Digital Adoption and Cyber Security: An Analysis of Canadian Businesses[*]

Joann Jasiak[†] Peter MacKenzie[‡] Purevdorj Tuvaandorj[§]

April 14, 2025

Abstract

This paper examines how Canadian firms balance the benefits of technology adoption against the rising risk of cyber security breaches. We merge data from the 2021 Canadian Survey of Digital Technology and Internet Use and the 2021 Canadian Survey of Cyber Security and Cybercrime to investigate the trade-off firms face when adopting digital technologies to enhance productivity and efficiency, balanced against the potential increase in cyber security risk. The analysis explores the extent of digital technology adoption, differences across industries, the subsequent impacts on efficiency, and associated cyber security vulnerabilities. We build aggregate variables, such as the Business Digital Usage Score and a cyber security incidence variable to quantify each firm's digital engagement and cyber security risk. A survey-weight-adjusted Lasso estimator is employed, and a debiasing method for high-dimensional logit models is introduced to identify the drivers of technological efficiency and cyber risk. The analysis reveals a digital divide linked to firm size, industry, and workforce composition. While rapid expansion of tools such as cloud services or artificial intelligence can raise efficiency, it simultaneously heightens exposure to cyber threats, particularly among larger enterprises.

Keywords: Cyber Security, Digital Technology, Survey Data, Lasso, Technological Efficiency, Post-Selection Inference

[*] The authors thank participants at the 58th Annual Conference of the Canadian Economics Association, the 39th Annual Meeting of the Canadian Econometrics Study Group (CESG), and the Bank of Canada Workshop on Understanding and Correcting for Non-Response Bias for their helpful comments. The authors also gratefully acknowledge access to data provided by the RDC and Statistics Canada (Project 21-MAPA YRK-721).

[†] York University, jasiakj@yorku.ca

[‡] York University, petem9@yorku.ca

[§] York University, tpujee@yorku.ca

1 Introduction

The digital transformation of business operations is reshaping the Canadian economy, offering opportunities for increased productivity and innovation. For businesses, adopting digital technologies is no longer optional; it is essential for maintaining competitiveness in a globalized market. However, with greater reliance on digital tools comes an increased exposure to cyber security risks. Understanding this *trade-off*—the efficiency gains from digital adoption versus the vulnerabilities it introduces is critical for Canadian businesses.

This study examines the extent of digital technology adoption and cyber security practices among Canadian firms, using a novel dataset that combines information from the 2021 Canadian Survey of Digital Technology and Internet Use (SDTIU) and the 2021 Canadian Survey of Cyber Security and Cybercrime (CSCSC). To our knowledge, this is the first study to analyze the most recent version of the CSCSC dataset, offering an opportunity to explore both digital adoption and cyber security. These surveys provide detailed data on firms' adoption of advanced technologies such as cloud computing, artificial intelligence, and enterprise management systems, as well as the frequency, severity, and impact of cyber security incidents on business operations.

The primary objective of this paper is to explore the trade-off between digital adoption and cyber security risk, by examining how technological advancement can enhance productivity while simultaneously increasing firms' vulnerability to cyber threats. Specifically, we analyze which types of firms are adopting digital technologies, how adoption levels vary across industries, and whether these technologies are increasing cyber security incidents. We also investigate the firm-specific factors such as size, industry, and employee characteristics that shape both technological efficiency and the ability to manage cyber security risks. Our methodological contributions include introducing the Business Digital Usage Score (BDUS) to quantify digital adoption, enabling a comparative analysis of firms' engagement with digital technologies.

Given the large number of firms and the high dimensionality of mostly qualitative explanatory variables in our dataset, we introduce high-dimensional logit models estimated via a Lasso-penalized maximum likelihood estimator with survey weights to ensure representativeness of Canadian firms. We use a debiasing method for inference on the selected model's coefficients and establish its asymptotic validity. To assess how closely firms operate to their technological usage frontier, we apply stochastic frontier analysis. Additionally, we employ k-means clustering to categorize firms by technological efficiency, facilitating the identification of distinct profiles of digital adoption and efficiency.

The literature on digital technology adoption emphasizes its role in improving firm productivity. Larger firms are often better positioned to implement these technologies, benefiting from economies of scale and greater access to resources (Leung et al., 2008). In contrast, smaller firms frequently encounter barriers, including high costs of implementation and limited technical expertise, which can hinder their ability to fully realize the potential benefits of digital adoption. Ferrari (2012) demonstrates that industry-specific differences in digital readiness significantly affect adoption rates, while Aghimien et al. (2021) highlight the role of regional policies in either enabling or constricting firms' technological advancement. Together, these studies emphasize the structural factors shaping disparities in digital technology adoption across firms.

In Canada, Bilodeau et al. (2019), using data from the 2017 CSCSC, show the widespread reliance of Canadian businesses on digital technologies. They report that about 92% of Canadian businesses used one or more digital technologies or services in 2017, with significant increases in the adoption of websites and social media integration since 2013. In addition, just over one-fifth of Canadian businesses reported being impacted by cyber security incidents that affected their operations, with 54% noting that these incidents prevented employees from carrying out day-to-day work and about 30% experiencing additional repair or recovery costs.

While digital adoption can boost productivity, it also introduces new risks, particularly related to cyber security. The Geneva Association, a leading international think tank of the insurance industry, defines cyber risks as breaches in confidentiality, availability, and data integrity, posing operational threats to firms that increasingly rely on interconnected digital systems (Eling et al., 2016). Cebula and Young (2010) expand on this definition, framing cyber risks as disruptions that extend beyond information technology (IT) systems to affect broader business stability. In Canada, the Toronto Public Library system experienced significant disruptions following a cyberattack, while the Nova Scotia Health Department faced operational delays and data breaches during a similar event (Bridge and Zoledziowski, 2024; Bousquet, 2023).

The adoption of advanced technologies such as Internet of Things (IoT) and enterprise management systems could increase cyber security risks. Blichfeldt and Faullant (2021) highlight that while these systems can improve productivity, their complexity often introduces integration challenges that, if poorly managed, can undermine business operations. Additionally, the widespread use of interconnected devices has expanded the potential attack surface for cybercriminals, requiring firms to invest more heavily in cyber security infrastructure.

Remote work adoption during the COVID-19 pandemic accelerated the reliance on digital tools but also introduced new risks and challenges. Hackney et al. (2022) find that firms effectively utilizing digital technologies during the pandemic demonstrated resilience in maintaining operations under lockdown conditions. However, Aczel et al. (2021) and Kitagawa et al. (2021) report that the rapid shift to remote work led to heightened employee burnout and increased risks of phishing attacks, underscoring the broader implications of accelerated digital adoption.

Cyber insurance is a tool for mitigating cyber security risks. According to the OECD (OECD, 2017), the cyber insurance market doubled in size between 2015 and 2020, fueled by firms' increasing awareness of cyber threats. However, Fitch Ratings (Fitch Ratings, 2021) highlights that high premiums and restrictive coverage policies hinder adoption, especially among smaller firms. Globally, cyber security spending is projected to surpass $170 billion by 2026 (Gartner, Inc., 2021).

Despite the growing demand for cyber insurance, firms may strategically underinvest in cyber security measures due to the unobservable nature of such investments. Ahnert et al. (2022) argue that firms may prioritize visible innovations over less transparent risk mitigation strategies, as clients are often unable to directly evaluate cyber security expenditures. This dynamic creates a paradox in which firms recognize the increasing risks of digital adoption but fail to allocate sufficient resources to mitigate them effectively.

The paper is organized as follows. Section 2 describes the datasets used in the study, the SDTIU and the CSCSC, as well as the construction of the associated scores and variables. Section 3 outlines the paper's main contribution and methodological framework: the survey-weighted debiased logit Lasso method. Section 4 presents the empirical results, discussing the key determinants of technological efficiency and cyber security vulnerabilities. We conclude in Section 5 with a discussion of the implications of our findings for Canadian businesses and policymakers. The appendix is divided into two parts: Appendix A provides additional technical details on the survey-weighted debiased logit Lasso model, and Appendix B outlines the questions that comprise the BDUS and the Cyber Security Incidence indicator.

2 Data Description and Variable Construction

The empirical analysis is based on a merged dataset from two Statistics Canada surveys: the 2021 SDTIU and the 2021 CSCSC. These surveys were merged using firm size and industry classifications from the North American Industry Classification System (NAICS), enabling an integrated study of digital adoption and cyber security risks among Canadian businesses.[1]

The SDTIU focuses on digital technology adoption, including metrics such as internet usage, e-commerce participation, and ICT adoption, while the CSCSC examines cyber security practices and the impact of cyber incidents on businesses. The merged dataset includes both qualitative variables, such as the implementation of specific technologies and cyber security measures, and quantitative variables, such as cyber security expenditure and incident-related costs. Together, the surveys cover numerous variables relevant to digital adoption and cyber security. The firms vary in size, ranging from small enterprises with fewer than 10 employees to large corporations with over 500 employees. The industries covered include manufacturing, retail, professional services, and information technology.

Both surveys apply stratified sampling by industry and firm size, with survey weights correcting for selection probabilities, non-responses, and sampling biases. These weights ensure that the results are representative of the broader population of Canadian enterprises. The SDTIU achieved a response rate of 73%, while the CSCSC had a response rate of 65%. The pre-matched weighted samples cover 327,567 enterprises for the SDTIU and 185,644 for the CSCSC. After merging, the final dataset comprises a weighted sample of 179,657 enterprises, ensuring comprehensive coverage across various business demographics.

Below, we introduce aggregate measures of digital technology usage: BDUS in Section 2.1, a separate measure of Business Technological Efficiency (based on k-means clustering) in Section 2.2, which examines whether firms encounter significant challenges in implementing the technologies they adopt, and an indicator for Cyber Security Incidence in Section 2.3.

2.1 Business Digital Usage Score

The BDUS is a quantitative variable constructed to evaluate the digital engagement of Canadian firms by measuring the technologies they have adopted. It condenses responses from the 2021 SDTIU into a single, interpretable score that reflects the cumulative utilization of digital tools across ten distinct domains, such as cloud computing, digital payment systems, artificial intelligence (AI), smart devices, online sales, government digital connectivity, fiber-optic internet, and company websites (see Appendix B for the exact questions).

[1] Methodological differences between the surveys include variations in enterprise size definitions (for small firms: SDTIU defines small firms as those with 5–49 employees, while CSCSC defines them as 10–49 employees) and potential differences in primary respondents' understanding of their enterprise's operations. Additionally, the surveys differ in target populations, including NAICS and enterprise size requirements.

For each of the ten domains, we create a binary variable equal to 1 if the firm reports using that technology. Summing these indicators yields a score between 0 and 10, with higher values indicating more extensive digital engagement. Although it is conceivable that advanced technologies (e.g., AI) have disproportionate impacts on firm productivity compared to simpler ones (e.g., a basic website), weighting them by perceived importance would introduce additional assumptions. Firms also have heterogeneous technological needs, some rely heavily on cloud computing for scalable data storage, while others benefit more from online payment systems or data analytics. Therefore uniform summation provides a transparent, easily replicable index of a firm's overall digital footprint.

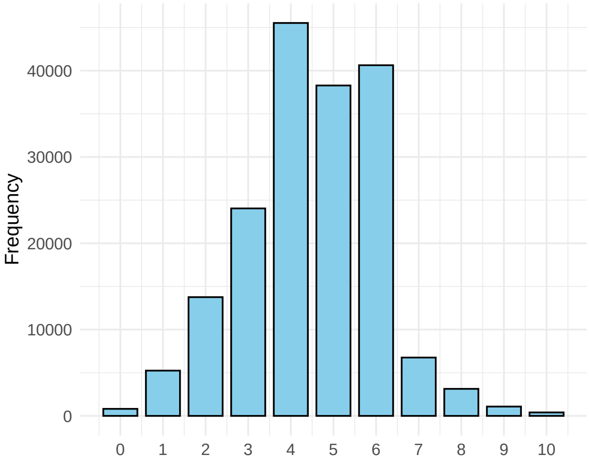

Figure 1: Histogram of Business Digital Usage Scores

Figure 1 shows the distribution of BDUS scores. The majority of firms fall within scores of 3 to 7, suggesting moderate overall adoption levels, with mild peaks at 4 and 6. At the extremes, a small proportion of firms report virtually no digital engagement (score = 0), whereas another small subset achieves near-comprehensive adoption (scores of 9 or 10). This dispersion highlights the heterogeneity of adoption patterns: some businesses implement only a handful of relevant technologies, while others adopt a wide array.

It is important to note that the BDUS does not measure the intensity of usage or the ease of integration. Rather, it provides a concise snapshot of whether certain well-known tools have been adopted in at least a basic form. We use the BDUS to investigate the characteristics of firms that adopt digital technologies through a stochastic frontier model in Section 4, measuring how close businesses are to their technological frontier. Additionally, we use the BDUS to assess whether having a more extensive digital profile correlates with cyber risk exposure (Table 1).

2.2 Business Technological Efficiency

While the BDUS captures which digital technologies firms adopt, it does not gauge how well these technologies are integrated. To address this, we construct a measure of Technological Efficiency by applying a k-means clustering algorithm to a set of SDTIU survey items about technological implementation challenges. Specifically, these survey items parallel the BDUS domains but ask whether the firm experiences difficulties using each technology, whereas the BDUS questions simply ask if firms use each technology. The exact questions used in this clustering exercise are listed in Appendix B.

