Digital Divide: Empirical Study of CIUS 2020*

Joann Jasiak† Peter MacKenzie‡ Purevdorj Tuvaandorj§

October 7, 2024

Abstract

As Canada and other major economies consider implementing "digital money" or Central Bank Digital Currencies, understanding how demographic and geographic factors influence public engagement with digital technologies becomes increasingly important. This paper uses data from the 2020 Canadian Internet Use Survey and employs survey-adapted Lasso inference methods to identify individual socio-economic and demographic characteristics determining the digital divide in Canada. We also introduce a score to measure and compare the digital literacy of various segments of Canadian population. Our findings reveal that disparities in the use of e.g. online banking, emailing, and digital payments exist across different demographic and socio-economic groups. In addition, we document the effects of COVID-19 pandemic on internet use in Canada and describe changes in the characteristics of Canadian internet users over the last decade.

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

*The authors thank Henry Kim, the participants of the 13th CEQURA Conference, 48th ACEA Conference, 57th Annual Conference of the Canadian Economics Association, Canadian Econometrics Study Group (CESG) 38th Annual Meeting and the seminar of CBDC Econ Working Group at Bank of Canada 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 was awarded Best Student Presentation for presenting this paper at the 48th ACEA Conference.

†York University, `jasiakj@yorku.ca`

‡York University, `petem9@yorku.ca`

§York University, `tpujee@yorku.ca`

1 Introduction

The digital divide represents a gap between those who can fully participate in the digital world and those who cannot. It is determined by the availability of digital infrastructure, such as high-speed internet, and the ability and willingness of individuals to engage with digital technologies, which depend on their socio-economic and demographic characteristics.

As Canada and other major economies explore the implementation of "digital money" or Central Bank Digital Currencies (CBDC), it becomes crucial to understand the extent of the digital divide in Canada and reveal the characteristics of Canadians affected by it. Limited individual engagement with digital technologies is an obstacle in the advancement of internet-based services, including digital banking, education, and emerging financial technologies such as CBDC. Individual connectivity is necessary to ensure a balanced growth of the economy and fair participation for Canadians in an increasingly digital society.

The Canadian government has taken significant steps to develop internet availability by investing heavily in digital infrastructure. As a result of the High-Speed Access for All: Canada's Connectivity Strategy, along with a \$1.7 billion investment from the 2019 federal budget, 94.15% of Canadians reported having internet access at home in 2020. However, access to the internet does not ensure the use of internet, and on-line services. Individuals must be able to afford internet services and devices, have access to them, and be willing to use them.

This paper uses the 2020 Canadian Internet Use Survey (CIUS) to investigate how socio-economic and demographic characteristics of Canadians influence their engagement with digital technologies. CIUS 2020 contains more information than the previous installments of CIUS conducted by Statistics Canada, especially on the use of internet for online banking and digital payments. It also includes more individual characteristics, such as visible minority status and Aboriginal identity. Our objective is to provide an updated and comprehensive study to inform Canadians and future policymakers given Canada's aging society and the increasing role of digital banking and cashless transactions.

Our first contribution is in applying survey-adapted Lasso inference methods to identify key socio-economic and demographic individual characteristics that determine the use of internet, e-mail, online banking and digital payments through virtual wallets and credit cards. The novel `svy LLasso` estimator for logit models (Jasiak and Tuvaandorj, 2023) allows us to analyze a large number of explanatory variables and their interactions, while incorporating survey weights to ensure the results are representative of the Canadian population. We also use multiple correspondence analysis (MCA) of qualitative socio-demographic variables from CIUS 2020 to reveal combined effects of multiple individual characteristics.

Limited usage of the internet by the individuals who can access and afford it is often caused by a poor level of digital literacy, defined as “the ability to use information and communication technologies to find, evaluate, create, and communicate information, requiring both cognitive and technical skills” (American Library Association, n.d.). Our second contribution is in approximating digital literacy in Canada by designing a composite score based on digital technology usage. This score allows us to compare and rank digital literacy across various population segments.

The COVID-19 pandemic has highlighted the importance of both digital access and digital literacy. To study the effects of the pandemic, we use cluster analysis to identify two distinct groups of Canadians: one that increased their use of digital technologies during the pandemic and another that was less inclined to embrace these changes. We then observe that the COVID-19 measures introduced by different provinces possibly influenced digital adoption rates.

In our empirical study, we reveal several interesting new results based on CIUS 2020 reflecting recent trends in Canadian society. We find that women are more likely than men to use email and score higher in digital literacy. Recent immigrants and visible minorities score high in digital literacy too. Among the recent immigrants the English-speaking ones use more online banking. We also observe that visible minorities are more frequently using virtual wallets. In contrast, certain groups of individuals such as those age 65 and over, those with low income, the unemployed, single older individuals, and those with only a high school education, especially residents of Manitoba and Maritime provinces — are becoming more disconnected from an increasingly digital society. While the CIUS data excludes First Nations on reserve, we find that off-reserve First Nations use less internet and score lower on digital literacy than non-Aboriginals. This raises concerns given Canada’s aging population and slow progress of truth and reconciliation initiatives, and emphasizes the need to address the digital divide affecting the disadvantaged groups.

Consistent with earlier research, we find that age, income, and education influence digital technology usage. These individual characteristics were found relevant in the past study of internet use and online activity level in Canada by Haight et al. (2014) based on CIUS 2010, and remain statistically significant, in addition to the gender and recent immigrant status. Similar findings have also been reported in empirical research conducted abroad (Reddick et al., 2020; Robinson et al., 2015; Cullen, 2001; Friedline et al., 2020). However in the past, women and recent immigrants were found to be less likely to use online activities in Canada (Haight et al., 2014) and the U.S. (Zickuhr and Smith, 2012). We observe that internet connectivity among women has increased, and the access gap between immigrants and Canadian citizens identified by Haight et al. (2014) has diminished. The high digital literacy of women, especially those employed, as documented in our study, may be attributed to their growing participation in STEM and technology-intensive fields. In addition, Canada’s high-skilled immigration policies and Federal Skilled Trades Program may be positively influencing the digital literacy and technology usage of immigrants.

The availability of the visible minority status in CIUS 2020 allows us to provide new results, which are mostly encouraging. However, visible minorities may not be accessing internet and digital payments equally in all provinces. Our study of interaction effects reveals that visible minority in Manitoba are less likely to use email, which points to the problem of regional disparities in Canada.

In Canada, the digital divide has traditionally been characterized by the rural-urban gap, with urban areas generally exhibiting higher levels of digital engagement compared to rural counterparts (Carson, 2013). Our results indicate that rural residents are not only less likely to use the internet but also emails and virtual wallets, with this issue concerning both Ontario and Quebec. However, we find no evidence of rural residents using less online banking and credit cards than the urban residents. Because of the limited scope of CIUS (2020), our study does not cover on-reserve Aboriginal communities, whose access to broadband internet use in Canada is discussed by (Koch, 2022) in the context of the aforementioned federal Connectivity Strategy in rural communities and First Nations reserves.

Recent studies have also explored the intra-urban divide, focusing on disparities driven by factors like education, income, and other socio-economic variables (Reddick et al., 2020; Dewan and Riggins, 2005; Wavrock et al., 2022). These studies show that significant inequalities persist, particularly among vulnerable groups such as low-income households and seniors. We complement these findings by studying the interactions of variables, including low income and age. In particular, low income or single Canadians who are more than 65 years old, or single and French-speaking Canadians are found in our study to be disadvantaged.

Koch (2022), Reddick et al. (2020), and Van Deursen and Van Dijk (2019) have underscored the importance of digital literacy and usage in understanding the full scope of the divide. Reddick et al. (2020) and Van Deursen and Van Dijk (2019) argue that digital literacy is a major obstacle in access to broadband internet in the U.S. and Netherlands respectively. Koch (2022) addresses the importance of designing government funded initiatives to improve the digital literacy in Canada. Our digital literacy score is a new instrument of analysis indicating clearly which segments of Canadian population need to be given priority in this respect: old, low income, with low educational attainment, First Nations, and residents of Maritime provinces.

The COVID-19 pandemic caused significant changes in the digital lives of Canadians (Koch, 2022; Wavrock et al., 2022; Engert and Huynh, 2022). With physical distancing measures and stay-at-home orders, Canadians increasingly turned to digital platforms for work, shopping, education, and social interaction (Aston et al., 2020; Deng et al., 2020). Education and income levels played a crucial role in determining internet access and the ability to fully participate in this digital shift (Wavrock et al., 2022), which was not homogeneous across Canada according to our results.

Our evidence shows that relatively less Canadians adopted new digital technologies during the pandemic in the Maritime provinces.

In the existing literature, the digital divide is conceptualized across three levels: access (material access to technology), usage (the ability to effectively engage with digital tools), and outcomes (the tangible benefits resulting from digital usage) (Van Deursen and Van Dijk, 2019; Ferreira et al., 2021). Early research predominantly concentrated on the first-level digital divide, which pertains to the supply of internet and physical access to technology (Cullen, 2001; Van Dijk and Hacker, 2003), whereas recent research has shifted focus toward the demand side—the factors influencing usage and outcomes.

Among the 5.85% of the Canadian population without internet access, both demand-side and supply-side barriers persist (Jordan, 2019). Demand-side barriers account for the majority of reasons and include lack of interest (50.83%), high service costs (26.08%), and the high cost of equipment (2.48%). Supply-side barriers, such as service unavailability, represent a smaller portion, accounting for 6.75% of non-adoption reasons. These barriers limit people's connectivity to the internet and digital technologies, with demand-side issues being the predominant obstacles.

On the demand side, personal characteristics and digital literacy, influence the preferences for digital technologies impacting their usage. For instance, Chen, Engert, Huynh, O'Habib, Wu and Zhu (2022) find that the use of debit and credit cards has generally increased since the pandemic; however, some subsets of Canadians continue to prefer cash for transactions (Henry et al., 2023; Engert and Huynh, 2022). This indicates that even when access is available, demand-side factors such as personal preferences and digital literacy affect the usage of digital technologies and the benefits derived from them.

As digital payment adoption rises, significant challenges persist for many First Nations communities, including limited internet access and difficulties maintaining access to cash (Chen, Engert, Huynh and O'Habib, 2022). The shift from traditional to digital banking has altered consumer behavior and accelerated the closure of physical bank branches as institutions prioritize digital optimization (Aversa et al., 2022). While this shift offers convenience, it raises concerns about financial exclusion (Kamdjoug et al., 2021), in particular given lower digital literacy of (off-reserve) First Nations documented in our study. Cash usage in Canada has declined sharply, with only one in three transactions now involving physical cash (Huynh, 2017). A concern of the Bank of Canada is the increasing interest in cryptocurrency which are not regulated and highly volatile. Despite global interest, cryptocurrency adoption so far remains low in Canada (Huynh et al., 2020; Adrian and Mancini-Griffoli, 2019). The data on cryptocurrency use is available in CIUS (2020). However, the sample of respondents is too small for valid inference and virtual wallets are explored instead in this paper. Our study provides reliable data-based insights that can help identifying groups to be disadvantaged in a future cashless economy, or with a fully digitalized banking system.

