characteristics $X_{i,t}$ and card characteristics $C_{i,b,t}$.[13] We also include FI-fixed effects (B_b). These fixed effects take into account any pricing policy set by FIs 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.

Now let's consider the household-level choice of credit cards. Although credit cards are common financial products, households rarely have cards with all seven of the FIs considered in our sample. The decision of household i to have a credit card with FI b is given by the following latent equation:

$$\begin{aligned} D^{*\,cc}_{i,b,t} &= \gamma_1^{cc} + \gamma_2^{cc} \cdot X_{i,t} + \gamma_3^{cc} \cdot heavy\_usage_{i,t} + \gamma_4^{cc} \cdot sophisticated_{i,t} + \\ &\quad \gamma_5^{cc} \cdot advice_{i,t} + \gamma_6^{cc} \cdot Close_{i,b,t} + +\gamma_7^{cc} \cdot Cult_{i,b,t} + \gamma_8^{cc} \cdot Close_{i,b,t} \cdot Cult_{i,b,t} + \varepsilon^{cc,D}_{i,b,t}. \end{aligned} \tag{4}$$

And we observe that household i has a credit card with FI b, that is, $D^{cc}_{i,b,t} = 1$, when the latent variable $D^{*\,cc}_{i,b,t} \geq 0$:

$$D^{cc}_{i,b,t} = \begin{cases} 1 & if \quad D^{*\,cc}_{i,b,t} \geq 0 \\ 0 & if \quad D^{*\,cc}_{i,b,t} < 0 \end{cases} . \tag{5}$$

In addition to demographic and product characteristics, we select control variables to be included in every equation based on broad perceptions about how households and banks make their economic decisions. For instance, the variable $heavy\_usage_{i,t}$ represents the importance for the household of using different electronic payment channels.[14] It could affect the household-level credit card choice because credit cards represent a convenient electronic method of payment. On the other hand, we do not include $heavy\_usage_{i,t}$ as an explanatory variable for fees paid because FIs may be able to discriminate on fees using observed demographics (age, income, etc.) of the household but not the potential channel usage, which should be a variable that is private information for households (especially for new clients). We also do not include this variable as a control

[13] 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 because they provide additional benefits to the card holders.

[14] 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 through a variety of payment channels (online, mobile, branches, ABM, etc.) in one month. We would expect that households with high usage of various payment channels would be more likely to use credit cards.

in the line of credit equations since a line of credit is not a method of payment.

Furthermore, we use indicators for financial sophistication and financial advice as variables that only affect credit card and line of credit choices, since they are relatively opaque for the FIs. Variable *sophisticated* is an indicator variable of financial sophistication, which is assumed to be equal to 1 when a household has more than 20% of the value of its total assets either in stock exchange assets or mutual funds. In addition, variable *advice* is an indicator variable equal to 1 if the household regularly uses a financial advisor.

Credit cards are risky products for FIs because of the risk of default on the card balance. A crucial variable used by FIs to control risk is the credit limit of the credit card. Typically, households can easily get applications approved for credit cards from most FIs, but FIs 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 FI b to household i in year t:

$$\begin{aligned} Limit^{cc}_{i,b,t} = \beta^{cc}_1 + \beta^{cc}_2 \cdot X_{i,t} + \beta^{cc}_3 \cdot C_{i,b,t} + \beta^{cc}_4 \cdot Close_{i,b,t} + \beta^{cc}_5 \cdot Cult_{i,b,t} + \beta^{cc}_6 \cdot Close_{i,b,t} \cdot Cult_{i,b,t} \\ + \beta^{cc}_7 \cdot N_{i,t} + \beta_8 \cdot Unempl_{i,t} + \beta_9 \cdot DiffPayDebt_{i,t} + B_b + \varepsilon^{cc,L}_{i,b,t}. \end{aligned} \tag{6}$$

In addition to variables included in the fees equation, we consider two variables that should affect the perceived riskiness of households by the FI and, therefore, should determine the total credit provided: an indicator for unemployment, and a variable (ranging between 0 and 9) ranking the household's perceived difficulty for paying debt. These variables should not affect the choice likelihood for credit cards. These exclusion restrictions are explained in greater detail in the Appendix.

4.3 Lines of Credit

Lines of credit are the second financial product that we consider. 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.

For lines of credit, we consider the following vector of product characteristics:

$$L_{i,b,t} = [fixed\_rate_{i,b,t}, secured_{i,b,t}]. \tag{7}$$

We assume that these rates are determined by the following equation:

$$\begin{aligned} Rates^{loc}_{i,b,t} = \alpha^{loc}_1 + \alpha^{loc}_2 \cdot X_{i,t} + \alpha^{loc}_3 \cdot L_{i,b,t} + \alpha^{loc}_4 \cdot Close_{i,b,t} + \alpha^{loc}_5 \cdot Cult_{i,b,t} + \alpha^{loc}_6 \cdot Close_{i,b,t} \cdot Cult_{i,b,t} \\ + \alpha^{loc}_7 \cdot Limit_{i,b,t} + \alpha^{loc}_8 \cdot N_{i,t} + \alpha^{loc}_9 \cdot Length_{i,t} + B_b + \varepsilon^{loc,R}_{i,b,t}. \end{aligned} \tag{8}$$

Variable *Rates* is the annual interest rate charged on the outstanding balance of the line of credit, which is expressed in logs. In addition to variables used in previous equations, we use three additional variables to characterize the line of credit: $fixed\_rate_{i,b,t}$ is an indicator variable for lines of credit with a fixed rate, $secured_{i,b,t}$ is an indicator variable for lines of credit that are secured with some form of collateral (such as a house), and $Length_{i,b,t}$ is a categorical variable that shows the length of the relationship in years between the FI that provides the service and the household. Variables used in Eq. (8) are also used in the credit limit equation for lines of credit:

$$\begin{aligned} Limit^{loc}_{i,b,t} =& \beta^{loc}_1 + \beta^{loc}_2 \cdot X_{i,t} + \beta^{loc}_3 \cdot L_{i,b,t} + \beta^{loc}_4 \cdot Close_{i,b,t} + \beta^{loc}_5 \cdot Cult_{i,b,t} \\ &+ \beta^{loc}_6 \cdot Close_{i,b,t} \cdot Cult_{i,b,t} + \beta^{loc}_7 \cdot Unempl_{i,t} + \beta^{loc}_8 \cdot DiffPayDebt_{i,t} \\ &+ \beta^{loc}_9 \cdot N_{i,t} + \beta^{loc}_{10} \cdot Length_{i,b,t} + B_b + \varepsilon^{loc,L}_{i,b,t}. \end{aligned} \tag{9}$$

In Eq. (9), we also use similar variables as those proposed to explain the credit limit equation for credit cards. As for the other financial products considered, the decision to have a line of credit with FI b depends on the following latent variable:

$$\begin{aligned} D^{*\,loc}_{i,b,t} =& \gamma^{loc}_1 + \gamma^{loc}_2 \cdot X_{i,t} + \gamma^{loc}_3 \cdot sophisticated_{i,t} + \gamma^{loc}_4 \cdot advice_{i,t} \\ &+ \gamma^{loc}_5 \cdot Close_{i,b,t} + \gamma^{loc}_6 \cdot Cult_{i,b,t} + \gamma^{loc}_7 \cdot Close_{i,b,t} \cdot Cult_{i,b,t} + T_t + \varepsilon^{loc,D}_{i,b,t} \end{aligned} \tag{10}$$

4.4 Geographic Presence of Financial Institution Branches

To account for the endogeneity of branch entry, which affects both the geographical proximity and competition variables in the household-branch level equations, 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).[15] We assume that each potential entrant independently decides whether or not to enter into every market, observing all the factors that enter into the other entrants' 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 FIs (see Ishii, 2005; Dick, 2007). In our empirical model, we consider firm-level controls such as the national and provincial sizes of the FI, which are able to include this effect.