Let $\{z_{i1}, z_{i2}, \ldots, z_{i10}\}$ be binary indicators for firm i, capturing whether it reports a challenge in each of the ten domains. A response of "Yes" indicates an operational problem in using the technology associated with that domain. We apply k-means clustering to these ten binary variables and determine that $k = 2$ is the optimal cluster count, resulting in two groups: *Digitally Efficient* and *Not Digitally Efficient*. Figure 2 displays the percentage of reported challenges in each domain for these two clusters. Firms in the latter cluster (52,505 businesses) report significantly more issues than those in the former (127,152 businesses). For example, 70.71% of *Not Digitally Efficient* firms cite difficulties with AI, compared to 23.58% in the *Digitally Efficient* group. Similarly, 57.55% of *Not Digitally Efficient* firms encounter challenges with cloud computing, versus only 3.21% among *Digitally Efficient* firms.

The high incidence of challenges among *Not Digitally Efficient* firms does not imply that they fail to adopt these tools. Some businesses report both a high BDUS score (indicating widespread adoption) and frequent operational issues, suggesting partial or suboptimal implementation. Even in the *Digitally Efficient* cluster, a non-trivial share of firms faces difficulties in at least one domain.

This k-means classification yields a binary variable, Technological Efficiency, which we use as a dependent variable in one of the logit models in Section 4. Whereas the BDUS measures the extent of adoption, the Technological Efficiency grouping reflects the firm's ability to use the technology effectively. In this sense, the two measures are complementary: the BDUS indicates how many digital tools a firm adopts, while the logit model using the Technological Efficiency variable reveals whether the adopted technologies are being used efficiently.

2.3 Cyber Security Incidence

To examine the cyber security challenges faced by Canadian businesses, we construct a set of variables based on survey responses related to the occurrence of cyber security incidents. The

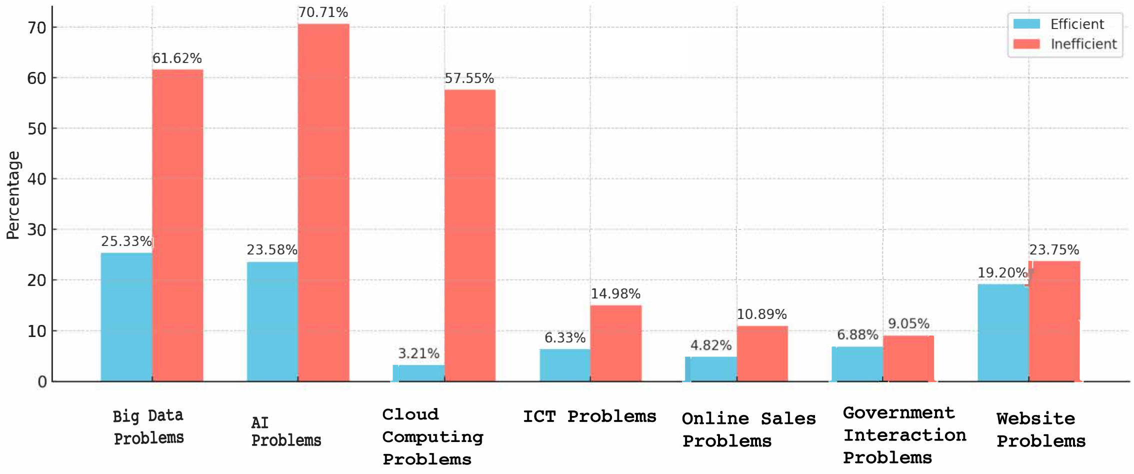

Figure 2: Percentage of Technological Problems Between Efficient and Inefficient Clusters

primary variable, *Cyber Security Incidence*, indicates whether a business was affected by any cyber security incident in 2021. This binary variable is coded as 1 if the business reported experiencing one or more types of cyber security incidents, and 0 otherwise. These incidents range from theft of assets, business data, or intellectual property to disruptions of business activities. Of the 179,656 surveyed firms, 39,524 (22%) reported at least one such incident, while the remaining 140,132 (78%) reported none.

Figure 3 shows the occurrence of these incident types among the 39,524 firms that experienced any cyber incident. The most frequently cited issues included incidents to steal money or demand ransom payment (37%), incidents with an unknown motive (39%), incidents to steal personal or financial information (33%), incidents to disrupt or deface the business or web presence (18%), and incidents to access unauthorised or privileged areas (20%). Fewer respondents reported incidents to steal or manipulate intellectual property or business data (9.8%) and incidents to monitor and track business activity (9.2%).

We use this binary *Cyber Security Incidence* variable in subsequent analyses for two primary reasons. It serves as a dependent variable in the cyber security incidence logit model (Section 4), where we identify which firm characteristics and digital adoption practices predict a higher likelihood of experiencing an incident. The *Cyber Security Incidence* variable is also

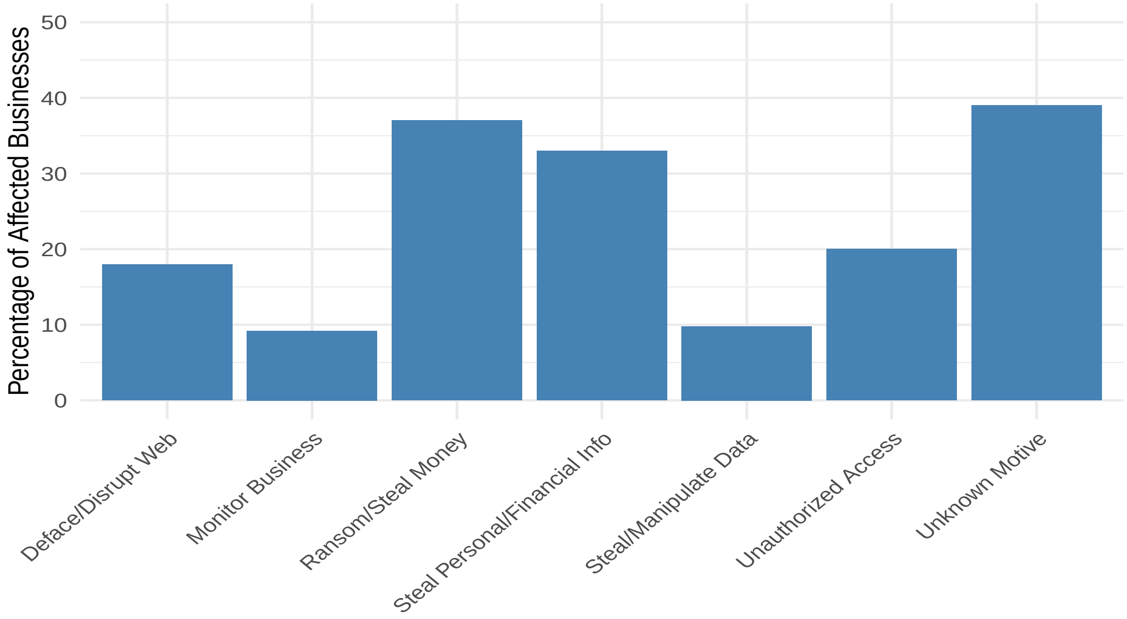

Figure 3: Types of Issues Reported by Businesses After Cyber Security Incident (Percentage of Affected Businesses)

included as a regressor in our stochastic frontier analysis to examine whether having experienced a breach correlates with digital technology usage. The rationale is that a prior cyber incident may influence firms' subsequent decisions or capabilities regarding digital adoption and security investments. A full list of cyber incidence types appears in Appendix B.

3 Estimation Methodology: Survey Weighted Debiased Lasso

To estimate models predicting technological efficiency and cyber security incidents, we employ a survey-weighted logistic Lasso (hereafter `svy LLasso`) estimation technique and introduce a debiasing method for inference. This approach is well-suited to our analysis, as it effectively handles high-dimensional data and incorporates a diverse set of variables, including digital technology usage, cyber security practices, and firm characteristics. Our primary Lasso models include over 50 independent variables, while second-order interaction models expand to more than 200 variables.

To accurately represent the population of Canadian businesses, we adapt the standard logistic Lasso method to incorporate Statistics Canada survey weights. To this end, we define the following general ℓ_1-penalized maximum likelihood estimator for a Generalized Linear Model with survey weights, where the `svy LLasso` estimator is a special case:

$$\hat{\theta} = \underset{\theta=(\alpha,\beta')'\in\mathbb{R}^{p+1}}{\operatorname{argmin}} \left(-L_n(\theta) + \lambda \sum_{j=1}^{p} |\beta_j| \right),$$

where:

- $\theta = (\alpha, \beta')'$ includes the intercept α and coefficients $\beta \in \mathbb{R}^p$,
- $L_n(\theta) = -n^{-1} \sum_{i=1}^{n} w_i g(y_i, x_i'\theta)$ is the survey-weighted log-likelihood, where $g(y, x'\theta)$ is the negative log-density function (see Appendix A.2 for a detailed description), x_i is the regressor vector for firm i, y_i is the outcome variable, and w_i is the survey weight,
- $\lambda \sum_{j=1}^{p} |\beta_j|$ is the penalty term, with tuning parameter λ, enforcing sparsity by shrinking less relevant coefficients to zero.

For our logit models, this approach selects the variables most predictive of technological efficiency and cyber security incidents while accounting for survey weights. This framework was adopted by Jasiak and Tuvaandorj (2023) and Jasiak et al. (2024), with the former developing an inference method different from the one we consider below.

Inference via Debiasing. For statistical inference on model parameters and average marginal effects (AME) (Appendix A.1), we adapt the debiased Lasso method from Zhang and Zhang (2014) and Javanmard and Montanari (2014) to the survey context described above. Given the ℓ_1-penalized maximum likelihood estimator $\hat{\theta}$, the debiasing (DB) method applies a one-step correction:

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

where $\hat{H}(\hat{\theta})$ and $S(\hat{\theta})$ are the negative Hessian and score function of the weighted log-likelihood function $L_n(\theta)$. The adjustment term $\hat{H}(\hat{\theta})^{-1} S(\hat{\theta})$ corrects the bias introduced by the ℓ_1-penalized variable selection, enabling reliable inference. This variant, which uses the standard Hessian, follows Xia et al. (2023), who consider standard GLMs, adapted here for survey weights. We estimate the asymptotic variance of $n^{1/2} S(\theta_0)$ using a sample information matrix $\hat{I}(\hat{\theta})$ (see Appendix A.2), where θ_0 represents the unknown true values of θ.

For a nonlinear parameter function $\rho(\theta)$ (an $r \times 1$ vector, possibly n-dependent), e.g., the AME, we define the debiased estimator:

$$\tilde{\rho} = \rho(\hat{\theta}) + \dot{\rho}(\hat{\theta})'\hat{H}(\hat{\theta})^{-1}S(\hat{\theta}), \quad \dot{\rho}(\theta) = \frac{\partial \rho(\theta)'}{\partial \theta}. \tag{3.1}$$

Asymptotic Validity. We establish the asymptotic validity of Wald-type inference in the following proposition. To keep the exposition simple, the underlying framework, definitions, and assumptions are provided in Appendix A.

Proposition 3.1 (Asymptotic Validity of Survey Debiasing Estimator). *Let Assumption 1 hold, and assume that*

- $\lambda = C\sqrt{\frac{\log p}{n}}$ *with* $C = O(1)$, $p \geq 1$,
- $p^2/n \to 0$, *and* $m_0 \log p\sqrt{\frac{p}{n}} \to 0$ *as* $n \to \infty$, *where* m_0 *is the number of non-zero coefficients of* θ_0,
- $\rho(\theta)$ *(with fixed* $r < p+1$*) is differentiable near* θ_0 *with a locally Lipschitz Jacobian* $\dot{\rho}(\theta)$, *and* $\lambda_{\min}(\dot{\rho}(\theta_0)'\dot{\rho}(\theta_0)) > \lambda_l > 0$, *where* $\lambda_{\min}$ *denotes the minimum eigenvalue.*

Then:

$$\left(\dot{\rho}(\hat{\theta})'\hat{H}(\hat{\theta})^{-1}\hat{I}(\hat{\theta})\hat{H}(\hat{\theta})^{-1}\dot{\rho}(\hat{\theta})\right)^{-1/2} n^{1/2}(\tilde{\rho} - \rho(\theta_0)) \xrightarrow{d} N(0, I_r).$$

The proof is provided in Appendix A.3. The order of the tuning parameter λ is standard in the literature (Bühlmann and van de Geer, 2011; Negahban et al., 2012; van de Geer et al., 2014; Hastie et al., 2015). The assumptions on the number of covariates p and model sparsity m_0 align with those in Xia et al. (2023). In particular, the condition $m_0 \log p\sqrt{p/n} \to 0$ is stronger than the condition $m_0 \frac{\log p}{\sqrt{n}} \to 0$ assumed by van de Geer et al. (2014). However, unlike van de Geer et al. (2014), no direct assumption is imposed on the sparsity of the inverse Hessian (or information matrix).

The assumption of a locally Lipschitz Jacobian $\dot{\rho}(\theta)$ is slightly stronger than the usual continuous differentiability condition required for testing nonlinear hypotheses (see, e.g., Newey and McFadden (1994, Section 9) and Hansen (2022*a*,*b*)). Under this assumption, the estimation error arising from the estimation of θ_0 and $\rho(\theta_0)$ becomes negligible.

Using this proposition, we make inference on the parameters θ_0 and $\rho(\theta_0)$. In sum, the debiasing approach above offers a straightforward and robust way to analyze high-dimensional survey data, ensuring both variable selection and valid inference in the presence of survey weights.