The paper is organized as follows: Section 2 describes the CIUS 2020 dataset. Section 3 lays out the paper's estimation and inference approach. Section 4 presents the `svyLLasso` results, MCA diagrams, and digital literacy approximations. Section 5 provides additional analyses examining the impact of COVID-19 on the digital divide and a comparison with the results of Haight et al. (2014) on CIUS 2010. We conclude in Section 6. The online appendix provides a description of the sampling and weighting scheme used in CIUS 2020, technical details of the methods used in the paper, further information on the digital literacy score, and additional estimation results.

2 Data description

This section describes the CIUS 2020 survey and the variables used in our empirical analysis. CIUS 2020 is the most relevant 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 overall response rate to the survey is 41.6%.

CIUS 2020 data are appropriately weighted using sample weights. Statistics Canada provides the weight variables, which are based on independent estimates for various age and sex groups in each province and account for survey non-response, among other factors (see Online Appendix Section A for the stratification scheme and survey weights). Properly weighting the data allows the sample of the Canadian population used in CIUS 2020 to accurately represent the entire population.

To assess the digital divide, we study the demographic and socio-economic characteristics of CIUS 2020 respondent, which appear as the explanatory variables in logit models of the use of the internet and selected internet-based services. Sections 2.1 and 2.2 describe these dependent and explanatory (independent) variables, respectively.

2.1 Dependent variables

We consider the logit models of internet use and of the use of internet-based services, which are internet use, email use, online-banking, virtual wallet, and credit card payments. The first variable (internet use) reveals the social connectivity of Canadians. The latter four dependent variables are chosen to examine the readiness of Canadians to transition towards digital financial technologies.

The following five questions from CIUS 2020 serve as the dependent variables for our logistic models:

- **Internet use (Model 1):** "During the past three months have you used the internet from any location?" This binary question, with responses *Yes* or *No*, helps determine the factors affecting whether a Canadian individual has access to the internet.
- **Online banking (Model 2):** "During the past three months have you conducted online banking?" This question gauges the demographic factors influencing a person's proficiency and trust in conducting online financial transactions.
- **Email use (Model 3):** "During the past three months have you sent and received emails?" The use of emails is a basic marker of digital literacy and provides insights into the user's familiarity with standard online communication tools.
- **Virtual wallet usage (Model 4):** "During the past twelve months have you used a virtual wallet to pay for goods over the internet?" This question identifies factors affecting whether Canadians use virtual wallets for payments.
- **Online credit card use (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?" This question provides insights into the trust and usability of online financial transactions among Canadians.

In Models 2–5, we categorize the possible responses as 1) *Yes*, 2) *No*, and 3) *Not stated*. We test the Independence of Irrelevant Alternatives (IIA) hypothesis to determine whether to include the *Not stated* category in Section 4.1.

Sample sizes

Model 1, concerning internet use, encompasses the full sample with 17,409 respondents. For Models 2 and 3, representing online banking and email use respectively, we excluded the *Not stated* responses based on the results of the Hausman test for IIA. Model 2 is thus based on 17,135 respondents after excluding 274 *Not stated* responses, while Model 3 comprises 17,268 respondents following the exclusion of 141 *Not stated* responses.

Models 4 and 5 are each based on data from 12,124 respondents. The reduction of the sample size, when compared to the full sample, arises from the sequential structure of CIUS design. The survey filters respondents based on their internet usage and other specific activities, such as expenditure on digital goods and services. As a result, certain alternatives have a probability of zero. Additionally, 307 responses were marked as *Not stated* and excluded based on the results of a Hausman test for IIA.

2.2 Explanatory (Independent) variables

The selected explanatory variables in the logistic regression models 1 to 5 provide a comprehensive profile of the respondents, capturing their socioeconomic, and demographic characteristics. These encompass income, education, employment status, Aboriginal identity, visible minority status, immigration status, age, gender, location, type of household, language spoken at home, and province. All variables are multi-categorical and detailed in the regression tables.

For many explanatory variables, a *Not stated* category exists as well and is retained in the regressions. Exclusion of this category could introduce bias, given that respondents who selected *Not stated* for one question often provided answers to others.

Each model omits the categories associated with a representative individual 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

We consider 41 explanatory (independent) categorical variables. Some of these variables are expected to have direct effects on the dependent variables of the model, while others are included to account for potential interaction effects. For instance, household type and income variables may exhibit cross-effects on dependent variables like internet use and online banking. Accounting for second-order interactions results in 674 control variables, a number that is relatively large compared to the sample size. However, there is no a priori guidance on which variables should enter the model.

In situations where a model contains many regressors, Lasso variable selection techniques are known to flexibly reduce the dimensionality of the data and select variables with higher predictive power for explaining the categorical dependent variables of interest. For these reasons, we adopt the logistic Lasso approach, well-suited for this problem. It possesses optimality properties under a sparsity assumption and leads to automatic variable selection (Belloni et al., 2014; Mullainathan and Spiess, 2017).

The survey weights play a crucial role in ensuring the generalizability of survey results to the entire Canadian population. However, existing Lasso-based estimation and inference methods, including the commonly used logit Lasso variable selection, require adjustment for survey weights. This paper employs a new logistic Lasso variable selection method for binary choice models in a survey environment, termed the `svy LLasso`, which is described below. The asymptotic properties of the `svy LLasso` estimator are given in Jasiak and Tuvaandorj (2023).

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).

Prior to inference being made on the coefficients, the `svy LLasso` estimator needs to be transformed to ensure valid results. 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.

An alternative transformation method considered is 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 labelled as C_α. See Online Appendix Section 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 shows the `svy LLasso` logit estimation results for Models 1-5. We analyze the logit models with interaction effects 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.

4.1 svy LLasso logit models

As stated in Section 2, the online banking, email use, virtual wallet, and credit card dependent variables have three categories: *Yes*, *No*, and *Not stated*. We use first the survey-weighted Hausman-McFadden test of the IIA hypothesis to see if we can remove the *Not stated* observations from the logit models. The online banking variable has a Hausman-McFadden statistic of 0.05 with a p-value of 1, which is strong evidence in favor of IIA. Hence, we use the restricted specification of Model 2, 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. Therefore, conventionally, the *Not stated* observations are removed from these models as well.

We report the empirical results, including the `svy LLasso` estimates and the test results based on the debiased Lasso estimates of the logit model coefficients and AMEs, $\tilde{\theta}^{DB}$ and $\widetilde{\text{AME}}^{DB}$ in Tables 1–5 below. Tables 2–6 in the Online Appendix Section D present the results of the selective inference and $C(\alpha)$ procedures, which are consistent with the debiased Lasso results.

The estimation of the internet use Model 1 reveals which explanatory variables influence a person's internet connectivity. The estimation of Models 2-5 shows which explanatory variables are essential for the use of internet-based devices. We consider evidence of a digital divide in Models 1-5 under the following conditions: a) When some explanatory variables are selected by the `svyLLasso` and statistically significant while others are not, the divide is between individuals with and without the characteristics represented by the selected variables. b) When some categories within an explanatory variable are selected by the `svyLLasso` and statistically significant while others are not, this suggests a divide between individuals belonging to the selected categories and those who do not. c) When all categories of an explanatory variable are selected by the `svyLLasso` and statistically significant, but the coefficients either have different signs or take noticeably different values, the digital divide is indicated by these distinctions in coefficient signs or magnitudes.

Model 1: Internet use. Table 1 presents the results based on the internet use model and reveals the explanatory variables influencing an individual's connectivity and usage of the internet.

The model reveals a rural-urban divide in internet access. Specifically, rural Canadians have a 1.7% lower probability of having used the internet in the prior three months compared to their urban counterparts. This disparity corroborates with the findings of *Canada's connectivity strategy*. Despite substantial federal investments to bolster rural internet access, this discrepancy persists, emphasizing the enduring challenges rural residents confront in bridging the digital divide.

All age group categories are selected by `svy LLasso` and are statistically significant. Younger age brackets, specifically those between 15 and 44, have positive coefficients and AME values. In contrast, those aged 55 and above have negative coefficients and AME values. Individuals in the *25-34* age category are 5.4 percentage points more likely to be internet users, while the eldest group (*65 and older*) is 8.2 percentage points less likely compared to the reference group of *45-54* years.

Several demographic factors are selected by `svy LLasso` and statistically significant. For instance, those who are employed, predominantly English speakers, university graduates, and high earners are more likely to use the internet. 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 have a lower likelihood of internet usage.

The results for internet connectivity are generally close to the findings of past research on internet connectivity in Canada (Haight et al., 2014; Friedline et al., 2020; Jordan, 2019). However, there are differences concerning the gender or immigration variables, which are not selected by `svy LLasso` or found to be statistically significant for internet use in our analysis.

Model 2: Online banking. Table 2 presents the findings from the online banking model, exploring the factors that influence Canadians' adoption and use of digital financial technologies. The online banking model is of particular significance, as the current online banking systems may share functional parallels with potential digital financial technologies, like a CBDC system.

The results indicate that younger, employed, high-income, and university-educated Canadians are more inclined to utilize online banking. Factors such as lower educational attainment, lower income, identification as a visible minority, and being aged 55 or older reduce the likelihood of online banking usage. As indicated by the absolute value of AMEs, the most impactful variables include the age category of *65 and older*, employment status, and educational attainment of *High* *school or less.*

Individuals in the *65 and older* age category display a notable divergence in behavior, being 15.4 percentage points less likely to use online banking than those in the *45-54* age group. Employment status is another prominent determinant; specifically, those employed exhibit a 10.6 percentage point heightened likelihood of using online banking relative to their unemployed counterparts. Educational credentials further accentuate the divide. Those whose highest educational achievement is *High school or less* are 11.4 percentage points less likely to engage in online banking than individuals with at least *Some post-secondary* education. Notably, the `svy LLasso` did not select variables such as *Location, Gender, Aboriginal identity*, and *Province* as influential determinants in the model.

Model 3: Email use. Table 3 presents the results for email usage, a metric that gauges Canadians' digital social and professional connectivity. Email is one of the most commonly used internet service among those that are connected. While there are similarities in the variables influencing both email usage and online banking, as seen in the selections by `svy LLasso` and the statistically significant explanatory variables, there are also intriguing distinctions.

One notable difference is the influence of location. While the *Rural* category is associated with a reduced likelihood of email use, it does not significantly affect online banking. A plausible interpretation is geographical necessity: rural Canadians might lean towards online banking due to their distance from physical bank branches. Additionally, rural employment might not demand as extensive email communication as certain urban jobs.

Gender dynamics offer another dimension of differentiation. The *Female* category is influential in the email use model, though its impact, as evidenced by the debiased Lasso AME, is relatively modest. The difference in email use based on gender could reflect occupational patterns, with women potentially occupying more office roles that necessitate email, in contrast to blue-collar roles that might be more prevalent among men.

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* is selected by `svy LLasso` and is statistically significant. This category 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 background influences email usage patterns. Those with a *University degree* or higher exhibit a stronger propensity for email use than those with only *Some post-secondary education.* Individuals with an education level of *High school or less* show a diminished likelihood.

The trend seen in educational attainment might arise from the nature of jobs accessible at different educational levels, often linking higher education qualifications to roles that require frequent email communication.

