A Model of Household Online Buying

by

Meyyappan Narayanan

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Management Sciences

Waterloo, Ontario, Canada, 2006

AUTHOR’S DECLARATION FOR ELECTRONIC SUBMISSION OF A THESIS

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Meyyappan Narayanan

ABSTRACT

A Model of Household Online Buying

The Internet has made profound changes in how people conduct their daily lives as well as how they buy goods and services. This study's objective is to shed light on the use and diffusion of online or electronic buying (e-buying). Canadian households have not adopted e-buying equally, as revealed by Statistics Canada's Household Internet Use Survey (HIUS) data of 1997 – 2003. We explore how e-buying varies across age groups, genders, education levels, income levels, and the nature of goods. We first develop a simple model for e-buying demand in the context of a utility-maximizing individual choosing between e-buying and conventional buying. We employ a parameter reflecting individual taste, so we can study the influence of individual-specific factors in e-buying adoption decisions. The taste parameter is distributed in a population in some unknown way, and we try different distributions in empirical tests. We use the literature in conjunction with the model to derive the model's implications in terms of variables available in the HIUS datasets. We employ Tobit and Poisson regression models for the empirical tests. The tests suggest that household e-buying is more when household income is more, when heads of households are more educated, and for homogeneous goods; but that household e-buying is less when heads of households are female. This understanding may help policy makers, businesses, and other interested parties find ways to promote Internet use and e-buying across all segments of society.

ACKNOWLEDGEMENTS

I express my heartfelt gratitude to my supervisor, Professor Brian Cozzarin, for his kindness, valuable advice, and encouragement throughout the course of this research. I also wish to express my sincere thanks to Professor Bonwoo Koo and Professor Scott A. Jeffrey, the readers of the thesis, for their valuable comments for the thesis.

I thank Ms. Carol Perry and Ms. A. Michelle Edwards, Ph.D., of University of Guelph for their timely help with the datasets of the Tri-University Data Resources (TDR).

I sincerely acknowledge the help and support of my friends Vadivananthan Visuvalingam and Sachin Jayaswal throughout the research.

I remain forever indebted to my parents Gandhimathi and S. M. Narayanan for instilling in me the value of education and endeavor. I thank my wife Seetha for her continued support.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 1
Introduction

1.1 Background to the Research

The Internet has enabled profound changes in how people conduct their daily lives as well as how they buy goods and services. The Internet brings a world of information to homes, with a few clicks. The main characteristic of the Internet is the almost instantaneous communication of information that fundamentally has changed how our socio-economic world operates. The Internet's main uses have been interpersonal communication through electronic mail (e-mail), online chatting, entertainment, world-wide access to information, online banking, online learning, working from home, electronic commerce (e-commerce, e-com), and electronic governance (e-governance). The Internet's diffusion has been rapid in the last ten years, especially in the developed world. For instance, in Canada the percentage of households with a computer at home increased from 45% in 1999 to 65% in 2003, and the percentage of households with any kind of connection increased from 29% to 54%. The percentage of households using e-mail increased from 26% to 52% and households with a high-speed connection increased from 3% to 34% over the same period.

The growth in online buying or electronic buying (e-buying) and online payment or electronic payment (e-paying) has also been noteworthy. The percentage of Canadian households with connections that e-buy (at least once in the past year) increased from 8% in 1999 to 29% in 2003 and those who e-buy and e-pay increased from 5% to 22%. E-commerce as a whole is growing rapidly all over the world, and retail e-commerce (household e-buying) in the U.S. was expected to grow from $45 billion in 2000, or 1.5% of total retail sales, to $269 billion in 2005, or 7.8% of total retail sales. With "click and mortar" sales[1] included, total retail sales affected by e-com in 2005 are expected to be $647 billion or 18.5% of total retail sales (Bakos, 2001). The Canadian e-commerce

[1] A "click and mortar" sale refers to a sale in which the household collects information about goods online and then buys the goods in a traditional retail outlet, also known as a "brick and mortar" store.

market was estimated to have increased from C$ 5.3 billion in 1998 to C$ 80 billion in 2003 (Michalak and Jones, 2003).