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

$$a_{i,m} = \begin{cases} 1 & \text{if} \quad \pi_{i,m} \geq 0 \\ 0 & \text{otherwise} \end{cases} . \tag{11}$$

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

The assumption of profitable market presence is clearly reasonable for the case of commercial banks, which are private companies that maximize profits. It also applies to credit unions, which typically follow 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 market m, then the profits from presence in the market are equal to:

$$\pi_{i,m} = \theta_0 + \theta_1 X_m + \theta_2 Z_i + \sum_{j,j \neq i} \theta_{ij} a_{j,m} + \varepsilon_m^{\pi,i}, \tag{12}$$

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.[16] 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 1. $\varepsilon_m^{\pi,i}$ is observed by all potential entrants, but not by the econometrician.

We also model competitive effects between FIs that enter the market, as 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 FI j on FI i's profit if j is present in the market. This is a flexible way to take into account firm-level unobserved effects that affect each FI's competitiveness against other FIs.

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:

$$\pi_{i,m}(a_{1,m}^*, ..., a_{i,m}^*, ..., a_{E,m}^*) \geq \pi_{i,m}(a_{1,m}^*, ..., a_{i,m}, ...a_{E,m}^*) \quad \text{for any } i \in E \text{ and any } a_{i,m}\ , \tag{13}$$

where $E = 7$ is the set of potential entrants. [17] After solving for all Nash equilibria, we assume that the most efficient equilibrium is selected (see identification section in the Appendix). We assume that each competitor only affects the other potential entrants' profit through the competitive term, given that our markets are isolated and, therefore, there are no network effects. Our model assumes that all FIs are competing in the same market and for the same customers.

[16]Given the low number of observations per market, we do not include parameters from our household-branch level equations in the entry model.

[17]In total we consider seven potential entrants in every market selected. The Big Six banks and credit unions in all provinces except Quebec and New Brunswick. In the latter two provinces, the potential entrants are the Big Six banks and Desjardins.

4.5 Covariance Matrices

Although we could allow for arbitrary correlation between the elements of the covariance matrices of the error terms of the equations in the empirical model, in practice we restrict the number of correlations for tractability reasons and 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 FIs for the products that they sell to clients in those markets. In addition, we assume that a correlation exists between the error term that affects the decision to buy product p from a FI (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 two products are uncorrelated. This is a relevant assumption that reduces the number of parameters to estimate and also greatly simplifies the calculation of the likelihood used in the estimation. More details are given in the Appendix.

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 household i has acquired from FI b in a given year t. If the household has a product with b, then we will observe fees/rates or credit limits. Therefore, for the two products $p \in \{cc, loc\}$, 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 FI 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. (13) 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 given FI b and 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:

$$\Pr(Fees^p_{i,b,t}, Limit^p_{i,b,t}, D^p_{i,b,t}, Close_{i,b,t}, N_{i,t}) = \left[\Pr(D^p_{i,b,t} = 0, Close_{i,b,t}, N_{i,t})\right]^{(1-D^p_{i,b,t})} \cdot \left[f(Fees^p_{i,b,t}, Limit^p_{i,b,t} / D^p_{i,b,t} = 1, Close_{i,b,t}, N_{i,t}) \cdot \Pr(D^p_{i,b,t} = 1, Close_{i,b,t}, N_{i,t})\right]^{D^p_{i,b,t}}. \tag{14}$$

In this equation, we denote f as the probability density function of the continuous variables $Fees^p_{i,b,t}$ and $Limit^p_{i,b,t}$. The other 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 lines of credit, we consider rates rather than fees. Eq. (14) can be rewritten using conditional probabilities, as follows:

$$\Pr(Fees^p_{i,b,t}, Limit^p_{i,b,t}, D^p_{i,b,t}, Close_{i,b,t}, N_{i,t}) = \left[\Pr(D^p_{i,b,t} = 0/Close_{i,b,t}, N_{i,t}) \cdot \Pr(Close_{i,b,t}, N_{i,t})\right]^{(1-D^p_{i,b,t})} \cdot$$
$$\left[f(Fees^p_{i,b,t}, Limit^p_{i,b,t}/D^p_{i,b,t} = 1, Close_{i,b,t}, N_{i,t}) \cdot \Pr(D^p_{i,b,t} = 1/Close_{i,b,t}, N_{i,t}) \cdot \Pr(Close_{i,b,t}, N_{i,t})\right]^{D^p_{i,b,t}}. \tag{15}$$

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 simulated methods. In the Appendix, we explain in detail the steps necessary to calculate Eqs. (14) and (15).

5 Empirical Results

5.1 Proximity Effects

We now discuss the estimated coefficients for the proximity indicators. In all cases, we show the estimates of the structural model and we compare them with the OLS or probit estimates.

Table 4 shows the estimated coefficients of the choice equations for the two financial products. For both credit cards and lines of credit, we observe a significant and large positive effect of geographical proximity on the likelihood of choosing a financial product. The marginal effect of going from geographically far to geographically close is an increase of 11% in choice probability at the sample mean for credit cards and 9% for lines of credit. These effect sizes are large compared with the average 14% choice rate for credit cards and 5% choice rate for lines of credit, representing a relative 70% and 180% increase, respectively. This positive effect is consistent with closer geographical proximity resulting in both reducing transportation costs and increased soft information provision.

We also observe a significant large effect of cultural proximity on choice likelihood. At the sample mean, the marginal effect of going from culturally far to culturally close is an increase in 10% for credit card choice and 6% for line of credit choice, representing a relative 70% and 120% increase, respectively. This positive effect is consistent with closer cultural proximity bringing both taste-based discrimination and increased soft information provision.

Table 4: Estimates for the Choice of Financial Products

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

	Credit cards		**Lines of credit**	
Variable	Structural	Probit	Structural	Probit
Geographical Proximity	0.973***	0.821***	0.926***	0.768***
	(0.165)	(0.0334)	(0.180)	(0.0496)
Cultural Proximity	0.799***	0.770***	0.558***	0.660***
	(0.166)	(0.0303)	(0.144)	(0.0462)
Proximity Interaction	-0.567***	-0.489***	0.614***	-0.413***
	(0.160)	(0.0374)	(0.167)	(0.0550)
Financial advisor	0.173	0.0474***	-0.0609	0.0897***
	(0.159)	(0.0146)	(0.149)	(0.0193)
Heavy usage	0.161	0.0225***		
	(0.116)	(0.00331)		
Sophisticated	-0.159	-0.00598	-0.190	0.00901
	(0.147)	(0.0162)	(0.149)	(0.0222)
Household variables	YES	YES	YES	YES
Provincial fixed effects	YES	YES	YES	YES
Year fixed effects	YES	YES	YES	YES
Number of household-year-bank product observations	59,157	59,157	59,157	59,157

Moving on to the interaction effects, there is a negative significant interaction effect of geographical and cultural proximities on credit card choice, with the marginal effect going from a 40% probability of choosing a credit card to a 35% probability at the sample mean. This non-zero interaction effect is consistent with both proximities' effect being partly driven by lowering the cost of soft information provision. More specifically, this large negative significant interaction effect suggests that soft information obtained via closer geographically proximity is a substitute for that obtained via closer cultural affinity, at least for credit card choice.

Next, we observe a positive significant interaction effect of geographical and cultural proximity on line of credit choice likelihood, with the marginal effect going from a 3% chance of choosing a line of credit to a 10% chance at the sample mean. This large positive interactive effect is again consistent with greater soft information provision playing a major role. However, in contrast to credit cards, this result suggests that soft information obtained via closer geographical proximity complements that obtained through closer cultural affinity for line of credit choice.[18]

The opposite signed interactive effects for the two products could be rationalized by a difference in an ex-ante need for information before offering each product. Indeed, while many banks

[18]Interestingly, naive probit estimates show a negative significant effect, which demonstrates the importance of controlling for endogenous entry and selection in our model.