Model 4: Virtual wallet. Some internet users make online payments using virtual wallets. Table 4 details the explanatory variables influencing the adoption of virtual wallets in Canada, a crucial variable for research on digital currencies in the country. While previous models assessed Canadians' internet connectivity and use of other digital technologies, the virtual wallet model will show what factors currently affect the uptake of digital payment methods.

Age emerges as an important determinant in virtual wallet use. All age group categories, excluding those aged *35-44* were selected by `svy LLasso` and statistically significant. Younger Canadians have the highest probability of using a virtual wallet. The age group, *15-24*, has an 11.2 percentage point increase in the likelihood of using a virtual wallet than the base age group of *45-54*. In contrast, older Canadians, especially those *65 and older*, demonstrate a decreased likelihood. The age group *65 and older* is the least 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 than the reference age group to use a virtual wallet.

The coefficient for *Visible minority* is chosen by `svy LLasso` and statistically significant. *Visible minority* has a positive AME on the use of a virtual 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 to use a virtual wallet 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.

Income and education stand out as influential variables. Specifically, Canadians earning $\$146,560$ or more are more likely to use a virtual wallet than those in the base income category of $\$52,204$–$\$92,485$. Additionally, individuals holding a *University degree* are 8 percentage points more likely to use a virtual wallet than those with *Some post-secondary* education.

Model 5: Credit card. Credit card payments are the most popular way to make purchases online. Table 5 presents findings on the explanatory variables influencing Canadians' use of credit cards for online transactions—a crucial understanding considering the anticipated card component of a potential CBDC. Given the frequent use of credit and debit cards in the current Canadian financial landscape, these insights are pivotal for the successful integration of a CBDC.

The model reveals some interesting results. Again, age is a significant determinant: younger Canadians, specifically those in the *15-24* age bracket, are 8.8 percentage points less likely to utilize a credit card for online purchases when compared to the reference group of *45-54*. Education emerges as another prominent factor. Individuals with a *High school or less* education level show a reduced likelihood for online credit card transactions. Those with a *University degree* are more inclined towards such transactions.

Economic and regional factors have an impact on whether Canadians use credit cards for online shopping. People residing in Quebec and those in the lowest income bracket are less likely to use credit cards for online purchases. On the other hand, English-speaking Canadians, those with a university degree, people who are employed, residents of family households without children under 18, and Ontarians are more likely to use credit cards for online shopping.

Summary of results. Our analysis identifies age, education, and income as key factors in digital adoption: younger, more educated, and higher-income individuals are more digitally engaged, revealing a socio-economic digital divide. Contrary to previous research, immigration status and gender did not have an impact on the digital divide, with immigrants and women displaying similar levels of digital engagement as other groups. Additionally, visible minorities are increasingly adopting new technologies like virtual wallets, and younger demographics prefer alternative digital payment methods over traditional ones, indicating a shifting digital landscape.

4.2 Interaction effects

To enhance the model's predictive ability, we include the relationships between the explanatory variables, incorporated as interaction terms in the logit Models 1-5 estimated by `svy LLasso`. First, we examine whether the second-order specification with interaction terms is more appropriate than the first-order specification in Models 1–5. Accordingly, we 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 with and without the interactions, using `R` package `polywog`. Table 6 reports the result.

The results show that a linear specification for Models 1, 4, and 5 results in smaller mean-squared errors, while a second-order specification might be preferred for Models 2 and 3 in terms of the prediction error.

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.

Tables 7 and 8 present the interaction results for extended Models 2 and 3 of online banking and email usage, respectively. The tables include only those variables which are statistically significant at the 5% level, as indicated by their coefficient p-values. We have omitted significant interaction variables involving *Not stated* responses due to interpretability concerns.

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	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
	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	Female	0.013	0.099	0.200	0.006	0.211
Aboriginal	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	–	0.065	0.836	0.004	0.842
	English and French	–	0.793	0.124	0.036	0.231
	French and Non-official	–	−0.533	0.544	−0.035	0.495
	English, French and Non-official	–	−1.434	0.193	$-0.118^{\bullet}$	0.067
Employment	Employed	0.514	0.574^{***}	0.000	0.032^{***}	0.000
Education	High school or less	−0.911	-0.971^{***}	0.000	-0.058^{***}	0.000
	University degree	0.451	0.519^{***}	0.000	0.027^{***}	0.000
Minority	Visible minority	−0.048	-0.352^{*}	0.037	-0.021^{*}	0.034
Household type	Family w/o 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
	$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	Non-landed immigrant	–	−0.258	0.176	−0.014	0.211
Province	NL	–	−0.31	0.111	$-0.019^{\bullet}$	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

Note: $n = 17,409$. $\tilde{\theta}^{DB}$ 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`. Comparison categories and variables with *Not stated* answers are not displayed for clarity and interpretability. Significance levels are marked as: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, $^{\bullet}$ $p < 0.1$.

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	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
	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	Female	–	0.089	0.107	0.014	0.123
Aboriginal	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	–	0.337•	0.090	0.050	0.122
	English and French	–	0.239	0.526	0.036	0.561
	French and Non-official	–	−0.280	0.648	−0.046	0.647
	English, French and Non-official	–	−0.127	0.858	−0.020	0.861
Employment	Employed	0.662	0.653***	0.000	0.106***	0.000
Education	High school or less	−0.637	−0.686***	0.000	−0.114***	0.000
	University degree	0.331	0.409***	0.000	0.062***	0.000
Minority	Visible minority	−0.135	−0.303**	0.003	−0.048**	0.005
Household type	Family w/o 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•	0.068
Income	$52,203 and lower	−0.265	−0.252***	0.001	−0.041**	0.002
	$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	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

Note: $n = 17,135$. $\tilde{\theta}^{DB}$ 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`. Comparison categories and variables with *Not stated* answers are not displayed for clarity and interpretability. Significance levels are marked as: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, • $p < 0.1$.

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	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
	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	Female	0.087	0.151*	0.021	0.015*	0.025
Aboriginal	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
	French and Non-official	–	−0.302	0.669	−0.032	0.656
	English, French and Non-official	–	−1.908*	0.019	−0.272***	0.001
Employment	Employed	0.411	0.457***	0.000	0.045***	0.000
Education	High school or less	−0.790	−0.851***	0.000	−0.088***	0.000
	University degree	0.750	0.828***	0.000	0.072***	0.000
Minority	Visible minority	−0.192	−0.346**	0.008	−0.035*	0.011
Household type	Family w/o 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
	$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	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	–	$-0.251^{\bullet}$	0.098	$-0.026^{\bullet}$	0.090
	QC	−0.154	−0.326*	0.044	$-0.033^{\bullet}$	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	$0.260^{\bullet}$	0.077	$0.024^{\bullet}$	0.073

Note: $n = 17,268$. $\tilde{\theta}^{DB}$ 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`. Comparison categories and variables with *Not stated* answers are not displayed for clarity and interpretability. Significance levels are marked as: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, $^{\bullet}$ $p < 0.1$.

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	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
	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	Female	–	−0.091	0.280	−0.010	0.277
Aboriginal	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	–	−0.410	0.116	−0.040	0.153
	English and French	–	0.105	0.829	0.012	0.822
	French and Non-official	–	−0.618	0.448	−0.055	0.534
	English, French and Non-official	–	−0.907	0.359	−0.073	0.495
Employment	Employed	–	0.020	0.853	0.002	0.852
Education	High school or less	–	−0.066	0.568	−0.007	0.568
	University degree	0.027	0.254**	0.009	0.028**	0.008
Minority	Visible minority	0.162	0.453***	0.001	0.052***	0.001
Household type	Family w/o 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
	$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	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

Note: $n = 12,124$. $\tilde{\theta}^{DB}$ 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`. Comparison categories and variables with *Not stated* answers are not displayed for clarity and interpretability. Significance levels are marked as: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, • $p < 0.1$.

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	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
	55–64	–	-0.022	0.830	-0.003	0.835
	65 and older	–	-0.053	0.653	-0.008	0.662
Gender	Female	–	-0.004	0.958	-0.001	0.959
Aboriginal	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	–	-0.044	0.844	-0.007	0.849
	English and French	–	-0.185	0.646	-0.030	0.644
	French and Non-official	–	-0.777	0.263	-0.141	0.202
	English, French and Non-official	–	$-1.352^{\bullet}$	0.088	-0.266^{*}	0.035
Employment	Employed	0.002	$0.148^{\bullet}$	0.083	$0.023^{\bullet}$	0.091
Education	High school or less	-0.411	-0.453^{***}	0.000	-0.073^{***}	0.000
	University degree	0.357	0.490^{***}	0.000	0.073^{***}	0.000
Minority	Visible minority	–	-0.235^{*}	0.044	-0.037^{*}	0.046
Household type	Family w/o 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
	$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	Non-landed immigrant	–	0.151	0.232	0.024	0.237
Province	NL	–	$-0.287^{\bullet}$	0.081	$-0.047^{\bullet}$	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	$0.037^{\bullet}$	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

Note: $n = 12,124$. $\tilde{\theta}^{DB}$ 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`. Comparison categories and variables with *Not stated* answers are not displayed for clarity and interpretability. Significance levels are marked as: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, $^{\bullet}$ $p < 0.1$.

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

Note: 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`. Models 1-5 are internet use, online banking, email use, virtual wallet, and credit card, respectively.

In the extended online banking Model 2, alongside the notable negative influence of the variables *High school or less* and *Visible minority*, an interesting pattern of the interaction variables emerges. Specifically, the age group *15-24* interacting with *Family without children under 18* has a significant positive effect on online banking use. This result highlights the distinct digital behavior of younger individuals in specific family settings, emphasizing the role of age and household composition in digital engagement.

The interaction between the *65 and older* age group and *Single* households is particularly revealing. This combination is significant, corroborates the finding in Section 4.1 and shows the digital divide disproportionately affects older, single individuals, especially those with lower incomes. This finding is crucial as it explores demographic effects combined with socioeconomic factors, suggesting a more complex picture where subsets of the older population are particularly at a disadvantage digitally. It is a pattern that warrants attention, as it highlights a segment of the population that might be struggling to keep pace with the rapid digitization of financial services. The Canadian population is also aging, which makes these findings even more important for policymakers.

In the email usage model, the interaction of the age category *65 and older* with the lower income category (*$52,203 and lower*) further underscores this concern. The significant negative impact on digital engagement among older, lower-income individuals indicates the challenges this demographic faces in accessing and utilizing digital technologies. It paints a picture of a group being left behind in the digital landscape, emphasizing the need for targeted interventions to bridge this gap.

