Term Paper Proposal
The Impact of Perceived Internet Security on Online Shopping Behaviours

Jiarui Cai, Jing Yang, Qinghui Yu

1 Introduction

1.1 Background

Thanks to a leap in the number of Internet-enabled devices and an improved logistic network, we are seeing rapid growth in the e-commerce business. Reported by Statistics Canada, Canadian companies sold more than \$136 billion in goods and services online in 2013, 11% up from \$122 billion of 2012 [1]. That is for a good reason - e-commerce is an efficient way for retailers to sell their goods. By eliminating rent, profit margin is higher and consumer surplus is boosted with the lower prices. Despite the rosy picture, some refuse to shop online. One prominent reason is, as pointed out by Statistics Canada, one-fifth (19%) of Canadian refuses to shop online due to security concerns [2]. Interestingly, in Miyazaki and Fernandez's work, they pointed out that the perceived level of risk by consumers has bears no correlation with online retailers' privacy and security reality [3]. This means that with respect to security, consumer's purchase decision relies not on actual risk, but on perceived ones.

1.2 Research Question

As seen from section 1.1, perception plays a large role on individuals' online shopping decision. Our research concerns the magnitude of such effect, specifically, it explores the marginal effects of various security perception proxies on the tendency for an individual to shop online. We hope that the study could shed some lights on policy and educational decisions undertaken by governments and institutions. If the effect is negative and large, there might be social benefits to tear down the insecurity façade put up by individuals. On the other hand, if the effect is negligible, resources might be dedicated elsewhere where they can be more productive. Results from the study would also be valuable for online retailers. If there is an observable effect, they can plan marketing strategies accordingly to attract new customers and gain competitive advantages.

2   Data

The data used in this study is taken from The Canadian Internet Use Survey (CIUS) in the form of a microdata file. The survey totals 22615 samples, with 132 variables each. Some important points to note is that the public use microdata is not *raw*, meaning that it had been edited by the Statistics Canada to eliminate any integrity loopholes. Confidential information such as income and age was suppressed to protect privacy of the respondents and was replaced with derived variable like income quintile and age groups.

3   Model Construction and Hypothesis

Because we want to study the tendency for an individual to shop online, the dependent variable of choice is the **number of orders made online over the past year** (*NumOrd*). The independent variable of interest is consumer's level of Internet security concern. We will use three binary variables as proxy for perception of security, namely **concerned about using credit card over the Internet** (sc_1), **experienced misuse of personal information on Internet** (sc_2) and **experienced a computer virus that result in loss of information or damage of software** (sc_3). As 93% of Canadian online shoppers pay for purchases with credit cards (Frederick, 2016), concern over the security of credit card payments will hold Canadians back from buying more online (Grau, 2008). Additionally, online shopping sites require other sensitive information such as email, phone number and address. Research have shown that concern for information privacy affects risk perceptions, trust, and willingness to transact (Van Slyke et al., 2006). Furthermore, we believe that if a consumer has experienced a computer virus, he will be more cautious of any website's security. Therefore, these binary variables are highly correlated with consumer's level of security concern when shopping online and serve as good proxy variables for perception of security. A starting point would be

$$NumOrd = \beta_0 + \beta_1 sc_1 + \beta_2 sc_2 + \beta_3 sc_3 + u \tag{1}$$

But this would be quite a bad model, because there are other factors that obviously affect *NumOrd*. Hence, we have identified the following variables to be considered:

- **Home Internet access** (*homea*): As home access allows consumers to browse and shop online at home, in their free time, we anticipate that those having home Internet access are likely to browse more frequently and thus make more purchases online.

- **Years of Internet usage** (*year*): People with a longer history of Internet usage is predicted to have more Internet knowledge. This knowledge is positively correlated with the number of Internet purchase (Case et al., 2001).

- **Characteristic of community** (*urban*): We predict that people who live near an urban center such as Toronto, Montréal and Vancouver would be placing less orders online compared to their rural counterparts because it is harder for rural household to order selected goods in physical stores (McKeown and Brocca, 2009).

- **Weekly average hours on Internet for personal use** (*hrs*): We believe this has a positive effect on number of online orders.

- **Age** (*age*): Children and senior people are predicted to shop less online, while people between these two groups generally consume more (McKeown and Brocca, 2009).

