Digital Divide: Empirical Study of CIUS 2020*

Joann Jasiak† Peter MacKenzie‡ Purevdorj Tuvaandorj§

January 19, 2023

Abstract

As Canada and other major countries investigate implementing "digital money" or Central Bank Digital Currencies (CBDC), important questions need to be answered relating to the effect of demographic and geographic factors on the population's digital literacy. This paper uses the Canadian Internet Use Survey (CIUS) 2020 and survey versions of Lasso inference methods to assess the digital divide in Canada and determine the relevant factors that influence it. We find that a significant divide in the use of digital technologies, e.g., online banking and virtual wallet, continues to exist across different demographic and geographic categories. We also create a digital divide score that measures the survey respondents' digital literacy and provide multiple correspondence analyses that further corroborate these findings.

Keywords: Digital divide, inference after selection, Lasso, logistic regression, marginal effect, survey sample.

*The authors thank Henry Kim and the participants of the 13th CEQURA Conference 2022 and 48th ACEA Conference for helpful comments. We acknowledge financial support of the Catalyzing Interdisciplinary Research Clusters Initiative, York University. This research is part of the project "Digital Currencies", approved and funded by the office of VPRI and participating faculties of York University. The authors acknowledge access to the data provided by RDC and Statistics Canada (Project 21-MAPA YRK-721). The second author has won the best student presentation at the 48th ACEA Conference.

†York University, `jasiakj@yorku.ca`.

‡York University, `petem9@yorku.ca`.

§York University, `tpujee@yorku.ca`.

1 Introduction

As the Bank of Canada (BoC) investigates transitioning to a Central Bank Digital Currency (CBDC), the effect of the "digital divide" becomes a crucial factor in this transition. The digital divide arises between those that have been able to adapt to new digital technologies and those who have not. In a similar vein, it refers to the difference between those connected and possessing enough digital literacy to use the internet and other online technologies and those who are either not connected to, or do not have enough digital literacy, to use the internet. The future utility of these new digital currencies and digital modes of payment in Canada depends on internet connectivity and digital literacy. The digital divide is a prime cause of the limited uptake of new financial technologies amongst communities with poor access to the internet (Maniff, 2020).

This paper examines the Canadian Internet Use Survey (CIUS) 2020 *i.* to study internet access and the use of the internet in relation to financial technologies in Canada, and *ii.* to assess the digital divide in Canada and what effects this divide could have on the use of financial services, digital currencies, and digital payments. Moreover, this paper contributes to the debate on the future digitalization of money in Canada by providing data-based arguments for (or against) policy options such as introducing the CBDC (Christodorescu et al., 2020; Maniff, 2020).

Our empirical study also provides information on the outcomes and progress of the Government of Canada *High-Speed Access for All: Canada's Connectivity Strategy* by examining the internet connectivity of Canadians and the factors associated with internet access and usage.[1]Through Canada's connectivity strategy, the Canadian government recognizes the importance of high-speed internet for the economic and social well-being of Canadians.

Specifically, we use three main data analysis techniques to investigate how gender, age, race, education, income, geographical location, aboriginal identity, immigration status, and language impact connectivity and contribute to the digital divide in Canada. Knowing what factors affect the digital divide will allow for a fact based assessment of the government's strategy and help policymakers to make informed decisions regarding the

[1]Available at: https://ised-isde.canada.ca/site/high-speed-internet-canada/en

allocation of future resources.

We first perform survey-weighted logistic Lasso variable selection which reduces the dimensionality of the data. The logistic Lasso estimates are not suitable for inference since the Lasso selects the variables that have higher predictive power and trades off bias for variance. To address this issue, we use post-selection methods that are provably valid and allow for inference on the Lasso logit coefficients.

Second, we perform multiple correspondence analysis (MCA) and present the results in graphical form. The variable category groupings produced by MCA will be compared and contrasted with the logistic Lasso regression results.

In addition, we create a digital literacy score to measure the survey respondents' digital literacy. We compute the score of digital inclusion/digital divide and study its distributional properties in the entire and sub-samples of individuals with different demographic characteristics pertaining to the social groups with different origin, gender, age, location, and education level. We then examine the main explanatory variables that make up the score of digital inclusion among the variables in the dataset.

Over the past number of years, the use of physical cash in Canada has decreased dramatically; now, cash is used only in one of every three transactions (Huynh, 2017). This decrease in cash use is primarily due to the increased use of debit and credit cards. Recently, a digital technology known as cryptocurrency has made waves in the financial world. Many believe these cryptocurrencies could replace traditional currencies as the primary means of payment worldwide. Despite this speculation, the actual use of cryptocurrencies in Canada remains relatively low (Huynh et al., 2020).

This paper is motivated by cryptocurrencies and distributed ledger technology (DLT), commonly known as blockchain. According to Barr et al. (2021), DLTs have three main features; *i*. a ledger stored in multiple locations, *ii*. mechanisms to determine the accuracy of the data, *iii*. cryptographic security. These features separate cryptocurrencies like Bitcoin and Ethereum from types of digital currencies like PayPal and prepaid cards. These types of digital currencies do not travel from buyer to seller directly but instead travel through a "storage facility" during the electronic journey from buyer to seller (Bank of Canada, 2020*b*). A possible reason for the slower than expected uptake in cryptocurrencies is that these currencies do not satisfy the traditional definition of money; that is, *i*. medium of exchange, *ii*. store of value, *iii*. unit of account. As a result, cryptocurrencies have so far behaved more as speculative assets in the market, and their stability of value has been questionable (Adrian and Mancini-Griffoli, 2019).

Despite the slower than expected uptake of cryptocurrencies, countries worldwide have noticed increasing interest in these currencies and decreasing use of physical cash. These events have led many countries to start researching and, in some cases, implementing CBDC systems (Christodorescu et al., 2020). A CBDC system can be implemented in several ways. For example, according to Bordo and Levin (2017), a CBDC system could consist of private individuals having an account directly with the central bank, or commercial banks could have specialized accounts that hold the CBDC for individuals.

Many countries are concerned that if cryptocurrency use increases to the extent that these currencies are used in place of banks' credit systems, this could weaken a country's ability to implement monetary policy properly. If private digital currencies decrease banks' role in the monetary system, governments would no longer be able to affect the interest rates in the economy by controlling the rates at which banks borrow and lend (Brunnermeier et al., 2019). A CBDC could be a countermeasure to a scenario like this. Implementing a CBDC would create a direct channel for monetary policy to work through thereby restoring power to the country's monetary authority and not requiring direct regulation in regards to new emerging cryptocurrencies (Brunnermeier et al., 2019).

Since 2014, China, through its central bank, the People's Bank of China (PBOC) has implemented a form of digital currency known as the digital Renminbi (DCEP) (Barr et al., 2021). This CBDC is used in five major cities in China. Besides its CBDC, digital forms of payment are used more than physical forms in China, with WeChat and Alipay's digital wallets being the primary forms of payment (Brunnermeier et al., 2019).

Even if most major countries are not currently using a CBDC, many have begun to study the effects of their implementation and use. For example, the United States Federal Reserve released a study on CBDC in January 2022. The UK has studied the use of CBDC and determined that they are not prepared to transition to an entirely cashless society citing digital inclusion concerns (Barr et al., 2021). If the UK did begin to use a

CBDC they would use it along with physical cash (Barr et al., 2021). Canada has also stated it currently has no plans to introduce a CBDC (Carmichael, 2020).

The BoC has, however, introduced a "contingency plan" to implement a CBDC if physical cash was no longer used at all or if private digital currencies were used more than the Canadian dollar (Bank of Canada, 2020*a*). Right now, however, neither of these two scenarios seems likely in the near future. Very few Canadians are using cryptocurrencies, and 39% of Canadians "would not be able to cope" if cash was no longer used in Canada (Huynh et al., 2020).

There are several potential advantages to using CBDCs compared to the traditional banking model used in Canada. First, a CBDC could be used and held by both individuals and businesses, thus potentially cutting out the commercial intermediary (Brainard, 2019). A safe and secure way to hold money would increase competition with banks for individual deposits. Financial institutions wield a great deal of market power. A CBDC, through direct competition for deposits with these institutions, could be a cheaper and simpler method than developing competition policies (Usher et al., 2021). Also, CBDC could make payments cheaper and faster than the traditional bank wire system (Barr et al., 2021).

It is important to remember when researching the development and use of CBDCs that they cannot be used by or assist individuals that do not have access to the internet or a fair amount of digital literacy (Barr et al., 2021). This is one of the reasons why research in Canada on internet access and usage is essential. Christodorescu et al. (2020) suggest that a "two-tier hierarchical trust infrastructure" with the country's central bank being the main authority and other financial institutions being the intermediary certificate authority would potentially allow for an offline capability of a CBDC, thereby making offline payments possible. However, this would still require the individual to have access to the internet at some point.

High-speed internet increases social progress and improves overall quality of life (Jordan, 2019). It is therefore vital to increase access to high-speed internet in Canada, not only in the case of implementing a CBDC but, more generally, to enhance quality of life and social progress. The government understands this need in Canada and, in the 2019 budget, the federal government made a 1.7 billion dollar commitment to connecting all Canadians to reliable high-speed internet.

To increase access to high-speed internet we must understand the factors that influence whether an individual has internet access. Friedline et al. (2020) found that rural communities of colour have the lowest fintech rates. There is also a significant rural/urban divide regarding high-speed internet access in Canada, with only 37% of rural households having access to high-speed internet, compared to 97% of urban homes. The digital divide is even more significant for indigenous communities, with just 24% having access to high-speed internet.

Haight et al. (2014) use CIUS 2010 to examine the digital divide in Canada. Using standard regression techniques and several demographic and geographic variables, the authors predict internet access and social networking site usage. The study's main finding is that the digital divide continues to exist in Canada, with income, education, immigration status, urban living, and age all having a statistically significant effect on internet usage.

A lot may have changed over the last ten years regarding the population's digital literacy. Compared to the previous installment considered by Haight et al. (2014), CIUS 2020 provides more refined categories of variables and offers an up-to-date assessment of the current digital divide in Canada. In addition, we account for the dimensionality of the variables and use Lasso variables selection techniques that result in more predictive power for explaining the categorical variables of interest.

We confirm the importance of individual characteristics such as the age, income and education. The novelty of our approach is in revealing a significant impact of the visible minority status on the use of virtual wallets and through the use of interaction variables determining that older single individuals are significantly impacted by the digital divide. To the authors' knowledge, we are the first to use these interaction variables. The use of these interaction variables facilitated by the Lasso approach allows us to show the complexity in the persisting digital divide.

This paper is organized as follows. Section 2 provides a description of the CIUS 2020 data. Section 3 lays out the estimation and testing approach of the paper. Section 4 reports the main results. We conclude in Section 5. Appendix A provides a description of the sampling and weighting scheme used in CIUS 2020. Appendix B describes the technical aspects of methods used in the paper. Further details on the digital literacy score is given in Appendix C and additional inference results are reported in Appendix D.

2 Data description

CIUS 2020 is the most current data source on Canadian internet usage and comprises $17,409$ observations on households across Canada. The survey includes answers from Canadians 15 years of age and older living in one of Canada's ten provinces. The survey has a cross-sectional design, which uses both landline and cellular phone numbers from Statistics Canada's dwelling frame. Statistics Canada uses stratified sampling at the census metropolitan area and census agglomeration level. The survey is filled out online by one member of the household who is 15 years of age or older and the overall response rate to the survey is 41.6%.

The data is appropriately weighted using sample weights. The weight variables are provided by Statistics Canada [see Appendix A for the stratification scheme and survey weights]. Properly weighting the data allows for the sample of the Canadian population used in CIUS 2020 to represent the whole population. However, the data excludes aboriginal Canadians living on reserves and Canadians living in the territories. The sample weight variable used in CIUS 2020 is based on independent estimates from Statistics Canada for each province's various age and sex groups.

A limitation to CIUS 2020 is that it is conducted off reserve. The data on internet use in Northern Canada will be forthcoming in the following Northern CIUS. Therefore, the analysis based on CIUS 2020 can be considered the first step in a long-term project exploring Canada's digital divide.

We perform survey-weighted logistic Lasso variable selection/inference for multiple categorical dependent variables to determine the relevant demographic and geographic factors impacting the digital divide. The dependent variables corresponding to different model specifications and the independent variables are described in Sections 2.1 and 2.2, respectively.

2.1 Dependent variables

We consider the following logistic regression models where the dependent variables come from five questions asked to survey respondents.

- Model 1: "During the past three months have you used the internet from any location?"
- Model 2: "During the past three months have you conducted online banking?"
- Model 3: "During the past three months have you sent and received emails?"
- Model 4: "During the past twelve months have you used a virtual wallet to pay for goods over the internet?"
- Model 5: "During the past twelve months did you use a credit card previously entered or entered at the time of purchase to pay for goods over the internet?"

The internet use variable is a binary variable where respondents answered *Yes* or *No*. The dependent variables 2-5 are not binary, with each variable having three categories; 1) *Yes*; 2) *No*; 3) *Not stated*. We test the hypothesis of Independence of Irrelevant Alternatives (IIA) to determine if the *Not stated* category should be included.

The internet use question is used in this analysis to better determine what factors affect whether a person in Canada has access to the internet. Online banking, email use, and credit card dependent variables are used to judge what demographic factors affect a person's digital literacy. Determining what factors play a role in Canadian's digital literacy and internet connectivity could improve policymakers' ability to focus their efforts effectively when trying to reduce the digital divide in Canada. The analysis of these variables could also help the BoC know what groups of people will be affected by transitioning to a CBDC and a cashless economy.

The virtual wallet question determines what factors affect whether Canadians use virtual wallets when making payments. As the BoC investigates implementing a CBDC, knowing the demographic factors that affect whether someone uses virtual wallets plays an important role. The virtual wallet dependent variable question may be restrictive because many people use virtual wallets to hold digital currencies as speculative assets rather than a liquid currency to spend online.

2.2 Independent variables

The independent variables in this analysis are the same for the logistic regression models 1 to 5. These variables are income, education, employment status, aboriginal identity, visible minority status, immigration status, age, gender, location, type of household, language spoken at home, and province. All of them have two or more categories which are reported in the regression tables.