4 Empirical Results

The empirical analysis examines the relationship between digital technology adoption and cyber security vulnerabilities among Canadian businesses. We begin by assessing whether broader digital adoption correlates with increased cyber risk (Section 4.1) using rank correlations between the BDUS and various security measures. Next, we estimate a Stochastic Frontier Analysis (SFA) model (Section 4.2), treating the BDUS as an "output" to identify which factors drive a firm toward its digital usage frontier. We then employ logit models with Lasso selection for two binary outcomes: Technological Efficiency (Section 4.3) and Cyber Security Incidence (Section 4.4), allowing us to determine whether the same firm characteristics or digital practices that foster greater adoption also enhance implementation efficiency or increase cyber vulnerability.

The sample sizes for the analyses are as follows: the `svy LLasso` results for business efficiency are based on a weighted sample of 175,428 businesses, excluding those with a BDUS score less than 2. The cyber security incidence model uses the full merged dataset, encompassing 179,657 businesses. The `svy LLasso` procedure is implemented using `R`, where we utilize the default value of the tuning parameter λ from the `R` package `glmnet` (see Appendix A.1).

4.1 Digital Technology Adoption and Cyber Security

We explore whether a broader digital presence for a firm correlates with heightened cyber risk. Table 1 reports polychoric and polyserial correlations between the BDUS (an ordinal score from 0 to 10) that measures the amount of technology a firm has adopted and various cyber security measures, including firms cyber security spending (continuous variable), a binary indicator for whether or not a firm experienced a cyber incident, and whether or not a firm paid ransom as a result of a cyber security breach. We use polychoric correlation for ordinal to binary comparisons and polyserial correlation for the ordinal to continuous comparison.

The positive and significant correlation between BDUS and both cyber security spending ($\rho = 0.156^{***}$) and experiencing an incident ($\rho = 0.083^{***}$) suggests that while digitally engaged firms invest more in security, they also face greater exposure to attacks. Firms with a higher BDUS are also more likely to have to pay a ransom as a result of a cyber security breach ($\rho = 0.266^{***}$). The absence of any cyber security measures correlates negatively with BDUS ($\rho = -0.068^{***}$), indicating that firms with minimal digital footprints may perceive fewer threats but also forgo basic protective actions.

Table 1: Polychoric/Polyserial Correlations Between BDUS and Cyber Security Measures

Cyber Security Measure	**Correlation**	**p-value**
Cyber security spending (numeric)	0.156	$< 0.001^{***}$
Experienced a cyber security incident (binary)	0.083	$< 0.001^{***}$
Firm implemented cloud storage (binary)	0.108	$< 0.001^{***}$
Firm paid ransom (binary)	0.266	$< 0.001^{***}$
Firm implemented no cyber security measures (binary)	-0.068	$< 0.001^{***}$

Notes: The table reports polychoric correlations for ordinal–binary comparisons and polyserial correlations for the ordinal BDUS and the continuous measure of cyber security spending. All correlations are statistically significant at the 1% level. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Sample size: 179,657.

These correlations provide preliminary evidence of a trade-off: as firms adopt more digital tools (higher BDUS), they may both allocate more resources to cyber security and become more frequent targets of attacks. This finding motivates the proceeding empirical analyses.

4.2 Digital Technology Adoption by Canadian Businesses

We estimate an SFA model to examine which firm-level and industry-level characteristics bring businesses closer to their "digital usage frontier." Treating BDUS (a quantitative score 0–10) as an "output" — that is, the extent to which a firm adopts digital tools — allows us to separate random variation from systematic shortfalls in adoption. The frontier approach is particularly useful because some firms may lag in adopting new technologies for reasons beyond mere chance (e.g., internal constraints or strategic decisions). In this specification, a positive coefficient indicates that the corresponding variable moves firms closer to the frontier of digital adoption.

Table 2 presents the maximum likelihood estimates for the logistic SFA model. We use a logistic specification to account for the bounded nature of BDUS, recognizing that moving from, for example, two to three adopted technologies represents a more substantial increase in digital capacity than moving from eight to nine. The explanatory variables are grouped into three categories: *Firm characteristics*, *Digital technologies*, and *Cyber security measures*. After controlling for industry fixed effects, the estimated mean efficiency score across all firms is 77%, suggesting that while most businesses integrate some digital tools, gaps remain.

Table 2 indicates that several firm characteristics and digital practices significantly influence how close a business is to its digital usage frontier. *Medium* (0.085^{***}) and *Large* (0.173^{***}) firms both exhibit higher BDUS levels than the *Small* reference category, while *Working from*

Table 2: Logistic Stochastic Frontier Model for BDUS

Variable	**Coef.**	**Std. Error**	**z-value**	**p-value**
Intercept	1.721	0.021	82.800	$< 0.001^{***}$
Firm characteristics				
Medium-sized firm	0.085	0.012	6.840	$< 0.001^{***}$
Large-sized firm	0.173	0.016	10.920	$< 0.001^{***}$
Working from home	0.147	0.015	9.630	$< 0.001^{***}$
Mining/Utilities/Construction	−0.095	0.025	−3.800	$< 0.001^{***}$
Manufacturing	0.043	0.015	2.910	0.004^{***}
Wholesale/Retail/Transport	−0.055	0.022	−2.530	0.011^{**}
Education/Health	−0.066	0.017	−3.820	$< 0.001^{***}$
Arts/Accommodation/Food	0.106	0.022	4.840	$< 0.001^{***}$
Other services	0.001	0.027	0.030	0.980
Digital technologies				
Blockchain	0.058	0.030	1.920	0.055^{*}
Open source technologies	0.090	0.014	6.380	$< 0.001^{***}$
Client information management	0.183	0.012	15.540	$< 0.001^{***}$
Paid advertising	0.039	0.017	2.260	0.024^{**}
Free advertising	0.075	0.014	5.440	$< 0.001^{***}$
Firm provides ICT training	0.030	0.029	1.020	0.310
Sales-related problems	0.097	0.040	2.450	0.014^{**}
Cyber security measures				
Gender in ICT roles (% female)	0.002	0.000	5.670	$< 0.001^{***}$
Gender in cyber security (% female)	−0.001	0.000	−0.460	0.644
Cyber security certification	0.014	0.013	1.020	0.306
Cyber security practices	−0.007	0.015	−0.450	0.655
Cyber security training	0.029	0.018	1.620	0.106
Cyber security insurance	−0.019	0.014	−1.410	0.157
Cyber security incidents	0.010	0.017	0.600	0.551
Variance parameters:				
$\ln(\sigma_v^2)$	−4.265	0.130	−32.920	$< 0.001^{***}$
$\ln(\sigma_u^2)$	−2.001	0.088	−22.630	$< 0.001^{***}$
σ_v	0.119	0.008	15.440	$< 0.001^{***}$
σ_u	0.368	0.016	22.630	$< 0.001^{***}$
σ_u/σ_v	3.103	0.021	151.180	$< 0.001^{***}$

Notes: Reference categories: Small firm (firm size) and Professional services (industry). All other variables are binary. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. A positive coefficient indicates that the variable moves the firm closer to its digital usage frontier. Sample size: 179,657.

home (0.147^{***}) also aligns positively with adoption. Among digital technologies, *Open source solutions* (0.090^{***}), *Client information management* (0.183^{***}), and *Online advertising* (*Paid*: 0.039^{**}; *Free*: 0.075^{***}) yield significantly higher BDUS.

Certain industry sectors register negative coefficients: *Mining/Utilities/Construction* (-0.095^{***}), *Wholesale/Retail/Transport* (-0.055^{**}), and *Education/Health* (-0.066^{***}). Whereas *Manufacturing* (0.043^{***}) and *Arts/Accommodation/Food* (0.106^{***}) both exceed the baseline category of *Professional services*.

For cyber security measures, the share of *Female employees in ICT roles* (0.002^{***}) is the only strongly significant predictor, indicating that a higher percentage of women in ICT correlates with greater digital adoption. The variance parameter estimates confirm that systematic inefficiency (σ_u) dominates random noise (σ_v), suggesting that much of the under-adoption of digital tools is rooted in structural constraints rather than chance.

These results imply that resource capacity (firm size), remote-work arrangements, and the availability of specific digital tools or managerial practices (e.g., open source, client information management) can propel Canadian firms closer to their feasible frontier of technology usage. Industries such as Construction or Education/Health appear to face adoption barriers, while manufacturing and consumer-facing sectors integrate digital tools more readily. Although extensive digital adoption may heighten cyber threats (Section 4.1), most of the cyber security variables analyzed here do not notably influence how far along a firm is on the adoption curve.

4.3 Technological Efficiency of Canadian Businesses

The SFA analysis, using BDUS as the dependent variable in Section 4.2, measures the extent of digital adoption but does not capture how effectively these tools are utilized. To address this distinction, we introduce a binary variable for *Technological Efficiency*, derived through k-means clustering. This clustering algorithm categorizes firms based on whether they report few or many operational difficulties in using adopted technologies. A firm is classified as "Technologically Efficient" if it experiences relatively few challenges across multiple domains (see Section 2 for details). We restrict the sample to businesses with BDUS ≥ 2, ensuring a baseline level of digital engagement prior to evaluating efficiency.

Table 3 presents the `svy LLasso` results for the probability of a firm being Technologically Efficient. The independent variables are grouped into three categories: *Firm characteristics*, *Digital technologies*, and *Cyber security measures*. A positive coefficient indicates that the variable increases the likelihood of efficient technology use, while a negative coefficient suggests a reduced likelihood. The final two columns report the AMEs and their corresponding p-values.

Table 3: Debiased Logit Lasso Estimation Results for Technological Efficiency

Variables	`svy LLasso`	$\tilde{\theta}^{DB}$	p-value	$\widetilde{\text{AME}}^{DB}$	p-value
Intercept	−0.060	0.105	0.606	0.122	0.606
Firm characteristics					
Medium firm	0.551	0.856	$< 0.001^{***}$	0.158	$< 0.001^{***}$
Large firm	0.366	1.237	$< 0.001^{***}$	0.102	$< 0.001^{***}$
Remote work	0.509	0.664	$< 0.001^{***}$	0.139	$< 0.001^{***}$
Female in ICT roles (1–20%)	.	0.486	0.254	0.023	0.720
Female in ICT roles (21–40%)	.	0.150	0.688	0.139	0.153
Female in ICT roles (41–60%)	0.211	1.066	0.087^{*}	0.123	0.379
Female in ICT roles (>60%)	.	0.915	0.273	−0.032	0.214
Foreign market	.	0.041	0.873	0.031	0.566
Mining/Utilities/Construction	−0.772	−1.108	$< 0.001^{***}$	0.020	0.458
Manufacturing	.	0.130	0.411	−0.065	0.044^{**}
Wholesale/Retail/Transport	−0.300	−0.416	0.024^{**}	−0.030	0.392
Education/Health	.	−0.193	0.354	0.113	0.006^{***}
Arts/Accommodation/Food	0.256	0.808	0.002^{***}	0.146	0.001^{***}
Other services	0.524	1.086	$< 0.001^{***}$	0.146	0.001^{***}
Digital technologies					
Blockchain usage	0.086	1.066	0.153	0.046	0.428
ICT training	0.045	0.314	0.374	0.070	0.327
Online orders	.	−0.211	0.166	0.006	0.884
AI	.	0.205	0.520	0.264	$< 0.001^{***}$
IoT	1.765	1.971	$< 0.001^{***}$	0.156	$< 0.001^{***}$
Computer network	0.661	0.989	$< 0.001^{***}$	−0.001	0.968
Customer relationship management	.	−0.008	0.965	−0.141	$< 0.001^{***}$
Electronic data interchange	−0.680	−0.883	$< 0.001^{***}$	−0.101	0.007^{**}
Enterprise resource planning	.	−0.636	0.005^{***}	−0.168	0.050^{*}
Big data usage	0.365	1.350	0.019^{**}	0.076	0.012^{**}
Open source technologies	0.288	0.519	0.007^{***}	−0.022	0.730
Advertising	−0.327	−0.440	0.010^{**}	0.110	$< 0.001^{***}$
Free advertising	0.431	0.740	$< 0.001^{***}$	0.073	0.025^{**}
Website	0.399	0.454	0.010^{**}	−0.006	0.852
Company apps	.	−0.041	0.839	−0.087	$< 0.001^{***}$
Social media	−0.287	−0.587	$< 0.001^{***}$	−0.027	0.240
Fiber optic	.	−0.179	0.198	−0.046	0.113
Online sales	−0.064	−0.300	0.083^{*}	−0.013	0.607
Client information management	0.064	0.094	0.498	−0.068	0.019^{**}
Cyber security measures					
Female in cyber security roles (1–20%)	.	−0.141	0.708	−0.012	0.815
Female in cyber security roles (21–40%)	.	−0.077	0.799	0.030	0.443
Female in cyber security roles (41–60%)	.	0.202	0.391	0.049	0.166
Female in cyber security roles (>60%)	0.062	0.329	0.123	−0.019	0.487

Variables	`svy LLasso`	$\tilde{\theta}^{DB}$	p-value	$\widetilde{\text{AME}}^{DB}$	p-value
Cyber security employees (1–2)	.	-0.127	0.445	-0.003	0.946
Cyber security employees (3+)	.	-0.020	0.941	-0.014	0.622
Cyber security insurance	.	-0.090	0.590	0.014	0.537
Employee monitoring	.	-0.087	0.574	-0.184	$< 0.001^{***}$

Notes: All numeric values are rounded to three decimals. $\tilde{\theta}^{DB}$ and $\widetilde{\text{AME}}^{DB}$ denote the debiased logit Lasso coefficient and AME estimates, respectively. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Reference categories: Small firm, 0% female in ICT roles, 0% female in cyber security roles, 0 cyber security employees, and Industry: Professional services. Sample size: 175,428.