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

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 w/o 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) × (High school or less)	–	1.144^{*}	0.026
Language × *Immigration*	(Non-official) × (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 w/o children under 18) ×($52,203 and lower)	–	-0.650^{*}	0.016
	(Single) × ($52,203 and lower)	–	-0.659^{*}	0.013

Note: $n = 17,135$. The coefficients shown in this table are found to be significant at the 5% level based on their estimated p-values. Comparison categories and variables with *Not stated* answers are not displayed for clarity and interpretability. Significance levels are marked as: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Table 8: Lasso Logistic Regression with Interactions for Email Use 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
Minority × *Province*	(Visible minority) × (MB)	–	-1.825^{**}	0.005

Note: $n = 17,268$. The coefficients shown in this table are found to be significant at the 5% level based on their estimated p-values. Comparison categories and variables with *Not stated* answers are not displayed for clarity and interpretability. Significance levels are marked as: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

4.3 Multiple correspondence analysis

The dependent and explanatory variables discussed in Sections 2.1 and 2.2 are categorical. Hence, the relationships between these variables can be analyzed using the multiple correspondence analysis (MCA) for associations between categorical variables. Its advantage is that the MCA displays graphically complex dependencies involving the interactions between groups of variables. We consider it as complementary to the logit estimation results, as the MCA results are less rigorous, although easy to interpret. In particular, the MCA allows us to explore further the interaction effects, including both the dependent and explanatory variables.

Figures 1 and 2 display the variable categories represented in two-dimensional space.

Internet use, email use and online banking. Figure 1 presents the associations between the explanatory and dependent variables appearing in Models 1-3 of internet use, email use, and online banking. The green-labelled variable categories are the dependent variables in our logit models and the supplemental variables in the MCA. The red-labelled categories are the explanatory variables in our logit models.

The groupings of variable categories illustrate graphically the underlying structure of the data. The most apparent grouping of variable categories is in the top left quadrant of the graph. This grouping includes individuals who did not use the internet, email or online banking 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. Our logistic regressions identified these explanatory variables as statistically significant.

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.

Virtual wallet and credit card use. Figure 2 illustrates the associations between the dependent and explanatory variables appearing in Models 4-5 of the virtual wallet and credit card use. 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 `svy LLasso`.

`svy LLasso` selected the variable *Visible minority.* 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.

4.4 Digital literacy score

The use the internet and internet-based services is determined not only by the demographic and socio-economic characteristics of an individual, but also by their digital ability, or digital literacy. The digital literacy is an outcome of various socio-economic characteristics. It is unobserved, i.e. latent as there is no CIUS question that provides direct information about the digital literacy of the respondents. We create a measure (score) of digital literacy and apply it to groups of individuals distinguished in the previous sections to assess the digital divide in a more rigorous way.

We analyze the distributional properties of the score of digital literacy in the entire sample. We also compare its values in the groups of individuals distinguished with respect to location, age,

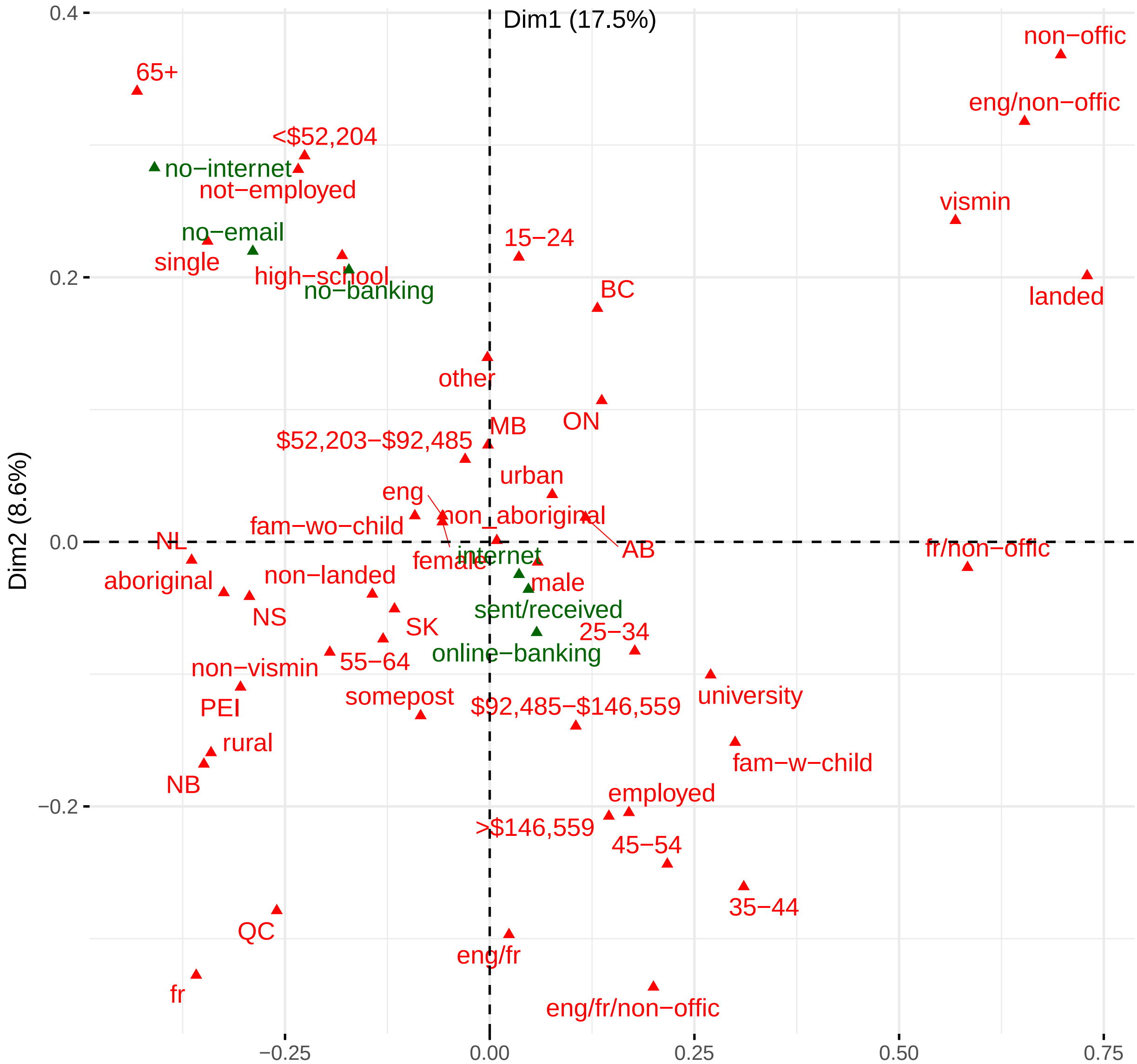


Figure 1: Coordinate plot for Internet Use, Email Use and Online Banking

Note: Figure 1 shows the multiple correspondence analysis coordinate plot for the dependent variables internet use, email use, and online banking (labelled in green). The explanatory variables in this analysis are labelled in red.

gender, education level and immigration status. We consider noticeable differences in the value of the digital literacy score as the evidence of digital divide between, or inside the groups.

The digital literacy score comes from survey respondents' answers to 10 questions in CIUS 2020 (see Online Appendix Section C for the list of 10 questions our score comprises). Respondents that answer *Yes* to these questions receive 1 point per *Yes* response [1].

[1] All relevant questions were asked to a subset of 12,431 CIUS 2020 respondents. After removing *Not stated* answers from this subset, we are left with 11,874 observations. We compared our digital literacy score with other

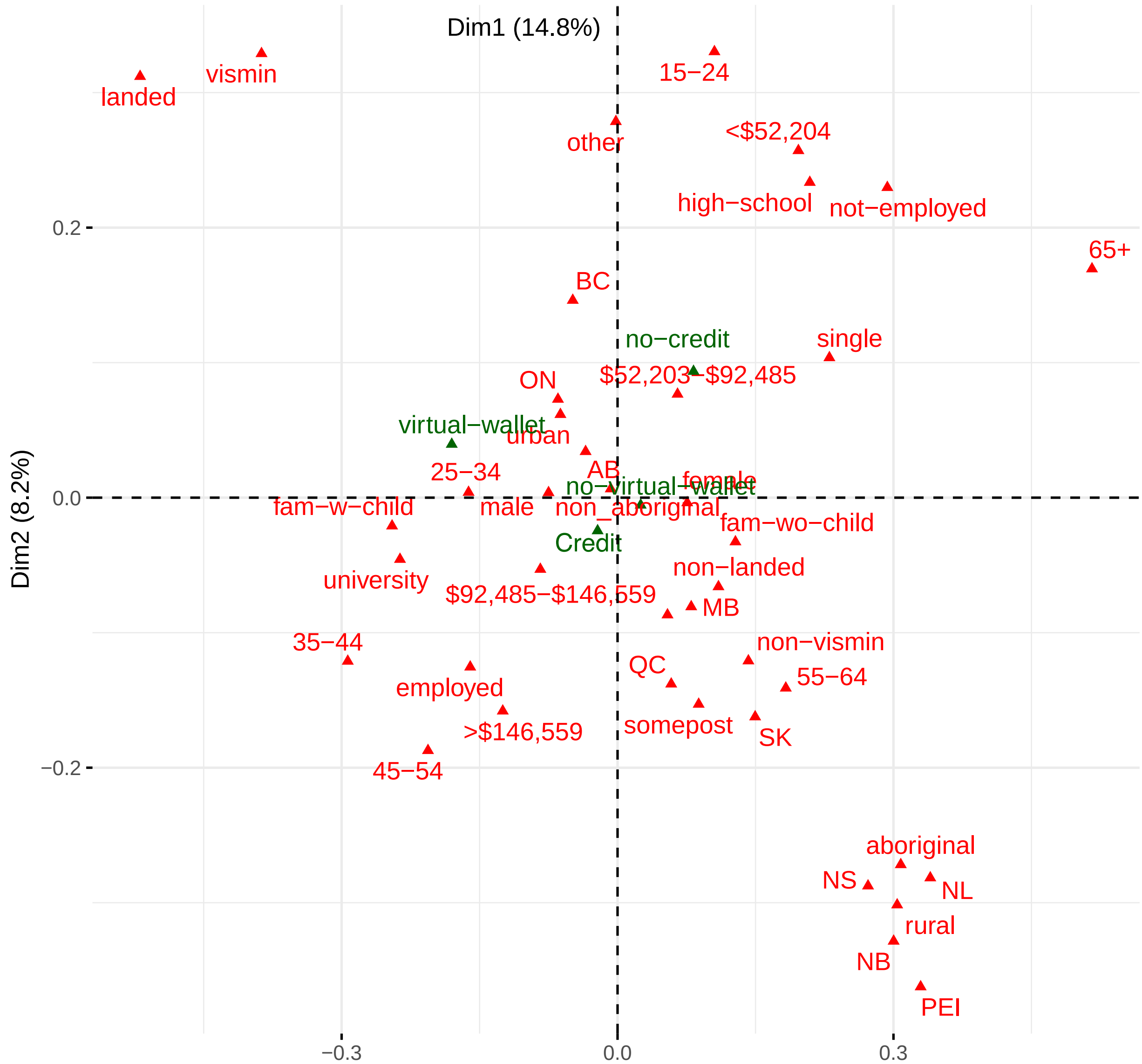


Figure 2: Coordinate plot for Virtual Wallet and Credit Card Use

Note: Figure 2 shows the multiple correspondence analysis coordinate plot for the dependent variables virtual wallet and credit card (labelled in green). The explanatory variables in this analysis are labelled in red.

The higher the score (out of 10), the higher the perceived digital literacy of the respondent. Next, we compute the average scores for the aforementioned groups of individuals and display these results in Table 9. The average score from all respondents in Table 9 is 7.11, and the standard deviation is equal to 2.15. Therefore, respondents answered an average of seven questions with *Yes*.

samples where *Not stated* answers were replaced by multiple imputation methods following Van Buuren (2018). Results were consistent across models.