Table 5: Estimates for Fees/Rates Equations

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

	Credit card fees		**Line of credit rates**	
Variable	Structural	OLS	Structural	OLS
Geographical Proximity	-0.481***	-0.0970	-0.231	0.506
	(0.157)	(0.188)	(0.143)	(0.308)
Cultural Proximity	0.187	-0.0418	0.195	0.435
	(0.141)	(0.171)	(0.146)	(0.285)
Proximity Interaction	0.0566	0.358*	-0.427***	-0.520
	(0.157)	(0.194)	(0.162)	(0.323)
Card protection	0.234	0.294***		
	(0.146)	(0.0705)		
Rewards	1.882***	1.727***		
	(0.222)	(0.0885)		
Limit (in logs)	0.437***	0.219***	0.0998	0.189***
	(0.150)	(0.0143)	(0.153)	(0.0444)
Fixed rate			0.325**	0.344***
			(0.141)	(0.0966)
Secured			0.0296	0.0186
			(0.142)	(0.0948)
Length of relationship			-0.0549	0.00377
			(0.143)	(0.0320)
Number of competitors	-0.148	-0.0460**	0.102	-0.00479
	(0.131)	(0.0183)	(0.151)	(0.0261)
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 observations	8,085	8,085	3,072	3,072

offer credit cards to new customers, controlling risk by setting low credit limits, lines of credit represent a more complex product with much higher credit limits, such that a bank is more reluctant to offer it to customers without additional information about their creditworthiness. Therefore, our results suggest that while some soft information from either geographical or cultural proximity could be enough for a bank to make a credit card offer, making them substitutes, obtaining additional soft information from multiple sources is needed for banks to make line of credit offers, making them complements.

Table 5 shows estimated coefficents for the fees/rates equations. Surprisingly, except for geographical proximity in the credit card fees equation, all other proximity indicators do not have a significant effect, suggesting that FIs are employing neither price discrimination based on market power nor marginal cost pricing, for consumer loan products. This does not imply that such pricing strategies are not being used, but the effects, if they do exist, are not large enough to be detected by our model.

Table 6: Estimates for Limit Equations

This table shows 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 of relationship is a categorical variable with values 1-7 (see Appendix for definitions). Household variables and bank, provincial and year fixed effects are used in all cases. Standard errors, in parentheses, obtained using bootstrap (for structural model).

	Credit card limits		**Line of credit limits**	
Variable	Structural	OLS	Structural	OLS
Geographical Proximity	-0.214	-0.0151	-0.228	-0.159
	(0.145)	(0.120)	(0.144)	(0.126)
Cultural Proximity	-0.0236	0.0277	-0.230	-0.0466
	(0.147)	(0.111)	(0.154)	(0.117)
Proximity Interaction	-0.183	-0.0453	-0.164	0.00628
	(0.149)	(0.125)	(0.143)	(0.132)
Card protection	-0.0087	0.115***		
	(0.141)	(0.0439)		
Rewards	0.313**	0.395***		
	(0.150)	(0.0491)		
Fixed rate			-0.424***	-0.351***
			(0.146)	(0.0390)
Secured			0.579***	0.704***
			(0.160)	(0.0366)
Difficulty debt	0.0238	0.0157**	-0.144	-0.0187**
	(0.134)	(0.00799)	(0.148)	(0.00759)
Employed	0.176	0.134*	0.0109	0.112
	(0.144)	(0.0812)	(0.154)	(0.0843)
Length of relationship			-0.0468	0.0109
			(0.146)	(0.0131)
Number of competitors	0.0810	0.00764	0.122	0.0194*
	(0.135)	(0.0124)	(0.141)	(0.0107)
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 observations	8,085	8,085	3,072	3,072

Table 6 shows estimated coefficients for the limits equations. Similarly to the fees/rates equations, the proximity indicators do not have significant effects, offering evidence for neither price discrimination based on market power nor marginal cost pricing. Results from the covariance matrix and the market presence model are mostly intuitive as well. A detailed discussion of these estimates can be found in the Appendix.

5.2 Discussion

While the theoretical mechanisms from Section 2 focus on marginal effects, we can only observe equilibrium choices made by households and banks because of limitations in our data.[19] Therefore, our estimated coefficients reflect equilibrium effects that integrate both supply and demand. In this section, we explore how measuring equilibrium effect could affect the interpretation of our results.

First, one concern is that the household product choice decisions should depend on the prices at which the products are being offered. This would imply that choice likelihood declines with higher offered prices. Because we do not know the offered prices for credit card and line of credit offers that households declined, we are unable to include prices as an explanatory variable in the choice equations. This omission could bias our estimated coefficients for the choice equations. In other words, we need to take into account the possibility that higher choice likelihood due to geographical and cultural proximity is not simply a product of lower prices. This is a potential bias from the demand side.

Let's consider this potential demand-side bias. If banks adopt a price discrimination strategy, where they charge higher prices for closer geographical and cultural proximity, our estimate of the proximity coefficients in the choice equations would be biased downward. If banks adopt a marginal cost pricing strategy, where they charge higher prices for households that are farther away geographically and culturally, then our estimate of the proximity coefficients in the choice equations would be biased upward. However, since the estimated coefficients for almost all proximity indicators in the pricing and limit equations are insignificant, banks may be adopting a uniform pricing strategy without taking proximity into account, which means that our estimates for the choice equation would not be biased. The only exception is that geographical proximity seems to significantly decrease annual credit card fees. Given the size of the overall effect and that we do not find an equivalent effect for line of credit rates, a pure pricing effect is unlikely to explain the entire geographical proximity effect on credit card choice.

On the other hand, banks may be more willing to make credit offers where the price they can charge is higher, meaning that observed choice likelihood would increase with higher offered prices. This would be a bias from the supply side. Given that we find almost no significant effect of proximity on pricing and limits, this channel is also unlikely to explain the full proximity effect on choice likelihood.

In summary, we observe equilibrium geographical and cultural proximity effects on choice likelihood for both credit cards and lines of credit, which are unlikely to be explained by a pure pricing effect. We find that the soft information mechanisms play an important role in determining the choice of banking products. Furthermore, soft information obtained via the two types

[19]For example, we do not observe prices on declined credit card and line of credit offers.

of proximity are likely substitutes for each other for credit card choice and complements of each other for line of credit choice.

5.3 The Effect of Bank Branch Closures

To better illustrate the effects of both proximities on financial product choice likelihood, we conduct a counterfactual experiment where banks close branches due to an increase in fixed entry costs. A large number of bank branch closures have been observed in all major economies over the last decade. For instance, in the United States, there were more than 11,000 net bank branch closures between 2012 and 2018 as consumers continued to migrate to online and mobile banking options.[20] This is also an observed phenomenon in Canada (see Figure C.1) and in Europe with more than 48,000 net bank branch closures since 2008.[21] One concern of policymakers is that these closures reduce access to financial services, especially for vulnerable households who are unable to use technology-based financial services due to factors such as age (Norris, 2001; Rice and Katz, 2003; Billon *et al.*, 2009). Our counterfactual experiment contributes to better understanding the implications of branch closures on consumer access to financial services. It complements the recent study by Nguyen (2019), who focuses on the effects of bank branch closures on small business lending.

The closure of bank branches affects the choice likelihood of financial products because households lose access to their nearby branch, reducing geographical proximity between households and banks. Given our results in Table 4, this would reduce the choice likelihood for financial products. To measure this effect, we estimate the bank branches that have closed in all of the markets in Canada as a result of the higher fixed entry cost, and then recalculate the variable $Close_{i,b,t}$ and estimate the choice likelihood using Eqs. (4) and (10).

We present the results for the case of all individuals in Figure 1. The extreme right-hand side of the figure shows the current observed case, where we normalize choice probability to 1 for both products considered. As expected, the choice likelihood of both credit cards and lines of credit decreases as banks close branches, as more and more households lose access to their nearby branch. The decrease is non-linear for both products, which can be explained by the fact that less-used branches are closed first as entry costs increase, while more-used branches are closed last. The sizes of the decreases, however, differ for the two products: credit card choice probability decreases by less than 10% as almost all branches close, while line of credit choice probability decreases by more than 30%.