Table 3 shows that several firm characteristics significantly increase the probability of using digital tools efficiently. *Medium* ($\widetilde{\text{AME}}^{DB} = 0.158^{***}$) and *Large* ($\widetilde{\text{AME}}^{DB} = 0.102^{***}$) firms exhibit higher AME relative to *Small* firms. *Remote work* ($\widetilde{\text{AME}}^{DB} = 0.139^{***}$) is also associated with increased efficiency. Among digital technologies, the *IoT* ($\widetilde{\text{AME}}^{DB} = 0.156^{***}$) and *Big data analytics* ($\widetilde{\text{AME}}^{DB} = 0.076^{**}$) both show strong positive effects, while advertising efforts, such as free advertising ($\widetilde{\text{AME}}^{DB} = 0.073^{**}$), also correlate with efficient usage. For industries, *Arts/Accommodation/Food* ($\widetilde{\text{AME}}^{DB} = 0.146^{***}$) and *Education/Health* show a positive marginal effect ($\widetilde{\text{AME}}^{DB} = 0.113^{***}$) compared to *Professional services*.

Firms endowed with more resources, due to scale (medium or large size) or operational flexibility (remote work) can allocate staff and capital to integrate digital systems more effectively. Real-time connectivity from IoT and big data appears to reinforce structured work flows, while advertising activities may align with better-organized digital platforms. Consumer-facing industries, such as Arts/Accommodation/Food may capitalize on digital marketing tools or consumer facing technologies such as reservation systems more readily than sectors facing heavier regulatory or operational barriers.

The variables that reduce the likelihood of efficient digital implementation include *Electronic data interchange* ($\widetilde{\text{AME}}^{DB} = -0.101^{***}$) and *Enterprise resource planning* ($\widetilde{\text{AME}}^{DB} = -0.168^{**}$). The AME values suggest that more complex systems can pose challenges with technological adoption. The *Manufacturing* industry ($\widetilde{\text{AME}}^{DB} = -0.065^{**}$) has a negative effect based on AMEs, while certain cyber security variables, such as *Employee monitoring* ($\widetilde{\text{AME}}^{DB} = -0.184^{***}$), also correlate negatively with Technological Efficiency.

These negative or insignificant AMEs point to organizational or regulatory factors that can undermine digital adoption benefits. Complex software solutions (EDI, ERP) often require robust training and IT resources; without sufficient support, firms may experience integration hurdles. Industries like Mining/Utilities/Construction may face specialized work flows or regulatory strictures that may obstruct rapid digital platform adoption. Certain cyber security practices (employee monitoring) can introduce procedural friction or negative employment sentiment that overshadows efficiency gains if not carefully managed.

4.4 Cyber Security Incidence

The determinants of *Cyber Security Incidence* variable are analyzed using a binary dependent variable introduced in Section 2. The *Cyber Security Incidence* variable equals 1 if a firm reported experiencing at least one cyber security incident during 2021, and 0 otherwise; among 179,656 surveyed firms, 32,371 (18.0%) reported an incident. Cyber incidents encompass a range of adverse events, including theft of business assets, data breaches, disruptions to business activities, intellectual property losses, and other cyber-related issues. Although a large number of independent variables are included in the analysis, relatively few emerge as statistically significant predictors, indicating that cyber risk is shaped by a limited subset of factors.

Table 4: `svy LLasso` Results for Cyber Security Incidence

Variables	`svy LLasso`	$\tilde{\theta}^{DB}$	p-value	$\widetilde{\text{AME}}^{DB}$	p-value
Intercept	-1.904	-2.989	$< 0.001^{***}$	.	.
Firm characteristics					
Medium firm	.	0.047	0.716	0.007	0.723
Large firm	.	0.370	0.038^{**}	0.057	0.030^{**}
Remote work	.	0.104	0.476	0.015	0.488
Female in ICT roles (1–20%)	.	-0.325	0.236	-0.042	0.291
Female in ICT roles (21–40%)	.	0.389	0.178	0.060	0.155
Female in ICT roles (41–60%)	.	0.123	0.781	0.018	0.783
Female in ICT roles (>60%)	.	-0.305	0.660	-0.040	0.694
Mining/Utilities/Construction	.	0.728	0.002^{***}	0.119	0.001^{***}
Manufacturing	.	0.464	0.002^{***}	0.071	0.001^{***}
Wholesale/Retail/Transport	.	0.398	0.034^{**}	0.059	0.032^{**}
Education/Health	.	0.058	0.771	0.008	0.777
Arts/Accommodation/Food	.	0.235	0.394	0.034	0.394
Other services	.	0.484	0.083^{*}	0.075	0.066^{*}
Digital technologies					
Blockchain usage	.	0.202	0.667	0.030	0.663
ICT training	.	0.361	0.168	0.055	0.151
Online orders	.	0.032	0.835	0.005	0.841
AI	.	-0.289	0.207	-0.038	0.255
IoT	.	0.148	0.336	0.021	0.345
Computer network	.	0.016	0.909	0.002	0.912

Variables	`svy LLasso`	$\tilde{\theta}^{DB}$	p-value	$\widetilde{\text{AME}}^{DB}$	p-value
Customer relationship management		0.129	0.419	0.019	0.426
Electronic data interchange	.	−0.176	0.293	−0.024	0.323
Enterprise resource planning	.	0.137	0.477	0.020	0.480
Big data usage	.	−0.414	0.307	−0.053	0.374
Open source technologies	.	−0.102	0.528	−0.014	0.550
Confidential cloud	0.161	0.466	$< 0.001^{***}$	0.067	0.002^{***}
Personal device	.	0.222	0.120	0.031	0.141
VPN	.	0.003	0.986	0.000	0.986
Payment services	.	−0.123	0.663	−0.017	0.682
Client information management	.	0.082	0.569	0.012	0.582
Website	.	−0.098	0.616	−0.014	0.624
Company apps	.	0.112	0.556	0.016	0.563
Social media	.	−0.173	0.255	−0.025	0.265
Online sales	.	0.072	0.663	0.010	0.673
Cyber security measures					
Anti malware	.	−0.236	0.323	−0.034	0.327
Web security	.	−0.115	0.502	−0.016	0.517
Email security	.	0.272	0.271	0.037	0.304
Network security	.	0.087	0.691	0.012	0.704
Data security	.	0.164	0.362	0.023	0.374
POS security	.	−0.115	0.502	−0.016	0.522
Software security	.	−0.251	0.171	−0.034	0.198
Hardware security	.	0.203	0.259	0.029	0.266
Password security	0.159	0.267	0.143	0.038	0.157
Access security	.	0.008	0.963	0.001	0.964
Female in cyber security roles (1–20%)	.	−0.042	0.904	−0.006	0.908
Female in cyber security roles (21–40%)	.	0.062	0.819	0.009	0.823
Female in cyber security roles (41–60%)	.	0.188	0.434	0.028	0.434
Female in cyber security roles (>60%)	.	−0.424	0.054^{*}	−0.056	0.082^{*}
Cyber security employees (1–2)	.	0.527	0.061^{*}	0.074	0.07^{*}
Cyber security employees (3+)	.	0.059	0.737	0.008	0.743
Cyber security insurance	.	−0.321	0.050^{**}	−0.043	0.073^{*}
Cyber consultant	.	−0.018	0.941	−0.003	0.943
Cyber information	0.295	0.214	0.450	0.030	0.459
Cyber training	.	0.195	0.254	0.028	0.259
Cyber policy	.	−0.108	0.507	−0.015	0.527
Cyber practice	0.051	0.254	0.331	0.036	0.346
Employee monitoring	.	0.242	0.122	0.035	0.123
Risk assessment	.	0.199	0.219	0.029	0.226
Cyber team: white employees only	.	−0.049	0.785	−0.007	0.794
Cyber team: minority employees only	.	0.424	0.152	0.065	0.131
Cyber certification	.	0.096	0.570	0.014	0.580
No cyber security measures	.	−0.179	0.634	−0.024	0.657

Notes: All numeric values are rounded to three decimals. $\tilde{\theta}^{DB}$ and $\widetilde{\text{AME}}^{DB}$ denote the debiased logit Lasso coefficient and AME estimates, respectively. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Reference categories: Small firm, 0% female in ICT roles, 0% female in cyber security roles, 0 cyber security employees, cyber team: diverse employees. Sample size: 179,657.

Table 4 displays the `svy LLasso` estimates for the probability of experiencing a cyber security incident. Among the variables with positive and statistically significant associations are being a large firm ($\widetilde{\text{AME}}^{DB} = 0.057^{**}$) and adopting confidential cloud solutions ($\widetilde{\text{AME}}^{DB} = 0.067^{***}$). In addition, certain industry categories Mining/Utilities/Construction, Manufacturing, Wholesale/Retail/Transport, and Other services exhibit statistically significant positive coefficients.

Two factors show negative and statistically significant effects on the likelihood of a cyber incident. Having over 60% female employees in cyber security roles ($\widetilde{\text{AME}}^{DB} = -0.056^{*}$) and holding cyber security insurance ($\widetilde{\text{AME}}^{DB} = -0.043^{*}$) are both associated with lower probabilities of experiencing an incident.

These results suggest that larger enterprises may face heightened cyber risks, potentially reflecting the extensive data infrastructures and more valuable assets characteristic of bigger firms. Reliance on confidential cloud solutions might increase exposure to attacks, as remote and cloud-based systems use electronic methods to store data. Lower incidence rates among firms with higher proportions of female cyber security personnel could indicate that diverse cyber teams are more adept at preventing or responding to threats. Likewise, the negative association of cyber security insurance with incidence likelihood hints that insured firms may adopt more robust preventive strategies to manage their risk profiles.

4.5 Interaction Effects

Using an adaptive Lasso approach with polynomial expansions, we investigate whether second-order (interaction) terms enhance the predictive accuracy of our *Technological Efficiency* and *Cyber Security Incidence* specifications. We estimate both a first-order (linear) model and a second-order model that includes all pairwise interactions among the explanatory variables. See Bühlmann and van de Geer (2011) for the theoretical background.

Table 5 reports mean-squared cross-validation (CV) errors under each specification, along with the penalty parameter λ_{cv}. For the *Technological Efficiency* model, allowing second-order terms substantially lowers the CV error, suggesting that interactions play an important role in explaining efficiency gains from digital technologies. In the *Cyber Security Incidence* model, the simpler first-order specification yields a slightly lower CV error, indicating that higher-order interactions do not improve predictions of cyber security incidence likelihood.

A second-order polynomial specification is justified for the *Technological Efficiency* model, therefore we re-estimate the `svy LLasso` model with interaction terms involving firm size, remote work, key technologies, and industry classifications. Table 6 shows the significant interaction terms selected by the `svy LLasso` estimator, along with their debiased parameter estimates $\tilde{\theta}^{DB}$ and p-values. Only the interaction coefficients that were selected by the Lasso and statistically significant at the 5% level are included in Table 6.

Table 5: Cross-Validation Results for Models With and Without Interaction Terms

Model	**Degree**	λ_{cv}	**CV Error**	**Preferred Specification**
Technological Efficiency	1	0.00261	0.92442	
	2	0.00029	0.34142	Second-order
Cyber Security Incidence	1	0.00685	0.87247	First-order
	2	0.03132	0.86486	

Notes: The table shows the mean-squared CV error from an adaptive Lasso specification with polynomial expansions of different degrees (1 = linear, 2 = second-order interactions). For each model, λ_{cv} denotes the penalty parameter that minimizes the CV error. Based on these metrics, the second-order polynomial is preferred for the Technological Efficiency model, while a first-order specification is preferred for the Cyber Security Incidence model. Sample size: 179,657.

Table 6: `svy LLasso` with Interactions: Technological Efficiency

Variables	**Lasso**	$\tilde{\theta}^{DB}$	**p-value**
Intercept	1.095	3.953	0.012**
Remote work	2.641	4.295	0.002***
Online orders	−2.008	−4.555	< 0.001***
IoT	2.249	4.873	0.009***
Computer network	−0.369	−3.173	0.006***
Fiber optic	−1.197	−3.019	0.009***
Medium firm × Remote work	−0.884	−1.428	0.040**
Medium firm × IoT	1.901	5.554	< 0.001***
Medium firm × Cyber security insurance	−0.470	−1.257	0.009***
Medium firm × Website	1.878	2.934	0.004***
Medium firm × Company apps	−2.170	−4.019	0.002***
Medium firm × Manufacturing	−2.471	−5.658	< 0.001***
Medium firm × Other services	−2.677	−8.226	< 0.001***
Remote work × Female in ICT roles (1–20%)	0.024	−16.870	0.008***
Remote work × CRM	0.340	2.496	0.018**
Remote work × Open source technologies	−1.689	−2.598	0.005***
Remote work × Website	−1.864	−3.230	0.003***