The first group of individuals we investigate is distinguished with respect to the location. The first two rows of the table show the average score out of 10 for survey respondents residing in urban and rural locations. Urban residents score higher in digital literacy than rural residents. This rural/urban divide is consistent with our `svy LLasso` and MCA results that show a divide between rural and urban residents regarding internet connectivity.

The age group variable shows one of the largest divides regarding digital literacy score. The oldest age group category *65 years and older* has the 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 focus on making online purchases and using digital technology, potentially skewing toward people in the middle age groups.

There is no major difference in the scores of males and females. The lack of a digital divide among genders is consistent with our `svy LLasso` results. However, in contrast to gender, we observe a digital literacy gap between *Aboriginal* and *non-Aboriginal* respondents. *Aboriginal* individuals, on average, score lower on the digital literacy scale. This result may reflect broader socioeconomic and geographical challenges faced by *Aboriginal* communities.

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

The lowest income category of individuals making $\$52,203$ *and lower* has the second 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. The debiased Lasso results also showed that employment, income, and education categories are relevant explanatory variables in almost every logit model specification.

Both groups distinguished with respect to immigration status and visible minority status show surprising results. The immigration status variable category *Landed immigrant* has a slightly higher digital literacy score than *Non-landed immigrant* (non-immigrant/non-recent immigrant). The variable *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 have slightly higher digital literacy scores than non-visible minorities and non-immigrants.

The digital literacy scores for provinces vary. The Maritime provinces—Newfoundland and Labrador (NL), Prince Edward Island (PEI), Nova Scotia (NS), and New Brunswick (NB)—along with Manitoba (MB) and Saskatchewan (SK), score the lowest. In contrast, British Columbia (BC), Ontario (ON), and Alberta (AB) have the highest digital literacy scores, with these provinces showing almost identical results. The MCA results from each plot consistently group the Maritime provinces with the rural category, which explains their lower scores on the digital literacy scale.

Individuals who have used a virtual wallet score the highest on our digital literacy score, with an average score of 8.31. Canadians currently using virtual wallets have very high digital literacy, much higher than the average Canadian.

5 Additional analyses and robustness checks

This section presents additional analyses and robustness checks for the digital divide. In Section 5.1, we investigate the effects of COVID-19 on the digital divide and also consider the influence of provincial safety (stringency) measures on technology adoption during the period covered in CIUS. In Section 5.2, we compare internet use and its determinants from CIUS 2010 to CIUS 2020 to study the evolution of the digital divide over the past decade.

5.1 Impact of COVID-19 on the digital divide

The onset of the COVID-19 pandemic has reshaped financial behaviors, notably transitioning from traditional to digital transaction and communication methods. We observed a decline in cash usage at the pandemic's onset, followed by an increased adoption of digital payments (Chen, Engert, Huynh and O'Habib, 2021; Chen, Engert, Huynh, O'Habib and Zhu, 2021). This shift is evident in the rise of mobile payment usage from 11% in November 2020 to 17% by April 2021. The analysis is based on questions posed to CIUS respondents regarding changes in their usage of various digital technologies during the COVID-19 pandemic.

To explore behavioral changes in the use of digital technology, we employ k-means clustering on CIUS survey questions related to online activities during the pandemic. Using the elbow method and silhouette scores, we determine the optimal number of clusters. Profiling these clusters and comparing them with demographics, we derive centroids for each cluster to better understand their characteristics. The k-means cluster analysis identifies two distinct clusters, which we label

Table 9: Digital Literacy Score

Variables	Categories	Digital Literacy Score
Location	Urban	7.18
	Rural	6.68
Age	15–24	7.12
	25–34	7.71
	35–44	7.61
	45–54	7.18
	55–64	6.66
	65 and older	6.06
Gender	Male	7.12
	Female	7.10
Aboriginal identity	Non-Aboriginal	7.12
	Aboriginal	6.86
Employment status	Employed	7.38
	Not employed	6.61
Education	High school or less	6.48
	Some post-secondary	6.99
	University degree	7.71
Visible minority status	Visible minority	7.33
	Not a visible minority	7.03
Household type	Family with children under 18	7.41
	Single	6.72
	Family w/o children under 18	6.98
	Other household type	7.15
Income	$52, 203 and lower	6.46
	$52,204–$92,485	6.84
	$92,486–$146,559	7.23
	$146, 560 and higher	7.82
Immigration status	Landed immigrant	7.28
	Non-landed immigrant	7.07
Province	NL	6.84
	PEI	6.88
	NS	6.81
	NB	6.85
	QC	7.03
	ON	7.15
	MB	6.99
	SK	6.94
	BC	7.20
	AB	7.21
Virtual wallet	Used virtual wallet	8.31
	No virtual wallet	6.94

Note: This table presents the average Digital Literacy Scores derived from 10 questions in CIUS 2020 data. Scores range from 0 to 10, with higher values indicating greater digital literacy. The table categorizes respondents based on various demographic and socioeconomic factors such as location, age, gender, Aboriginal identity, employment status, education, visible minority status, household type, income, immigration status, province, and virtual wallet use.

as *Digital Adopters* and *Digital Resisters*.

In the digital literacy score distributions for the identified clusters, we observe the *Digital Adopters* cohort has a median digital literacy score of 8.0, with an interquartile range from 7.0 to 9.0, indicating a consistent, higher proficiency in digital literacy within this group. Conversely, the *Digital Resisters* cohort displays a more dispersed distribution, with a median score of 6.0 and an interquartile range spanning from 4.0 to 7.0. The heterogeneity in this cluster indicates a broader spectrum of digital engagement behaviors; see Figure 3 for a graphical representation of these distributions.

Figure 5 depicts the demographic composition of the *Digital Adopters* cluster. The demographics with the highest representation in the *Digital Adopters* cluster include those who are employed, have a university education, are part of a family with children, and have an income of $162,800 or more. In contrast, the demographics with the lowest representation in the cluster are the unemployed, individuals with high school education, singles, and those with an income of $44,119 or less.

Figure 3: Digital Literacy Score By Cluster

Note: Figure 3 shows a boxplot of the *Digital Adopters* and *Digital Resisters* digital literacy scores. The black bar inside the boxplot shows the average score for each group.

We compare the percentages of individuals in each cluster engaging in online banking, email, online credit card usage, and virtual wallet usage. The *Digital Adopters* demonstrate a higher inclination to use these digital tools compared to the *Digital Resisters*. A grouped bar chart in Figure 4 visually represents this distinction, clearly emphasizing the differences in digital engagement between the clusters. *Digital Adopters*, on average, have a much higher digital literacy score.

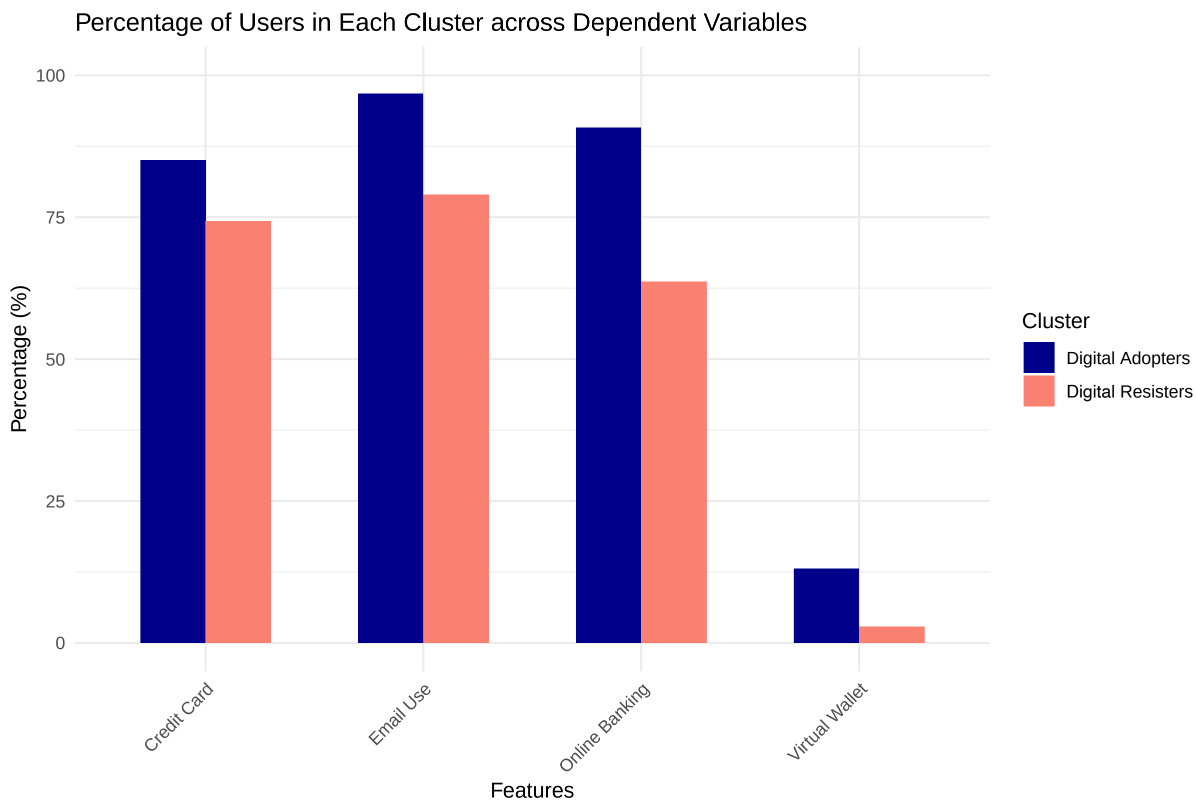


Figure 4: Percentage of Respondents in Each Cluster Using Credit Card Online, Email, Online Banking, and Virtual Wallets

Incorporation of the stringency dataset

To analyze the impact of governmental responses to COVID-19 on digital adoption, we use the stringency dataset developed by Cheung et al. (2021). This dataset, adapted from methodology developed by Oxford University's Blavatnik School of Government for the Oxford COVID-19 Government Response Tracker, measures the stringency of containment restrictions across Canadian provinces. We use the timeline from January 1, 2020, until the end of CIUS data collection on March 3, 2021. Figure 7 shows the breakdown of the average stringency for each province over the specified time frame.