Some independent variables included an answer category *Not stated*. Unlike the case with the dependent variables, for these independent variables, we have still included this category in the regression. Many respondents who answered *Not stated* to one question answered many others; therefore, removing their answers may bias results. There are $12,124$ observations for the logit models 4 and 5. In the email use and online banking models there are $17,268$ and $17,135$ observations, and the internet use model includes all $17,409$ observations in the survey.

The categories associated with a representative individual are omitted in each model as the comparison category for the logistic regression. That representative individual has the following characteristics - urban, age 45-54, male, non-aboriginal, english and non-official language speaker, not employed, some post-secondary education, not a visible minority, family household with children under 18, income of $\$52,204$–$\$92,485$, landed immigrant (recent immigrant), and from the province Alberta.

3 Survey-weight adjusted logit Lasso inference

The survey weights play an important role as they ensure that the results of the survey can be generalized to the entire population of Canadians. However, the existing Lasso-based estimation and inference methods, including the widely-used logit Lasso variable selection, are not directly adjustable for survey weights. To overcome this gap in the literature, this paper uses a survey-weighted logistic Lasso (`svy LLasso` hereafter) variable selection for binary choice models which extends the logistic Lasso, to a survey environment. The inference procedures based on `svy LLasso` estimator are further studied by Jasiak and Tuvaandorj (2023).

There are 41 independent categorical variables, many of which are expected to have negligible or no effect on the dependent variables considered. Moreover, some of the independent variables may interact with one another; for example, household type and income variables may have a cross-effect on the dependent variables such as internet use and online banking. Taking into account the second-order interactions gives 674 control variables which are large relative to the sample size. Yet, there is no a priori guidance on which variables should enter the model. Due to these reasons, we take the logistic Lasso approach, which is well-suited for this problem, known to have optimality properties under sparsity assumption and offers an automatic variables selection (Belloni et al., 2014; Mullainathan and Spiess, 2017)

Let θ denote the parameter vector of the logistic regression including the slope parameters β and intercept α. The (non-negative) tuning parameter used in the Lasso is denoted by λ. A survey-weighted logistic Lasso is based on minimizing the weighted negative log-likelihood function $L(\theta)$ subject to ℓ_1 penalty on the parameter vector:

$$\min_{\theta=(\alpha,\beta')'\in\mathbb{R}^{p+1}} \left(-L(\theta) + \lambda \sum_{j=1}^{p} |\beta_j|\right), \tag{3.1}$$

where $L(\theta) = n^{-1} \sum_{i=1}^{n} w_i(y_i x_i'\theta - \log(1+\exp(x_i'\theta)))$, $x_i'\theta = \alpha + \tilde{x}_i'\beta$, and $(y_i, x_i')' \in \mathbb{R}^{p+1}$, $i = 1, \ldots, n$, are the pairs of dependent and independent observations with the corresponding strictly positive survey weights $w_i, i = 1, \ldots, n$. The sampling scheme used in CIUS 2020 is akin to simple stratified sampling (Cameron and Trivedi, 2009), so we treat w_i as given, and $\{(y_i, x_i')'\}_{i=1}^{n}$ as independent.

Note that, as is standard in the Lasso literature, only the "slope" parameters in $\beta = (\beta_1, \ldots, \beta_p)'$ are penalized in (3.1). We fit the model (3.1) using the `R` package `glmnet`. For the tuning parameter λ, we use the package's default value chosen by 10-fold cross validation with the loss function "`auc`" (area under the ROC curve).

The logistic Lasso estimates are not suitable for inference since the Lasso selects the variables that have higher predictive power and trades off bias for variance. Due to its computational and conceptual simplicity, we use a survey-version of the debiased Lasso (DB) method proposed by Zhang and Zhang (2014), Javanmard and Montanari (2014) and Xia et al. (2020) as the main inferential tool for the logit coefficients and the average marginal effects (AMEs) after variable selection by `svy LLasso`. It is based on the following one-step estimator constructed from the initial `svy LLasso` estimator $\hat{\theta}$:

$$\tilde{\theta}^{DB} \equiv \hat{\theta} + H(\hat{\theta})^{-1} S(\hat{\theta}),$$

where $H(\cdot)$ and $S(\cdot)$ are the (sample) Hessian and the score functions for the full parameter vector in the logistic model. The one-step (or DB) estimator removes the bias of the initial `svy LLasso` estimator and has an asymptotic normal distribution, thus facilitating standard t-ratio-based inference.

In addition, we consider the survey-logit versions of the "selective inference" (SI) procedure proposed by Lee et al. (2016) and Taylor and Tibshirani (2018), and the $C(\alpha)$ (or Neyman orthogonalization) method after Lasso variable selection proposed by Belloni et al. (2016) to make inference on the model parameters and AMEs. The former method is based on a one-step estimator denoted as $\tilde{\theta}^{SI}$ and the test statistic in the latter is labeled as C_{α}. See Appendix B.1 for a brief description of these methods and Jasiak and Tuvaandorj (2023) for further theoretical analyses.

4 Empirical results

This section reports the empirical results. Section 4.1 presents the results from the weight-adjusted Lasso logit estimation of models 1 to 5. `svy LLasso` estimates and the test results based on the debiased Lasso estimates of the model coefficients and AMEs, $\tilde{\theta}^{DB}$ and $\widetilde{\mathrm{AME}}^{DB}$, are reported in Tables 1-5 below. The outcomes of the selective inference and $C(\alpha)$ test results are generally consistent with the debiased Lasso test results, thus are relegated to Tables 11-15 in Appendix D.

An analysis of possible interaction effects is provided in Section 4.2. We report the outcomes of the multiple correspondence analysis in Section 4.3 and present the digital divide score in Section 4.4.

As stated in Section 2, the online banking, email use, digital wallet, and credit card dependent variables originally had three categories: *Yes*, *No*, and *Not stated*. We use the survey-weighted Hausman-McFadden test for IIA hypothesis to see if we can remove the *Not stated* observations from the logistic regression. The online banking model has Hausman-McFadden statistic of 0.05 with a p-value of 1. Therefore the results for the model show strong evidence in favour of IIA, so we use the restricted model specification removing the *Not stated* observations from the model. The dependent variables email use, virtual wallet, and credit card use have Hausman-McFadden statistics $-0.95, -0.77$, and -1.29. Negative values of the statistic are viewed as evidence in favour of the null hypothesis (Hausman and McFadden, 1984), we proceed with the restricted model with *Not stated* removed.

4.1 svy LLasso regressions

Model 1: Internet use. The results from the internet use model reported in Table 1 display the significant variables that determine whether or not someone is connected and using the internet. Unlike the other regression models which measure respondents ability to use digital technologies, the internet use model directly determines what variables affect whether or not someone is connected to the internet.

Concerning the location variable, the estimation result shows that those living in rural areas across Canada are less likely to have used the internet in the previous three months. The corresponding AME is equal to -0.017, meaning the probability of internet connectivity decreases by 1.7% for a rural resident compared to an urban. This rural/urban divide in access to the internet is consistent with *Canada's connectivity strategy* findings and the impetus for the significant investments from the federal government to improve internet connectivity in rural communities. The persistence of this rural/urban divide in internet connectivity despite the large investments already made by the federal government highlights the challenges rural residents face with internet connectivity and the digital divide.

`svy LLasso` selects all age group categories. Comparing the five age categories to the omitted age group *45-54*, we see that the three younger age groups comprising ages *15-44* have positive coefficients and the two older age categories comprising *55-64* and *65 and* *older* have negative coefficients. Older age groups were less likely to have been connected to the internet than the omitted age category. The age group categories with the largest absolute value in AME are the age group category *25-34* and the oldest (*65 and older*). The second youngest age group category is 5.4 percentage points more likely to be connected and the oldest age group category is 8.2 percentage points less likely to be connected to the internet than the base age category *45-54*. The AME for both of these age categories are highly significant.

The estimation results indicate a correlation between various demographic factors and internet usage in Canada. Specifically, individuals who were employed, speak English as their primary language, possess university degrees, and have high incomes are found to have a higher likelihood of internet usage within the past three months. Conversely, individuals residing in the province of Quebec, who are older, have a high school education or less, identify as a visible minority, are single, and have low incomes, are found to have a lower likelihood of internet usage. These findings suggest the existence of a persistent digital divide even in terms of basic internet access within Canada.

The results for internet connectivity are generally consistent with and reinforce the findings of past research on internet connectivity in Canada (Haight et al., 2014; Friedline et al., 2020; Jordan, 2019) Younger, highly educated, high income Canadians are most likely to use the internet.

Model 2: Online banking. The online banking model is used to measure Canadians' ability to use digital technology, specifically digital financial technology. In theory, there would likely be many similarities between the online banking system currently used by the prominent Canadian banks and a system designed by the Canadian government or BoC to run a CBDC system.

The results for the dependent variable online banking reported in Table 2 show that younger, employed, high-income and university-educated Canadians are most likely to use online banking. Low education attainment, low income, being a visible minority and being 55 or older negatively affected the likelihood of an individual using online banking. The variables with the largest (in absolute value) AMEs on whether a person uses online banking are the age category *65 and older*, whether or not a person is employed, and if their educational attainment was a *High school or less*.

People in the oldest age category are found to be 15.4 percentage points less likely than the base age category *45-54* to use online banking according to the debiased Lasso AME estimates. An employed person was found to be 10.6 percent more likely than an unemployed person to use online banking. People with low educational attainment of a *High school or less* are 11.4 percentage points less likely than someone with *Some post-secondary* education to use online banking. The variables *Location, Gender, Aboriginal identity*, and *Province* were not selected by `svy LLasso`.

Model 3: Email use. Tables 3 reports the logit regression models for the dependent variable email use. The email use variable is used in the same way as the online banking variable to determine what factors affect Canadians' digital literacy.

In Table 3, we see many of the same variables selected by `svy LLasso` as in the internet use and online banking models. However, there are some interesting differences between the models. The category *Rural* of the location variable is selected and the coefficient is negative in the email use model but not selected in the online banking model. This difference in the two models may be due to geography. Rural Canadians may use online banking if they live far from a bank location. Canadians living in rural locations may also be less likely than urban residents to have employment requiring extensive use of email.

The category *Female* is selected in the email use model and not the online banking one. However, the variable's debiased Lasso AME is quite small. A possible explanation for this is the number of women working office jobs compared to men, who may be more likely to work blue-collar jobs where email is not as frequently required. `svy LLasso` chooses more variables in the email use model than the previous two, with every variable having at least one of its categories chosen.

In Table 3, the variable with the largest estimated AME (in absolute value) is the language variable category *English, French, and Non-official language*. However, despite the large AME estimate, the variable is not selected by `svy LLasso`. The oldest age category, *65 and older*, has the second largest AME, those *65 and older* are 10 percentage points less likely than those in the age group *45-54* to send and receive emails.

Educational attainment has a significant effect on the email use model. Those holding a *University degree* or higher were much more likely than those with *Some post-secondary education* to use email and those with a *High school or less* were much less likely to use email. Email is commonly used in jobs that require a higher degree of education. Workplace requirements could explain the large difference we see in the likelihood of email use depending on a person's educational attainment.

Model 4: Virtual wallet. Whether or not someone has made payments with money from a virtual wallet is one of the most relevant variables in our model concerning research on digital currencies in Canada. The previous regression models have been used to measure Canadians' internet connectivity and digital literacy. The virtual wallet model will show what factors currently affect the uptake of digital forms of payment in Canada.

The logistic lasso regression results in Table 4 show the coefficients of the variables selected by `svy LLasso`. Rural Canadians are less likely to use a digital wallet than urban residents. All age group categories, excluding those aged *35-44* were selected. Younger Canadians have the highest probability of using a virtual wallet. The age group, *15-24*, has a 11.2 percentage point increase in the likelihood of using a virtual wallet than the base age group *45-54*. The older age group categories both have negative coefficients. The age group *65 and older* is the least less likely to use a virtual wallet compared to the age group *45-54*. The debiased Lasso AME for the oldest age group shows that those *65 and older* are 8.3 percentage points less likely to use a virtual wallet than those *45-54*.

The coefficient for *Visible minority* is chosen by `svy LLasso` and has a positive AME on the use of a digital wallet. This result is striking, considering the variable category *Visible minority* in previous results has either not been selected by `svy LLasso` or had a negative effect on the dependent variable. The AME shows that a person identifying as a visible minority is 5.2 percentage points more likely than a person who is not a visible minority. The positive *Visible minority* coefficient might reflect the increased use of foreign cryptocurrencies like Alipay and WeChat pay by visible minorities in Canada. The only significant income category is the highest. Canadians with income equal to or higher than $146,560 are found to have a higher probability of using a virtual wallet than the base income category ($52,204–$92,485).

The age and income variables have the largest AMEs in absolute value. The education variable *University degree* is choosen by `svy LLasso` and has a positive coefficient. Those with a *University degree* are shown to be 2.8 percentage points more likely to use a virtual wallet than those with *Some post-secondary* education. In contrast to previous logistic lasso regression results, the variable employment was not selected by `svy LLasso`.

Model 5: Credit card. The implementation of CBDC would likely involve a payment card component similar to the debit and credit cards Canadians use now. Understanding what factors influence whether someone uses a credit card to make purchases online is essential in the context of implementing a CBDC.

The logistic Lasso regression results in Table 5 show that `svy LLasso` has selected fewer variables than the previous models with the exception of the virtual wallet model. The youngest age group is the only statistically significant category for the age group variable. The age group category *15-24* has a negative and significant coefficient. The AME shows that the youngest age group is 8.8 percentage points less likely to use a credit card for online purchases than the base age category *45-54*. The youngest age category being less likely to use a credit card than those in *45-54* is reasonable given that many people do not use credit cards until later in life. `svy LLasso` selected both categories of the education variable with low educational attainment of a *High school or less* having a negative coefficient and high educational attainment of a *University degree* with a positive coefficient.

`svy LLasso` selected the lowest income category and the province of Quebec. Both have negative coefficients meaning that people with low income or from Quebec are less likely to use credit cards for online purchases than the comparison categories. The credit card model results show that Canadians from Quebec, with low educational attainment, low income, and that speak French are less likely to use a credit card for an online purchase. In contrast, Canadians that speak English, are employed, have a university degree, live in a family household without children under eighteen, and from Ontario are more likely to use credit card for online purchases.