Variables	**Lasso**	$\tilde{\theta}^{DB}$	**p-value**
Remote work × Social media	2.361	3.824	$< 0.001^{***}$
ICT training × Foreign market	0.298	7.453	0.002^{***}
ICT training × CIM	−0.240	−4.743	0.002^{***}
Female in ICT roles (1–20%) × Open source	0.136	9.104	0.005^{***}
Large firm × CRM	−1.116	−4.636	0.004^{***}
Online orders × Computer network	1.278	2.872	0.004^{***}
Online orders × ERP	−0.047	−3.533	0.002^{***}
Online orders × Social media	0.939	3.102	0.002^{***}
Online orders × Manufacturing	0.453	3.207	0.001^{***}
Foreign market × EDI	1.704	5.321	0.003^{***}
Foreign market × CIM	−0.622	−2.730	0.035^{**}
IoT × Computer network	−0.890	−3.443	$< 0.001^{***}$
IoT × CIM	0.663	2.285	0.032^{**}
IoT × Manufacturing	−2.497	−4.684	0.002^{***}
Computer network × CRM	−2.041	−2.983	0.006^{***}
Computer network × Advertising	−0.191	−3.003	0.018^{**}
Computer network × Fiber optic	1.525	3.576	$< 0.001^{***}$
Computer network × Other services	−0.815	−5.782	0.039^{**}
CRM × CIM	−0.105	−2.083	0.015^{**}
CRM × Advertising	−1.041	−2.367	0.035^{**}
EDI × Advertising	−4.335	−5.309	$< 0.001^{***}$
EDI × Online sales	−2.325	−3.303	0.006^{***}
ERP × Manufacturing	0.953	2.583	0.038^{**}
ERP × Wholesale/Retail/Transport	2.113	3.094	0.034^{**}
CIM × Mining/Utilities/Construction	1.256	6.701	$< 0.001^{***}$
CIM × Education/Health	−4.227	−6.218	$< 0.001^{***}$
Advertising × Free advertising	2.001	2.886	0.021^{**}
Advertising × Website	1.479	4.818	0.029^{**}
Advertising × Education/Health	−5.083	−9.202	$< 0.001^{***}$
Free advertising × Website	−0.161	−6.123	0.013^{**}
Free advertising × Social media	−0.298	−3.753	0.010^{**}
Free advertising × Wholesale/Retail/Transport	−3.066	−3.263	0.038^{**}
Free advertising × Education/Health	3.234	6.648	$< 0.001^{***}$
Website × Online sales	1.572	6.424	0.011^{**}
Website × Mining/Utilities/Construction	−1.338	−4.640	$< 0.001^{***}$
Company apps × Social media	5.323	10.627	$< 0.001^{***}$
Company apps × Manufacturing	−0.974	−3.853	0.030^{**}
Company apps × Arts/Accommodation/Food	−0.883	−9.426	0.033^{**}
Social media × Wholesale/Retail/Transport	−2.163	−4.058	$< 0.001^{***}$
Fiber optic × Arts/Accommodation/Food	0.388	4.972	0.049^{**}
Fiber optic × Other services	2.051	9.419	$< 0.001^{***}$
Online sales × Mining/Utilities/Construction	2.554	8.200	$< 0.001^{***}$
Online sales × Manufacturing	−0.384	−2.825	0.007^{***}
Online sales × Other services	−1.933	−9.947	$< 0.001^{***}$

Notes: Numeric values are rounded to three decimal places. $\tilde{\theta}^{DB}$ denotes the debiased logit Lasso coefficient estimate. Coefficients statistically significant at the 5% level based on their p-values are reported. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Variable names are abbreviated for brevity: CRM (Customer Relationship Management), CIM (Client Information Management), EDI (Electronic Data Interchange), ERP (Enterprise Resource Planning). Sample size: 175,428.

Medium-sized firms exhibit positive and statistically significant interactions with the adoption of the IoT (*Medium firm* × *IoT*: $\tilde{\theta}^{DB} = 5.554^{***}$) and firm website use (*Medium firm* × *Website*: $\tilde{\theta}^{DB} = 2.934^{***}$). Remote work arrangements positively interact with social media (*Remote work* × *Social media*: $\tilde{\theta}^{DB} = 3.824^{***}$) and customer relationship management software (*Remote work* × *CRM*: $\tilde{\theta}^{DB} = 2.496^{**}$).

The largest statistically significant positive interaction occurs between company-specific applications and social media use (*Company apps* × *Social media*: $\tilde{\theta}^{DB} = 10.627^{***}$). The presence of female employees in ICT roles (1–20%) interacts positively and significantly with open-source technology adoption (*Female in ICT roles (1–20%)* × *Open source*: $\tilde{\theta}^{DB} = 9.104^{***}$). Online sales and the Mining, Utilities, and Construction industry have a positive statistically significant interaction (*Online sales* × *Mining/Utilities/Construction*: $\tilde{\theta}^{DB} = 8.200^{***}$). Additionally, free advertising positively interacts with firms in the Education and Health industry (*Free advertising* × *Education/Health*: $\tilde{\theta}^{DB} = 6.648^{***}$).

The interaction between EDI and advertising is negative and statistically significant (*Electronic data interchange* × *Advertising*: $\tilde{\theta}^{DB} = -5.309^{***}$). Similarly, medium-sized firms show a negative and statistically significant interaction with company-specific applications (*Medium firm* × *Company apps*: $\tilde{\theta}^{DB} = -4.019^{***}$). Remote work arrangements negatively interact with open-source technologies (*Remote work* × *Open source*: $\tilde{\theta}^{DB} = -2.598^{***}$). Online sales exhibit a negative interaction with firms in the Wholesale, Retail, and Transport industries (*Online sales* × *Wholesale/Retail/Transport*: $\tilde{\theta}^{DB} = -4.058^{***}$).

Medium-sized firms benefit from adopting IoT solutions and online platforms, likely due to greater resource availability compared to smaller firms. Remote work enhances the efficiency of communication-oriented technologies, such as CRM and social media. The remote work variable itself also exhibits a strong, statistically significant positive effect on firm technological efficiency. Workforce diversity in ICT roles is positively correlated with technological efficiency, especially when adopting open-source systems. Industry-specific interactions produce varied effects depending on the technology and sector: online sales positively interact with mining, utilities, and construction, while advertising has a positive effect in the education and health sectors. Conversely, interactions like online sales with wholesale, retail, and transport are negatively associated with firm technological efficiency.

5 Conclusion

This paper contributes new evidence on how Canadian businesses navigate the trade-off between digital technology adoption and heightened cyber security risk. Using data from Statistics Canada's SDTIU and CSCSC surveys, we construct a BDUS to gauge overall adoption levels and then evaluate how effectively businesses use these tools by modeling their technological efficiency. In addition, we employ a survey-weight-adjusted Lasso estimator and introduce a debiasing method for high-dimensional logit models to identify the drivers of technological efficiency and cybersecurity risk.

The stochastic frontier analysis suggests that larger firms, remote work arrangements, and specific advanced technologies (e.g., open-source solutions, client information management systems) can push a firm closer to its digital "frontier." At the same time, a portion of businesses lag behind feasible adoption levels, as indicated by the high ratio of inefficiency to noise in the frontier estimations.

Firms must balance efficiency gains against growing cyber vulnerabilities. Firms that adopt more sophisticated digital technologies or store sensitive data in the cloud often face elevated risk exposure. Our `svy LLasso` model on cyber incidence confirm that large firms and those using cloud-based services are more likely to report cyber security incidents. However, the predictive power of cyber risk does not improve with second-order interaction effects, suggesting that firm size and core technological choices are the primary drivers of cyber exposure. The two main variables that decreased the likelihood of cyber incidents were firms having cyber security insurance and firms that had a high representation of females in cyber security roles.

When it comes to technological efficiency simple linear relationships fail to capture the complexity of how organizational choices, workforce composition, and industry shape digital outcomes. By allowing second-order (interaction) terms, the `svy LLasso` approach shows that certain combinations of variables such as *medium firms* adopting IoT, or *female ICT representation* interacting with advanced tools like AI can be particularly conducive to efficiency improvements. On the other hand, friction in implementing complex software like EDI or ERP can negate some of these benefits.

The analysis demonstrates the importance of firm size and industry. While small firms are sometimes more "locally efficient," the resource advantages of larger organizations may facilitate deeper or more comprehensive integration of technologies. Industries also differ substantially. In resource- and asset-intensive sectors such as Mining or Construction, strong positive interactions emerge between targeted digital solutions and improved operational processes, whereas compliance-heavy fields like Education and Health exhibit more negative or complex relationships.

Gender composition in ICT roles has meaningful consequences for digital adoption outcomes. Although our results do not prove a causal mechanism, the recurring positive coefficients on interactions involving a share of female ICT staff and advanced technologies suggest that even partial gender diversity in technical teams can amplify the returns to adopting new tools. This pattern is also seen in broader research suggesting that heterogeneity in skill sets and perspectives can catalyze creative problem-solving.

The results highlight the balance firms must strike between achieving efficiency gains from digital technologies and managing increased cyber risks. Larger, digitally advanced firms approach their efficiency frontier yet face elevated cyber vulnerability, especially when security practices fall short. Remote work arrangements enhance both digital adoption and technological efficiency without increasing cyber risk exposure. Female representation in ICT and cyber security roles consistently improves outcomes across adoption, efficiency, and cyber security. Certain cyber security practices, particularly obtaining cyber security insurance and ensuring gender diversity within cyber security roles significantly reduce incident likelihood. Reliance on cloud-based services, notably confidential cloud storage, emerges as a risk factor increasing cyber vulnerability. These results suggest policymakers should implement targeted digital strategies tailored by industry and firm size to boost technological efficiency, while simultaneously establishing baseline cyber security practices such as insurance coverage and workforce diversity to mitigate cyber threats effectively.

References

Aczel, B., Kovacs, M., Van Der Lippe, T. and Szaszi, B. (2021), 'Researchers Working from Home: Benefits and Challenges', *PloS one* **16**(3), e0249127.

Aghimien, D., Aigbavboa, C., Meno, T. and Ikuabe, M. (2021), 'Unravelling the Risks of Construction Digitalisation in Developing Countries', *Construction Innovation* **21**(3), 456–475.

Ahnert, T., Brolley, M., Cimon, D. A. and Riordan, R. (2022), 'Cyber Security and Ransomware in Financial Farkets', *Available at SSRN 4057505* .

Bilodeau, H., Lari, M. and Uhrb, M. (2019), 'Cyber Security and Cybercrime Challenges of Canadian Businesses, 2017', *Juristat: Canadian Centre for Justice Statistics* pp. 1–18.

Blichfeldt, H. and Faullant, R. (2021), 'Performance Effects of Digital Technology Adoption and Product & Service Innovation–A Process-Industry Perspective', *Technovation* **105**, 102275.

Bousquet, T. (2023), '100,000 Current and Past Nova Scotia Health, IWK, and Public Service Employees Had Their Payroll Information Stolen in MOVEit Breach', https://www.halifaxexaminer.ca/government/province-house/100000-current-past-nova-scotia-health-employees-had-their-payroll-information-stolen-in-moveit-breach/.

Bridge, S. and Zoledziowski, A. (2024), '1 Million Books and 4 Months Later, Toronto's Library Recovers from a Cyberattack', https://www.cbc.ca/news/canada/toronto/toronto-library-ransomware-recovery-1.7126412.

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

Cebula, J. J. and Young, L. R. (2010), A Taxonomy of Operational Cyber Security Risks, Technical report, Software Engineering Institute, Carnegie Mellon University.

Eling, M., Schnell, W. and Sommerrock, F. (2016), *Ten Key Questions on Cyber Risk and Cyber Risk Insurance*, The Geneva Association, Zurich. Available at https://www.genevaassociation.org.

Ferrari, A. (2012), *Digital Competence in Practice: An Analysis of Frameworks*, Vol. 10, Luxembourg: Publications Office of the European Union.

Fitch Ratings (2021), ‘Sharply Rising Cyber Insurance Claims Signal Further Risk Challenges’.
URL: *https://www.fitchratings.com/research/insurance/sharply-rising-cyber-insurance-claims-signal-further-risk-challenges-15-04-2021*

Gartner, Inc. (2021), ‘Gartner Forecasts Worldwide Security and Risk Management Spending to Exceed $150 Billion in 2021’.
URL: *https://www.gartner.com/en/newsroom/press-releases/2021-05-17-gartner-forecasts-worldwide-security-and-risk-managem*

Hackney, A., Yung, M., Somasundram, K. G., Nowrouzi-Kia, B., Oakman, J. and Yazdani, A. (2022), ‘Working in the Digital Economy: A Systematic Review of the Impact of Work-from-Home Arrangements on Personal and Organizational Performance and Productivity’, *Plos one* **17**(10), e0274728.

Hansen, B. (2022*a*), *Econometrics*, Princeton University Press.

Hansen, B. (2022*b*), *Probability and Statistics for Economists*, Princeton University Press.

Hastie, T., Tibshirani, R. and Wainwright, M. (2015), *Statistical Learning with Sparsity: The Lasso and Generalizations*, CRC press.

Jasiak, J., MacKenzie, P. and Tuvaandorj, P. (2024), ‘Digital Divide: Empirical Study of CIUS 2020’, *ArXiv preprint arXiv:2301.07855* .

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.

Kitagawa, R., Kuroda, S., Okudaira, H. and Owan, H. (2021), ‘Working from home and productivity under the covid-19 pandemic: Using survey data of four manufacturing firms’, *PLoS One* **16**(12), e0261761.

Leung, D., Meh, C. and Terajima, Y. (2008), ‘Productivity in Canada: Does Firm Size Matter?’, *Bank of Canada Review* **2008**(Autumn), 7–16.

Negahban, S. N., Ravikumar, P., Wainwright, M. J. and Yu, B. (2012), 'A Unified Framework for High-Dimensional Analysis of M-Estimators with Decomposable Regularizers', *Statistical Science* **27**(4), 538 – 557.
URL: *https://doi.org/10.1214/12-STS400*

Newey, W. K. and McFadden, D. (1994), Large Sample Estimation and Hypothesis Testing, *in* R. F. Engle and D. L. McFadden, eds, 'Handbook of Econometrics, Volume 4', Amsterdam, chapter 36, pp. 2111–2245.