This study's objective is to shed light on the use and diffusion of e-buying. Statistics Canada collected data on household Internet behaviour from 1997 – 2003 with the Household Internet Use Survey (HIUS). These data enable the study of determinants of use and diffusion of e-buying in Canadian households.

1.2 Research Questions and Research Objective

Canadian consumers have not adopted e-buying equally when presented with the technological opportunity of e-buying. Individual-specific factors may explain why some consumers e-buy while others do not. To explore issues concerning factors that influence e-buying adoption by the public, we choose as our question for this research, "how does e-buying intensity vary across age groups, gender, education levels, income levels, and the nature of goods?"

1.3 Justification for the Research

The HIUS data, which we use for empirical hypotheses tests, include data pertaining to variables such as the categories of goods ordered by the sample households; individual-specific (or household-specific) details such as age group, gender, and education level of the heads of the sample households; the income levels of the sample households; the types of connections; whether the sample households ordered any goods online in the past twelve months, and if so, the number of orders placed and their dollar value. These data enable research into the determinants and the diffusion and use and e-buying by Canadian households.

New technologies often bring benefits in the form of incremental and transformational improvements. Hence, it is often desirable that new technologies and innovations diffuse quickly in economies. Though new technologies may be beneficial, some technologies may not be accepted at all by users, and many take more time to diffuse than desired. Some diffuse fast. The technological diffusion rate and the determinants of adoption may vary from country to country. In view of these complexities, it is of interest to study various aspects of diffusion of technologies in economies; knowledge from such studies may help with diffusion of similar technologies in the future.

The research question that we have adopted will help us understand what individual-specific factors drive households to adopt Internet use and e-buying. Such understanding will help policy makers and administrators, businesses, and other interested parties find ways to promote Internet use and e-buying across all segments of society. For example, if it is found that households with an older person as the head of household e-buy less, then businesses may target such households with appropriate strategies to increase e-buying, which may be mutually beneficial.

1.4 Method

We first develop a simple theoretical model for e-buying demand in the context of a utility-maximizing individual faced with the choice of e-buying or conventional buying. In developing the model, we employ a parameter that reflects individual taste so we can study the influence of individual-specific factors in the adoption decisions of e-buying. The taste parameter is distributed in a population in some way, which we need to identify. Since the parameter's distribution is modeled in general form, we can try different distributions in empirical tests.

We then use the literature in conjunction with the model to derive the model's implications in terms of variables available in the HIUS 1999 - 2003 datasets. We then empirically test the hypotheses developed, using regression methods. We use e-buying activity measures as dependent variables. Since these are limited dependent variables, we employ Tobit and Poisson regression models.

1.5 Outline of the Thesis

Chapter 2 reviews literature pertaining to the Internet's impacts on trade (economic and social), diffusion of technologies, determinants of Internet use (socio-economic and behavioural), and e-commerce and tax. Chapter 3 deals with the theoretical and empirical models. We note the assumptions of the theoretical model and then develop it; we then derive its implications, compare them to the literature, and develop hypotheses for empirical tests. We then present the empirical model equation and briefly mention the regression methods, Tobit and Poisson. Chapter 4 describes the HIUS data and the methods for empirical tests that we use. We note some details of the SHAZAM procedures of Tobit and Poisson regressions. Chapter 5 presents the empirical results and analysis, and notes issues associated with the theoretical model and empirical tests. Chapter 6 concludes by revisiting the research objective, summarizing our approach, revisiting the results, recording the limitations of the research, and identifying some scope for future research.

Chapter 2
Literature Review

2.1 Introduction

A thorough review of the literature on the diffusion of the Internet and its applications follows. This section of the literature review opens with a review of models of diffusion of technologies, chiefly the S-curve model, followed by a review of the literature on alternative models of diffusion. At a more basic level, we review the literature on determinants of Internet use, especially socio-economic and behavioural. A review of the considerable volume of literature on determinants of applications use in organizations will help us in studying determinants of the application adoption by individuals. Taxation has been a relevant concern of researchers and policy makers in studies of adoption of the Internet and electronic commerce. Hence, we close the literature review with a discussion of research on the importance of taxation on the Internet and e-commerce adoption decisions by individuals.