Further remarks. `svy LLasso` selected at least one of the age variable categories in every regression specification. The younger age categories were more likely to be connected to the internet and use services such as online banking, email, and a virtual wallet than the comparison age group *45-54*. The oldest age group, *65 years and older*, was less likely to use the internet and other digital services.

The credit card model was the only case where this relationship between the dependent variable and the age group categories did not hold. In this model, the youngest age group category was selected by `svy LLasso` and found to have a negative effect on credit card usage. Although not consistent with the other models designed to measure a person's digital literacy, this result was not surprising. Younger people are generally less likely to make purchases requiring a credit card than older people who must show good credit scores and credit history to make large purchases such as cars and homes.

`svy LLasso` selected at least one category from the variables *Employment, Education,* and *Income* in almost all of the models. Higher educational attainment is found to have positive effects on the probability of using the internet and having a high degree of digital literacy. People with higher income and education were more likely to be connected to the internet and have sufficient digital literacy to effectively use it. Those employed were also more likely to use the internet and conduct online banking and email. Being employed was not selected by `svy LLasso` in the virtual wallet model and was selected in the credit card model, but was not significant.

`svy LLasso` did not select the variable *Immigration status* in any of the regression models. The lack of significance of the immigration status variable was not expected. It was assumed that new immigrants to Canada would have a more challenging time accessing the internet and may have a lower degree of digital literacy than Canadian-born residents. The variables lack of significance revealed in our models could be due to Canadian government immigration policies (such as the Global Skills Strategy) helping highly skilled workers immigrate to Canada. It is also possible that most Canadian immigrants may have used the internet and other services like email during the process of becoming a Canadian citizen.

As mentioned in Section 3, the DB test results of this section, and the $C(\alpha)$ and SI test results reported in Appendix D are based on `svy LLasso` estimates with a λ chosen by cross-validation. `svy LLasso` estimates with a fixed λ yielded qualitatively similar results which are available upon request.

Table 1: Lasso Logistic Regression Results for Internet Use Dependent Variable

Variables	Categories	`svy LLasso`	$\tilde{\theta}^{DB}$	p-value	$\widetilde{\text{AME}}^{DB}$	p-value
Intercept		3.428	3.246***	0.000	–	–
Location	Urban (omitted)	–	–	–	–	–
	Rural	−0.225	−0.287***	0.001	−0.017***	0.001
Age	15–24	0.627	1.235***	0.000	0.054***	0.000
	25–34	0.161	0.683**	0.007	0.033*	0.014
	35–44	0.038	0.548*	0.016	0.027*	0.035
	45–54 (omitted)	–	–	–	–	–
	55–64	−0.721	−0.527**	0.003	−0.032*	0.014
	65 and older	−1.570	−1.262***	0.000	−0.082***	0.000
Gender	Male (omitted)	–	–	–	–	–
	Female	0.013	0.099	0.200	0.006	0.211
Aboriginal identity	Non-aboriginal (omitted)	–	–	–	–	–
	Aboriginal	–	−0.497*	0.021	−0.032*	0.011
Language	English	0.354	0.598*	0.037	0.035*	0.044
	French	–	0.246	0.435	0.013	0.464
	Non-official language	–	0.065	0.836	0.004	0.842
	English and French	–	0.793	0.124	0.036	0.231
	English and Non-official language (omitted)	–	–	–	–	–
	French and Non-official language	–	−0.533	0.544	−0.035	0.495
	English, French and Non-official language	–	−1.434	0.193	−0.118·	0.067
Employment	Employed	0.514	0.574***	0.000	0.032***	0.000
	Not employed (omitted)	–	–	–	–	–
Education	High school or less	−0.911	−0.971***	0.000	−0.058***	0.000
	Some post-secondary (omitted)	–	–	–	–	–
	University degree	0.451	0.519***	0.000	0.027***	0.000
Visible minority	Visible minority	−0.048	−0.352*	0.037	−0.021*	0.034
	Not a visible minority (omitted)	–	–	–	–	–
Household type	Family with children under 18 (omitted)	–	–	–	–	–
	Family without children under 18	–	−0.029	0.872	−0.002	0.875
	Single	−0.596	−0.665***	0.000	−0.043***	0.001
	Other household type	–	0.149	0.635	0.008	0.656
Income	$52,203 and lower	−0.536	−0.475***	0.000	−0.028***	0.000
	$52,204–$92,485 (omitted)	–	–	–	–	–
	$92,486–$146,559	–	0.092	0.469	0.005	0.486
	$146,560 and higher	0.359	0.547***	0.001	0.028***	0.001
Immigration	Landed immigrant (omitted)	–	–	–	–	–
	Non-landed immigrant	–	−0.258	0.176	−0.014	0.211
Province	NL	–	−0.31	0.111	−0.019·	0.091
	PEI	–	−0.272	0.155	−0.017	0.135
	NS	–	−0.298	0.123	−0.018	0.104
	NB	–	−0.101	0.586	−0.006	0.585
	QC	−0.296	−0.448*	0.026	−0.027*	0.034
	ON	0.039	−0.018	0.911	−0.001	0.913
	MB	–	−0.501*	0.013	−0.032**	0.006
	SK	–	−0.413*	0.037	−0.026*	0.024
	BC	0.031	0.095	0.602	0.005	0.613
	AB (omitted)	–	–	–	–	–

Notes: $n = 17,409$. The comparison category for each variable is labeled omitted in paranthesis. $\tilde{\theta}^{DB}$ and $\widetilde{\text{AME}}^{DB}$ denote the debiased Lasso estimates of the logit parameter and AME respectively. " – " denotes the variables not selected by `svy LLasso` or "not computed" because the variable category is used as a comparison. Significance codes are: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

4.2 Interaction effects

The inclusion of interaction terms in `svy LLasso` can improve the model's ability to capture complex relationships between variables. We examine whether the second-order specification with interaction terms could be more appropriate than the first-order specification in models 1 to 5. To this end, we use the `R` package `polywog` to compare the mean-squared 10-fold cross-validation (CV) error of the adaptive Lasso estimator (see Bühlmann and van de Geer (2011) for a detailed treatment) for both specifications.

Table 6 reports the result. The linear specification is selected for models 1, 4, and 5, while the second-order specification is chosen for models 2 and 3. The difference between the two specifications is minimal for models 1 and 3. For models 2 and 3, after fitting the second-order model with 674 variables by `svy LLasso` we make inference on the coefficients using the debiased Lasso procedure.

Table 7 and 8 display the interaction results for the online banking and email use models, respectively. Only the variables deemed significant at the 5% level based on the estimated p-values for the coefficients are displayed. We did not display the significant interactions variables that involve *Not stated* answers because of the lack of interpretability. In addition to the two non-constant variables *High school or less* and *Visible minority* which had negative effects on the online banking, four interaction variables, the age group category *15-24* interacted with *Family without children under 18*, *65 and older* and *Single*, and *Female* interacted with *English* and *Employed* are both selected by `svy LLasso` and significant. The signs of the selected coefficients appear to be reasonable.

It is clear that that most variables that are significant at 5% level are not significant at 1%. The age group category *15-24*, when interacted with *Family without children under 18* and *Other household type*, has highly significant positive effects, and when interacted with *English and French*, has a negative effect on online banking. Interestingly, *English* interacted with *High school or less* and *Non-landed immigrant* appear to have highly significant positive effects.

In contrast to the online banking model with interactions, only the interaction of the age category *65 and older* and the income category *$52,203 and lower* is both selected by `svy LLasso` and significant, and two interaction variables *(Rural)*×*(65 and older)* and *(Visible*

Table 2: Lasso Logistic Regression Results for Online Banking Dependent Variable

Variables	Categories	`svy LLasso`	$\tilde{\theta}^{DB}$	p-value	$\widetilde{\text{AME}}^{DB}$	p-value
Intercept		1.120	0.625^{**}	0.009	–	–
Location	Urban (omitted)	–	–	–	–	–
	Rural	–	−0.092	0.154	−0.015	0.167
Age	15–24	–	0.045	0.721	0.007	0.734
	25–34	0.414	0.637^{***}	0.000	0.092^{***}	0.000
	35–44	0.267	0.540^{***}	0.000	0.079^{***}	0.000
	45–54 (omitted)	–	–	–	–	–
	55–64	−0.071	-0.324^{***}	0.000	-0.052^{***}	0.001
	65 and older	−0.718	-0.873^{***}	0.000	-0.154^{***}	0.000
Gender	Male (omitted)	–	–	–	–	–
	Female	–	0.089	0.107	0.014	0.123
Aboriginal identity	Non-aboriginal (omitted)	–	–	–	–	–
	Aboriginal	–	−0.248	0.105	−0.040	0.106
Language	English	0.000	0.509^{**}	0.005	0.081^{**}	0.007
	French	–	0.598^{**}	0.005	0.086^{*}	0.012
	Non-official language	–	$0.337^{\cdot}$	0.090	0.050	0.122
	English and French	–	0.239	0.526	0.036	0.561
	English and Non-official language (omitted)	–	–	–	–	–
	French and Non-official language	–	−0.280	0.648	−0.046	0.647
	English, French and Non-official language	–	−0.127	0.858	−0.020	0.861
Employment	Employed	0.662	0.653^{***}	0.000	0.106^{***}	0.000
	Not employed (omitted)	–	–	–	–	–
Education	High school or less	−0.637	-0.686^{***}	0.000	-0.114^{***}	0.000
	Some post-secondary	–	–	–	–	–
	University degree	0.331	0.409^{***}	0.000	0.062^{***}	0.000
Visible minority	Visible minority	−0.135	-0.303^{**}	0.003	-0.048^{**}	0.005
	Not a visible minority (omitted)	–	–	–	–	–
Household type	Family with children under 18 (omitted)	–	–	–	–	–
	Family without children under 18	0.078	0.305^{***}	0.000	0.048^{***}	0.001
	Single	−0.166	−0.137	0.137	−0.022	0.167
	Other household type	–	0.372^{*}	0.042	$0.054^{\cdot}$	0.068
Income	\$52,203 and lower	−0.265	-0.252^{***}	0.001	-0.041^{**}	0.002
	\$52,204–\$92,485 (omitted)	–	–	–	–	–
	\$92,486–\$146,559	–	0.123	0.132	0.019	0.153
	\$146,560 and higher	0.086	0.252^{**}	0.004	0.039^{**}	0.006
Immigration	Landed immigrant (omitted)	–	–	–	–	–
	Non-landed immigrant	–	−0.082	0.463	−0.013	0.486
Province	NL	–	−0.015	0.905	−0.002	0.909
	PEI	–	−0.068	0.604	−0.011	0.615
	NS	–	−0.079	0.540	−0.012	0.552
	NB	–	−0.068	0.603	−0.011	0.614
	QC	–	−0.101	0.460	−0.016	0.474
	ON	–	−0.032	0.748	−0.005	0.758
	MB	–	-0.383^{**}	0.004	-0.063^{**}	0.003
	SK	–	−0.114	0.377	−0.018	0.389
	BC	–	−0.010	0.931	−0.002	0.934
	AB (omitted)	–	–	–	–	–

Notes: $n = 17,135$.

Table 3: Lasso Logistic Regression Results for Email Use Dependent Variable

Variables	Categories	`svy LLasso`	$\tilde{\theta}^{DB}$	p-value	$\widetilde{\text{AME}}^{DB}$	p-value
Intercept		1.960	1.964***	0.000	–	–
Location	Urban	–	–	–	–	–
	Rural	−0.158	−0.207**	0.005	−0.021**	0.007
	15–24	0.390	0.658***	0.000	0.058***	0.000
	25–34	0.444	0.742***	0.000	0.063***	0.000
	35–44	0.294	0.585***	0.000	0.051***	0.000
	45–54 (omitted)	–	–	–	–	–
	55–64	−0.425	−0.343**	0.004	−0.035**	0.009
	65 and older	−1.036	−0.899***	0.000	−0.100***	0.000
Gender	Male (omitted)	–	–	–	–	–
	Female	0.087	0.151*	0.021	0.015*	0.025
Aboriginal identity	Non-aboriginal (omitted)	–	–	–	–	–
	Aboriginal	–	−0.473**	0.008	−0.051**	0.004
Language	English	0.402	0.301	0.179	0.030	0.207
	French	–	−0.118	0.640	−0.012	0.644
	Non-official	−0.047	−0.225	0.353	−0.023	0.357
	English and French	–	0.426	0.327	0.037	0.395
	English and Non-official language (omitted)	–	–	–	–	–
	French and Non-official language	–	−0.302	0.669	−0.032	0.656
	English, French and Non-official language	–	−1.908*	0.019	−0.272***	0.001
Employment	Employed	0.411	0.457***	0.000	0.045***	0.000
	Not employed (omitted)	–	–	–	–	–
Education	High school or less	−0.790	−0.851***	0.000	−0.088***	0.000
	Some post-secondary (omitted)	–	–	–	–	–
	University degree	0.750	0.828***	0.000	0.072***	0.000
Visible minority	Visible minority	−0.192	−0.346**	0.008	−0.035*	0.011
	Not a visible minority (omitted)	–	–	–	–	–
Household type	Family with children under 18 (omitted)	–	–	–	–	–
	Family without children under 18	–	−0.055	0.644	−0.005	0.655
	Single	−0.456	−0.571***	0.000	−0.062***	0.000
	Other household type	–	−0.052	0.824	−0.005	0.828
Income	$52,203 and lower	−0.383	−0.323***	0.000	−0.033***	0.000
	$52,204–$92,485 (omitted)	–	–	–	–	–
	$92,486–$146,559	–	0.088	0.371	0.008	0.391
	$146,560 and higher	0.329	0.441***	0.000	0.040***	0.000
Immigration	Landed immigrant (omitted)	–	–	–	–	–
	Non-landed immigrant	0.016	0.147	0.304	0.015	0.311
Province	NL	–	−0.240	0.120	−0.025	0.111
	PEI	–	−0.174	0.265	−0.018	0.260
	NS	–	−0.387*	0.012	−0.041**	0.008
	NB	–	−0.251·	0.098	−0.026·	0.090
	QC	−0.154	−0.326*	0.044	−0.033·	0.050
	ON	0.164	0.069	0.577	0.007	0.582
	MB	–	−0.466**	0.004	−0.050**	0.002
	SK	–	−0.364*	0.021	−0.038*	0.015
	BC	0.236	0.260·	0.077	0.024·	0.073

Notes: $n = 17,268$.