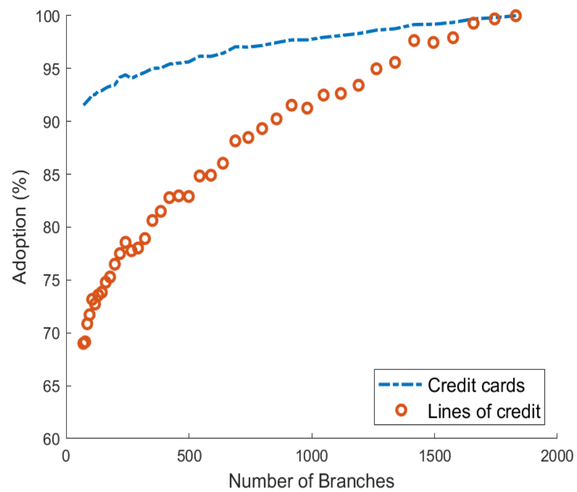
This large difference can be explained by the effects of the proximity interaction term. For credit cards, the proximity interaction term has a negative effect, suggesting that as bank branches

[20] See https://www.spglobal.com/marketintelligence/en/news-insights/latest-news-headlines/49360224

[21] See https://www.reuters.com/article/us-europe-banks-closures/eu-banks-close-branches-cut-jobs-as-customers-go-online-idUSKCN1BN2BV)

Figure 1: Effects of Branch Closures on Choice of Financial Products for all Households

This figure shows the relative change of choice likelihood of financial products for all households in our sample, as the number of total bank branches decreases. The graph is generated by imposing an increase in fixed entry costs to all FIs, which reduces the market presence of FIs across markets. We then obtain predicted choice likelihoods by product. The observed case for 2006 is at the extreme right-hand side of the graph. For comparison reasons, we normalize to 1 the choice indicators for every product.



close, soft information obtained from geographical proximity can be substituted by soft information obtained from cultural proximity, which is relatively easy to do since almost three quarters of all household-bank pairs are culturally close.

For lines of credit, however, the estimated effect of proximity interaction is positive, suggesting that in many cases, soft information from both cultural and geographical proximity is needed for banks to make credit offers. Therefore, the effects of geographical proximity and proximity interaction reinforce each other, which gives a total effect that is much larger than for credit cards.

The much smaller decrease for credit card choice likelihood is also consistent with the fact that non-local competitors, including banks with no local branches and non-bank competitors such as retailers and airlines, have a strong presence in the credit card market. In practice, households can acquire credit cards from banks and other providers even if bank branches are closed. Indeed, as shown in Table D.4 in the Appendix, more than 30% of all credit cards held by sample CFM households are from non-bank (and non-credit union) providers, while less than 5% of all lines of credit held are from non-depositionary institutions.

6 Conclusion

In this paper we quantify the joint effects of both geographical and cultural proximity of bank branches on the household-level choice probability of banking services in a culturally segmented country such as Canada. Several strands of the literature are related to our work: soft information in banking and structural models in industrial organization and urban economics. Ours is the first paper, to our knowledge, that analyzes both measures of proximity together. Our methodology takes into account the endogeneity of the market-presence decision, and we estimate the joint effect of both geographical and cultural proximity on choice and outcomes—rates and credit limits—of credit cards and lines of credit.

We observe significant geographical and cultural proximity effects on household choice for both credit cards and lines of credit, which are unlikely to be explained by a pure pricing effect. We show that soft information mechanisms play an important role in determining the choice of banking products. Moreover, soft information obtained via the two types of proximity are likely substitutes for each other in credit card choices and complements of each other for lines of credit. Hence, complementarity between soft information sources is more likely for products that require more ex-ante screening, since FIs tend to need much more information for line of credit offers than credit card offers.

Our results demonstrate that despite the emergence of Internet and mobile banking, physical branches and proximity still matter for driving consumer choice of financial services. Therefore, despite the rise of Internet and mobile banking, and similar to the staying power of automatic teller machines through a pandemic (Chen and Felt, 2022), physical bank branches are here to stay.

References

AGARWAL, S., CHOMSISENGPHET, S. and LIU, C. (2010). The importance of adverse selection in the credit card market: Evidence from randomized trials of credit card solicitations. *Journal of Money, Credit and Banking*, **42** (4), 743–754.

— 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. (2016). Diversification of geographic risk in retail bank networks:Evidence from bank expansion after the Riegle-Neal Act. *The RAND Journal of Economics*, **47** (3), 529–572.

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 — (2019). Search frictions and market power in negotiated-price markets. *Journal of Political Economy*, **127** (4), 1550–1598.

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

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

BECKER, G. S. (1957). *The economics of discrimination*. University of Chicago press.

BELLUCCI, A., BORISOV, A. and ZAZZARO, A. (2013). Do banks price discriminate spatially? evidence from small business lending in local credit markets. *Journal of Banking & Finance*, **37** (11), 4183–4197.

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.

CALOMIRIS, C. W. and HABER, S. H. (2014). *Fragile by design: The political origins of banking crises and scarce credit*. Princeton University Press.

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

CHEN, H. and FELT, M.-H. (2022). *Canadians Access to Cash Before and During the COVID-19 Pandemic*. Tech. rep., Bank of Canada.

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.

CORNELL, B. and WELCH, I. (1996). Culture, information, and screening discrimination. *Journal of political Economy*, **104** (3), 542–571.

D'ACUNTO, F., GHOSH, P., JAIN, R. and ROSSI, A. G. (2021). How costly are cultural biases? evidence from fintech. *Evidence from FinTech (October 25, 2021)*.

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

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

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.

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.

FISMAN, R., PARAVISINI, D. and VIG, V. (2017). Cultural proximity and loan outcomes. *American Economic Review*, **107** (2), 457–92.

FORD, C. (2019). Characteristics of recent home buyers. *HousingEconomics. com*.

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.

— and — (1993). Simulation-based inference: A survey with special reference to panel data models. *Journal of Econometrics*, **59** (1), 5–33.

GRINBLATT, M. and KELOHARJU, M. (2001). How distance, language, and culture influence stockholdings and trades. *The Journal of Finance*, **56** (3), 1053–1073.

HAUSWALD, R. and MARQUEZ, R. (2006). Competition and strategic information acquisition in credit markets. *The Review of Financial Studies*, **19** (3), 967–1000.

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*.

KNYAZEVA, A. and KNYAZEVA, D. (2012). Does being your banks neighbor matter? *Journal of Banking & Finance*, **36** (4), 1194–1209.

LIBERTI, J. M. and PETERSEN, M. A. (2019). Information: Hard and soft. *Review of Corporate Finance Studies*, **8** (1), 1–41.

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.

NGUYEN, H.-L. Q. (2019). Are credit markets still local? evidence from bank branch closings. *American Economic Journal: Applied Economics*, **11** (1), 1–32.

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

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. (2022). Cultural affinity, regulation, and market structure: Evidence from the Canadian retail banking industry. *American Economic Journal: Microeconomics*, **14** (1), 451–89.

PETERSEN, M. A. and RAJAN, R. G. (1994). The benefits of lending relationships: Evidence from small business data. *The journal of finance*, **49** (1), 3–37.

REHBEIN, O., ROTHER, S. *et al.* (2020). *Distance in bank lending: The role of social networks*. Tech. rep., University of Bonn and University of Mannheim, Germany.

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.

STATISTICS CANADA (2006). 2006 Census Community Profiles. https://www12.statcan.gc.ca/census-recensement/2006/index-eng.cfm (accessed November 20, 2020).

STIGLITZ, J. E. and WEISS, A. (1981). Credit rationing in markets with imperfect information. *The American economic review*, **71** (3), 393–410.

WILKINS, R. and KHAN, S. (2010). Postal code conversion file plus (pccf+) version 5g users guide.

Appendices

A Variable Definitions

- Proximity Indicators:
  - Close: Geographical Proximity Indicator variable equal to 1 if the FI has a presence in a 5-km radius around the household.
  - Cult: Cultural Proximity Indicator variable equal to 1 if the FI has the same cultural origin as the household's language spoken at home.
  - N: Number of FIs with a presence in a 5-km radius around the centroid of the market.
- 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 a loyalty program that provides miles, points, and so on.
- Lines of credit:
  - Rate: Annual interest rate (in %) charged on the outstanding balance of the line of credit.
  - Fixed rate: Indicator equal to 1 if the interest rate charged on outstanding balance 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: 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.
- Household demographic variables:
  - Age: Age in years of the head of the house.
  - Assets: Total assets of household in Canadian dollars (in logs). This includes total balance in accounts, value of bonds, mutual funds, stock, real estate, other liquid assets, illiquid assets, and so on.
  - 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.
  - Heavy usage payments: Total number of payment transactions per month, including ATM, phone payment, online, and mobile payment transactions.
  - Income: Total annual income of the household.
  - Own house: Indicator equal to 1 if the house is owned by the household.