2.2 Impacts of the Internet

2.2.1 Economic Impacts of the Internet on Trade

2.2.1.1 Cost Reduction

Bakos (1997) investigates how the Internet impacts the marketplace through reduced buyer search costs. The paper develops theoretical models of search costs in differentiated markets and also theoretical models of incentives for buyers, sellers, and intermediaries to invest in electronic marketplaces. The paper shows that, if search costs are low enough in electronic markets, buyers will buy products that best serve their needs, resulting in a socially optimal allocation. The paper argues further that buyers gain from rent redistribution through lower prices, in addition to the social benefits of lower search costs and better matching of buyers and sellers.

2.2.1.2 Creation of Value

Borenstein and Saloner (2001) give an elaborate account of how value is created and distributed in e-commerce, from the viewpoint of standard microeconomics. The earliest effect of the Internet on markets is incremental and rests on cost reduction. In due course, drastic shifts arise, i.e. through e-commerce new opportunities are possible, resulting in business restructuring.

The Internet creates value by drastically reducing the cost of transferring information. If the product itself is information, the possibilities for value creation are significant. When the transaction involves a tangible good with information and the cost of moving the good is low in comparison to the value of the good, the reach of the Internet is much greater than traditional alternatives. An example is a unique item that a buyer may value highly. The main effect of increased geographical market reach is the improvement in the matching of the buyer and the good, especially in inefficient markets. An example is the used consumer durable market.

Even when the cost of transportation is high, the Internet causes an increasing separation of movement of goods and information. Information from the Internet has many valuable characteristics: first, it is inexpensive; second, communication can be asynchronous—an especially valuable characteristic in the face of a globalizing economy; third, greater interactivity and search capability drastically reduce the costs of customization of service.

Value is created either on the cost side or on the demand side. On the cost side, the cost reduction in distribution can be very large. In business to consumer (B2C) commerce, in-store operations costs may be reduced, by reducing shoplifting[2], rent, selling costs, etc. In addition, online buyers enjoy effectively lower taxes. The Internet is also causing changes in the production process. In the service sector, some functions can be out-sourced to regions with a cost or time-difference advantage.

[2] The authors state that this is as high as 3 per cent of the retailer's sales.

On the demand side, improved matching of supplier and buyer occurs through improved information, better access to goods, and ease of customization. Still, issues of fit, feel, quality, real appearance of goods, etc. remain. A solution may be the hybrid store that can serve as a showroom and handle online delivery. However, the showroom concept suffers from the "free-rider" problem: the situation of customers checking goods in the showroom and then buying the goods online from another seller who sells only online. Such a seller is likely to have low overhead.

2.2.1.3 Distribution of Value

As per Borenstein and Saloner (2001), online markets seem to be marked by price dispersion and increasing market concentration. Sustainable price dispersion depends on "stickiness." Since most customers value non-financial aspects such as trustworthiness and after-sales support, they may "stick" with certain sellers. Economies of scale in shipping and handling may force them to stick to a few sellers. Finally, retailers create stickiness through website customization, design, etc.

It is not clear how much price dispersion will survive in equilibrium. By tracking online consumers' behaviour, sellers engage in price discrimination. This is actually a kind of price dispersion, but for the same product among different customers. On the other hand, technology provides applications that enable buyers to compare online prices. This ability can counter price dispersion. However, the balance of the two in equilibrium is difficult to predict. The equilibrium market structure, which is still not clear, will decide how the value created by e-commerce is distributed among competing firms, customers, and intermediaries.

Vulkan (2003) provides another detailed treatment of the creation and distribution of value in e-commerce. In "automated e-commerce," automated software known as "agent technologies" conducts product searches, price negotiations, intermediation services, etc. on behalf of buyers, sellers, and intermediaries. These programs can result in improved transaction times, a higher quality of search, a better match of buyer and good, etc.

Moreover, business can be carried out in ways previously impractical, including auctions in which people from all over the world participate. The extensive use of economic tools such as game theoretic models is possible in e-commerce, especially in auctions, because of the automation of commerce. These tools not only reduce cost but also improve the match of buyer and good. E-commerce can also seamlessly integrate various functions of a firm, such as buying, operations, support functions, and sales. As regards distribution of value, automated e-commerce also incorporates personalization technologies that track an online buyer, identify her preferences, recommend products, and permit customization. This ability can blunt the buyer’s price comparison ability, thereby enabling the seller to engage in “dynamic pricing” and charge a different and usually higher price. However, the buyer also has software to help him or her. In effect, the individual consumer becomes a separate market of his or her own that many firms compete to capture. The resulting dynamics are complex, so how value is distributed is difficult to predict.