Table 4: Lasso Logistic Regression Results for Virtual Wallet Dependent Variable

Variables	Categories	`svy LLasso`	$\tilde{\theta}^{DB}$	p-value	$\widetilde{\text{AME}}^{DB}$	p-value
Intercept		−2.038	−2.650***	0.000	–	–
Location	Urban (omitted)	–	–	–	–	–
	Rural	−0.220	−0.609***	0.000	−0.057***	0.000
Age	15–24	0.300	0.867***	0.000	0.112***	0.000
	25–34	0.207	0.619***	0.000	0.075***	0.000
	35–44	–	0.334**	0.005	0.039**	0.003
	45–54 (omitted)	–	–	–	–	–
	55–64	−0.308	−0.608***	0.000	−0.057***	0.000
	65 and older	−0.548	−1.009***	0.000	−0.083***	0.000
Gender	Male (omitted)	–	–	–	–	–
	Female	–	−0.091	0.280	−0.010	0.277
Aboriginal identity	Non-aboriginal (omitted)	–	–	–	–	–
	Aboriginal	–	0.040	0.872	0.004	0.869
Language	English	–	0.129	0.596	0.014	0.597
	French	–	0.132	0.653	0.015	0.640
	Non-official language	–	−0.410	0.116	−0.040	0.153
	English and French	–	0.105	0.829	0.012	0.822
	English and Non-official language (omitted)	–	–	–	–	–
	French and Non-official language	–	−0.618	0.448	−0.055	0.534
	English, French and Non-official language	–	−0.907	0.359	−0.073	0.495
Employment	Employed	–	0.020	0.853	0.002	0.852
	Not employed (omitted)	–	–	–	–	–
Education	High school or less	–	−0.066	0.568	−0.007	0.568
	Some post-secondary (omitted)	–	–	–	–	–
	University degree	0.027	0.254**	0.009	0.028**	0.008
Visible minority	Visible minority	0.162	0.453***	0.001	0.052***	0.001
	Not a visible minority (omitted)	–	–	–	–	–
Household type	Family with children under 18 (omitted)	–	–	–	–	–
	Family without children under 18	–	0.064	0.547	0.007	0.543
	Single	–	0.033	0.797	0.004	0.793
	Other household type	–	0.121	0.621	0.014	0.606
Income	$52,203 and lower	–	0.080	0.551	0.009	0.541
	$52,204–$92,485 (omitted)	–	–	–	–	–
	$92,486–$146,559	–	0.155	0.203	0.017	0.189
	$146560 and higher	0.233	0.563***	0.000	0.066***	0.000
Immigration	Landed immigrant (omitted)	–	–	–	–	–
	Non-landed immigrant	–	0.129	0.372	0.014	0.380
Province	NL	–	−0.270	0.178	−0.027	0.214
	PEI	–	−0.311	0.131	−0.030	0.169
	NS	–	−0.282	0.163	−0.028	0.198
	NB	–	−0.065	0.757	−0.007	0.760
	QC	–	−0.126	0.532	0.013	0.539
	ON	–	0.043	0.759	0.005	0.756
	MB	–	−0.420*	0.032	−0.040·	0.058
	SK	–	−0.215	0.270	−0.022	0.298
	BC	–	0.080	0.636	0.009	0.627
	AB (omitted)	–	–	–	–	–

Notes: $n = 12,124$.

Table 5: Lasso Logistic Regression Results for Credit Card Use Dependent Variable

Variables	Categories	`svy LLasso`	$\tilde{\theta}^{DB}$	p-value	$\widetilde{\text{AME}}^{DB}$	p-value
Intercept		1.334	1.100***	0.000	–	–
Location	Urban (omitted)	–	–	–	–	–
	Rural	–	−0.125	0.134	−0.020	0.140
Age	15–24	−0.363	−0.522***	0.000	−0.088***	0.000
	25–34	–	0.055	0.630	0.008	0.644
	35–44	–	0.135	0.188	0.020	0.213
	45–54 (omitted)	–	–	–	–	–
	55–64	–	−0.022	0.830	−0.003	0.835
	65 and older	–	−0.053	0.653	−0.008	0.662
Gender	Male (omitted)	–	–	–	–	–
	Female	–	−0.004	0.958	−0.001	0.959
Aboriginal identity	Non-aboriginal (omitted)	–	–	–	–	–
	Aboriginal	–	0.198	0.306	0.029	0.347
Language	English	0.216	0.019	0.928	0.003	0.933
	French	−0.192	−0.679**	0.006	−0.116**	0.005
	Non-official language	–	−0.044	0.844	−0.007	0.849
	English and French	–	−0.185	0.646	−0.030	0.644
	English and Non-official language (omitted)	–	–	–	–	–
	French and Non-official language	–	−0.777	0.263	−0.141	0.202
	English, French and Non-official language	–	−1.352·	0.088	−0.266*	0.035
Employment	Employed	0.002	0.148·	0.083	0.023·	0.091
	Not employed (omitted)	–	–	–	–	–
Education	High school or less	−0.411	−0.453***	0.000	−0.073***	0.000
	Some post-secondary (omitted)	–	–	–	–	–
	University degree	0.357	0.490***	0.000	0.073***	0.000
Visible minority	Visible minority	–	−0.235*	0.044	−0.037*	0.046
	Not a visible minority (omitted)	–	–	–	–	–
Household type	Family with children under 18 (omitted)	–	–	–	–	–
	Family without children under 18	0.035	0.335***	0.000	0.051***	0.000
	Single	–	0.317**	0.002	0.046**	0.005
	Other household type	–	0.161	0.430	0.024	0.463
Income	\$52,203 and lower	−0.073	−0.286**	0.004	−0.046**	0.005
	\$52,204–\$92,485 (omitted)	–	–	–	–	–
	\$92,486–\$146,559	–	0.097	0.306	0.015	0.328
	\$146,560 and higher	–	0.084	0.393	0.013	0.415
Immigration	Landed immigrant (omitted)	–	–	–	–	–
	Non-landed immigrant	–	0.151	0.232	0.024	0.237
Province	NL	–	−0.287·	0.081	−0.047·	0.072
	PEI	–	0.078	0.637	0.012	0.655
	NS	–	−0.045	0.783	−0.007	0.788
	NB	–	−0.012	0.944	−0.002	0.946
	QC	−0.112	−0.042	0.798	−0.007	0.810
	ON	0.029	0.241*	0.042	0.037·	0.051
	MB	–	0.035	0.829	0.005	0.837
	SK	–	0.022	0.891	0.003	0.895
	BC	–	0.211	0.131	0.031	0.161
	AB (omitted)	–	–	–	–	–

Notes: $n = 12,124$.

Table 6: Order selection

	CV error				
Models	1	2	3	4	5
1st order	0.395	0.955	0.644	0.684	0.942
2nd order	0.396	0.944	0.643	0.692	0.945
sample size	17409	17135	17268	12124	12124

Notes: The table reports the mean-squared 10-fold cross-validation error for first-order model with 41 covariates and the second-order model with 674 covariates based on adaptive Lasso estimator obtained using the `R` package `polywog`.

minority)×(MB) have highly significant effects on the email use. Moreover, similarly to the online banking model with interactions, the language and age group category together have significant cross-effects.

Overall, the second-order interaction terms illustrate the complex relationship present between the use of digital technologies and the different demographic characteristics of the user, and point toward the continued presence of digital divide in Canada.

4.3 Multiple correspondence analysis

Multiple correspondence analysis is used to show the association measures between various categorical variables in the dataset. We also calculate and study correlations between the quantitative scores evaluated from subsamples of individuals distinguished with respect to their individual characteristics. The coordinate plots which represent the variable categories in two dimensional space are provided in Figures 1 and 2.

Internet use, email use and online banking. Figure 1 is a plot of the variable categories from the internet use, email use, and online banking regression models. The groupings of variable categories show the underlying structure of the data. The green labeled variable categories are the supplemental variables in the MCA and the dependent variables in our regression models. The red labeled categories are the explanatory variables in our regression models.

The most apparent grouping of variable categories is in the top left quadrant of the graph. This grouping includes people who did not use the internet, email or online bank-

Table 7: Lasso Logistic Regression with Interactions for Online Banking Dependent Variable

Variables	Categories	`svy LLasso`	$\tilde{\theta}^{DB}$	p-value
Intercept		1.013	3.151^{**}	0.006
Language	English	–	-2.663^{**}	0.007
	High school or less	-0.597	-1.554^{*}	0.012
	Visible minority	-0.114	-1.292^{*}	0.050
Location × *Immigration*	(Rural) × (Non-landed immigrant)	–	-0.991^{*}	0.029
Location × *Province*	(Rural) × (QC)	–	0.769^{*}	0.050
	(Rural) × (ON)	–	0.577^{*}	0.035
Age × *Language*	(15-24) × (English)	–	-1.465^{*}	0.049
	(15-24) × (English and French)	–	-5.084^{**}	0.006
Age × *Employment*	(15-24) × (Employed)	–	0.681^{*}	0.024
Age × *Education*	(15-24) × (University degree)	–	1.514^{*}	0.010
Age × *Household type*	(15-24) × (Family without children under 18)	0.291	1.177^{***}	0.000
	(15-24) × (Single)	–	1.087^{*}	0.025
	(15-24) × (Other household type)	–	2.096^{**}	0.006
	(65 and older) × (Single)	-0.065	-0.857^{*}	0.044
Gender × *Language*	(Female) × (English)	0.068	0.752^{*}	0.047
Gender × *Employment*	(Female) × (Employed)	0.153	0.342^{*}	0.017
	(Female) × (University degree)	–	-0.378^{*}	0.013
Language × *Education*	(English) × (High school or less)	–	1.643^{***}	0.001
Language × *Income*	(English) × ($146,560 and higher)	–	1.405^{*}	0.019
Language × *Immigration*	(English) × (Non-landed immigrant)	–	1.480^{***}	0.001
Language × *Education*	(French) × (High school or less)	–	1.331^{*}	0.014
Language × *Household type*	(French) × (Single)	–	-1.802^{*}	0.012
Language × *Immigration*	(French) × (Non-landed immigrant)	–	1.254^{*}	0.042
Language × *Education*	(Non-official language) × (High school or less)	–	1.144^{*}	0.026
Language × *Immigration*	(Non-official language) × (Non-landed immigrant)	–	0.963^{*}	0.044
Language × *Employment*	(French and Non-official) × (Employed)	–	3.790^{*}	0.041
Employment × *Income*	(Employed) × ($146,560 and higher)	–	-0.464^{*}	0.036
Household type × *Income*	(Family without children under 18) × ($52,203 and lower)	–	-0.650^{*}	0.016
	(Single) × ($52,203 and lower)	–	-0.659^{*}	0.013

Notes: $n = 17,135$. The coefficients shown in this table are found to be significant at the 5% level based on their estimated p-values.

Table 8: Lasso Logistic Regression with Interactions for Email Dependent Variable

Variables	Categories	`svy LLasso`	$\tilde{\theta}^{DB}$	p-value
Intercept		1.936	4.597**	0.002
Age	55-64	−0.321	−2.188*	0.035
Language	English	–	−2.799*	0.028
	French	–	−5.688*	0.033
	English, French and Non-official	–	−33.857***	0.001
Location × Age	(Rural) × (35-44)	–	0.750*	0.039
	(Rural) × (65 and older)	–	0.745**	0.008
Location × Language	(Rural) × (English, French and Non-official)	–	32.644*	0.041
Age × Immigration	(25-34) × (Non-landed immigrant)	–	−1.195*	0.035
Age × Province	(25-34) × (MB)	–	1.766*	0.023
Age × Language	(55-64) × (English)	–	1.945*	0.022
	(55-64) × (French)	–	2.074*	0.026
Age × Province	(55-64) × (MB)	–	1.439*	0.022
Age × Language	(65 and older) × (English)	–	1.705*	0.048
Age × Income	(65 and older) × ($52,203 and lower)	−0.223	−0.697*	0.040
Language × Income	(English) × ($146,560 and higher)	–	1.857*	0.022
Language × Province	(French) × (MB)	–	6.461*	0.018
Visible minority × Province	(Visible minority) × (MB)	–	−1.825**	0.005

Notes: $n = 17,268$. The coefficients shown in this table are found to be significant at the 5% level based on their estimated p-values.

ing in the last three months. Grouped with these dependent variable categories are the explanatory categories *65 years and older, Not employed, Single, High school or less*, and people who earn less than $\$52,204$ a year. These explanatory variables were all statistically significant in our logistic regressions and were chosen by `svy LLasso`.

In the lower right quadrant of the plot we see another grouping. The dependent variable categories of people who used the internet, email and online baking are in this quadrant grouped relatively close to the variables *University degree*, income of $\$92,485 - \$146,559$, income greater than $\$146,559$*, Families with children under 18, Employed*, and age group categories *45-54, 35-44, and 25-34*. In Tables 1, 2, and 3, these variables are all statistically significant and have positive coefficients. `svy LLasso` also selected these variable categories.

The other relevant variable groupings seen in Figure 1 are in the top right quadrant, where we see *Non-official language* speakers, *Visible minority*, and *Landed immigrant* grouped together. This grouping of categories makes sense as many new immigrants to Canada are visible minorities and would likely speak a non-official Canadian language.

Virtual wallet and credit card use. Figure 2 is a plot of the variable categories from the virtual wallet and credit card regression models. The dependent variable categories *Used virtual wallet* and *Did not use credit card* have obvious groupings of explanatory variable categories around them. On the other hand, the dependent variable categories *Did not use virtual wallet* and *Used credit card* are not as well represented in two-dimensional space. These dependent variable categories are grouped in the middle of the plot along with explanatory variables with relatively low contribution to the dimensions of the plot.

In the top right quadrant of the plot, the dependent variable category *Did not use credit card* is grouped with the explanatory variable categories $\$52,203$–$\$92,485$, *Single, High school or less, Not employed*, income less than $\$52,204$, *15-24*, and *65 and older*. In Table 5, we see that `svy LLasso` has selected the lowest age group category *15-24* and *High school or less*. The MCA grouping around *No credit card* usage is relatively consistent with the variables selected by `svy LLasso`.