We compare the distribution of *Digital Adopters* across provinces with either above or below-average stringency restrictions, utilizing the violin plot illustrated in Figure 8. Provinces with above-average stringency measures demonstrate a concentrated prevalence of *Digital Adopters*,

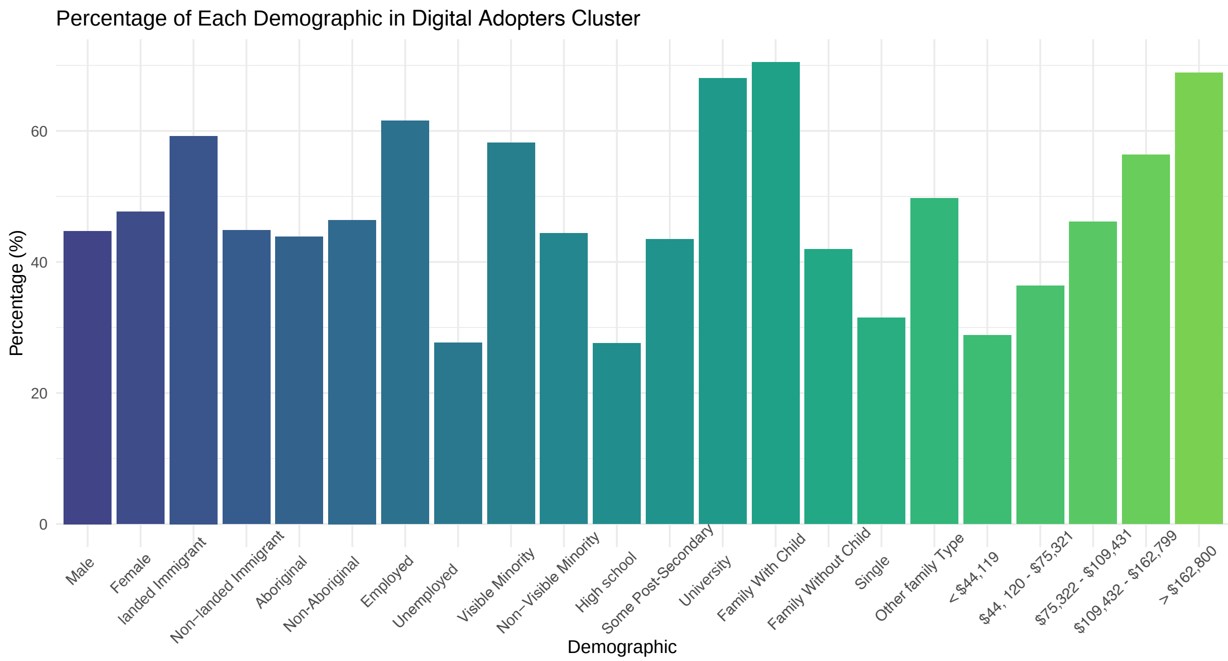


Figure 5: Demographics of Digital Adopters Cluster

Note: The bar graph in this figure shows the percentage of the *Digital Adopters* in each specified demographic group.

suggesting a potential correlation between increased stringency and higher digital adoption. Conversely, provinces with below-average stringency measures exhibit a broader and lower prevalence of Digital Adopters, indicating varied digital adoption rates under less stringent conditions.

To quantify this relationship, we calculate the (Pearson's) correlation between average stringency and the percentage of individuals in the *Digital Adopters* cluster. The test yields a correlation coefficient 0.327 with a p-value 0.357, suggesting that the correlation could be attributed to chance variation rather than a true underlying relationship.

5.2 Comparison with CIUS 2010

As the use of digital technologies in Canada increases, it is crucial to compare our findings with prior data and research. In this section, we trace the change in the digital divide by comparing internet use in CIUS 2010 and CIUS 2020 data. In particular, we shed light on whether the digital divide has grown or narrowed over time. This comparison allows us to identify persisting gaps as well as areas where progress has been made. Understanding these trends is vital for policymakers and stakeholders in crafting strategies and interventions to bridge the digital divide effectively.

Haight et al. (2014) conduct a study using the internet use variable from the CIUS 2010 data.[2] According to Haight et al. (2014), 80% of Canadians aged 16 and above were connected to the

[2] We limit our comparison with Haight to internet use, as not all variables used by Haight are available in the online CIUS 2010 dataset.

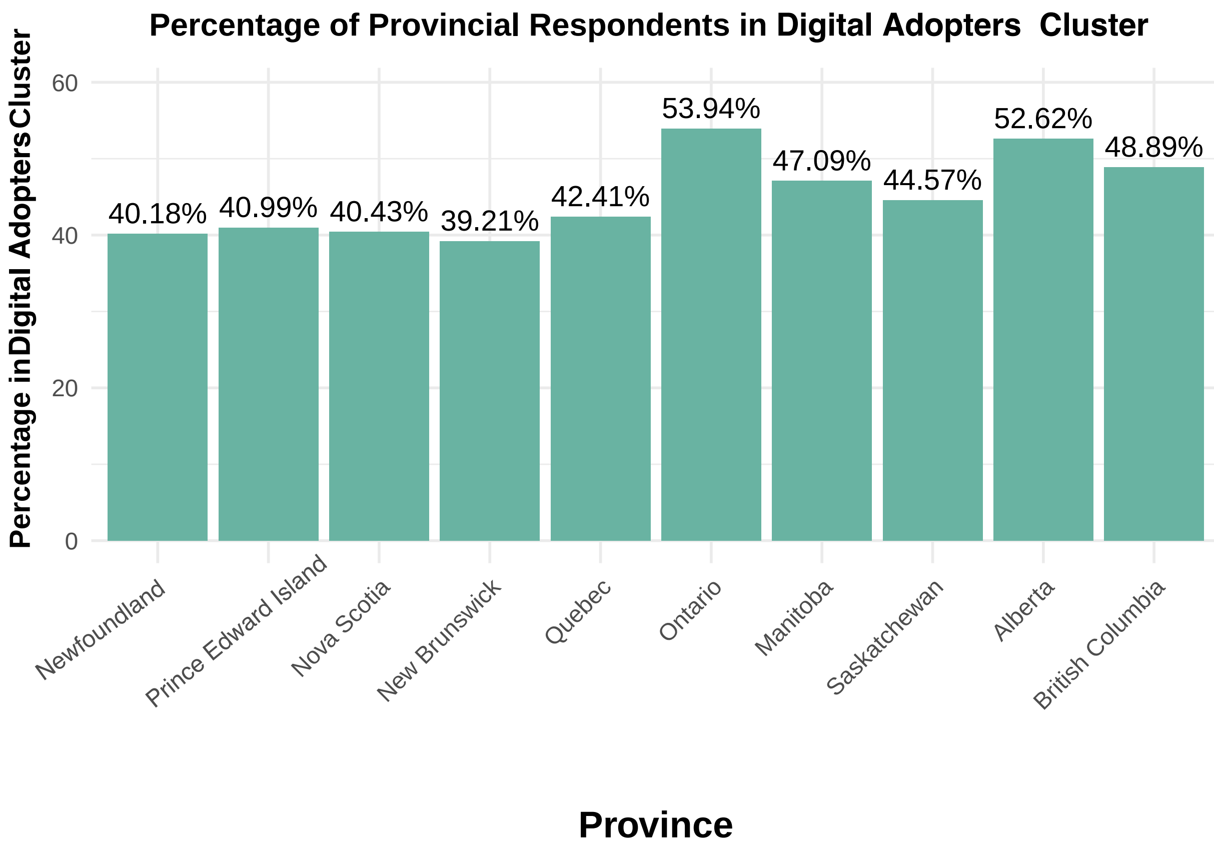


Figure 6: Percentage of Digital Adopters from each Province

Note: Figure 6 shows a bar graph of the percentage of each province in the *Digital Adopters* cluster.

internet in 2010, which represents a notable increase compared to previous years, with figures of 73% in 2007, 68% in 2005, 64% in 2003, and a mere 51% in 2000.

After accounting for the survey weights, the proportion of respondents who were connected to and use the internet (in the past three months) is 92.2% in the online CIUS 2020 data, a notable 11.9% increase over the 10-year span.[3]

To examine this pattern further, we estimate a survey-weighted logit model for the internet use dependent variable, closely mimicking the specification of Haight et al. (2014) using the online version of CIUS 2010. We then estimate the same model using comparable variables in CIUS 2020 data. Since the latter dataset does not include immigration status or high school graduation information used by Haight et al. (2014), we estimate the model without these variables. Similarly to Haight et al. (2014), we use the age variable without grouping it into different age cohorts. However, note that the quintiles of the income variable slightly differ between CIUS 2010 and

[3]The online CIUS 2020 and CIUS 2010 data are available at:
https://abacus.library.ubc.ca/dataset.xhtml?persistentId=hdl:11272.1/AB2/NUVBX2
https://abacus.library.ubc.ca/dataset.xhtml?persistentId=hdl:11272.1/AB2/YUIPZ7

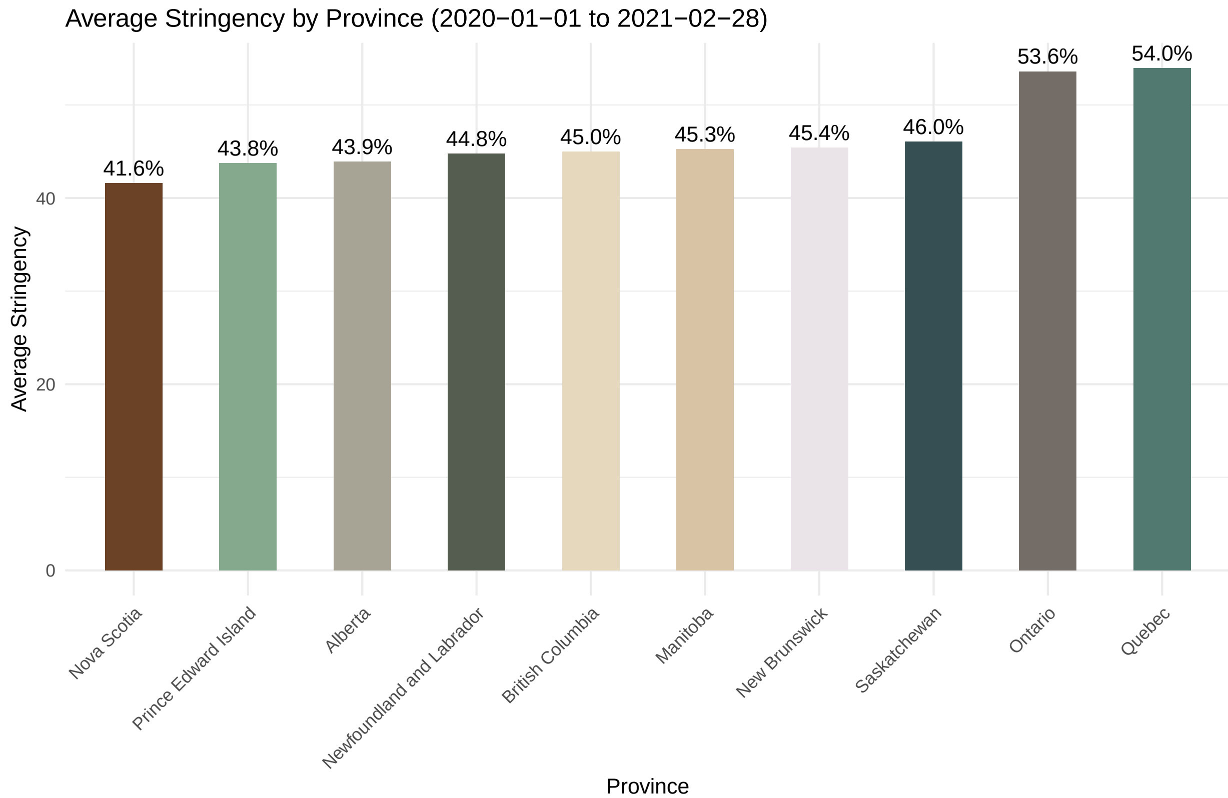


Figure 7: COVID-19 Stringency Index by Province

Note: The bar graph in this figure displays the average COVID-19 stringency restrictions for each province from January 1, 2020, to March 3, 2021. It is based on the stringency dataset developed by Cheung et al. (2021), which follows the methodology of Oxford University's Blavatnik School of Government, as used in the Oxford COVID-19 Government Response Tracker.