2.2.1.4 Internet Pricing

Brynjolfsson and Smith (2000) analyze the pricing behaviour in Internet trade to verify the claim that e-commerce is frictionless. Despite such a claim, research has shown that price sensitivity can be lower in Internet trade, that trust can cause decreased price sensitivity in Internet trade, that the provision of better information to customers can increase loyalty, that Internet prices for used cars are greater than their physical auction price, and so on.

To further explore the issue, these researchers studied the pricing behaviour of online retail outlets engaged in the trade of two homogenous goods–books and CDs, and compared their pricing behaviour to that of conventional retailers. They used data of 8500 pricing observations spread over 15 months in 1998-99 pertaining to 41 online and conventional retail outlets. In their analysis, they considered 4 Internet outlets, 8 hybrid outlets, and 4 conventional outlets each for books and CDs. Some of the retailers were replaced with others in the course of the study. The price data are for 20 book titles and

20 CD titles. The Internet and hybrid retailers sampled covered more than 96% of the web-page hits for both goods.

The above researchers conducted statistical and econometric tests on the data. The important results are as follows: Internet retailers have a much larger selection of goods; the mean price of Internet stores was significantly lower than that of conventional stores, 15.5% for books and 16.1% for CDs ($p < 0.001$); the lowest Internet price was less than the lowest conventional store price 92% of the time for books and 84.6% of the times for CDs ($p < 0.05$); the mean Internet price (including shipping, handling, and taxes) was lower than the mean conventional price by 9% for books and by 13% for CDs ($p < 0.001$); the lowest Internet price (including shipping, handling, and taxes) was lower than the lowest conventional price (including shipping, handling, and taxes) around 83% of the times for both books and CDs ($p < 0.05$ or better); the price adjustments were as low as \$ 0.05 for books and \$ 0.01 for CDs in Internet trade whereas they were \$ 0.35 for books and \$1.00 for CDs in conventional trade; the standard deviation of price adjustments on the Internet was lower than that in conventional trade for both books and CDs; and there is less dispersion in weighted[3] prices on the Internet than in conventional outlets for both books and CDs, though the unweighted price dispersion on the Internet was as high as 47%, with the average being 33% for books and 25% for CDs.

The above findings suggest that Internet trade is more efficient than conventional trade, but not completely efficient. It is not friction-less, but it has less friction. Friction comes from factors such as the importance of location of a retailer in the physical world or in virtual space, lack of information, retailers making positive economic profit, lack of price competition due to retailer heterogeneity, price dispersion, level of prices, inability to adjust price in accordance with supply and demand, etc. Internet trade has less friction mainly because (1) it is easier to access a particular online retailer than a particular physical store, (2) it is easier to get information about retailers and goods in online trade than in physical trade, (3) online prices[4] and price dispersions are generally lower than in

[3] Weighting is as per market share of the retailers.
[4] If price is more than what it would be under perfect competition, it contributes to friction.

physical stores, and (3) it is possible to adjust prices more often and minutely in online trade than in physical stores. Internet trade has some friction mainly because (1) not all online sellers are equally accessible to buyers (since not all can afford high advertising costs), (2) there is information asymmetry between sellers and buyers, (3) retailers can charge excess prices through price discrimination, and (4) there is heterogeneity among online retailers. The evidence for less friction in online trade are as follows: (1) the study finds that Internet price adjustments are much smaller, suggesting lower menu costs and frequent adjustments; frequent adjustments help to keep prices in line with structural adjustments in supply and demand, (2) the price dispersion weighted by market share was actually lower in Internet trade, implying market concentration in Internet trade and dominance by some heavily branded retailers; in fact, the study found that the market shares of the dominant firms were much higher in Internet trade than in conventional trade, and (3) price is generally lower in Internet trade, suggesting that the Internet lowers costs. The evidence for presence of some friction in online trade is from the authors' finding that online retailers with high market shares, e.g., Amazon.com, actually charged higher prices (suggesting that trust and brand loyalty are important in Internet trade and that these cause heterogeneity among online retailers); in fact, the Internet seems to strengthen the importance of trust and loyalty.