The top left quadrant of the plot has the dependent variable category *Used virtual wallet*. The explanatory variables grouped around *Used virtual wallet* are *Urban, 25-34, ON and AB*. In Table 4, the explanatory variables selected by `svy LLasso` are all the age group categories, *Rural*, *Visible minority*, the highest income category, and *University degree*. The grouping around the *Used virtual wallet* is mostly consistent with the variable categories selected by the Lasso.

`svy LLasso` selected the variable *Visible minority* and although it is not in the close grouping of variables around virtual wallet, it is in the same quadrant of the graph. *Visible minority* is closely grouped with *Landed immigrant*, which is consistent with Figure 1. The other explanatory variables selected by `svy LLasso` but not grouped with *Used virtual wallet* are grouped together in the bottom left quadrant of the graph. The highest income category is grouped with *Employed* and the age group category *45-54*, likely due to the fact the people with high income tend to be in the older segment of the working age population and people with high incomes are typically employed.

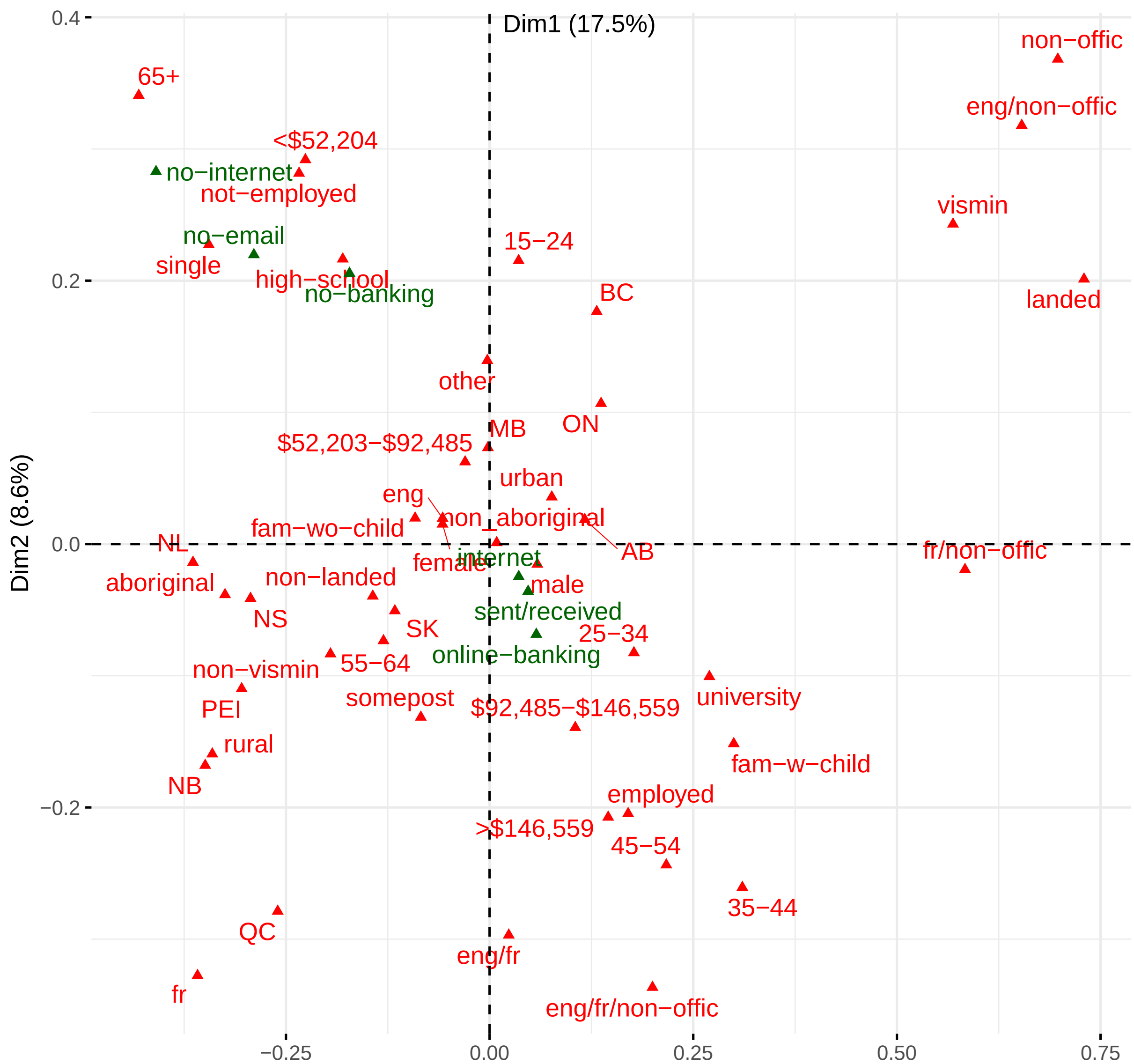

Figure 1: Coordinate plot for Internet Use, Email Use and Online Banking

4.4 Digital literacy score

We design and compute a measure (score) of digital inclusion/digital divide and study its distributional properties in the entire sample and subsamples of individuals with different demographic characteristics pertaining to the social groups with different origin, gender, age, location, and education level.

The digital literacy score is based on the answers of survey respondents to 10 questions from CIUS 2020. Respondents that answer *Yes* to these questions are given 1 point per

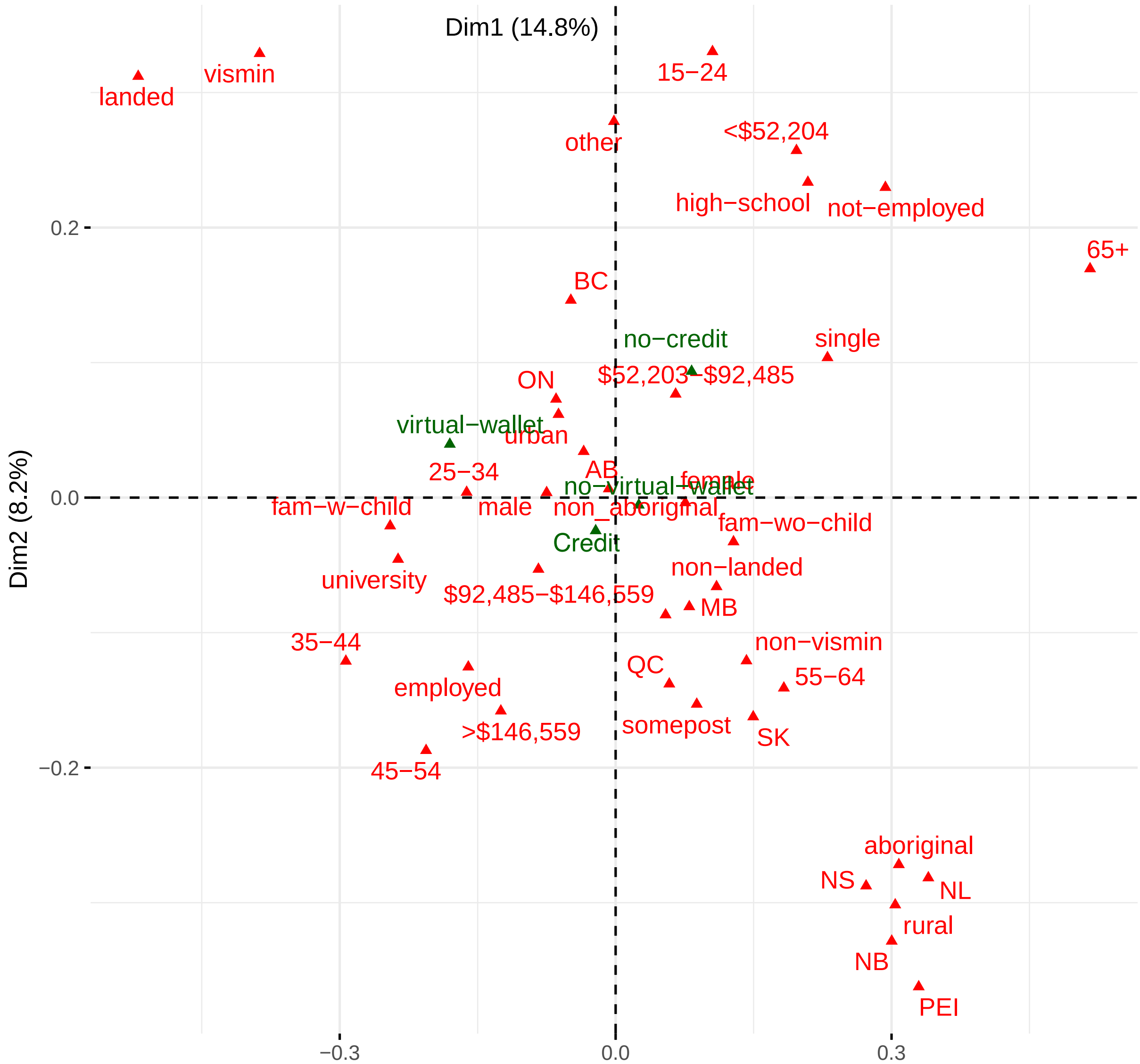

Figure 2: Coordinate plot for Virtual Wallet and Credit Card Use

Yes response. The higher the score (out of 10), the higher the perceived digital literacy of the respondent [See Appendix C for the list of 10 questions our score comprises]. We take the average scores for respondents grouped using variables from our analysis and display these results in Table 9. For example the first two rows of the table show the average score out of 10 for respondents of the survey that reside in urban and rural locations.

The average score from the respondents in Table 9 is 6.88, and the standard deviation is equal to 0.50. Therefore respondents answered an average of just under seven questions with *Yes*. The first characteristic variable we investigate is the location of respondents. Urban residents score slightly higher than the average respondent, and rural slightly lower. This rural/urban divide is consistent with our `svy LLasso` and MCA results that show a divide, albeit sometimes minor, between rural and urban residents regarding internet connectivity and digital literacy.

The age group variable shows one of the most significant divides regarding digital literacy score. The oldest age group category *65 years and older* has the second lowest digital literacy score in our study. The youngest age group also scores relatively lower than the three middle age categories. Due to the type of questions that make up the digital literacy score, younger respondents may have been less likely to answer *Yes* to these questions. Many of the questions have to do with making purchases online and using digital technology that may be skewed towards people in the middle age groups.

There is no significant difference between the scores of males and females. The lack of digital divide among gender is consistent with our `svy LLasso` results, where `svy LLasso` only selects the gender variable in the email use regression model. Similarly to the gender variable *Aboriginal identity* does not seem to significantly affect a respondent's digital literacy score. The small difference in the scores of aboriginals and non-aboriginals is surprising since we know from previous research that aboriginal people are often marginalized when it comes to internet connectivity and digital technology. A possible reason for this disconnect between the results of our score and previous examples of under-utilization of digital technologies in indigenous communities is that CIUS 2020 was an off-reserve survey and only included the 10 Canadian provinces, not the territories. It may be that the most prominent digital divide between *Aboriginal* and *Non-aboriginal* Canadians comes from the aboriginal person's on/off reserve status.

Employment status, educational attainment and income all show significant discrepancies between their varibale categories' digital literacy scores. *Employed* people scored an average of one point higher than *Non-employed*. People with low educational attainment of a *High school or less* score the second lowest only to people *65 years and older* on our digital literacy score test. Educational attainment of a *University degree* shows an average of almost two point difference in their digital literacy score compared to those with a *High* *school or less.*

The lowest income category of people making $\$52,203$ *and lower* has the lowest digital literacy score. The digital literacy score increases as income categories increase with the highest income category having the highest digital literacy score. These results are very consistent with the Lasso inference results. `svy LLasso` selected employment status, income, and education variables and the debiased Lasso results showed they affect the dependent variables in almost every regression specification.

Both immigration status and visible minority status have surprising results. The immigration status variable category *Landed immigrant* is found to have slightly higher digital literacy score than *Non-landed immigrant* (non-immigrant/non-recent immigrant). The variable visible minority status also shows that the category *Visible minority* scores higher on our digital literacy score than the category *Non-visible minority.*

From our MCA results, we know that the variable categories *Landed immigrant* and *Visible minority* are grouped together, suggesting that many recent immigrants are also visible minorities. New immigrants to Canada often have to use the internet and online resources when applying to immigrate to Canada and become citizens. These requirements could explain why visible minorities and recent immigrants in our study have slightly higher digital literacy scores than non-visible minorities and non-immigrants.

The digital literacy scores for each province are relatively similar. The maritime provinces, Newfoundland and Labrador (*NL*), Prince Edward Island (*PEI*), Nova Scotia (*NS*) and New Brunswick (*NB*), score the lowest, while British Columbia (*BC*) scores the highest. In each plot's MCA results, we saw that the maritime provinces were often grouped together with the location variable category *rural.* It is then consistent that these provinces would score slightly lower than others on the digital literacy score. Ontario (*ON*), British Columbia (*BC*), and Alberta (*AB*) score the highest out of the provinces and have almost identical scores.

People who have used a virtual wallet score the highest on our digital literacy score with an average score of 9.5. Non-virtual wallet users' scores are practically equivalent to our analysis's average digital literacy score. The similarity between non-virtual wallet users and the average score suggests that the only Canadians currently using digital wallet technologies are those with very high digital literacy, much higher than the average Canadian. For Canadians to use a newly implemented CBDC, they would likely have to have a much higher degree of digital literacy than they currently possess.

5 Concluding remarks

This paper used different methods, which include a survey-weighted Lasso variable selection/inference techniques, multiple correspondence analysis, and a digital literacy score, to assess the degree of the digital divide in Canada. All methods show consistent results.

Younger working-age Canadians who are employed with high incomes and a high degree of educational attainment have, on average, the highest digital literacy and utilize digital technologies the most. Although somewhat significant, the difference between rural and urban residents does not seem to be the driving factor any longer in the Canadian digital divide. Instead, the leading cause of the digital divide seems to be from the difference in economic class.

These results imply that to implement a CBDC in Canada, significant work and investment is needed to close the digital divide. If a CBDC were implemented today in Canada, people with lower incomes and education would have difficulty adapting to the new monetary system and payment methods. People from lower socioeconomic classes would be negatively impacted by the disappearance of cash, leading to further societal disadvantages. In order to improve this divide between socioeconomic classes concerning digital literacy and digital financial technologies, the government should focus investments not just in rural Canada but also in lower-income areas, irrespective of where they are located. For a CBDC to be beneficial in Canada, each Canadian needs to be able to understand and use it.