OECD (2017), *Enhancing the Role of Insurance in Cyber Risk Management*, OECD Publishing, Paris. Available at OECD iLibrary.
URL: *https://doi.org/10.1787/9789264282148-en*

van de Geer, S., Bühlmann, P., Ritov, Y. and Dezeure, R. (2014), 'On Asymptotically Optimal Confidence Regions and Tests for High-Dimensional Models', *The Annals of Statistics* **42**(3), 1166 – 1202.
URL: *https://doi.org/10.1214/14-AOS1221*

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

Xia, L., Nan, B. and Li, Y. (2023), 'Debiased Lasso for Generalized Linear Models with a Diverging Number of Covariates', *Biometrics* **79**(1), 344–357.

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

A Technical Appendix

A.1 Additional details on the implementation.

Tuning Parameter Selection. In the empirical analysis and the small Monte Carlo simulations below, the logit Lasso model `svyLLasso` is fitted using the `R` package `glmnet`. For the tuning parameter λ, we use the package's default value, determined by 10-fold cross-validation with the loss function `auc` (area under the ROC curve).

Average marginal effect in the logit model. We report the marginal effects (ME) for each variable along with the coefficient estimates for logit models in the tables. Let $\Lambda(z) = \exp(z)/(1+\exp(z))$ be the logistic distribution function. For a dummy regressor $\tilde{x}_{ij}$, where $j = 1, \ldots, p$ and $i = 1, \ldots, n$, the ME is defined as:

$$\mathrm{ME}_{ij}(\theta) \equiv \Lambda(x_i'\theta)|_{\tilde{x}_{ij}=1} - \Lambda(x_i'\theta)|_{\tilde{x}_{ij}=0}.$$

Given the survey weights $\{w_i\}_{i=1}^n$ corresponding to the observations $\{(y_i, x_i')'\}_{i=1}^n$, the AME of the j-th regressor is given by:

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

where θ_0 is the true parameter value, and the expectation is with respect to the regressors' distribution. An estimator for AME_j is:

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

where $\hat{\theta} = (\hat{\alpha}, \hat{\beta}')'$ is the estimated parameter vector, e.g., the `svy LLasso` estimator. The debiased logit Lasso estimator of AME_j is then constructed using the one-step iteration provided in the equation (3.1).

A.2 Post-Selection Inference for Survey-GLM

Consider the density of a scalar outcome variable y_i given a $(p+1)\times 1$ vector of covariates x_i (which includes a constant) specified as

$$f(y_i|x_i, \theta_0) = \exp(y_i x_i'\theta_0 - a(x_i'\theta_0))c(y_i), \quad i = 1, \ldots, n,$$

where θ_0 is the true value of the parameter vector $\theta \in \mathbb{R}^{p+1}$, and $a(\cdot)$ and $c(\cdot)$ are known functions. The combined SDTIU and CSCSC data were collected using a stratified sampling scheme, wherein units within each stratum are sampled independently with equal probability. In line with this, we treat w_i as fixed (see Wooldridge, 2001), and assume $\{(y_i, x_i')'\}_{i=1}^n$ to be independent.

Let $g(y, x'\theta) \equiv -\log f(y, x'\theta)$ and define the weighted log-likelihood function as:

$$L_n(\theta) \equiv -n^{-1} \sum_{i=1}^n w_i g(y_i, x_i'\theta). \tag{A.1}$$

The score function, the sample information and negative Hessian matrices corresponding to (A.1) are defined as

$$S(\theta) \equiv \frac{\partial L_n(\theta)}{\partial \theta} = -n^{-1} \sum_{i=1}^n w_i x_i \dot{g}(y_i, x_i'\theta), \quad \dot{g}(y,t) \equiv \frac{\partial g(y,t)}{\partial t}, \tag{A.2}$$

$$\hat{I}(\theta) \equiv n^{-1} \sum_{i=1}^n w_i^2 x_i x_i' \dot{g}(y_i, x_i'\theta)^2, \tag{A.3}$$

$$\hat{H}(\theta) \equiv -\frac{\partial^2 L_n(\theta)}{\partial \theta \partial \theta'} = n^{-1} \sum_{i=1}^n w_i x_i x_i' \ddot{g}(y_i, x_i'\theta), \quad \ddot{g}(y,t) \equiv \frac{\partial^2 g(y,t)}{\partial t^2}. \tag{A.4}$$

Moreover, we define $H(\theta_0) \equiv \mathrm{E}[\hat{H}(\theta_0)]$ and $I(\theta_0) \equiv \mathrm{E}[\hat{I}(\theta_0)]$.

We will using the following notations in the assumptions and the proof of Proposition 3.1 below. Let $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the smallest and the largest eigenvalue of a symmetric matrix A, respectively. For a real matrix $A = (a_{ij})$, let $\|A\|_\infty \equiv \max_{i,j} |a_{ij}|$, and $\|A\| = \sqrt{\mathrm{tr}(A'A)}$ and $\|A\|_2 = \sqrt{\lambda_{\max}(A'A)}$ denote its Frobenius and spectral norms, respectively. The sub-Gaussian norm of a random variable X is defined as $\|X\|_{\psi_2} \equiv \sup_{m \geq 1} m^{-1/2} (\mathrm{E}[|X|^m])^{1/m}$. The sub-Gaussian norm for the random vector is defined as $\|X\|_{\psi_2} \equiv \sup_{\|b\|=1} \|X'b\|_{\psi_2}$.

We establish the asymptotic validity of the debiasing method under the following assumptions imposed directly on the negative log-density function $g(y,t)$ which are similar to the assumptions employed in van de Geer et al. (2014) and Xia et al. (2023). Let $X = [x_1, \ldots, x_n]'$.

Assumption 1 (Asymptotic validity)**.**

(a) $\{(y_i, x_i')'\}_{i=1}^n$ *are independent with* $\max_{1 \leq i \leq n} a_i < C_u < \infty$ *a.s. where*

$$a_i \in \{\|x_i\|_{\psi_2}, \|x_i\|_\infty, \|X\theta_0\|_\infty\}.$$

Moreover, w_i *is non-random with* $0 < C_l < w_i < C_u$ *for all* n, i*.*

(b) *For* $A \in \{H(\theta_0), I(\theta_0), \mathrm{E}[n^{-1}X'X]\}$*, there exist positive constants* λ_l *and* λ_u *such that* $0 < \lambda_l \leq \lambda_{\min}(A) \leq \lambda_{\max}(A) \leq \lambda_u < \infty$*.*

(c) *The function* $g(y,t) \equiv a(t) - yt - \log c(y)$ *is convex in* $t \in \mathbb{R}$ *for all* y*, and twice differentiable with respect to* t *for all* (y,t)*. There exist a positive definite matrix* H *and* $\eta > 0$ *such that* $\lambda_{\min}(H) > \lambda_l > 0$ *and*

$$n^{-1}\sum_{i=1}^{n} \mathrm{E}[w_i(g(y_i, x_i'\theta) - g(y_i, x_i'\theta_0))] \geq \|H^{1/2}(\theta - \theta_0)\|^2 \tag{A.5}$$

for all $\|X(\theta-\theta_0)\|_\infty < \eta$*. Furthermore,* $\ddot{g}(y,t)$ *is Lipschitz with some constant* $L_0 > 0$*:*

$$\max_{t_0 \in \{x_i'\theta_0\}} \sup_{\max(|t-t_0|, |\tilde{t}-t_0|) \leq \eta} \sup_{y \in \mathcal{Y}} \frac{|\ddot{g}(y,t) - \ddot{g}(y,\tilde{t})|}{|t-\tilde{t}|} \leq L_0, \tag{A.6}$$

and

$$\max_{t_0 \in \{x_i'\theta_0\}} \sup_{y \in \mathcal{Y}} |\dot{g}(y,t_0)| \leq C_u, \tag{A.7}$$

$$\max_{t_0 \in \{x_i'\theta_0\}} \sup_{|t-t_0| \leq \eta} \sup_{y \in \mathcal{Y}} |\ddot{g}(y,t)| \leq C_u. \tag{A.8}$$

For discussions of these assumptions, we refer to Jasiak and Tuvaandorj (2023).

Simulations. We conduct a simulation experiment to verify the robustness of the debiased logit Lasso inference. We first generate $N = 10,000$ draws from a standard logit model as follows:

$$y_i \sim \text{Bernoulli}(\pi_i), \tag{A.9}$$

where $\theta_0 = (1, 1, 1, 0_{1\times(p-2)})'$, $\tilde{x}_{ij} \sim$ i.i.d. Bernoulli(0.5) for $j = 1, \dots, p$ and $i = 1, \dots, N$, $x_i = (1, \tilde{x}_i')'$, and $\pi_i = x_i'\theta_0$.

The population is then stratified into four strata of sizes 1,000, 2,000, 3,000, and 4,000. From each stratum, we draw 50 and 100 observations with replacement, yielding stratified samples of size $n = 200$ and $n = 400$, respectively. Observation weights are $w_i = 0.1, 0.2, 0.3, 0.4$, corresponding to the four strata. To evaluate the impact of regressor dimensionality, we set p such that $\frac{p}{n} \in \{0.01, 0.025, 0.05, 0.1, 0.25, 0.5\}$ for each $n \in \{200, 400\}$. The true AME for $\theta_{(2)} = \beta_1$ is 0.11.

We assess the empirical size of the tests by separately testing two null hypotheses:

$$H_0 : \theta_{(2)} = 1, \quad H_0 : \text{AME}_2 = 0.11. \tag{A.10}$$

The empirical sizes of the DB test and the standard survey t-test (t_{svy}) at the 5% nominal level are presented in the table below. The standard survey logit t_{svy} test overrejects by a wide margin, while the DB test exhibits reasonably accurate null rejection rates for both hypotheses in most cases, confirming its robustness to regressor dimensionality.

Table 7: Empirical rejection frequencies of the tests for $H_0 : \theta_{(2)} = 1$ and $H_0 : \text{AME}_2 = 0.11$ at 5% level. Standard stratified sampling.

Tests	$p = 2$	$p = 5$	$p = 10$	$p = 20$	$p = 50$	$p = 100$
$H_0 : \theta_{(2)} = 1$, $n = 200$						
DB	5.0	4.4	3.7	3.1	4.5	3.3
t_{svy}	6.2	6.4	8.0	8.7	36.0	94.9
$H_0 : \text{AME}_2 = 0.11$, $n = 200$						
DB	5.4	5.3	4.6	3.7	3.5	1.4
t_{svy}	5.7	7.7	7.4	8.2	50.9	93.3

Tests	$p = 4$	$p = 10$	$p = 20$	$p = 40$	$p = 100$	$p = 200$
$H_0 : \theta_{(2)} = 1$, $n = 400$						
DB	4.8	4.4	6.0	3.7	5.6	3.9
t_{svy}	5.0	5.1	6.3	15.9	40.4	98.3
$H_0 : \text{AME}_2 = 0.11$, $n = 400$						
DB	4.5	4.9	5.8	5.0	4.6	3.3
t_{svy}	5.3	6.9	9.1	10.8	46.8	93.7

Notes: $n = 200, 400$. DB and t_{svy} denote the debiased Lasso and standard survey-weighted t tests respectively. 1000 simulation replications.

A.3 Proof of Proposition 3.1

We first prove the following lemma, which establishes the asymptotic distribution of a studentized quantity involving the expected Hessian and information matrices, as well as the score function, all evaluated at the true parameters.

Lemma A.1. *Let Assumption 1 hold and* $p^{1+\delta_0}/n \to 0$ *for some* $0 < \delta_0 \leq 1$*. Then, as* $n \to \infty$

$$\left(\dot{\rho}(\theta_0)'H(\theta_0)^{-1}I(\theta_0)H(\theta_0)^{-1}\dot{\rho}(\theta_0)\right)^{-1/2}\dot{\rho}(\theta_0)'H(\theta_0)^{-1}n^{1/2}S(\theta_0) \xrightarrow{d} N(0, I_r).$$

Proof of Lemma A.1. Let $s_i(\theta_0) \equiv w_i x_i \dot{g}(y_i, x_i'\theta_0)$, $X_{ni} \equiv n^{-1/2}\dot{\rho}(\theta_0)'H(\theta_0)^{-1}s_i(\theta_0)$ and $\Sigma_n \equiv \operatorname{Var}[\sum_{i=1}^n X_{ni}] = \dot{\rho}(\theta_0)'H(\theta_0)^{-1}I(\theta_0)H(\theta_0)^{-1}\dot{\rho}(\theta_0)$. Let $\nu_n \equiv \lambda_{\min}(\Sigma_n)$. We will verify the conditions of the multivariate Lindeberg-Feller CLT (see e.g. Theorem 9.3 of Hansen (2022*b*)). First note that $\mathrm{E}[X_{ni}] = 0$ because $\mathrm{E}[s_i(\theta_0)|x_i] = -\mathrm{E}[x_i w_i(y_i - \dot{a}(x_i'\theta_0))|x_i] = 0$. Moreover, we have

$$\begin{aligned}
\nu_n &= \min_{\tau \in \mathbb{R}^r\setminus\{0\}} \frac{\tau'\dot{\rho}(\theta_0)'H(\theta_0)^{-1}I(\theta_0)H(\theta_0)^{-1}\dot{\rho}(\theta_0)\tau}{\tau'\tau} \\
&\geq \min_{\tau \in \mathbb{R}^r\setminus\{0\}} \frac{\tau'\dot{\rho}(\theta_0)'H(\theta_0)^{-1}I(\theta_0)H(\theta_0)^{-1}\dot{\rho}(\theta_0)\tau}{\tau'\dot{\rho}(\theta_0)'\dot{\rho}(\theta_0)\tau} \min_{\tau \in \mathbb{R}^r\setminus\{0\}} \frac{\tau'\dot{\rho}(\theta_0)'\dot{\rho}(\theta_0)\tau}{\tau'\tau} \\
&\geq \lambda_{\min}(H(\theta_0)^{-1}I(\theta_0)H(\theta_0)^{-1})\lambda_{\min}(\dot{\rho}(\theta_0)'\dot{\rho}(\theta_0)) \\
&\geq \lambda_{\min}(H(\theta_0)^{-1})\lambda_{\min}(I(\theta_0))\lambda_{\min}(H(\theta_0)^{-1})\lambda_{\min}(\dot{\rho}(\theta_0)'\dot{\rho}(\theta_0)) \\
&= \frac{\lambda_{\min}(I(\theta_0))}{(\lambda_{\max}(H(\theta_0))^2}\lambda_{\min}(\dot{\rho}(\theta_0)'\dot{\rho}(\theta_0)) \\
&\geq \lambda_l^2/\lambda_u^2.
\end{aligned}$$

where the first inequality follows from the extremal property of $\lambda_{\min}(\cdot)$, the second inequality is the eigenvalue product inequality (Hansen (2022*a*)) and the last inequality is by Assumption 1(b). Next, we will verify the Lindeberg condition: for $\delta = \frac{2}{\delta_0} > 0$ and any $\epsilon > 0$

$$\frac{1}{\nu_n^2}\sum_{i=1}^n \mathrm{E}[\|X_{ni}\|^2 1(\|X_{ni}\| \geq (\epsilon\nu_n^2)^{1/2})] \leq \frac{1}{\nu_n^{2+\delta}\epsilon^{\delta/2}}\sum_{i=1}^n \mathrm{E}[\|X_{ni}\|^{2+\delta}] \to 0. \tag{A.11}$$