    - Sophisticated investor: Indicator equal to 1 if 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.
    - Uses financial advisor: Indicator equal to 1 if the household regularly uses a financial advisor.

- Market-level variables (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 headquarters of the FI in 1972.

B Computational Details

B.1 Simulated Maximum Likelihood

Fermanian and Salanie (2004) show that we can estimate the conditional density function,

$$f(Fees^p_{i,b,t}, Limit^p_{i,b,t}/D^p_{i,b,t} = 1, Close_{i,b,t}, N_{i,t}), \tag{16}$$

using a simple non-parametric (kernel density) estimator. These estimators are relatively standard, and they are usually available in most statistical packages, such as Stata or Matlab. To estimate Eq. (16), we need to generate a large number of simulation draws. In Box 1, we explain in detail the steps necessary to calculate Eq. (16) and other conditional probabilities in Eq. (15) for a given product p.

To calculate $\Pr(D^p_{i,b,t} = 1/Close_{i,b,t}, N_{i,t})$ and $\Pr(D^p_{i,b,t} = 0/Close_{i,b,t}, N_{i,t})$, we use a simple frequency estimator to compute these probabilities, as they are discrete.

A probability that requires significant computation is

$$\Pr(Close_{i,b,t}, N_{i,t}), \tag{17}$$

which is calculated by solving the market presence equilibrium using Eq. (13). This term represents the predicted probability of observing the presence of FI b and a number of FIs N in a circle around household i. Because there is no closed form solution for this predicted probability of market presence, we need to numerically estimate it, that is, for each draw, we have to numerically solve for all Nash equilibria in every market using Eq. (13) and choose the most profitable one (equilibrium selection rule). This approach is 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 FIs decide to be present in given market m (a census subdivision), taking into account the demographic characteristics of the market and the market-presence decision of other FIs. Using the market-presence equilibrium, the variables $Close_{i,b,t}$ and $N_{i,t}$ in Eq. (17) can be calculated for every household i and FI b.

Note that because error terms across products are uncorrelated, the likelihood function is separable for each of the two products considered. In addition, because 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. (15) 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 of the equations of our model:

$$\max_{\alpha,\beta,\gamma,\theta} \sum_{p\in\{acc,cc,loc\}} \sum_{i=1}^{M} \sum_{b=1}^{7} \log \widehat{\Pr}(Fees^p_{i,b,t}, Limit^p_{i,b,t}, D^p_{i,b,t}, Close_{i,b,t}, N_{i,t}), \tag{18}$$

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. The asymptotic distribution of this maximum likelihood estimator has been studied by Gourieroux and Monfort (1990). We have a total of 59,157 household-bank-year product observations.

In the next box, we explain in detail the steps necessary to calculate Eq. (16) and other conditional probabilities in Eq. (15) for given product p:

BOX 1: Algorithm to Simulate Conditional Probabilities:

1. Select a large number of simulation draws S
2. Generate a set of independent random draws $\Gamma = \{\varepsilon_s^{cc}, \varepsilon_s^{loc}, \varepsilon_s^{\pi}\}_{s=1}^{s=S}$
3. Transform the set Γ in another set $\widetilde{\Gamma}$ that is distributed following the variance-covariance matrix Σ^p
4. Calculate $D_{i,b,t}^{*\,p}$ for all $\widetilde{\Gamma}$. Find the set of draws $\widetilde{\Gamma}_D$ such that $D_{i,b,t}^{*\,p} > 0$
5. Solve the entry equilibrium for all $\widetilde{\Gamma}$ using Eq. (13). Find the set of draws $\widetilde{\Gamma}_{E,N}$ such that $E_{i,b,t}, N_i$ is an equilibrium
6. Determine the subset of $\widetilde{\Gamma}_{\cap} = \widetilde{\Gamma}_D \cap \widetilde{\Gamma}_{E,N}$.
7. Calculate $Fees_{i,b,t}^p$ and $Limit_{i,b,t}^p$ for the set of errors $\widetilde{\Gamma}_{\cap}$
8. Estimate $f(Fees_{i,b,t}^p, Limit_{i,b,t}^p / D_{i,b,t}^p = 1, Close_{i,b,t}, N_i)$ with a kernel density estimator using values from the previous stage.

Using this algorithm, we construct the simulated likelihood from Eq. (18), which we maximize using a state-of-the-art optimizer (MATLAB/KNITRO) and the computing cluster of the Bank of Canada and IMF.

B.2 Structure of Covariance Matrices

The structure of the covariance matrix for credit cards is as follows:

$$\sum^{cc} = \begin{array}{c} \\ \varepsilon^{\pi,1} \\ \ldots \\ \varepsilon^{\pi,7} \\ \varepsilon^{cc,D} \\ \varepsilon^{cc,F} \\ \varepsilon^{cc,L} \end{array} \begin{array}{c} \begin{array}{cccccc} \varepsilon^{\pi,1} & \ldots & \varepsilon^{\pi,7} & \varepsilon^{cc,D} & \varepsilon^{cc,F} & \varepsilon^{cc,L} \end{array} \\ \begin{bmatrix} 1 & & & & & \\ \ldots & \ldots & & & & \\ 0 & \ldots & 1 & & & \\ \rho^{cc}_{\pi,D} & \ldots & \rho^{cc}_{\pi,D} & 1 & & \\ \rho^{cc}_{\pi,F} & \ldots & \rho^{cc}_{\pi,F} & \rho^{cc}_{D,F} & \sigma^{cc}_F & \\ \rho^{cc}_{\pi,L} & \ldots & \rho^{cc}_{\pi,L} & \rho^{cc}_{D,L} & 0 & \sigma^{cc}_L \end{bmatrix} \end{array} \tag{19}$$

And the structure of the covariance matrix for lines of credit is as follows:

$$\sum^{loc} = \begin{array}{c} \\ \varepsilon^{\pi,1} \\ \ldots \\ \varepsilon^{\pi,7} \\ \varepsilon^{loc,D} \\ \varepsilon^{loc,L} \\ \varepsilon^{loc,R} \end{array} \begin{array}{c} \begin{array}{cccccc} \varepsilon^{\pi,1} & \ldots & \varepsilon^{\pi,7} & \varepsilon^{loc,D} & \varepsilon^{loc,L} & \varepsilon^{loc,R} \end{array} \\ \begin{bmatrix} 1 & & & & & \\ \ldots & \ldots & & & & \\ 0 & \ldots & 1 & & & \\ \rho^{loc}_{\pi,D} & \ldots & \rho^{loc}_{\pi,D} & 1 & & \\ \rho^{loc}_{\pi,L} & \ldots & \rho^{loc}_{\pi,L} & \rho^{loc}_{D,L} & \sigma^{loc}_L & \\ \rho^{loc}_{\pi,R} & \ldots & \rho^{loc}_{D,R} & \rho^{loc}_{D,R} & 0 & \sigma^{loc}_R \end{bmatrix} \end{array} \tag{20}$$

B.3 Identification

We identify the parameters in our market presence model in two ways. First, we use exclusion restrictions in the profit function (variables that affect the profit of one FI but not the profit of the rest of the FIs). This is a well-known approach that is used in the literature to identify static entry games (see Berry (1992) and 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 FIs. 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.* (2016) 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 there than in more distant markets. We use the geographical presence of the banks in 1972 to construct this variable because that year was 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.[22] We expect significant inertia in the subsequent expansion of FIs over the decades, therefore this variable should be correlated with the geographic presence in 2006. In addition, this variable varies across markets for most FIs considered.

The other variables that we use are the total asset size of every FI (which does not vary across markets) and regional (provincial) size of every FI (which varies across provinces but not across markets within a province). Total asset size includes all geographical markets, including international markets and any business line (such as investment or wholesale banking). The Big Six 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. In addition, 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 rural markets, were probably covered by FIs much earlier than rural markets. Therefore, regional size can also be considered an exogenous variable to a certain extent.