Connecting Canadians to the internet is no longer sufficient to improve the digital divide. For a CBDC to be a valuable tool to all Canadians, significant investments need to be made in education and industry so Canadians who already have internet access can learn how to utilize it properly. To some, switching from a cash economy to a cashless economy through the use of a CBDC would be an easy transition. To others, it would likely be impossible without significant training and investment. As shown in our analysis, with

Table 9: Digital Literacy Score

Variables	Categories	Digital Literacy Score
Location	Urban	6.96
	Rural	6.58
Age	15–24	7.00
	25–34	7.62
	35–44	7.57
	45–54	7.04
	55–64	6.55
	65 and older	5.97
Gender	Male	6.82
	Female	6.88
Aboriginal identity	Non-aboriginal	6.86
	Aboriginal	6.65
Employment status	Employed	7.22
	Not employed	6.28
Education	High school or less	6.00
	Some post-secondary	6.74
	University degree	7.56
Visible minority status	Visible minority	7.16
	Not a visible minority	6.80
Household type	Family with children under 18	7.43
	Single	6.43
	Family without children under 18	6.72
	Other household type	6.89
Income	$52, 203 and lower	6.26
	$52,204–$92,485	6.64
	$92,486–$146,559	7.06
	$146, 560 and higher	7.37
Immigration status	Landed immigrant	7.17
	Non-landed immigrant	6.81
Province	NL	6.70
	PEI	6.66
	NS	6.72
	NB	6.49
	QC	6.89
	ON	6.94
	MB	6.85
	SK	6.83
	BC	6.93
	AB	7.03
Virtual wallet	Used virtual wallet	8.32
	No virtual wallet	6.72

Notes: Digital Literacy Score shows the average score out of 10 based on respondents answers grouped by location, age, gender, aboriginal identity, employment status, education, visible minority status, household type, income, immigration status, province, and virtual wallet use.

the current state of the digital divide in Canada the implementation of a CBDC would potentially increase the already apparent divide in digital literacy and the use of digital technologies between high and low socioeconomic classes in Canada.

References

Adrian, T. and Mancini-Griffoli, T. (2019), 'The Rise of Digital Money, International Monetary Fund', *Annual Review of Financial Economics* **13**.

Bank of Canada (2020*a*), 'Contingency Planning for a Central Bank Digital Currency'. **URL:** *https://www.bankofcanada.ca/2020/02/contingency-planning-central-bank-digital-currency/*

Bank of Canada (2020*b*), 'Understanding Money in the 21st Century'. **URL:** *https://www.bankofcanada.ca/2020/01/understanding-money-21st-century/*

Barr, M. S., Harris, A., Menand, L. and Thrasher, K. (2021), 'Should Central Banks Use Distributed Ledger Technology and Digital Currencies to Advance Financial Inclusion?', *Available at SSRN 3849051* .

Belloni, A., Chernozhukov, V. and Hansen, C. (2014), 'High-Dimensional Methods and Inference on Structural and Treatment Effects', *Journal of Economic Perspectives* **28**(2), 29–50.

Belloni, A., Chernozhukov, V. and Wei, Y. (2016), 'Post-Selection Inference for Generalized Linear Models with Many Controls', *Journal of Business & Economic Statistics* **34**(4), 606–619.

Bordo, M. D. and Levin, A. T. (2017), Central Bank Digital Currency and the Future of Monetary Policy, Technical report, National Bureau of Economic Research.

Brainard, L. (2019), Digital Currencies, Stablecoins, and the Evolving Payments Landscape, Technical report, Speech at "The Future of Money in the Digital Age," sponsored by the Peterson Institute for International Economics and Princeton University, Bendheim Center for Finance, Washington, DC, October16.

Brunnermeier, M. K., James, H. and Landau, J.-P. (2019), The Digitalization of Money, Technical report, National Bureau of Economic Research.

Bühlmann, P. and van de Geer, S. (2011), *Statistics for High-Dimensional Data: Methods, Theory and Applications*, Springer Science & Business Media.

Cameron, A. C. and Trivedi, P. K. (2009), *Microeconometrics: Methods and Evaluations*, Cambridge University Press.

Carmichael, K. (2020), Timothy Lane on the Future of Digital Currencies, Technical report, Centre for International Governance Innovation.

Christodorescu, M., Gu, W. C., Kumaresan, R., Minaei, M., Ozdayi, M., Price, B., Raghuraman, S., Saad, M., Sheffield, C., Xu, M. and others. (2020), 'Towards a Two-Tier Hierarchical Infrastructure: an Offline Payment System for Central Bank Digital Currencies', *arXiv preprint arXiv:2012.08003* .

Dufour, J.-M., Trognon, A. and Tuvaandorj, P. (2016), Generalized $C(\alpha)$ Tests in Estimating Functions with Serial Dependence, *in* W. K. Li, D. Stanford and H. Yu, eds, 'Time Series Methods and Applications: the McLeod Festschrift', Springer, Berlin and New York, pp. 151–178.

Friedline, T., Naraharisetti, S. and Weaver, A. (2020), 'Digital Redlining: Poor Rural Communities' Access to Fintech and Implications for Financial Inclusion', *Journal of Poverty* **24**(5-6), 517–541.

Haight, M., Quan-Haase, A. and Corbett, B. A. (2014), 'Revisiting the Digital Divide in Canada: The Impact of Demographic Factors on Access to the Internet, Level of Online Activity, and Social Networking Site Usage', *Information, Communication & Society* **17**(4), 503–519.

Hausman, J. and McFadden, D. (1984), 'Specification Tests for the Multinomial Logit Model', *Econometrica: Journal of the Econometric Society* pp. 1219–1240.

Huynh, K. P. (2017), How Canadians Pay for Things, Technical report, Bank Of Canada.

Huynh, K. P., Nicholls, G. and Nicholson, M. W. (2020), 2019 Cash Alternative Survey Results, Technical report, Bank Of Canada.

Jasiak, J. and Tuvaandorj, P. (2023), Penalized Likelihood Inference with Survey Data, Technical report, York University.

Javanmard, A. and Montanari, A. (2014), ‘Confidence Intervals and Hypothesis Testing for High-Dimensional Regression’, *The Journal of Machine Learning Research* **15**(1), 2869–2909.

Jordan, B. (2019), High-Speed Access for All: Canada’s Connectivity Strategy, Technical report, Government of Canada.

Lee, J. D., Sun, D. L., Sun, Y. and Taylor, J. E. (2016), ‘Exact Post-Selection Inference, with Application to the Lasso’, *The Annals of Statistics* **44**(3), 907–927.

Maniff, J. L. (2020), ‘Motives Matter: Examining Potential Tension in Central Bank Digital Currency Designs’, *Federal Reserve Bank of Kansas City: Omaha, NE, USA* .

Mullainathan, S. and Spiess, J. (2017), ‘Machine Learning: An Applied Econometric Approach’, *Journal of Economic Perspectives* **31**(2), 87–106.

Neyman, J. (1959), Optimal Asymptotic Tests of Composite Statistical hypotheses, *in* U. Grenander, ed., ‘Probability and Statistics, the Harald Cramér Volume’, Almqvist and Wiksell, Uppsala, Sweden, pp. 213–234.

Smith, R. J. (1987), ‘Alternative Asymptotically Optimal Tests and their Application to Dynamic Specification’, **LIV**, 665–680.

Taylor, J. and Tibshirani, R. (2018), ‘Post-selection Inference for l1-Penalized Likelihood Models’, *Canadian Journal of Statistics* **46**(1), 41–61.

Usher, A., Reshidi, E., Rivadeneyra, F. and Hendry, S. (2021), The Positive Case for a CBDC, Technical report, Bank of Canada.

Wooldridge, J. M. (2001), ‘Asymptotic Properties of Weighted M-Estimators for Standard Stratified Samples’, *Econometric Theory* **17**(2), 451–470.

Xia, L., Nan, B. and Li, Y. (2020), ‘A Revisit to De-biased Lasso for Generalized Linear Models’, *arXiv preprint arXiv:2006.12778* .

Zhang, C.-H. and Zhang, S. S. (2014), 'Confidence Intervals for Low Dimensional Parameters in High Dimensional Linear Models', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **76**(1), 217–242.

A Sampling and weighting methodology in CIUS 2020

Sampling. The collection of CIUS 2020 is based on a stratified design employing probability sampling. The stratification is done at the province/census metropolitan area (CMA) and census agglomeration (CA) level where each of the ten provinces were divided into strata/geographic areas.

CIUS 2020 uses a frame that combines landline and cellular telephone numbers from the Census and various administrative sources with Statistics Canada's dwelling frame. Records on the frame are groups of one or several telephone numbers associated with the same address.

Each record in the survey frame was assigned to a stratum within its province. A simple random sample without replacement of records (the groups of telephone numbers) was next selected in each stratum. CIUS 2020 only selects one respondent randomly from each eligible household to complete an electronic questionnaire or to respond to a telephone interview.

The number of respondents for the 2020 CIUS was 17,409, which is 41.6% of the sample size 41,817.

Weighting. Each record within a stratum has an equal probability of selection given by

$$\frac{\text{the number of records sampled in the stratum}}{\text{the number of records in the stratum from the survey frame}}.$$

A short description of the survey weights calculation is as follows.[2]

1. The initial weight is the inverse of an adjusted version of the probability of selection given above.

2. The person weight is equal to *Initial Household weight* × *Factor 1* × *Number of Eligible Household Members (capped at 5)*, where *Factor 1* involves an adjustment for non-response among others.

3. The final person weight w_i is an adjusted version of the person weight above.

[2] Further details of the weighting procedure can be found in Section 10 of Microdata user Guide, CIUS 2020 at https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=4432#a2

B Technical appendix

B.1 Inference with survey logistic Lasso

Since the CIUS 2020 data were collected using a stratified sampling scheme which is close to simple stratified sampling where the units within each stratum are sampled independently with equal probability, we treat w_i as constant and given (Wooldridge (2001), Section 3), and $\{(y_i, x_i')'\}_{i=1}^n$ as independent.

From the weighted log-likelihood function $L(\theta) = n^{-1}\sum_{i=1}^n w_i(y_i x_i'\theta - \log(1+\exp(x_i'\theta)))$, the score function, the information and negative Hessian matrices can be obtained respectively as

$$S(\theta) = \frac{\partial L(\theta)}{\partial \theta} = n^{-1}\sum_{i=1}^n w_i x_i (y_i - \Lambda(x_i'\theta)), \tag{B.1}$$

$$I(\theta) = n^{-1}\sum_{i=1}^n w_i^2 x_i x_i' \Lambda(x_i'\theta)(1-\Lambda(x_i'\theta)), \tag{B.2}$$

$$H(\theta) = -\frac{\partial^2 L(\theta)}{\partial\theta\partial\theta'} = n^{-1}\sum_{i=1}^n w_i x_i x_i' \Lambda(x_i'\theta)(1-\Lambda(x_i'\theta)), \tag{B.3}$$

where $\Lambda(z) = \exp(z)/(1+\exp(z))$ is the logistic CDF. We report the marginal effects for each variable along with the coefficient estimates in the regression tables which is defined as follows. For a dummy regressor $\tilde{x}_{ij}, j = 1, \dots, p; i = 1, \dots, n$, the marginal effect (ME) is $\text{ME}_{ij}(\theta) \equiv \Lambda(x_i'\theta)|_{\tilde{x}_{ij}=1} - \Lambda(x_i'\theta)|_{\tilde{x}_{ij}=0}$. The average marginal effect (AME) of the j-th regressor is defined as

$$\text{AME}_j = \text{AME}_j(\theta_0) \equiv \text{E}\left[\frac{1}{\sum_{i=1}^n w_i}\sum_{i=1}^n w_i \text{ME}_{ij}(\theta_0)\right],$$

where θ_0 denotes the true value of θ and the expectation is taken with respect to the distribution of the regressors. An estimator of AME_j is

$$\widehat{\text{AME}}_j(\hat{\theta}) \equiv \frac{1}{\sum_{i=1}^n w_i}\sum_{i=1}^n w_i \left(\Lambda(x_i'\hat{\theta})|_{\tilde{x}_{ij}=1} - \Lambda(x_i'\hat{\theta})|_{\tilde{x}_{ij}=0}\right),$$

where $\hat{\theta} = (\hat{\alpha}, \hat{\beta}')'$ is an estimator of θ_0 e.g. `svy LLasso` estimator. Note also that

$$\frac{\partial \widehat{\mathrm{AME}}_j(\hat{\theta})}{\partial \theta} \equiv \frac{1}{\sum_{i=1}^n w_i} \sum_{i=1}^n w_i \left\{ \left[x_i \Lambda(x_i'\hat{\theta})(1 - \Lambda(x_i'\hat{\theta})) \right] |_{\tilde{x}_{ij}=1} - \left[x_i \Lambda(x_i'\hat{\theta})(1 - \Lambda(x_i'\hat{\theta})) \right] |_{\tilde{x}_{ij}=0} \right\}.$$

Debiased Lasso

The debiased Lasso method of Zhang and Zhang (2014), Javanmard and Montanari (2014) and Xia et al. (2020) is based the one-step estimator constructed from the initial Lasso estimator $\hat{\theta}$:

$$\tilde{\theta} \equiv \hat{\theta} + H(\hat{\theta})^{-1} S(\hat{\theta}).$$

The standard errors for the parameters are calculated using the distributional approximation: as $n \to \infty$

$$(\tau' H(\hat{\theta})^{-1} I(\hat{\theta}) H(\hat{\theta})^{-1} \tau)^{-1/2} n^{1/2} \tau'(\tilde{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, 1),$$

where $\tau \in \mathbb{R}^{p+1}$ is a fixed vector with $\tau'\tau = 1$, and θ_0 denotes the true parameter vector. To obtain a confidence interval for $\mathrm{AME}_j, j = 2, \ldots, p+1$, define a one-step estimator

$$\widetilde{\mathrm{AME}}_j \equiv \widehat{\mathrm{AME}}_j(\hat{\theta}) + \frac{\partial \widehat{\mathrm{AME}}_j(\hat{\theta})}{\partial \theta'} H(\hat{\theta})^{-1} S(\hat{\theta}).$$

Then, under some regularity conditions as $n \to \infty$

$$\left(\frac{\partial \widehat{\mathrm{AME}}_j(\hat{\theta})}{\partial \theta'} H(\hat{\theta})^{-1} I(\hat{\theta}) H(\hat{\theta})^{-1} \frac{\partial \widehat{\mathrm{AME}}_j(\hat{\theta})}{\partial \theta} \right)^{1/2} n^{1/2} (\widetilde{\mathrm{AME}}_j - \mathrm{AME}_j) \xrightarrow{d} \mathcal{N}(0, 1).$$

$C(\alpha)$/Orthogonalization method

We follow Belloni et al. (2016) who develop a $C(\alpha)$-type subvector inference procedure in a sparse high-dimensional generalized linear model by constructing an estimating equation orthogonalized against the direction of the nuisance parameter estimation (see Neyman (1959) for the $C(\alpha)$ test), and consider a survey version of their statistic.