- **Gender** (*gender*): Male consumers make more online purchases and spend more money online than females (Zhou et al., 2007).

- **Income** (*income*): A high income will increase online shopping tendency (Zhou et al., 2007), and thus increases the number of online orders.

- **Education** (*coll, univ*): Education is included to control for omitted variable bias as education has a positive effect on number of online purchases and the better-educated respondents is predicted to be swayed less by perception of Internet security (Hui and Wan, 2007).

homea and *gender* are binary variables while education level is dummy coded into two binary variables, namely trade college (coll) and university degree (uni) (high school or less is used as the base group). The rest are categorical variables, each with multiple numerical ranges (see appendix). The mean of each range is assigned to the observations within that range. For ranges with only upper or lower bound, we will take the upper or lower bound respectively. This allows for interpretation of the beta estimates of the these variables as marginal effects of incremental move to the next range.

In light of the above, we have came up with the regression

$$NumOrd = \beta_0 + \beta_1 sc_1 + \beta_2 sc_2 + \beta_3 sc_3 + \beta_4 homea + \beta_5 year + \beta_6 urban$$
$$+ \beta_7 hours + \beta_8 age + \beta_9 gender + \beta_{10} log(income) + \beta_{11} coll + \beta_{12} univ + u \tag{2}$$

A potential term in the error that is likely to be correlated with the regressors and thus a source of endogeneity is an individual's marginal propensity to consume (MPC). MPC is positively correlated with *income* and *NumOrd*. MPC may also be correlated with *age*, *gender* and education as they affect consumption patterns and preferences. Quality of previous online shopping experience is another possible endogeneity. This has a positive effect on both *NumOrd* and concern of online credit card payment (sc_1), so omitted variable bias may be of issue here. As a result, the beta estimates of the model may be biased upwards. Therefore, it is necessary to test for endogeneity using the Hausman Test.

Functional misspecification and heteroskedasticity could happen as well. There is no guarantee that our model would be linear in its regressors. It is also naïve to assume that the variance of error term is constant

for all values in the input space. We have to test for both using RESET test and White test respectively. If any problem of functional form were to arise, we can introduce some non-linearities or take log form of the dependent variable. If heteroskedasticity is present, we can use robust error or generalized least squares to preserve goodness-of-fit of our model.

4   Existing Literature

Internet Privacy and Security: An Examination of Online Retailer Disclosures (Miyazaki and Fernandez, 2000) examines the effect disclosures of privacy and security practices by online retailers on consumer's perceptions of risk and purchase intentions. Contrary to the expectation that the prevalence of online privacy and security statements are negatively correlated with risk perceptions, it showed no effect on neither privacy nor security perceptions. However, the percentage of privacy and security statements is positively related to online purchase tendencies. **Consumer Perceptions of Privacy and Security Risks for Online Shopping** (Miyazaki and Fernandez, 2001) explores risk perceptions among consumers of varying levels of Internet experience and how these perceptions relate to online shopping activity. By applying multiple regression analysis, as the researchers expected, perceived risk and concerns toward online shopping is negatively related to online purchase rates, but this impact is relatively insignificant. **Consumer patronage and risk perceptions in Internet shopping** (Forsythe and Shi, 2003) confirms these findings. Forsythe and Shi discusses the types of risks perceived by Internet shoppers and browsers and the relationship between those risks and online patronage. Results suggests that although Internet shoppers perceive several risks in Internet shopping, these perceived risks may not significantly influence Internet patronage behaviors in an extensive and systematic way.

References

[1] Digital technology and Internet use, 2013

   http://www.statcan.gc.ca/daily-quotidien/140611/dq140611a-eng.htm

[2] Individual Internet use and E-commerce

   http://www.statcan.gc.ca/daily-quotidien/111012/dq111012a-eng.htm

[3] Anthony D. Miyazaki and Ana Fernandez *Internet Privacy and Security: An Examination of Online Retailer Disclosures.* Journal of Public Policy & Marketing Vol.19, No.1, 2000.

[4] Jaz Frederick. *2016 Canada Ecommerce Market.*

   Retrieved Jan 25, 2017 from http://www.pfsweb.com/blog/2016-canada-ecommerce-market/

[5] Grau, J. *Canada B2C E-Commerce: A Work in Progress.* New York: eMarketer. 2008.