First, note that

$$\begin{aligned}
\|\dot{\rho}(\theta_0)'H(\theta_0)^{-1}x_i\|^{2+\delta} &\leq \|\dot{\rho}(\theta_0)\|^{2+\delta}\left(\|H(\theta_0)^{-1}x_i\|^2\right)^{1+\delta/2} \\
&\leq r^{1+\delta/2}\|\dot{\rho}(\theta_0)\|_2^{2+\delta}\left(\lambda_{\max}(H(\theta_0)^{-1}H(\theta_0)^{-1})\|x_i\|^2\right)^{1+\delta/2} \\
&\leq r^{1+\delta/2}\lambda_u^{2+\delta}\left(\frac{\|x_i\|^2}{(\lambda_{\min}(H(\theta_0)))^2}\right)^{1+\delta/2} \\
&\leq r^{1+\delta/2}\lambda_u^{2+\delta}\frac{(p+1)^{1+\delta/2}C_u^{2+\delta}}{\lambda_l^{2+\delta}}.
\end{aligned} \tag{A.12}$$

where the first inequality is by Cauchy-Schwarz, the second inequality is by the inequality

$\|\dot\rho(\theta_0)\| \le r^{1/2}\|\dot\rho(\theta_0)\|_2$ and the extremal property of $\lambda_{\max}(\cdot)$, the third inequality is by the eigenvalue product inequality (Hansen (2022*a*), Appendix B), and the last inequality is by Assumption 1(a) and (b). Thus, using $|w_i|^{2+\delta}|\dot g(y_i, x_i'\theta_0)|^{2+\delta} \le C_u^{4+2\delta}$ and (A.12), we have

$$\begin{aligned}\sum_{i=1}^n \|X_{ni}\|^{2+\delta} &\le \frac{1}{n^{1+\delta/2}}\sum_{i=1}^n \|\dot\rho(\theta_0)'H(\theta_0)^{-1}x_i\|^{2+\delta}|w_i|^{2+\delta}|\dot g(y_i, x_i'\theta_0)|^{2+\delta}\\ &\le \frac{1}{n^{\delta/2}} r^{1+\delta/2}\lambda_u^{2+\delta}\frac{(p+1)^{1+\delta/2}C_u^{2+\delta}}{\lambda_l^{2+\delta}}C_u^{4+2\delta}\\ &\le \left(\frac{(p+1)^{1+\delta_0}}{n}\right)^{1/\delta_0} r^{1+\delta/2}\lambda_u^{2+\delta}\frac{C_u^{6+3\delta}}{\lambda_l^{2+\delta}}\\ &\to 0.\end{aligned}$$

This verifies (A.11) and the result follows. □

Proof of Proposition 3.1 By the mean value expansion,

$$S(\theta_0) = S(\hat\theta) + \hat H(\theta^*)(\hat\theta - \theta_0) = S(\hat\theta) + \hat H(\hat\theta)(\hat\theta - \theta_0) + R, \tag{A.13}$$

where θ^* is the mean-value between $\hat\theta$ and θ_0, and $R = [R_1, \ldots, R_{p+1}]'$ with

$$R_j \equiv n^{-1}\sum_{i=1}^n (\ddot g(y_i, x_i'\theta^*) - \ddot g(y_i, x_i'\hat\theta))w_i x_{ij} x_i'(\theta_0 - \hat\theta). \tag{A.14}$$

Note that since $\dot\rho(\theta)$ is locally Lipschitz in a neighborhood of θ_0, with probability approaching 1 $\|\dot\rho(\bar\theta) - \dot\rho(\hat\theta)\| \le B_0\|\bar\theta - \hat\theta\|$ for some $B_0 = O(1)$. Also, since

$$n^{1/2}(\rho(\hat\theta) - \rho(\theta_0)) = \dot\rho(\bar\theta)' n^{1/2}(\hat\theta - \theta_0), \tag{A.15}$$

where $\bar\theta$ is a mean-value between $\hat\theta$ and θ_0, we have

$$\begin{aligned} n^{1/2}\|\dot\rho(\hat\theta) - \dot\rho(\bar\theta)\|\|\hat\theta - \theta_0\| &= n^{1/2}B_0\|\hat\theta - \bar\theta\|\|\hat\theta - \theta_0\| = O_p(n^{1/2}m_0\lambda^2)\\ &= o_p(1), \end{aligned} \tag{A.16}$$

where the last line is by $n^{1/2}m_0\lambda^2 = n^{-1/2}m_0C^2\log p \le C^2 m_0(p/n)^{1/2}\log p = o(1)$. Then,

$$\begin{aligned} &n^{1/2}(\tilde\rho - \rho(\theta_0))\\ &= n^{1/2}(\rho(\hat\theta) - \rho(\theta_0)) + \dot\rho(\hat\theta)'\hat H(\hat\theta)^{-1}n^{1/2}S(\hat\theta) \end{aligned}$$

$$
\begin{aligned}
&= n^{1/2}\dot{\rho}(\bar{\theta})'(\hat{\theta}-\theta_0) + n^{1/2}\dot{\rho}(\hat{\theta})'\hat{H}(\hat{\theta})^{-1}S(\theta_0) - n^{1/2}\dot{\rho}(\hat{\theta})'(\hat{\theta}-\theta_0) - n^{1/2}\dot{\rho}(\hat{\theta})'\hat{H}(\hat{\theta})^{-1}R \\
&= n^{1/2}\dot{\rho}(\hat{\theta})'\hat{H}(\hat{\theta})^{-1}S(\theta_0) - n^{1/2}\dot{\rho}(\hat{\theta})'\hat{H}(\hat{\theta})^{-1}R + o_p(1),
\end{aligned}
$$

where the first equality is by the definition of $\tilde{\rho}$, the second equality is by (A.13) and (A.15), and the third is by (A.16). Below, the proof will be completed in three steps: the first two steps establish

$$
\begin{aligned}
&\dot{\rho}(\hat{\theta})'\hat{H}(\hat{\theta})^{-1}\hat{I}(\hat{\theta})\hat{H}(\hat{\theta})^{-1}\dot{\rho}(\hat{\theta}) - \dot{\rho}(\theta_0)'H(\theta_0)^{-1}I(\theta_0)H(\theta_0)^{-1}\dot{\rho}(\theta_0) = o_p(1), \\
&n^{1/2}\dot{\rho}(\hat{\theta})'\hat{H}(\hat{\theta})^{-1}S(\theta_0) - n^{1/2}\dot{\rho}(\theta_0)'H(\theta_0)^{-1}S(\theta_0) = o_p(1),
\end{aligned}
\tag{A.17}
$$

and the third step verifies $n^{1/2}\dot{\rho}(\hat{\theta})'\hat{H}(\hat{\theta})^{-1}R = o_p(1)$. It will then follow that

$$
\begin{aligned}
&\left[\dot{\rho}(\hat{\theta})'\hat{H}(\hat{\theta})^{-1}\hat{I}(\hat{\theta})\hat{H}(\hat{\theta})^{-1}\dot{\rho}(\hat{\theta})\right]^{-1/2} n^{1/2}(\tilde{\rho}-\rho(\theta_0)) \\
&= \left[\dot{\rho}(\hat{\theta})'\hat{H}(\hat{\theta})^{-1}\hat{I}(\hat{\theta})\hat{H}(\hat{\theta})^{-1}\dot{\rho}(\hat{\theta})\right]^{-1/2} \left[n^{1/2}\dot{\rho}(\hat{\theta})'\hat{H}(\hat{\theta})^{-1}S(\theta_0) - n^{1/2}\dot{\rho}(\hat{\theta})'\hat{H}(\hat{\theta})^{-1}R + o_p(1)\right] \\
&= \left[\dot{\rho}(\theta_0)'H(\theta_0)^{-1}I(\theta_0)H(\theta_0)^{-1}\dot{\rho}(\theta_0)\right]^{-1/2} n^{1/2}\dot{\rho}(\theta_0)'H(\theta_0)^{-1}S(\theta_0) + o_p(1).
\end{aligned}
$$

Finally, applying Lemma A.1 of and Slutsky's lemma give the desired result.

Step 1: $\dot{\rho}(\hat{\theta})'\hat{H}(\hat{\theta})^{-1}\hat{I}(\hat{\theta})\hat{H}(\hat{\theta})^{-1}\dot{\rho}(\hat{\theta}) - \dot{\rho}(\theta_0)'H(\theta_0)^{-1}I(\theta_0)H(\theta_0)^{-1}\dot{\rho}(\theta_0) = o_p(1)$.
First, by the triangle inequality

$$
\begin{aligned}
&\|\dot{\rho}(\hat{\theta})'\hat{H}(\hat{\theta})^{-1}\hat{I}(\hat{\theta})\hat{H}(\hat{\theta})^{-1}\dot{\rho}(\hat{\theta}) - \dot{\rho}(\theta_0)'H(\theta_0)^{-1}I(\theta_0)H(\theta_0)^{-1}\dot{\rho}(\theta_0)\|_2 \\
&\le \|\dot{\rho}(\hat{\theta})'\left[\hat{H}(\hat{\theta})^{-1}\hat{I}(\hat{\theta})\hat{H}(\hat{\theta})^{-1} - H(\theta_0)^{-1}I(\theta_0)H(\theta_0)^{-1}\right]\dot{\rho}(\hat{\theta})\|_2 \\
&\quad + \|\dot{\rho}(\hat{\theta})'H(\theta_0)^{-1}I(\theta_0)H(\theta_0)^{-1}(\dot{\rho}(\hat{\theta}) - \dot{\rho}(\theta_0))\|_2 \\
&\quad + \|(\dot{\rho}(\hat{\theta}) - \dot{\rho}(\theta_0))'H(\theta_0)^{-1}I(\theta_0)H(\theta_0)^{-1}\dot{\rho}(\theta_0)\|_2.
\end{aligned}
\tag{A.18}
$$

Consider the first term on the right-hand side of (A.18). By Cauchy-Schwarz inequality,

$$
\begin{aligned}
&\|\dot{\rho}(\hat{\theta})'\left[\hat{H}(\hat{\theta})^{-1}\hat{I}(\hat{\theta})\hat{H}(\hat{\theta})^{-1} - H(\theta_0)^{-1}I(\theta_0)H(\theta_0)^{-1}\right]\dot{\rho}(\hat{\theta})\|_2 \\
&\le \|\hat{H}(\hat{\theta})^{-1}\hat{I}(\hat{\theta})\hat{H}(\hat{\theta})^{-1} - H(\theta_0)^{-1}I(\theta_0)H(\theta_0)^{-1}\|_2\|\dot{\rho}(\hat{\theta})\|_2^2,
\end{aligned}
\tag{A.19}
$$

After rearranging and using the triangle and Cauchy-Schwarz inequalities

$$
\|\hat{H}(\hat{\theta})^{-1}\hat{I}(\hat{\theta})\hat{H}(\hat{\theta})^{-1} - H(\theta_0)^{-1}I(\theta_0)H(\theta_0)^{-1}\|_2
$$

$$
\begin{aligned}
&= \|(\hat{H}(\hat{\theta})^{-1} - H(\theta_0)^{-1})\hat{I}(\hat{\theta})\hat{H}(\hat{\theta})^{-1} + H(\theta_0)^{-1}(\hat{I}(\hat{\theta})\hat{H}(\hat{\theta})^{-1} - I(\theta_0)H(\theta_0)^{-1})\|_2,\\
&\leq \|\hat{H}(\hat{\theta})^{-1} - H(\theta_0)^{-1}\|_2\|\hat{I}(\hat{\theta})\|_2\|\hat{H}(\hat{\theta})^{-1}\|_2 + \|H(\theta_0)^{-1}\|_2\|\hat{I}(\hat{\theta})\hat{H}(\hat{\theta})^{-1} - I(\theta_0)H(\theta_0)^{-1}\|_2.
\end{aligned} \tag{A.20}
$$