The second strategy that we use to identify the market-presence model is related to the existence of multiple equilibria. Given the assumptions of our model, multiple equilibria are possible, in contrast 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. (17) would not be defined in the presence of multiple equilibria. A number of solutions have been proposed in the literature to solve this problem. We use the recent approach from Bajari *et al.* (2010), which includes an equilibrium selection rule and allows us to point-identify the parameters of the model.[23] To identify the equilibrium that we will use, we compute all possible equilibria using Eq. (13) and we then select the most efficient with probability 1.[24]

Regarding the rest of the equations of the model for the outcome equations, our model is 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 affects the

[22]More precisely, using the geographical presence of every FI in 1972 (see Canadian Bankers Association, 1972), we determine the market share of every FI in every province, and we use this market share to generate a weighted measure of distance to headquarters. Because there have been a significant number of mergers in Canada since 1972, the geographical presence of an FI is generated using the geographical presence of other FIs that were acquired by the FI between 1972 and 2006.

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

[24]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.

demand for a credit card but not the fees paid. Intuitively, this assumes that FIs may be able to discriminate on fees using observed demographics (e.g., age, income, etc.) of the household but do not use the potential channel usage, which should be 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, although they do not affect the rates, fees or limits of these products. Again, this implies that these are variables that are relatively opaque for FIs. We also use risk variables that affect the limits granted for financial products by FIs. Credit limits granted by FIs are highly dependent on the riskiness of the clients, therefore unemployment and difficulty to pay the debt should be particularly related to these limits but not with the demand for these products.

C Additional Figures and Tables

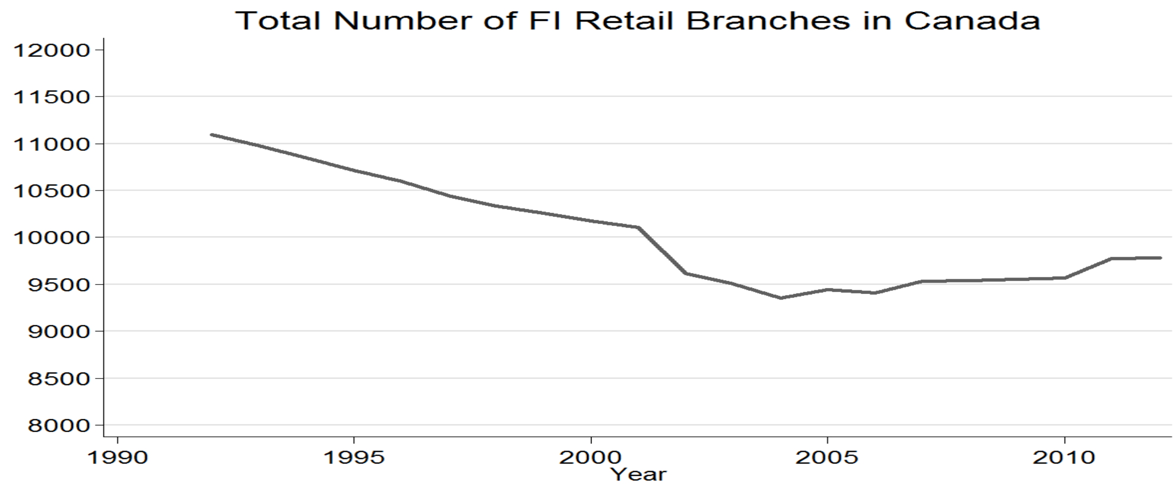
Table C.1: Summary Statistics for Markets Considered in Estimation

This table shows summary statistics of markets considered in estimation for all of Canada and by province. Source: Stats Canada.

Variable	Canada	BC	MB	NB	NFL	NS	ON	QC	SK
Population:									
mean	8,213	9,973	2,819	4,991	2,402	4,641	17,131	6,495	5,371
min	205	340	275	395	215	440	365	305	210
p25	760	1,390	670	860	370	1,385	3,175	780	345
p50	1,960	3,050	1,155	1,200	650	4,220	8,625	1,565	635
p75	7,035	8,335	2,725	1,590	3,180	7,720	17,535	2,960	1,163
max	199,385	122,175	40,705	66,690	13,385	11,405	155,995	144,595	199,385
Per capita income:									
mean	23,200	24,600	22,868	19,472	15,654	20,270	26,337	21,062	20,960
min	0	16,715	15,349	13,836	0	15,858	15,431	13,353	0
p25	20,240	21,898	20,311	17,842	15,712	18,390	22,664	18,474	18,719
p50	23,090	23,642	21,892	19,410	17,773	20,134	25,821	21,055	22,552
p75	26,767	26,496	24,688	21,661	18,652	20,929	29,405	23,110	25,513
max	47,906	38,508	38,311	24,384	27,079	27,443	39,947	38,234	32,131
Number of businesses:									
mean	645	869	313	297	127	378	1,130	379	506
min	0	0	0	24	0	61	0	0	0
p25	93	134	104	34	21	171	227	43	94
p50	227	363	152	79	35	239	652	104	162
p75	628	664	410	137	137	599	1,378	222	283
max	14,249	9,985	2,906	3,918	663	1,214	7,918	8,516	14,249
Proportion of French population (in %):									
mean	19.3	1.6	6.0	20.5	0.2	1.3	10.5	94.1	2.5
min	0.0	0.0	0.0	0.0	0.0	0.3	0.0	12.9	0.0
p25	0.7	0.9	0.8	1.1	0.0	1.0	1.0	95.0	0.0
p50	1.7	1.5	1.6	3.3	0.0	1.1	1.7	98.3	0.9
p75	6.3	2.1	2.7	14.6	0.2	1.6	6.1	99.2	2.2
max	100.0	7.7	86.7	89.0	1.4	2.8	100.0	100.0	38.9
Unemployment rate (in %):									
mean	9.3	8.7	5.0	13.6	29.1	13.7	8.4	13.9	7.5
min	0.0	0.0	0.0	2.7	0.0	5.7	0.0	0.0	0.0
p25	4.7	5.7	2.2	9.4	18.3	9.5	5.5	8.7	0.0
p50	7.4	8.0	4.2	13.5	25.0	11.2	7.3	12.7	5.5
p75	11.6	10.8	7.1	16.1	38.9	16.7	9.1	17.7	10.9
max	60.0	26.9	19.2	25.8	60.0	37.2	36.6	40.0	38.1

Figure C.1: Changes in Total Number of Financial Institution Branches

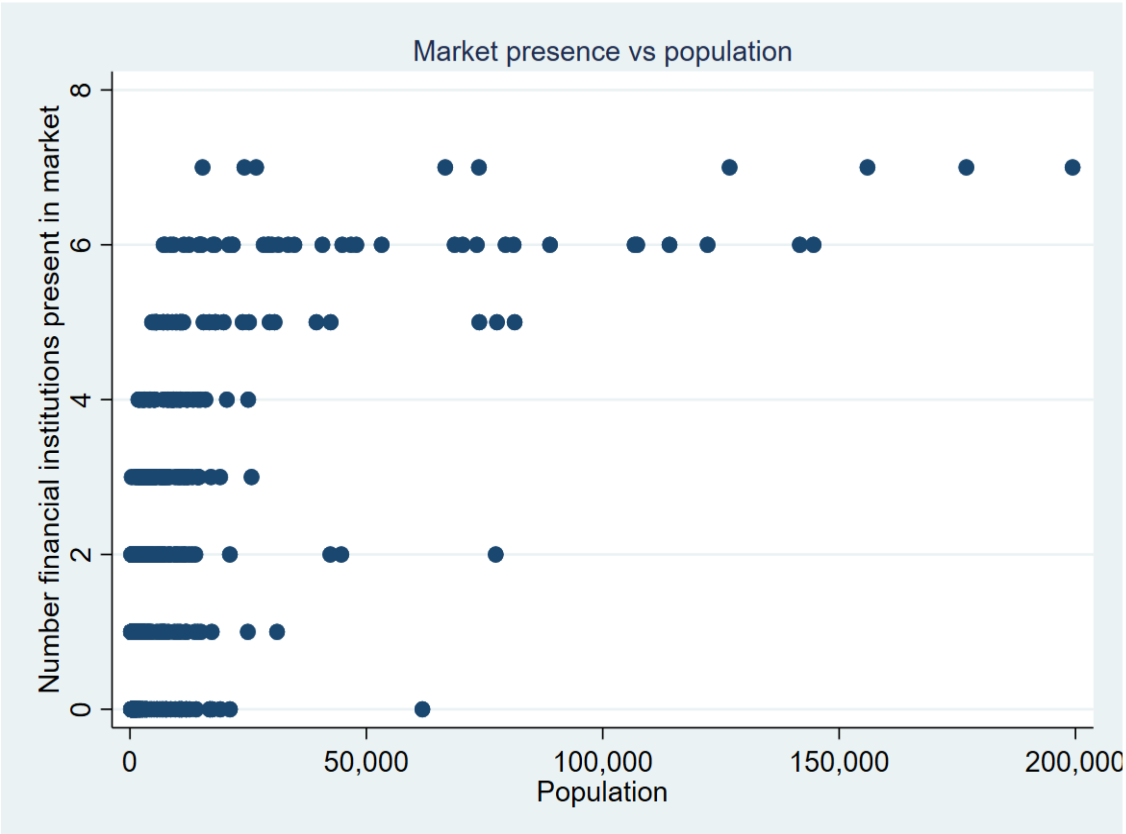
The total number of FI branches in Canada shows a steady decrease in the 1990s and early 2000s, with stabilization after the mid-2000s.