[6] C. Van Slyke, J. Shim, R. Johnson, J. Jiang *Concern for Information Privacy and Online Consumer Purchasing.* Journal of the Association for Information Systems, 7 (6), pp. 415–444, 2006.

[7] Thomas Case, O. Maxie Burns, Geoffrey Dick *Drivers of Online Purchasing Among U.S. University Students.* Association for Information Systems, 2001.

[8] Lawrence McKeown, Josie Brocca *Internet shopping in Canada: An examination of data, trends and patterns.* Statistics Canada – Catalogue no. 88F0006X, no. 5, 2009.

[9] Lina Zhou, Liwei Dai, Dongsong Zhang *Online Shopping Acceptance Model — A Critical Survey of Consumer Factors in Online Shopping.* Journal of Electronic Commerce Research, VOL 8, NO.1, 2007.

[10] Tak-Kee Hui and David Wan *Factors affecting Internet shopping behaviour in Singapore: gender and educational issues.* International Journal of Consumer Studies, Volume 31(Issue 3), Pp.310-316, 2007

[11] Anthony D. Miyazaki and Ana Fernandez *Consumer Perceptions of Privacy and Security Risks for Online Shopping.* Journal of Consumer Affairs, Vol.35, No.1, 2001.

[12] Sandra M Forsythe and Bo Shi *Consumer patronage and risk perceptions in Internet shopping.* Journal of Business Research, Volume 56, Issue 11, 2003.

5 Appendix

5.1 Summary Statistics

- Note that the microdata survey file has categorical answers to each question, for example: Skip for *hrs*. The usual summary statistics like **mean** and **variance** are not meaningful in this situation. So what we have included is a categorical distribution of sample data for each variable.

Distribution	Sample count	Sample %	Pop. count	Pop. %
Age				
16 - 24	1,814	8.0	4,070,691	14.5
25 - 34	3,198	14.1	4,737,890	16.9
35 - 44	3,520	15.6	4,579,161	16.3
45 - 54	4,007	17.8	5,260,552	18.7
55 - 64	4,420	19.5	4,453,169	15.9
65+	5,656	25.0	4,955,537	17.7
Gender				
male	10,135	44.8	13,821,744	49.3
female	12,480	55.2	14,235,256	50.7
Income				
≤ 25000	4,764	21.0	4,046,202	14.4
25000 - 45000	4,823	21.3	5,026,763	17.9
45000 - 70000	4,598	20.3	5,744,897	20.5
70000 - 100000	4,271	19.0	6,190,810	22.1
≥ 100000	4,159	18.4	7,048,327	25.1
coll	9,371	41.4	10,983,610	39.1
univ	4,504	20.0	6,646,198	23.7
homea				
Yes	4,438	19.6	6,741,058	24.0
No	5,644	25.0	7,471,812	26.6
Valid skip	11,827	52.3	12,965,443	46.2
Don't know	285	1.3	344,612	1.2
Refusal	2	0.0	1,685	0.0
Not stated	419	1.8	532,390	1.9

	Sample count	Sample %	Pop. count	Pop. %
year				
< 1	379	1.7	475,710	1.7
1-2	604	2.7	650,181	2.3
2-5	2,147	9.5	2,466,608	8.8
5-10	4,967	22.0	6,763,283	24.1
> 10	9,450	41.8	12,957,236	46.2
Skip	5,005	22.0	4,652,779	16.6
Don't know	46	0.2	70,660	0.3
Refusal	2	0.0	3,516	0.0
Not stated	15	0.1	17,028	0.1
hrs				
<5	7,921	35.0	9,808,900	35.0
5-10	4,670	20.7	6,173,017	22.0
10-20	2,983	13.2	4,249,900	15.1
20-30	1,114	5.0	1,752,417	6.2
30-40	386	1.7	649,461	2.3
>40	409	1.8	640,303	2.3
Valid skip	5,005	22.1	4,652,779	16.6
Don't know	94	0.4	94,488	0.3
Refusal	4	0.0	3,597	0.0
Not stated	29	0.1	32,139	0.1
Total	22,615	100	28,057,000	100