Consider testing a scalar component θ_1 (e.g. the i-th element β_i of β) of θ. Partition the parameters as $\theta = (\theta_1, \theta_2')'$, $\theta_1 \in \mathbb{R}$, $\theta_2 \in \mathbb{R}^p$. Also partition the quantities in (B.1) and

(B.3) as

$$
\begin{aligned}
S(\theta) &= [S_1(\theta)', S_2(\theta)']', \;\; S_1(\theta) \in \mathbb{R}, \;\; S_1(\theta) \in \mathbb{R}^p, \\
H(\theta) &= \begin{bmatrix} H_{11}(\theta) & H_{12}(\theta) \\ H_{21}(\theta) & H_{22}(\theta) \end{bmatrix}, \;\; H_{11}(\theta) \in \mathbb{R}, \;\; H_{21}(\theta) = H_{12}(\theta)' \in \mathbb{R}^p, \;\; H_{22}(\theta) \in \mathbb{R}^{p\times p}.
\end{aligned}
$$

Consider the restriction $H_0 : \theta_1 = \theta_{01}$ and let $\tilde{\theta}^* = (\theta_{01}', \tilde{\theta}_2^{*\prime})'$, where $\tilde{\theta}_2^*$ is the logistic Lasso or Post-logistic Lasso (Belloni et al., 2016) estimator of θ_2. The survey $C(\alpha)$ statistic is then defined as

$$
C_\alpha(\theta_{01}) \equiv n\, S(\tilde{\theta}^*)' D(\tilde{\theta}^*) \left(D(\tilde{\theta}^*)' I(\tilde{\theta}^*) D(\tilde{\theta}^*) \right)^{-1} D(\tilde{\theta}^*)' S(\tilde{\theta}^*), \tag{B.4}
$$

where $D(\theta) \equiv [I_{k_1}, -H_{22}(\theta)^{-1} H_{21}(\theta)]'$. Here $D(\tilde{\theta}^*)'S(\tilde{\theta}^*) = S_1(\tilde{\theta}^*) - H_{12}(\tilde{\theta}^*) H_{22}(\tilde{\theta}^{**})^{-1} S_2(\tilde{\theta}^*)$ is the effective score function obtained by orthgogalizing the score function of the parameters of interest against the score function of the nuisance parameters. Under $H_0 : \theta_1 = \theta_{01}$ and appropriate regularity conditions $C_\alpha(\theta_{01}) \xrightarrow{d} \chi_1^2$ as $n \to \infty$.

For testing the restriction $H_0 : \psi(\theta_0) = 0$ on a scalar nonlinear parameter $\psi(\theta)$, following Dufour et al. (2016) and Smith (1987) consider the $C(\alpha)$ statistic:

$$
C_\alpha(\psi_0) \equiv n \left(\frac{\partial \psi(\tilde{\theta}^*)}{\partial \theta'} H(\tilde{\theta}^*)^{-1} I(\tilde{\theta}^*) H(\tilde{\theta}^*)^{-1} \frac{\partial \psi(\tilde{\theta}^*)}{\partial \theta} \right)^{-1} \left(\frac{\partial \psi(\tilde{\theta}^*)}{\partial \theta'} H(\tilde{\theta}^*)^{-1} S(\tilde{\theta}^*) \right)^2, \tag{B.5}
$$

where the auxiliary estimate $\tilde{\theta}^*$ satisfies $\psi(\tilde{\theta}^*) = 0$. Let $\mathrm{AME}_{(1)}$ be the AME with respect to a dummy regressor with a coefficient θ_1. Then, $\mathrm{AME}_{(1)} = 0$ if $\theta_1 = 0$, and it follows that $\mathrm{AME}_{(1)}(\tilde{\theta}^*) = 0$ for $\tilde{\theta}^* = (0, \hat{\theta}_2')'$, where $\hat{\theta}_2$ is the `svy LLasso` estimate. It is easy to see that the statistic in (B.5) for the hypothesis $H_0 : \mathrm{AME}_{(1)} = 0$ is numerically identical to the statistic in (B.4) for the hypothesis $H_0 : \theta_1 = 0$.

Selective inference

We also consider the survey-logit version of the "selective inference" method proposed by Lee et al. (2016) and Taylor and Tibshirani (2018). This method makes inference on the coefficients selected by the Lasso i.e. the target parameters determined from the data which are random. This feature makes the selective inference method conceptually different from the debiased Lasso and $C(\alpha)$ methods where the target parameters are the population parameters.

The key ingredient in this method is the one-step estimator which updates the estimates of the (non-zero) coefficients selected by the survey logistic Lasso, denoted as $\hat{\theta}_M$:

$$\tilde{\theta}_M \equiv \hat{\theta}_M + H_M(\hat{\theta}_M)^{-1} S_M(\hat{\theta}_M), \tag{B.6}$$

where $H_M(\cdot)$ and $S_M(\cdot)$ are the Hessian and the score functions of the logistic model corresponding to `svy LLasso` selected coefficients. Then, a test statistic constructed from the conditional distribution of the one-step estimator (B.6) given Lasso selection events is used to test a hypothesis on the `svy LLasso` selected coefficients. We refer to Jasiak and Tuvaandorj (2023) for further details of the method in a survey setting.

B.2 Multiple correspondence analysis

The Multiple Correspondence Analysis (MCA) is an analog of principal component analysis (PCA) for multiple categorical variables. MCA may provide a useful summary and visualization of survey data (with categorical variables) by revealing the variables that contribute the most to the variation in the data, identifying a set of observations with similar characteristics in their survey response and quantifying the degree of associations between different categories.

The MCA process works by taking J categorical variables, each having K_j categorical levels with the sum of these levels being equal to K. Given I observations we denote the indicator matrix as X. This indicator matrix is used to perform correspondence analysis. The correspondence analysis gives two different sets of factor scores, one set for the rows of the matrix and the other for the columns.

We denote the total of this table as N and set $Z \equiv X/N$. The vector r contains the sums of the rows of the matrix Z and the vector c the sums of the columns of Z. To compute the MCA, we need to define diagonal matrices $D_r \equiv \text{diag}(r)$ and $D_c \equiv \text{diag}(c)$.

To find the factor scores we use the singular value decomposition

$$D_r^{-1/2}[Z - rc']D_c^{-1/2} = P\Delta Q'.$$

The matrix Δ is the diagonal matrix of singular values and the matrix of eigenvalues is $\Lambda = \Delta^2$. We find the row and column factor scores as

$$F = D_r^{-1/2}P\Delta \quad \text{and} \quad G = D_c^{-1/2}Q\Delta.$$

In this paper, the MCA is done using what is called the Burt matrix defined as $B = X'X$. Using the Burt matrix gives the exact same factors as is the case with the indicator matrix X but also gives a better approximation of the captured inertia.

Various plots accompanying the MCA can be used to visualize a global pattern within the data. The coordinate plots which represent the variable categories in two dimensional space are provided in Figures 1 and 2.

C Details on the digital literacy score

The calculation of the digital literacy score is based on the responses to the following 10 questions all of which have the following answers: *Yes, No, Valid skip, Don't know, Refusal, Not stated*. The questions 1-7 are "During the past three months, which of the following activities, related to communication, have you done over the Internet?" followed by:

1. "Have you used social networking websites or apps?"
2. "Have you made online voice calls or video calls?"
3. "Have you researched for information about community events?"
4. "Have you accessed the news?"
5. "Have you found locations and directions?"
6. "Have you researched for information on health?"
7. "Have you researched for information about goods or services?"

The remaining questions 8-10 are:

8. "During the past 12 months, how did you pay for the goods and services ordered over the Internet? Did you use an online payment service?"

9. "During the past 12 months, which of the following software related activities have you carried out using any device? Have you copied or moved files or folders?"

10. "Have you carried out any of the following to manage access to your personal data over the Internet during the past 12 months? Have you checked that the website where you provided personal data was secure e.g., https sites, safety logo or certificate?"

11,874 out of 17,409 respondents answered all the questions and the remaining respondents had at least one question unanswered (missing values). Figure 3 plots the weighted histogram and Table 10 below reports the weighted descriptive statistics of the scores of 11874 respondents who answered all the questions.

Table 10: Descriptive Statistics for Digital Literacy Scores

Weighted estimates						
Mean	Stdev	Skewness	Kurtosis	1st Quartile	Median	3rd Quartile
7.11	2.15	-0.85	0.26	6.00	8.00	9.00

D $C(\alpha)$ and selective inference results

This section presents the outcomes of the $C(\alpha)$ and SI for the logit coefficients and their AMEs for models 1 to 5. The results are very similar to the results reported in Section 4. Therefore, these additional results further corroborate the findings from the `svy LLasso` and debiased Lasso results.

Since the SI is made only on the coefficients chosen by `svy LLasso`, it tends to have less coefficient and AME estimates that are significant based on the p-values than in the case of the debiased Lasso results. Besides the small differences in significance, the selective inference and debiased Lasso results were relatively consistent. The C_α test statistics and corresponding p-values are also consistent with the `svy LLasso` and debiased Lasso results.

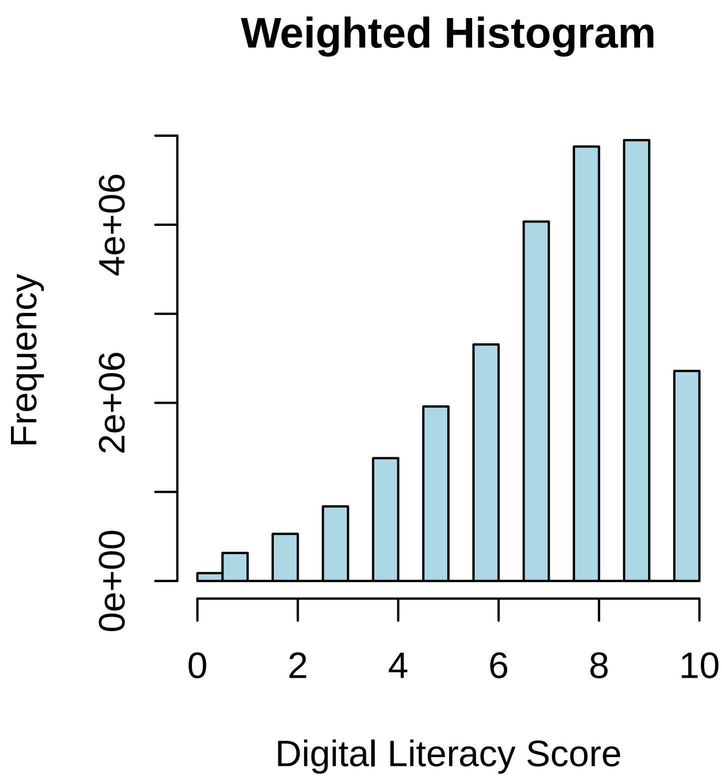

Figure 3: Weighted Histogram of Digital Literacy Scores

Table 11: Lasso Logistic Regression Results for Internet Use Dependent Variable

Variables	Categories	C_α	p-value	$\tilde{\theta}^{SI}$	p-value	$\widetilde{\text{AME}}^{SI}$	p-value
Intercept		80.184^{***}	0.000	2.930^{***}	0.000	–	–
Location	Urban (omitted)	–	–	–	–	–	–
	Rural	12.198^{***}	0.000	−0.307	0.998	−0.018	0.996
Age	15–24	13.647^{***}	0.000	1.207	0.140	0.056	0.123
	25–34	6.954^{**}	0.008	0.67	0.326	0.036	0.323
	35–44	5.687^{*}	0.017	0.555	0.628	0.032	0.627
	45–54 (omitted)	–	–	–	–	–	–
	55–64	19.400^{***}	0.000	−0.543	0.999	−0.032	0.996
	65 and older	176.970^{***}	0.000	−1.296	1.000	−0.081	1.000
Gender	Male (omitted)	–	–	–	–	–	–
	Female	1.620	0.203	0.093	0.756	0.005	0.757
Aboriginal identity	Non-aboriginal (omitted)	–	–	–	–	–	–
	Aboriginal	5.304^{*}	0.021	–	–	–	–
Language	English	$3.628^{\cdot}$	0.057	0.427^{**}	0.002	0.025^{**}	0.003
	French	0.609	0.435	–	–	–	–
	Non-official language	0.043	0.836	–	–	–	–
	English and French	2.361	0.124	–	–	–	–
	English and Non-official language (omitted)	–	–	–	–	–	–
	French and Non-official language	0.367	0.544	–	–	–	–
	English, French and Non-official language	1.692	0.193	–	–	–	–
Employment	Employed	21.370^{***}	0.000	0.609^{***}	0.000	0.034^{***}	0.000
	Not employed (omitted)	–	–	–	–	–	–
Education	High school or less	189.022^{***}	0.000	−0.888	1.000	−0.053	1.000
	Some post-secondary (omitted)	–	–	–	–	–	–
	University degree	12.502^{***}	0.000	0.624^{***}	0.000	0.032^{***}	0.000
Visible minority	Visible minority	4.379^{*}	0.036	−0.275	0.520	−0.016	0.515
	Not a visible minority (omitted)	–	–	–	–	–	–
Household type	Family with children under 18 (omitted)	–	–	–	–	–	–
	Family without children under 18	0.026	0.872	–	–	–	–
	Single	18.073^{***}	0.000	−0.665	1.000	−0.043	1.000
	Other household type	0.226	0.635	–	–	–	–
Income	\$52,203 and lower	30.504^{***}	0.000	−0.526	1.000	−0.031	1.000
	\$52,204–\$92,485 (omitted)	–	–	–	–	–	–
	\$92,486–\$146,559	0.524	0.469	–	–	–	–
	\$146,560 and higher	9.420^{**}	0.002	0.491^{**}	0.004	0.025^{**}	0.003
Immigration	Landed immigrant (omitted)	–	–	–	–	–	–
	Non-landed immigrant	1.829	0.176	–	–	–	–
Province	NL	2.540	0.111	–	–	–	–
	PEI	2.023	0.155	–	–	–	–
	NS	2.383	0.123	–	–	–	–
	NB	0.297	0.586	–	–	–	–
	QC	5.130^{*}	0.0240	−0.201	0.859	−0.011	0.832
	ON	0.017	0.895	0.183	0.469	0.011	0.470
	MB	6.162^{*}	0.013	–	–	–	–
	SK	4.347^{*}	0.037	–	–	–	–
	BC	0.250	0.617	0.264	0.604	0.015	0.604
	AB (omitted)	–	–	–	–	–	–

Notes: $n = 17,409$. C_α denotes the $C(\alpha)$ statistic which tests simultaneously the statistical significance of the coefficient and of its AME. $\tilde{\theta}^{SI}$ and $\widetilde{\text{AME}}^{SI}$ denote the selective inference one-step Lasso estimates of the logit parameter and AME respectively.