Source: Canadian Financial Services

Figure C.2: Market Presence and Population

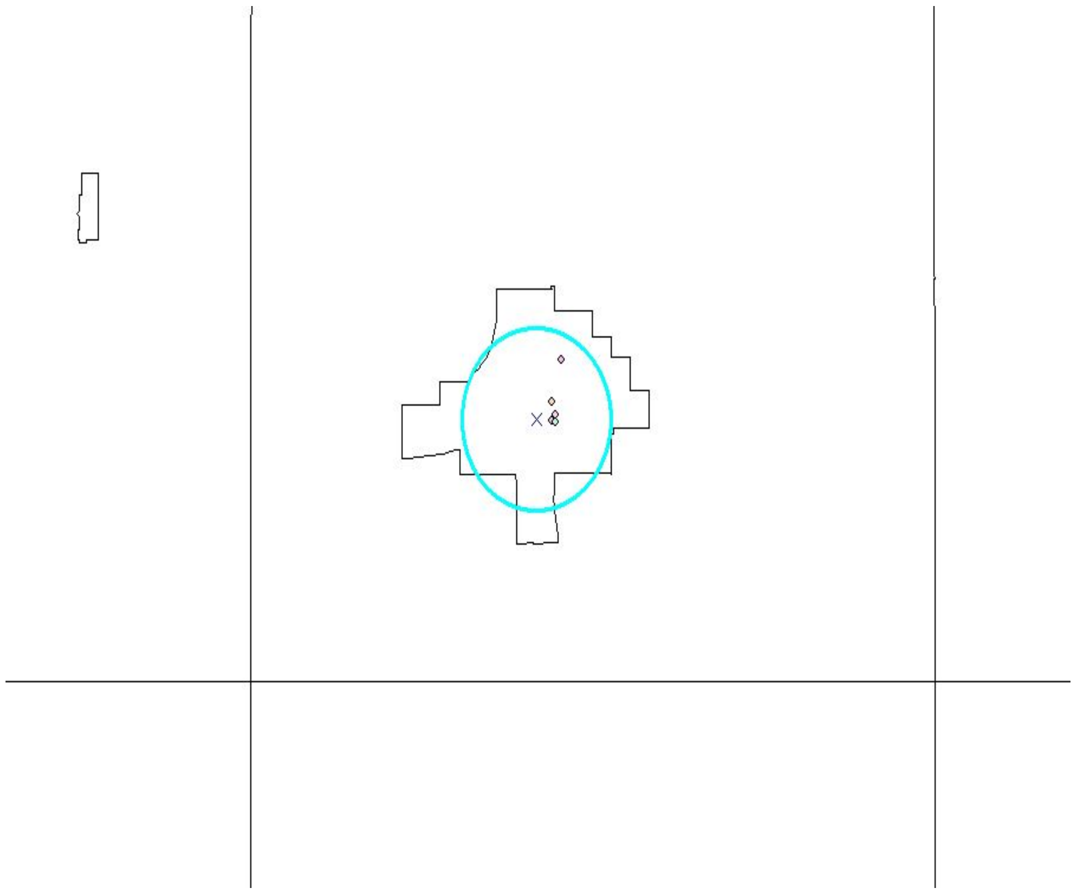
This figure of the number of entrants vs. population shows a clear positive correlation between the two. We also see that most markets have at most six FIs present.



Source: Statistics Canada, Authors' Calculations

Figure C.3: Example of Market: Moose Jaw, Saskatchewan

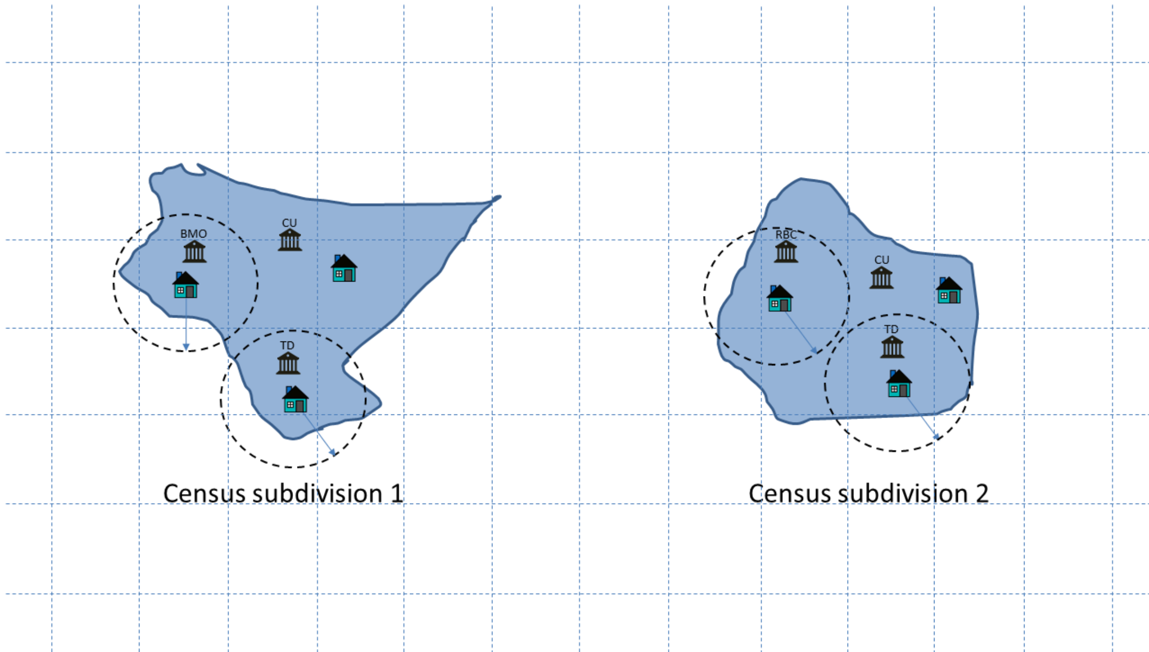
This figure shows the branches of various FIs located near Moose Jaw, Saskatchewan. Each diamond represents a branch, and the colours represent different FIs. Branches identified are in a 5-km radius of the centroid of the market. We use exact latitude-longitude information for branches and market centroids.



Source: Statistics Canada, Canadian Financial Services

Figure C.4: Computing Geographical Proximity

This diagram shows the computation of geographical proximity indicators between CFM households and bank branches. In Census subdivision 1, the leftmost household is close to the BMO branch but not close to the CU or TD branches. However, the bottom household in the same census subdivision is close to the TD branch, but not to BMO and CU branches.



D Additional Results Discussion

D.1 Covariance Matrix and Other Effects

The presence of an extra competitor in a geographical area close to the location of the household does not have a significant effect on fees, rates or limits, though their effect size is much different between structural and OLS estimates due to the large estimates from the covariance matrix.

Table D.1: Estimates of Elements of the Covariance Matrix

This table shows estimates of the elements of the variance-covariance matrix used in our econometric model. Standard errors, in parentheses, have been generated using bootstrap.