For the first summand of (A.20), by Lemma A.2 of Jasiak and Tuvaandorj (2023)

$$
\|\hat{H}(\hat{\theta})^{-1} - H(\theta_0)^{-1}\|_2\|\hat{I}(\hat{\theta})\|_2\|\hat{H}(\hat{\theta})^{-1}\|_2 = o_p(1). \tag{A.21}
$$

For the second factor in the second summand of (A.20), using the triangle and Cauchy-Schwarz inequalities

$$
\begin{aligned}
&\|\hat{I}(\hat{\theta})\hat{H}(\hat{\theta})^{-1} - I(\theta_0)H(\theta_0)^{-1}\|_2\\
&= \|(\hat{I}(\hat{\theta}) - I(\theta_0))(\hat{H}(\hat{\theta})^{-1} - H(\theta_0)^{-1}) + (\hat{I}(\theta_0) - I(\theta_0))H(\theta_0)^{-1} + I(\theta_0)(\hat{H}(\theta_0)^{-1} - H(\theta_0)^{-1})\|_2\\
&\leq \|\hat{I}(\hat{\theta}) - I(\theta_0)\|_2\|\hat{H}(\hat{\theta})^{-1} - H(\theta_0)^{-1}\|_2 + \|\hat{I}(\theta_0) - I(\theta_0)\|_2\|H(\theta_0)^{-1}\|_2\\
&\quad + \|I(\theta_0)\|_2\|\hat{H}(\theta_0)^{-1} - H(\theta_0)^{-1}\|_2\\
&\xrightarrow{p} 0,
\end{aligned} \tag{A.22}
$$

where the last line is by Lemma A.2 of Jasiak and Tuvaandorj (2023) and the CMT. From Lemma B.3 of Jasiak and Tuvaandorj (2023), $\|\hat{\theta} - \theta_0\| = O_p(m_0^{1/2}\lambda) = O_p\left(\left(\frac{m_0 \log p}{n}\right)^{1/2}\right) = o_p(1)$. Since $\dot{\rho}(\theta)$ is locally Lipschitz in a neighborhood of θ_0, with probability approaching 1, we have for $B_0 = O(1)$ $\|\dot{\rho}(\hat{\theta}) - \dot{\rho}(\theta_0)\| \leq B_0\|\hat{\theta} - \theta_0\|$. Thus,

$$
\|\dot{\rho}(\hat{\theta}) - \dot{\rho}(\theta_0)\|_2 \leq r^{1/2}\|\dot{\rho}(\hat{\theta}) - \dot{\rho}(\theta_0)\| = O_p\left(\left(\frac{m_0 \log p}{n}\right)^{1/2}\right). \tag{A.23}
$$

By the triangle inequality and (A.23)

$$
\|\dot{\rho}(\hat{\theta})\|_2 \leq \|\dot{\rho}(\hat{\theta}) - \dot{\rho}(\theta_0)\|_2 + \|\dot{\rho}(\theta_0)\|_2 = O_p(1). \tag{A.24}
$$

Therefore, the quantity in (A.19) is $o_p(1)$. Consider the second term on the right-hand side of (A.18). By the triangle inequality and (A.23),

$$
\begin{aligned}
&\|\dot{\rho}(\hat{\theta})'H(\theta_0)^{-1}I(\theta_0)H(\theta_0)^{-1}(\dot{\rho}(\hat{\theta}) - \dot{\rho}(\theta_0))\|_2\\
&\leq \|\dot{\rho}(\hat{\theta})\|_2\|H(\theta_0)^{-1}I(\theta_0)H(\theta_0)^{-1}\|_2\|\dot{\rho}(\hat{\theta}) - \dot{\rho}(\theta_0)\|_2\\
&\xrightarrow{p} 0.
\end{aligned}
$$

Similarly, for the third term on the right-hand side of (A.18)

$$\begin{aligned}
&\|(\dot\rho(\hat\theta)-\dot\rho(\theta_0))'H(\theta_0)^{-1}I(\theta_0)H(\theta_0)^{-1}\dot\rho(\theta_0)\|_2\\
&\le \|\dot\rho(\hat\theta)-\dot\rho(\theta_0)\|_2\|H(\theta_0)^{-1}I(\theta_0)H(\theta_0)^{-1}\|_2\|\dot\rho(\theta_0)\|_2\\
&\xrightarrow{p} 0.
\end{aligned}$$

Step 2: $n^{1/2}\dot\rho(\hat\theta)'\hat H(\hat\theta)^{-1}S(\theta_0)-n^{1/2}\dot\rho(\theta_0)'H(\theta_0)^{-1}S(\theta_0)=o_p(1)$**.**
Remark that from Assumption 1, $|\dot g(y_i,x_i'\theta_0)|\le C_u$, $|w_i|\le C_u$ and $\|x_i\|^2\le(p+1)C_u^2$ a.s. for all i. Using the independence assumption,

$$\mathrm{E}[\|S(\theta_0)\|_2^2]=\mathrm{E}[\|S(\theta_0)\|^2]=n^{-2}\,\mathrm{E}\left[\sum_{i=1}^n w_i^2\|x_i\|^2\dot g(y_i,x_i'\theta_0)^2\right]\le n^{-1}(p+1)C_u^6.$$

By Markov's inequality,

$$\|S(\theta_0)\|_2=O_p\left(\sqrt{\frac{p}{n}}\right). \tag{A.25}$$

Now rewrite

$$\begin{aligned}
&n^{1/2}\dot\rho(\hat\theta)'\hat H(\hat\theta)^{-1}S(\theta_0)-n^{1/2}\dot\rho(\theta_0)'H(\theta_0)^{-1}S(\theta_0)\\
&\quad=n^{1/2}(\dot\rho(\hat\theta)-\dot\rho(\theta_0))'\hat H(\hat\theta)^{-1}S(\theta_0)+n^{1/2}\left(\dot\rho(\theta_0)'\hat H(\hat\theta)^{-1}S(\theta_0)-\dot\rho(\theta_0)'H(\theta_0)^{-1}S(\theta_0)\right).
\end{aligned}\tag{A.26}$$

For the first term of (A.26),

$$\begin{aligned}
\|n^{1/2}(\dot\rho(\hat\theta)-\dot\rho(\theta_0))'\hat H(\hat\theta)^{-1}S(\theta_0)\|_2 &\le n^{1/2}\|\dot\rho(\hat\theta)-\dot\rho(\theta_0)\|_2\|\hat H(\hat\theta)^{-1}\|_2\|S(\theta_0)\|_2\\
&=n^{1/2}O_p\left(\sqrt{\frac{m_0\log p}{n}}\right)O_p(1)\,O_p\left(\sqrt{\frac{p}{n}}\right)\\
&=O_p\left(\sqrt{\frac{p\,m_0\log p}{n}}\right)\\
&=o_p(1),
\end{aligned}\tag{A.27}$$

where the first inequality is by Cauchy-Schwarz, the first equality uses

$$\|\hat H(\hat\theta)^{-1}\|_2=O_p(1).$$

as shown in the proof Lemma 3.5 of Jasiak and Tuvaandorj (2023), (A.23) and (A.25), and the last equality holds because $m_0(\log p)p/n\le m_0(\log p)(p/n)^{1/2}(p^2/n)^{1/2}\to 0$ by the assumption of the proposition. For the second term of (A.26), we have

$$\begin{aligned}
n^{1/2}\|\dot{\rho}(\theta_0)'\hat{H}(\hat{\theta})^{-1}S(\theta_0)-\dot{\rho}(\theta_0)'H(\theta_0)^{-1}S(\theta_0)\|_2 &\leq n^{1/2}\|\dot{\rho}(\theta_0)\|_2\|\hat{H}(\hat{\theta})^{-1}-H(\theta_0)^{-1}\|_2\|S(\theta_0)\|_2\\
&= n^{1/2}O_p\left(\sqrt{\frac{p}{n}}+m_0\lambda\right)O_p\left(\sqrt{\frac{p}{n}}\right)\\
&= O_p\left(\sqrt{\frac{p^2}{n}}+\sqrt{p}\,m_0\lambda\right)\\
&= o_p(1), \qquad \text{(A.28)}
\end{aligned}$$

where the first inequality is by Cauchy-Schwarz, the first equality is by by Lemma A.2 of Jasiak and Tuvaandorj (2023) and (A.25), and the last equality holds because $p^2/n \to 0$ and $p^{1/2}\,m_0\lambda = Cm_0(p/n)^{1/2}(\log p)^{1/2} \leq Cm_0(p/n)^{1/2}2\log p \to 0$ by the assumption of the proposition. It follows from (A.26), (A.27) and (A.28) that

$$n^{1/2}\dot{\rho}(\hat{\theta})'\hat{H}(\hat{\theta})^{-1}S(\theta_0)-n^{1/2}\dot{\rho}(\theta_0)'H(\theta_0)^{-1}S(\theta_0)=o_p(1).$$

Step 3: $n^{1/2}\dot{\rho}(\hat{\theta})'\hat{H}(\hat{\theta})^{-1}R=o_p(1)$**.**

By Cauchy-Schwarz, $n^{1/2}\|\dot{\rho}(\hat{\theta})'\hat{H}(\hat{\theta})^{-1}R\|_2 \leq n^{1/2}\|\dot{\rho}(\hat{\theta})\|_2\|\hat{H}(\hat{\theta})^{-1}R\|_2$. Remark from (A.24) that $\|\dot{\rho}(\hat{\theta})\|_2=O_p(1)$. To show $n^{1/2}\|\hat{H}(\hat{\theta})^{-1}R\|_2=o_p(1)$, note that

$$\begin{aligned}
\max_{1\leq j\leq p+1}|R_j| &\leq n^{-1}\sum_{i=1}^{n}|\ddot{g}(y_i,x_i'\theta^*)-\ddot{g}(y_i,x_i'\hat{\theta})||w_i|\max_{1\leq j\leq p+1}|x_{ij}||x_i'(\theta_0-\hat{\theta})|\\
&\leq n^{-1}\sum_{i=1}^{n}L_0|x_i(\theta^*-\hat{\theta})|C_u^2|x_i'(\theta_0-\hat{\theta})|\\
&\leq L_0C_u^2n^{-1}\sum_{i=1}^{n}|x_i'(\theta_0-\hat{\theta})|^2\\
&= L_0C_u^2O_p(m_0\lambda^2)\\
&= O_p(m_0\lambda^2), \qquad \text{(A.29)}
\end{aligned}$$

where the first inequality is by Assumption 1(c), and the first equality uses Lemma B.3 of Jasiak and Tuvaandorj (2023). Since $\|H(\theta_0)\|=O(1)$ and $\|\hat{H}(\hat{\theta})-H(\theta_0)\|=o_p(1)$, $\|\hat{H}(\hat{\theta})\|=O_p(1)$. Therefore,

$$\begin{aligned}
n^{1/2}\|\hat{H}(\hat{\theta})^{-1}R\|_2 &\leq n^{1/2}\|\hat{H}(\hat{\theta})^{-1}\|_2\|R\|_2\\
&\leq n^{1/2}\hat{H}(\hat{\theta})^{-1}(p+1)^{1/2}\|R\|_\infty
\end{aligned}$$

$$
\begin{aligned}
&= O_p((n(p+1))^{1/2} m_0 \lambda^2) \\
&= o_p(1), \qquad \text{(A.30)}
\end{aligned}
$$

where the first equality holds by using (A.29) and the second equality follows on noting that $(n(p+1))^{1/2} m_0 \lambda^2 = (n(p+1))^{1/2} m_0 C^2 (\log p)/n \leq (2p/n)^{1/2} m_0 C^2 \log p = o(1)$.

B Survey Questions Used for Variable Construction

This appendix provides the survey questions used to construct the Cyber Score, the k-means clustering variables, and the Business Digital Usage Score (BDUS). Each set of questions corresponds to specific aspects of digital technology adoption and cyber security challenges.

B.1 Cyber Security Incidence Variable Construction

The Cyber Security Incidence variable is equal to 1 if a firm answers "Yes" to one of the following questions and 0 otherwise.

> *To the best of your knowledge, which cyber security incidents impacted your business in 2021? Select all that apply.*

- Incidents to disrupt or deface the business or web presence.
- Incidents to steal personal or financial information.
- Incidents to steal money or demand ransom payment.
- Incidents to steal or manipulate intellectual property or business data.
- Incidents to access unauthorised or privileged areas.
- Incidents to monitor and track business activity.
- Incidents with an unknown motive.

B.2 Questions Used for k-means Clustering

The k-means clustering variables were derived from responses to the following survey questions, which identify challenges businesses face in utilizing various digital and financial technologies. Each affirmative response indicates an inefficiency or challenge:

- Does your business face challenges with online transaction processing?
- Does your business face challenges with digital marketing?
- Does your business face challenges with data analytics?
- Does your business face challenges with integrating digital technologies into business operations?
- Does your business face challenges with big data?
- Does your business face challenges with artificial intelligence?
- Does your business face challenges with cloud computing?
- Does your business face challenges with ICT infrastructure?
- Does your business face challenges with government connectivity?
- Does your business face challenges with website operations?

B.3 Questions Used for Business Digital Usage Score (BDUS)

The BDUS is based on responses to questions about the adoption of specific digital technologies. A “Yes” response to any of the following indicates that the business has utilized the respective technology:

- Does your business use online transaction processing systems?
- Does your business use digital marketing platforms?
- Does your business use data analytics tools?
- Does your business integrate digital technologies into business operations?
- Does your business utilize big data technologies?
- Does your business employ artificial intelligence tools?
- Does your business use cloud computing services?
- Does your business maintain ICT infrastructure?

- Does your business interact with government systems digitally?
- Does your business operate its own website?

The BDUS score is computed as the total number of “Yes” responses to these questions, with higher scores indicating greater digital engagement.