CIUS 2020, and the internet use variable in CIUS 2010 is an indicator of whether respondents have used the internet in the past 12 months, as opposed to the 3 months in CIUS 2020.

Table 10 reports the survey-logit estimation results. The coefficient estimates for the internet use variable in CIUS 2010 are qualitatively similar to those reported in Haight et al. (2014). The AME estimates suggest that a digital divide in Canada has narrowed across several crucial demographic dimensions. The effect of income across the quintiles 2–5 (Income 1–5 denote the dummy variables for the income brackets) relative to the income quintile 1 has decreased. The gap between individuals with a high school degree or less, or a university degree relative to those with some post-secondary education appears to have narrowed. Also, the effect of whether a respondent is currently a student or not (Student is the corresponding dummy) is absent in CIUS 2020, in contrast with the highly significant AME estimate of 0.127 in CIUS 2010 data.

Regarding rural vs. urban dummies, the coefficient estimates point toward a persistent gap. Interestingly, the negative effect of age on internet use seems to have decreased but still persists, and gender has no effect on internet use in both datasets.

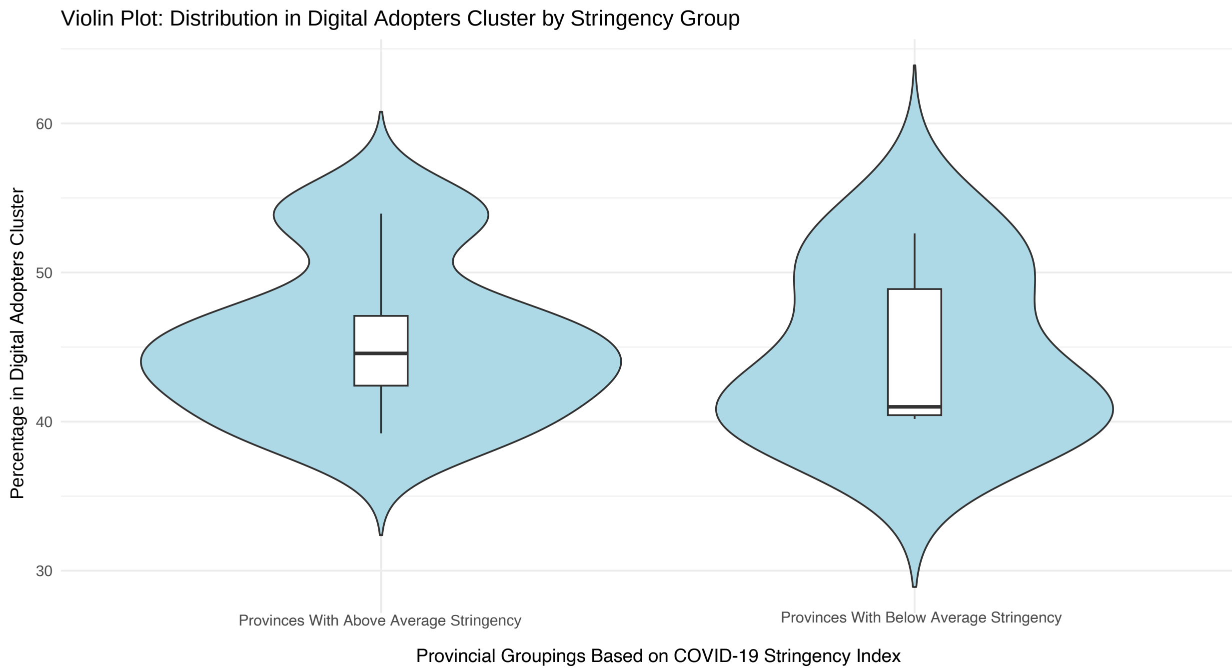


Figure 8: COVID-19 Stringency Index and Percentage of Observations in the Digital Adopters Cluster

Note: This figure shows a violin plot of two different distributions. The provinces have been divided into two groups: one with above-average provincial COVID-19 stringency restrictions and the other with below-average. The plot on the left shows the distribution of *Digital Adopters* in the above-average group, and the one on the right shows the distribution in the below-average group.

To explain the 11.9% internet use differential between CIUS 2010 and CIUS 2020 data, we perform the Oaxaca-Blinder decomposition, which decomposes the gap in the internet use rates in CIUS 2010 and CIUS 2020 data into portions due to differences in coefficients and characteristics (regressors).

Table 11 and Figure 9 display the twofold Oaxaca-Blinder decomposition, where CIUS 2020 model coefficients are used as the reference coefficients. The decomposition results suggest that we can attribute approximately -0.4% of the 11.9% difference to group differences in characteristics (i.e., age, education, gender), and the remaining 12.3% to differences in coefficients. This result is expected since the regressors in both models are comparable, and the coefficients exhibit some variations, as reported in Table 10.

The variable-by-variable twofold decomposition reveals that two key factors, namely the percentage of individuals in the second income bracket and the survey respondents who are currently enrolled as students, mainly drive the variation in internet usage. Not surprisingly, the change in the intercept exceeds zero by a large margin and is highly significant, as many dummy variables, e.g., employment status, family type, province, and age cohort, are omitted in the survey-logit specifications.

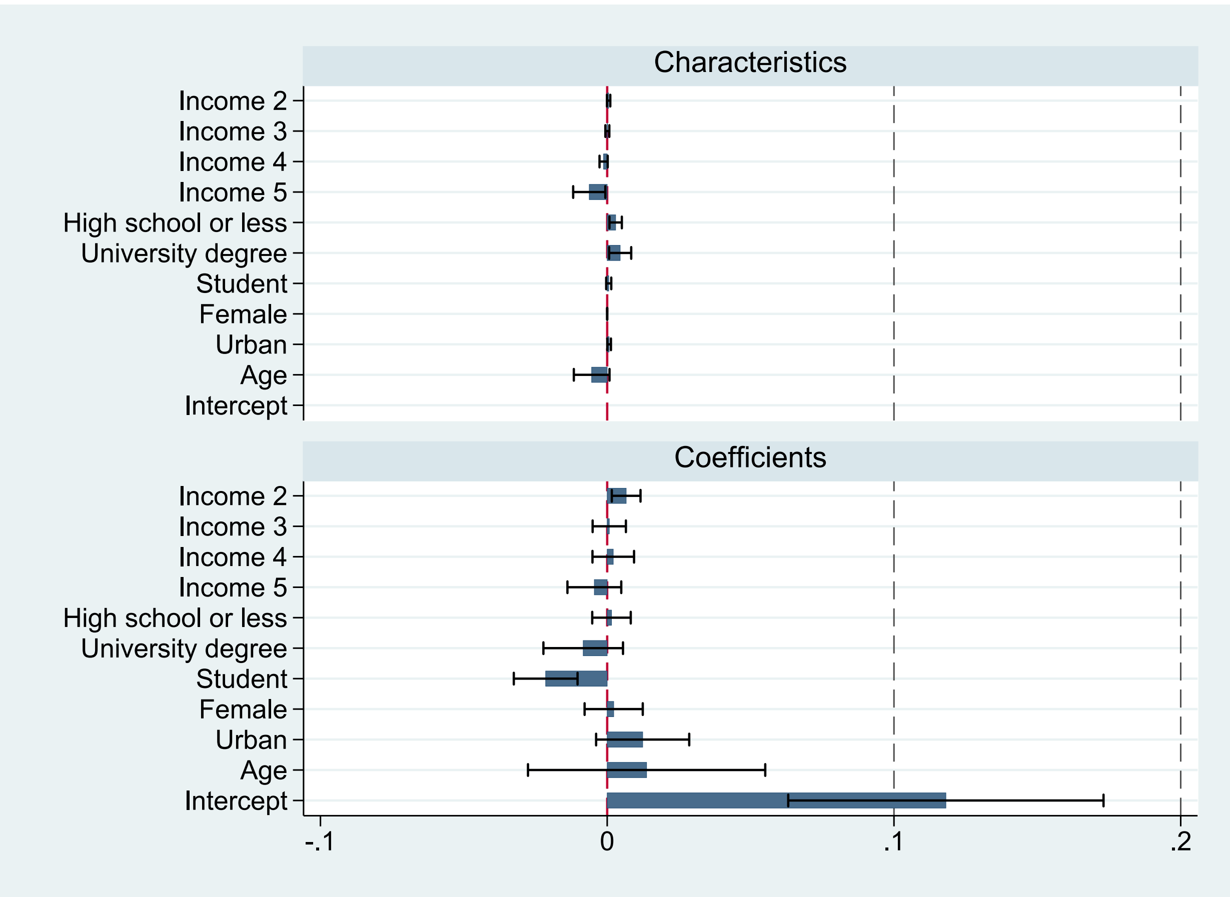
Table 10: Survey-Logit Estimates for Internet Use Dependent Variable

		Coefficient estimates	
Variables	Categories	CIUS 2010	CIUS 2020
Intercept		3.620***(0.156)	4.764***(0.277)
Location	Rural (omitted)	–	–
	Urban	0.188***(0.059)	0.330***(0.075)
Age	Age	−0.695***(0.030)	−0.659***(0.045)
Gender	Female	0.035 (0.066)	0.078 (0.074)
	Male (omitted)	–	–
Education	High school or less	−1.040***(0.065)	−0.996***(0.077)
	Some post-secondary (omitted)	–	–
	University degree	0.770***(0.144)	0.509***(0.156)
	Student	1.222***(0.261)	−0.362 (0.285)
Income	Income 1 (omitted)	–	–
	Income 2	0.397***(0.084)	0.717***(0.086)
	Income 3	0.904***(0.090)	0.940***(0.112)
	Income 4	1.207***(0.106)	1.309***(0.145)
	Income 5	1.888***(0.136)	1.671***(0.184)
	Observations	22,623	17,409

		AME estimates	
Variables	Categories	CIUS 2010	CIUS 2020
Intercept		–	–
Location	Rural (omitted)	–	–
	Urban	0.020** (0.006)	0.019***(0.004)
Age	Age	−0.072***(0.002)	−0.039***(0.002)
Gender	Female	0.004 (0.007)	0.005 (0.004)
	Male (omitted)	–	–
Education	High school or less	0.197***(0.014)	0.098***(0.011)
	Some post-secondary (omitted)	–	–
	Student	0.127***(0.028)	−0.021 (0.017)
	University degree	0.080***(0.014)	0.030***(0.009)
Income	Income 1 (omitted)	–	–
	Income 2	−0.108***(0.007)	−0.059***(0.005)
	Income 3	0.041***(0.009)	0.042***(0.005)
	Income 4	0.094***(0.009)	0.055***(0.007)
	Income 5	0.126***(0.010)	0.077***(0.009)
	Observations	22,623	17,409

Note: This table reports the survey-weighted logit estimates for comparable variables between CIUS 2010 and CIUS 2020. The CIUS 2020 model coefficients are used as the reference coefficients. The top panel reports the coefficient estimates while the bottom panel reports the corresponding AME estimates. The standard errors are in parantheses. Income 1–5 are the dummy variables for the income quintiles. Significance levels are marked as: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Figure 9: Regressor-by-Regressor Variation in the Oaxaca-Blinder Decomposition



Note: This figure plots the regressor-by-regressor variations in the Oaxaca-Blinder decomposition of the difference in the estimated conditional probabilities of internet use in CIUS 2010 and CIUS 2020 data based on survey-weighted logit model. The CIUS 2020 model coefficients are used as the reference coefficients. Income 2–5 are the dummy variables for the income quintiles.