Variable	Credit cards	Lines of credit
σ_F or σ_R	2.739***	2.530***
	(0.313)	(0.281)
σ_L	1.904***	2.316***
	(0.224)	(0.282)
$\rho_{\pi,D}$	-0.0971	2.449***
	(0.146)	(0.317)
$\rho_{\pi,F}$ or $\rho_{\pi,R}$	-0.698***	-4.568***
	(0.157)	(0.506)
$\rho_{\pi,L}$	-0.380**	-3.886***
	(0.157)	(0.429)
$\rho_{D,F}$ or $\rho_{D,R}$	1.352***	4.607***
	(0.209)	(0.519)
$\rho_{D,L}$	4.691***	-3.886***
	(0.495)	(0.429)

Estimates for the covariance matrix are shown in Table D.1. The correlation term $\rho_{\pi,D}$ is positive and large for lines of credit, suggesting that FIs preferentially enter markets where unobserved variables that positively affect entry also increase the choice likelihood for lines of credit. On the other hand, the correlation terms $\rho_{\pi,F}$, $\rho_{\pi,R}$ and $\rho_{\pi,L}$ are negative and significant for both credit cards and lines of credit, suggesting that FIs preferentially enter markets where unobservable variables that increase the fees/rates paid by households are negatively correlated with unobserved variables that increase entry profit. Finally, the correlation terms $\rho_{D,F}$, $\rho_{D,R}$ $\rho_{D,L}$ are almost all positive and significant, suggesting that unobserved variables that positively affect household choice of consumer loan products also positively affect their prices and limits. The only exception is LOC limits, where the correlation between unobserved variables that increase LOC limits and unobserved variables increasing LOC choice is negative.

Some other estimates offer intuitive results and are in line with OLS and structural estimates. For example, we find that certain product characteristics significantly affect prices and limits. Credit cards with rewards have almost triple the fees and 31% higher limits than cards with no rewards, consistent with the intuition that reward cards tend to be premium cards offered to higher-income households. Credit cards with higher limits also tend to have higher fees, such that a 1% increase in limits is associated with a 0.42% increase in fees. This is again consistent with premium cards having both higher fees and limits.

For lines of credit, a fixed rate LOC has both 33% higher rates and 42% lower limits than a variable rate LOC because the FI needs compensation for the additional risk it is taking. Furthermore, a secured LOC has limits that are 58% higher than a non-secured LOC since the FI is willing to offer more credit due to lower risk with the presence of collateral.

D.2 Estimates of Market Presence Model

Tables D.2 and D.3 show the estimates for the market presence model. Most variables have been divided by their mean to facilitate the comparison, and the variance of the error term in Eq. (12) is normalized to 1. The coefficients of the demographic variables in Table D.2 mostly follow expectations on the sign but present important differences by type of institution. Credit unions have a lower coefficient on business activity and on income per capita, which indicates a focus of credit unions on less economically attractive areas. The coefficient for unemployment is also larger for credit unions. These results show that the Big Six are particularly focused on markets that may be more economically attractive, consistent with results from Perez-Saiz and Xiao (2022).[25]

For competitive effects, four of the five competitive effects have an intuitive negative sign. Interestingly, the effect of the presence of Desjardins on the profits of the Big Six is positive. This result shows that there is a positive complementarity effect on the Big Six when Desjardins enters markets, other effects being constant. This suggests that the two types of institutions may compete for different types of markets, as discussed in the previous paragraph.

Table D.3 shows results for provincial and individual FI effects. The coefficient for Desjardins is positive and relatively large. The other banks have a negative or lower effect. This shows that Desjardins faces lower entry costs than the Big Six in general, so they are present in markets that are less attractive.[26]

Finally, as we would expect, distance is negatively and significantly related to profits. This gives regional players who expand to areas close to large population centres where they have their headquarters or main centres of activity an advantage.

These results suggest that CU/Desjardins 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 interpretation is that credit unions do not need to focus solely on the goal of maximizing profit, which means 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 FI, which is similar to how local farmers' markets are able to thrive. Furthermore, they may be more nimble than larger national banks and they may tailor their product offerings to the specific town that 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 in 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.

Moreover, credit unions and Desjardins face provincial prudential regulations that are different from their federal counterparts. The existence of different regulatory authorities in Canada at provincial and federal levels could also affect the effective implementation of the regulation and supervision of the industry (see for instance Agarwal *et al.* (2014), who show that state regulators

[25]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 good source of variation to identify this effect (see Table C.1).

[26]Surprisingly, CU has a lower coefficient than most of the Big Six banks.

tend to be more lenient than federal regulators in the United States).[27]

[27]In most provinces, deposit insurance limits are higher than the federal limit, but the effectiveness of bank supervision suggests that the effect of this channel may not be large. There are also restrictions to credit unions that wish to access emergency liquidity facilities because they must be deemed important by the regulator for financial stability.

Table D.2: Estimates of Market Presence Model (I)

This table shows estimates of the parameters of the profit function (Eq. (12)) in the market presence model. Demographic variables have been normalized by their mean. We have estimated separate effects of the demographic variables for all Big Six banks and for credit unions/Desjardins. "Competitive effect of X on Y" is the effect on the profit of FI Y if FI X is present in the market. Standard errors, in parentheses, obtained using bootstrap.

Variable	Market presence model
Panel A: Competitive effects:	
Competitive effect of BIG6 on BIG6	-0.0654
	(0.2142)
Competitive effect of CU on BIG6	-0.5315
	(0.2326)
Competitive effect of BIG6 on CU	-0.0838
	(0.2603)
Competitive effect of Desj on BIG6	0.4449
	(0.1894)
Competitive effect of BIG6 on Desj	-0.8761
	(0.1887)
Panel B: Demographic variables:	
Intercept	-0.0644
	(0.2538)
Population BIG6	-0.2281
	(0.2980)
Population CU/Desj	0.6792
	(0.2993)
Income per capita BIG6	0.2603
	(0.2689)
Income per capita CU/Desj	-0.0880
	(0.2215)
Unemployment BIG6	0.2424
	(0.2465)
Unemployment CU/Desj	-0.5184
	(0.2276)
Business activity BIG6	0.5116
	(0.3667)
Business activity CU/Desj	-0.1911
	(0.2593)
Proportion French BIG5/CU	-0.2851
	(0.2444)
Proportion French NBC/Desj	0.3046
	(0.2110)

Table D.3: Estimates of Market Presence Model (II)

This table shows estimates of the parameters of the profit function (Eq. (12)) in the market presence model. Some indicator variables are omitted due to perfect multicollinearity. Size and distance to headquarters are normalized by their mean. Standard errors, in parentheses, obtained using bootstrap.

Variable	Market presence model
Panel C: Provincial effects:	
British Columbia	0.1580
	(0.2453)
Manitoba	0.5877
	(0.2273)
New Brunswick	-0.8346
	(0.1718)
Newfoundland and Labrador	0.1955
	(0.2070)
Nova Scotia	0.9817
	(0.2716)
Quebec	0.3993
	(0.2169)
Saskatchewan	-0.1094
	(0.1957)
Panel D: Firm-level effects:	
National size	-0.1374
	(0.2691)
Regional size	0.2050
	(0.2652)
BMO	-0.3944
	(0.2311)
BNS	0.0152
	(0.2652)
CIBC	0.0974
	(0.2502)
CU	-0.4536
	(0.1861)
Desj	0.5411
	(0.1645)
NBC	-1.4373
	(0.2783)
distance to historical HQ	-0.7293
	(0.3489)
distance to historical HQ (square)	0.1252
	(0.3501)

D.3 Financial Product by Institution Type

Table D.4 classifies credit cards and lines of credit in our CFM sample by the offering institution type. While the majority of both financial products held are affiliated with banks and credit unions, there is a significant percentage of credit cards that are affiliated with non-depository institutions, such as retailers, gas stations, etc. This discrepancy between products is consistent with our estimation results, which show that credit card choice likelihood is much less dependent on the presence of a nearby bank branch than line of credit choice likelihood.

Table D.4: Financial Product by Institution Type

This table classifies credit cards and lines of credit held by households in our CFM sample by the offering institution type.

Institution Type	**Credit Cards**	**Lines of Credit**
Banks and credit unions	68.2%	95.9%
Retailers	22.0%	0.1%
Others	9.7%	4.0%