Table 12: Lasso Logistic Regression Results for Online Banking Dependent Variable

Variables	Categories	C_α	p-value	$\tilde{\theta}^{SI}$	p-value	$\widetilde{\text{AME}}^{SI}$	p-value
Intercept		10.569**	0.001	1.049***	0.000	–	–
Location	Urban (omitted)	–	–	–	–	–	–
	Rural	2.036	0.154	–	–	–	–
Age	15–24	0.127	0.721	–	–	–	–
	25–34	23.173***	0.000	0.639***	0.000	0.094***	0.000
	35–44	22.858***	0.000	0.513***	0.000	0.078***	0.000
	45–54 (omitted)	–	–	–	–	–	–
	55–64	12.714***	0.000	−0.331	0.978	−0.054	0.977
	65 and older	97.324***	0.000	−0.897	1.000	−0.161	1.000
Gender	Male (omitted)	–	–	–	–	–	–
	Female	2.592	0.107	–	–	–	–
Aboriginal identity	Non-aboriginal (omitted)	–	–	–	–	–	–
	Aboriginal	2.631	0.105	–	–	–	–
Language	English	7.723**	0.005	0.083	0.794	0.013	0.795
	French	8.042**	0.005	–	–	–	–
	Non-official language	2.883·	0.090	–	–	–	–
	English and French	0.402	0.526	–	–	–	–
	English and Non-official language (omitted)	–	–	–	–	–	–
	French and Non-official language	0.208	0.648	–	–	–	–
	English, French and Non-official language	0.032	0.858	–	–	–	–
Employment	Employed	70.108***	0.000	0.659***	0.000	0.108***	0.000
	Not employed (omitted)	–	–	–	–	–	–
Education	High school or less	129.578***	0.000	−0.668	1.000	−0.112	1.000
	Some post-secondary (omitted)	–	–	–	–	–	–
	University degree	24.964***	0.000	0.449***	0.000	0.069***	0.000
Visible minority	Visible minority	8.596**	0.003	−0.352	0.999	−0.058	0.999
	Not a visible minority (omitted)	–	–	–	–	–	–
Household type	Family with children under 18 (omitted)	–	–	–	–	–	–
	Family without children under 18	11.558***	0.001	0.259*	0.033	0.042*	0.035
	Single	2.252	0.133	−0.191	0.972	−0.031	0.964
	Other household type	4.138*	0.042	–	–	–	–
Income	$52,203 and lower	11.843***	0.001	−0.280	1.000	−0.046	1.000
	$52,204–$92,485 (omitted)	–	–	–	–	–	–
	$92,486–$146,559	2.270	0.132	–	–	–	–
	$146560 and higher	7.792**	0.005	0.184·	0.073	0.029·	0.075
Immigration	Landed immigrant (omitted)	–	–	–	–	–	–
	Non-landed immigrant	0.538	0.463	–	–	–	–
Province	NL	0.014	0.905	–	–	–	–
	PEI	0.270	0.604	–	–	–	–
	NS	0.376	0.540	–	–	–	–
	NB	0.270	0.603	–	–	–	–
	QC	0.547	0.460	–	–	–	–
	ON	0.103	0.748	–	–	–	–
	MB	8.498**	0.004	–	–	–	–
	SK	0.780	0.377	–	–	–	–
	BC	0.008	0.931	–	–	–	–
	AB (omitted)	–	–	–	–	–	–

Notes: $n = 17,135$.

Table 13: Lasso Logistic Regression Results for Email Use Dependent Variable

Variables	Categories	C_α	p-value	$\tilde{\theta}^{SI}$	p-value	$\widetilde{\text{AME}}^{SI}$	p-value
Intercept		50.422***	0.000	1.722***	0.000		
Location	Rural	8.218**	0.004	−0.246	0.992	−0.025	0.990
Age	15–24	12.364***	0.000	0.664**	0.001	0.060**	0.001
	25–34	17.75***	0.000	0.737***	0.000	0.065***	0.000
	35–44	14.221***	0.000	0.589***	0.001	0.054***	0.001
	55–64	11.781***	0.001	−0.367	0.999	−0.037	0.997
	65 and older	94.367***	0.000	−0.929	1.000	−0.104	0.999
Gender	Male (omitted)	–	–	–	–	–	–
	Female	5.024*	0.025	0.152$^{\cdot}$	0.061	0.015$^{\cdot}$	0.064
Aboriginal identity	Non-aboriginal (omitted)	–	–	–	–	–	–
	Aboriginal	6.942**	0.008	–	–	–	–
Language	English	1.278	0.258	0.337**	0.002	0.033**	0.003
	French	0.219	0.640	–	–	–	–
	Non-official language	0.881	0.348	−0.188	0.449	−0.019	0.447
	English and French	0.959	0.327	–	–	–	–
	English and Non-official language (omitted)	–	–	–	–	–	–
	French and Non-official language	0.183	0.669	–	–	–	–
	English, French and Non-official language	5.507*	0.019	–	–	–	–
Employment	Employed	23.347***	0.000	0.465**	0.002	0.046**	0.002
	Not employed (omitted)	–	–	–	–	–	
Education	High school or less	184.115***	0.000	−0.816	1.000	−0.085	1.000
	University degree	49.181***	0.000	0.873***	0.000	0.076***	0.000
Visible minority	Visible minority	7.085**	0.008	−0.322	0.977	−0.033	0.974
	Not a visible minority (omitted)	–	–	–	–	–	–
Household type	Family without children under 18	0.213	0.644	–	–	–	–
	Single	24.825***	0.000	−0.533	1.000	−0.058	1.000
	Other household type	0.049	0.824	–	–	–	–
Income	\$52,203 and lower	17.168***	0.000	−0.381	1.000	−0.039	1.000
	\$52,204–\$92,485 (omitted)	–	–	–	–	–	–
	\$92,486–\$146,559	0.800	0.371	–	–	–	–
	\$146,560 and higher	11.925***	0.001	0.402***	0.001	0.037***	0.000
Immigration	Landed immigrant (omitted)	–	–	–	–	–	–
	Non-landed immigrant	1.044	0.307	0.122	0.842	0.012	0.844
Province	NL	2.420	0.120	–	–	–	–
	PEI	1.244	0.265	–	–	–	–
	NS	6.260*	0.012	–	–	–	–
	NB	2.730$^{\cdot}$	0.098	–	–	–	–
	QC	4.091*	0.043	−0.192	0.793	−0.019	0.779
	ON	0.191	0.662	0.278**	0.006	0.027**	0.006
	MB	8.440**	0.004	–	–	–	–
	SK	5.343*	0.021	–	–	–	–
	BC	2.417	0.120	0.439**	0.002	0.041**	0.002

Notes: $n = 17,268$.

Table 14: Lasso Logistic Regression Results for Virtual Wallet Dependent Variable

Variables	Categories	C_α	p-value	$\tilde{\theta}^{SI}$	p-value	$\widetilde{\text{AME}}^{SI}$	p-value
Intercept		70.764***	0.000	−2.222***	0.000	–	–
Location	Urban (omitted)	–	–	–	–	–	–
	Rural	23.924***	0.000	−0.627	1.000	−0.061	1.000
Age	15–24	29.036***	0.000	0.702**	0.002	0.083**	0.003
	25–34	25.121***	0.000	0.504***	0.001	0.057**	0.001
	35–44	7.735**	0.005	–	–	–	–
	45–54 (omitted)	–	–	–	–	–	–
	55-64	14.361***	0.000	−0.747	1.000	−0.070	1.000
	65 and older	21.482***	0.000	−1.144	1.000	−0.096	1.000
Gender	Male (omitted)	–	–	–	–	–	–
	Female	1.165	0.280	–	–	–	
Aboriginal identity	Non-aboriginal (omitted)	–	–	–	–	–	
	Aboriginal	0.026	0.872	–	–	–	–
Language	English	0.281	0.596	–	–	–	–
	French	0.203	0.653	–	–	–	–
	Non-official language	2.469	0.116	–	–	–	–
	English and French	0.047	0.829	–	–	–	–
	English and Non-official language (omitted)	–	–	–	–	–	–
	French and Non-official language	0.574	0.448	–	–	–	–
	English, French, and Non-official language	0.842	0.359	–	–	–	–
Employment	Employed	0.034	0.853	–	–	–	–
	Not employed (omitted)	–	–	–	–	–	–
Education	High school or less	0.326	0.568	–	–	–	–
	Some post-secondary (omitted)	–	–	–	–	–	–
	University degree	6.847**	0.009	0.254	0.391	0.027	0.393
Visible minority	Visible minority	13.365***	0.000	0.242*	0.044	0.026$^{\cdot}$	0.053
	Not a visible minority (omitted)	–	–	–	–	–	–
Household type	Family with children under 18 (omitted)	–	–	–	–	–	–
	Family without children under 18	0.362	0.547	–	–	–	
	Single	0.066	0.797	–	–	–	–
	Other household type	0.244	0.621	–	–	–	–
Income	$52,203 and lower	0.356	0.551	–	–	–	–
	$52,204–$92,485 (omitted)	–	–	–	–	–	–
	$92,486–$146,559	1.620	0.203	–	–	–	–
	$146,560 and higher	24.149***	0.000	0.506***	0.000	0.057***	0.000
Immigration	Landed immigrant (omitted)	–	–	–	–	–	–
	Non-landed immigrant	0.796	0.372	–	–	–	–
Province	NL	1.811	0.178	–	–	–	–
	PEI	2.279	0.131	–	–	–	–
	NS	1.947	0.163	–	–	–	–
	NB	0.095	0.757	–	–	–	–
	QC	0.39	0.532	–	–	–	–
	ON	0.094	0.759	–	–	–	–
	MB	4.605	0.032	–	–	–	–
	SK	1.219	0.270	–	–	–	–
	BC	0.224	0.636	–	–	–	–
	AB (omitted)	–	–	–	–	–	–

Notes: $n = 12,124$.

Table 15: Lasso Logistic Regression Results for Credit Card Use Dependent Variable

Variables	Categories	C_α	p-value	$\tilde{\theta}^{SI}$	p-value	$\widetilde{\text{AME}}^{SI}$	p-value
Intercept		21.706***	0.000	1.193***	0.000	–	–
Location	Urban (omitted)	–	–	–	–	–	–
	Rural	2.250	0.134	–	–	–	–
Age	15–24	16.767***	0.000	−0.572	1.000	−0.098	1.000
	25–34	0.233	0.630	–	–	–	–
	35–44	1.735	0.188	–	–	–	–
	45–54 (omitted)	–	–	–	–	–	–
	55–64	0.046	0.830	–	–	–	–
	65 and older	0.202	0.653	–	–	–	–
Gender	Male (omitted)	–	–	–	–	–	–
	Female	0.003	0.958	–	–	–	–
Aboriginal identity	Non-aboriginal (omitted)	–	–	–	–	–	–
	Aboriginal	1.046	0.306	–	–	–	–
Language	English	0.001	0.980	0.278*	0.011	0.044*	0.014
	French	8.287**	0.004	−0.408	0.972	−0.067	0.967
	Non-official	0.039	0.844	–	–	–	–
	English and French	0.211	0.646	–	–	–	–
	English and Non-official language (omitted)	–	–	–	–	–	–
	French and Non-official	1.252	0.263		–	–	–
	English, French, and Non-official language	2.916$^{\cdot}$	0.088	–	–	–	–
Employment	Employed	3.006$^{\cdot}$	0.083	0.151	0.934	0.024	0.934
	Not employed (omitted)	–	–	–	–	–	–
Education	High school or less	33.289***	0.000	−0.470	1.000	−0.077	1.000
	Some post-secondary (omitted)	–	–	–	–	–	–
	University degree	29.228***	0.000	0.480***	0.000	0.072***	0.000
Visible minority	Visible minority	4.063*	0.044	–	–	–	–
	Not a visible minority (omitted)	–	–	–	–	–	–
Household type	Family with children under 18 (omitted)	–	–	–	–	–	–
	Family without children under 18	13.824***	0.000	0.219	0.542	0.035	0.543
	Single	9.465**	0.002	–	–	–	–
	Other household type	0.623	0.430		–	–	–
Income	\$52,203 and lower	8.466**	0.004	−0.291	0.976	−0.047	0.975
	\$52,204–\$92,485 (omitted)	–	–	–	–	–	–
	\$92,486–\$146,559	1.049	0.306	–	–	–	–
	\$146,560 and higher	0.728	0.393	–	–	–	–
Immigration	Landed immigrant (omitted)	–	–	–	–	–	–
	Non-landed immigrant	1.430	0.232		–	–	–
Province	NL	3.045$^{\cdot}$	0.081	–	–	–	–
	PEI	0.223	0.637	–	–	–	
	NS	0.076	0.783	–	–	–	–
	NB	0.005	0.944	–	–	–	
	QC	0.076	0.783	−0.119	0.633	−0.019	0.620
	ON	4.051*	0.044	0.155	0.456	0.024	0.457
	MB	0.046	0.829	–	–	–	–
	SK	0.019	0.891	–	–	–	–
	BC	2.277	0.131	–	–	–	–
	AB (omitted)	–	–	–	–	–	–

Notes: $n = 12,124$.