Table 11: Twofold Oaxaca-Blinder Decomposition of Internet Use Difference in CIUS 2010 and CIUS 2020

Variables	Categories	Survey-logit	p-value
	CIUS 2020	0.922^{***}	0.000
	CIUS 2010	0.803^{***}	0.000
Decomposition	Difference	0.119^{***}	0.000
	Characteristics	−0.004	0.270
	Coefficients	0.123^{***}	0.000
	Characteristics		
Location	Rural (omitted)	–	–
	Urban	0.001^{*}	0.040
Age	Age	−0.005	0.088
Gender	Male (omitted)	–	–
	Female	0.000	0.778
Education	High school or less	0.003^{**}	0.008
	Student	0.001	0.235
	Some post-secondary (omitted)	–	–
	University degree	0.005^{*}	0.021
Income	Income 1 (omitted)	–	–
	Income 2	0.001	0.089
	Income 3	0.000	0.883
	Income 4	−0.001	0.085
	Income 5	-0.006^{*}	0.029
	Coefficients		
Location	Rural (omitted)	–	–
	Urban	0.012	0.135
Age	Age	0.014	0.515
Gender	Male (omitted)	–	–
	Female	0.002	0.661
Education	High school or less	0.002	0.662
	Student	-0.022^{***}	0.000
	Some post-secondary (omitted)	–	–
	University degree	−0.008	0.238
Income	Income 1 (omitted)	–	–
	Income 2	0.007^{**}	0.010
	Income 3	0.001	0.807
	Income 4	0.002	0.569
	Income 5	−0.004	0.349
	Intercept	0.118^{***}	0.000
	Observations	40,032	

Note: This table reports the regressor-by-regressor variations in the Oaxaca-Blinder decomposition of the difference in the internet use rates in CIUS 2010 and CIUS 2020 data based on survey-weighted logit model. Income 2–5 are the dummy variables for the income quintiles. Significance levels are marked as: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

6 Conclusion

This paper examines the CIUS 2020 data using the `svy LLasso` estimator of logit models with internet use, online banking, email, virtual wallet, and online credit card usage as the dependent variables. We find large disparities in digital literacy across different demographic groups within Canada. Younger Canadians, particularly those aged 15 to 24, are choosing to use emerging financial technologies like virtual wallets, moving away from traditional methods such as online banking and credit cards. Individuals with high incomes, high educational attainment, and stable employment are the most engaged with digital technologies. They consistently utilize a broader array of digital services and are more likely to adopt new innovations. Our findings on women, visible minorities, and immigrants were contrary to previous findings. Women, especially those in the workforce, are more likely to use email and exhibit higher digital literacy. Visible minorities and recent immigrants are also showing strong digital engagement, with new immigrants often matching or even exceeding Canadian-born citizens.

Regions such as Manitoba and the Maritime provinces face significant digital barriers. These areas are home to populations with lower digital literacy and engagement, particularly among seniors and lower-income individuals. Prioritizing these provinces in national strategies is crucial to ensuring the benefits of a digital economy are shared equitably across all Canadians.

Our comparison with CIUS 2010 data shows the evolution of the digital divide over the past decade, with increased overall digital connectivity but a persistent urban rural divide. Older Canadians, especially seniors, exhibit lower engagement with digital technologies, reflecting a digital divide along age lines. This divide is most acute among individuals who are older and single or older with low income, emphasizing the intersectionality of age and income in digital exclusion.

The COVID-19 pandemic highlighted the essential role of digital access and literacy. While we observed a positive correlation between stricter provincial public health measures and digital adoption, it was not statistically significant suggesting these public health policies did not alter the already established patterns of the digital divide. Even though the use of digital technologies increased during the pandemic, pre-existing disparities continued to shape digital engagement, despite the varying strictness of restrictions across provinces.

Given the current state of the digital divide, the potential introduction of a CBDC could disproportionately disadvantage individuals from lower socioeconomic classes who may struggle to adapt to new digital monetary systems. There is a need for targeted investments in digital literacy and infrastructure—not only in rural areas but also in lower-income urban communities. Providing internet access alone is insufficient; comprehensive strategies that include education and support are essential to equip all Canadians to participate fully in an increasingly digital economy. As Canada moves toward a potential cashless society, these efforts are crucial to prevent the widening of the digital divide.

Limitations and areas for future research. The CIUS 2020 data used in our analysis covers only the ten Canadian provinces, excluding the territories and Aboriginal reserves. This omission may underrepresent the true extent of the digital divide, particularly in remote and rural areas such as on reserves where unique socioeconomic and geographical challenges affect digital connectivity. Additionally, although the proportion of unbanked individuals in Canada is exceedingly small—nearly 99% of respondents in the 2017 Methods-of-Payment Survey reported having a bank account—the unbanked population predominantly resides in the Northern territories and reserves, areas not captured in our data. While CIUS includes participants who self-identify as Aboriginal, it does not gather information directly from First Nations reserves. This could skew perceptions of digital inclusivity among Indigenous populations. Comparisons with the 2017 Aboriginal People's Survey show similar internet usage rates among Indigenous respondents, but it is essential to approach these findings with an understanding of the unique challenges faced by these communities. Future studies can explore the territories and on-reserve Aboriginal communities using the Norther Canadian Internet Use Survey (NCIUS).

References

Adrian, T. and Mancini-Griffoli, T. (2019), 'The Rise of Digital Money, International Monetary Fund', *Annual Review of Financial Economics* **13**, 57–77.

American Library Association (n.d.), 'Digital Literacy'.
URL: *https://literacy.ala.org/digital-literacy/*

Aston, J., Vipond, O., Virgin, K. and Youssouf, O. (2020), Retail E-Commerce and COVID-19: How Online Shopping Opened Doors While Many were Closing, Technical report, Statistics Canada.

Aversa, J., Shaker, R., Cleave, E. and Salooja, N. (2022), 'Determinants of Bank Closures: Exploring the Relationship between Neighborhood Characteristics and Bank Branch Locations', *Papers in Applied Geography* **8**(4), 393–413.

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.

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.

Carson, S. A. (2013), 'The Remote Rural Broadband Deficit in Canada', *Journal of Rural and Community Development* **8**(2).

Chen, H., Engert, W., Huynh, K. P. and O'Habib, D. (2021), An Exploration of First Nations Reserves and Access to Cash, Technical report, Bank of Canada Staff Discussion Paper.

Chen, H., Engert, W., Huynh, K. P. and O'Habib, D. (2022), Identifying Financially Remote First Nations Reserves, Technical report, Bank of Canada Staff Discussion Paper.

Chen, H., Engert, W., Huynh, K. P., O'Habib, D., Wu, J. and Zhu, J. (2022), Cash and COVID-19: What Happened in 2021, Technical report, Bank of Canada.

Chen, H., Engert, W., Huynh, K. P., O'Habib, D. and Zhu, J. (2021), Cash and COVID-19: The Impact of the Second Wave in Canada, Technical report, Bank of Canada Staff Discussion Paper.

Cheung, C., Lyons, J., Madsen, B., Miller, S. and Sheikh, S. (2021), The Bank of Canada COVID-19 Stringency Index: Measuring Policy Response Across Provinces, Technical report, Bank of Canada Staff Analytical Notes.

Cullen, R. (2001), 'Addressing the Digital Divide', *Online Information Review* **25**(5), 311–320.

Deng, Z., Morissette, R. and Messacar, D. (2020), Running the Economy Remotely: Potential for Working From Home During and After COVID-19, Technical report, Statistics Canada.

Dewan, S. and Riggins, F. J. (2005), 'The Digital Divide: Current and Future Research Directions', *Journal of the Association for Information Systems* **6**(12), 298–337.

Engert, W. and Huynh, K. P. (2022), Cash, COVID-19 and the Prospect for a Canadian Digital Dollar, Technical report, Bank of Canada Staff Discussion Paper.

Ferreira, D., Vale, M., Carmo, R. M., Encalada-Abarca, L. and Marcolin, C. (2021), 'The Three Levels of the Urban Digital Divide: Bridging Issues of Coverage, Usage and Its Outcomes in VGI Platforms', *Geoforum* **124**, 195–206.

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.

Henry, C. S., Engert, W., Sutton-Lalani, A., Hernandez, S., McVanel, D. and Huynh, K. P. (2023), 'Unmet Payment Needs and a Central Bank Digital Currency', *Journal of Digital Banking* **8**(3), 242–255.

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', *ArXiv preprint https://arxiv.org/abs/2304.07855* .

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.

Kamdjoug, J. R. K., Wamba-Taguimdje, S.-L., Wamba, S. F. and Kake, I. B. (2021), 'Determining Factors and Impacts of the Intention to Adopt Mobile Banking App in Cameroon: Case of SARA by Afriland First Bank', *Journal of Retailing and Consumer Services* **61**, 102509.

Koch, K. (2022), 'The Territorial and Socio-Economic Characteristics of the Digital Divide in Canada', *Canadian Journal of Regional Science* **45**(2), 89–98.

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.

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

Reddick, C. G., Enriquez, R., Harris, R. J. and Sharma, B. (2020), 'Determinants of Broadband Access and Affordability: An Analysis of a Community Survey on the Digital Divide', *Cities* **106**, 102904.

Robinson, L., Cotten, S. R., Ono, H., Quan-Haase, A., Mesch, G., Chen, W., Schulz, J., Hale, T. M. and Stern, M. J. (2015), 'Digital Inequalities and Why They Matter', *Information, Communication & Society* **18**(5), 569–582.

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

Van Buuren, S. (2018), *Flexible Imputation of Missing Data, Second Edition*, Chapman and Hall.

Van Deursen, A. J. A. M. and Van Dijk, J. A. G. M. (2019), 'The First-Level Digital Divide Shifts from Inequalities in Physical Access to Inequalities in Material Access', *New Media & Society* **21**(2), 354–375.

Van Dijk, J. and Hacker, K. (2003), 'The Digital Divide as a Complex and Dynamic Phenomenon', *The Information Society* **19**(4), 315–326.

Wavrock, D., Schellenberg, G. and Schimmele, C. (2022), Canadians' Use of the Internet and Digital Technologies Before and During the COVID-19 Pandemic, Technical report, Statistics Canada.

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.

Zickuhr, K. and Smith, A. (2012), 'Digital Differences', Pew Research Center.
URL: *https://www.pewresearch.org/internet/2012/04/13/digital-differences/*