Apart from the reason of retailer heterogeneity, Internet retailers may charge higher prices to uninformed customers. It is unclear whether the price dispersion caused by heterogeneities arising from information asymmetry and customer loyalty will hold in equilibrium. Thus, whether Internet prices are lower than physical market prices is an empirical question rather than a theoretical one. These imply that though there is lower friction in Internet trade, it is not frictionless as claimed.

2.2.1.5 Business to Consumer E-Commerce

Bakos (2001) examines the likely impact of e-commerce on Business to Consumer (B2C) commerce. Retail e-commerce is increasing, but some goods are more dominant than others. The projected shares of different product categories in total online retail sales for

2005 for the U.S. are as follows: consumables (health and beauty aids, general supplies, beverages, etc.), 18%; apparel, 16%; computers and electronics, 12.4%; automobiles, 12.2%; leisure travel, 12.1%; books, music, videos, and software, 9.6%.

When searching for goods, buyers incur opportunity costs of time spent, cost of travel to stores, magazine costs, etc. Sellers incur costs in finding buyers (e.g., market research, advertising, sales calls, etc.). Search engines such as Google, price comparison agents such as Pricewatch, and agents such as R-U-sure.com that monitor consumer behaviour to find the right product at the right price help reduce buyer search costs. The cost of acquiring information about sellers also is reduced by agents such as eBay or Bizrate. Seller search costs are reduced by cost-effective communication of product information, targeted advertising, one-on-one marketing, etc.

Lower buyer search costs encourage price competition. Online markets may have lower entry costs or smaller efficient scales, leading to many sellers in equilibrium with lower prices and profits. Online buyers expect discounts in the range of 20 to 30 per cent for items priced $30 to $500. The “friction-free” market dynamics are not favourable to sellers. Since products are not truly homogenous, sellers may resort to differentiation to offset lower profits due to lower buyer search costs. Online product differentiation and offerings are facilitated because physical shelf space is not required. Information-rich products lend themselves to cost-effective customization. Customization is facilitated by technologies that track the behaviour of consumers. Merchants may also create high switching costs by using superior user interfaces, offering desired products based on knowledge about customers, etc. Generally, online buyers have less price sensitivity. However, sellers may not have a greater ability to charge a premium, since buyers have been found to shop at Amazon.com and then actually buy at lower-priced Buy.com. Customer information, the ability to differentiate products, and lower menu costs mean an increased ability to price-discriminate. This ability may offset price competition brought on by reduced search costs. On the whole, the evidence on price dispersion is mixed.

New price discovery mechanisms, e.g. new types of auctions, are employed in e-commerce. These mechanisms may help to obtain more efficient markets, in some cases improving welfare in general. But an efficient market does not result when information asymmetry is present. When sellers are better informed, they may increase profits through price discrimination. Electronic markets usually emphasize product information rather than price information, in the interest of sellers. The new patterns of price discovery change the market microstructure, that is, how prices are set by buyers and sellers. The microstructure affects both value creation and distribution.

New types of intermediaries develop in e-commerce markets. These intermediaries help online markets mainly by aggregating services and products. Distribution is likely to be transformed, especially for information goods. In the case of tangible goods (e.g., Dell computers), traditional intermediaries such as wholesalers may be eliminated. However, new types of intermediaries develop even in the case of tangible goods (e.g., FedEx and UPS, who have expertise in logistics and have economies of scale). In short, intermediaries that provide physical inventory will disappear and those that provide information services will flourish.

In the case of information goods, lower transaction costs result in new strategies such as bundling, site licensing, and per-use fees. These are basically strategies of aggregating or disaggregating. Aggregating a large number of information goods produces higher profits to sellers and a wider distribution of goods to buyers. Aggregation may change the shape of the demand curve. When the valuations of goods are not perfectly correlated, the average valuation of the bundle is usually around the mean valuation. Hence, bundling and pricing the bundle just below the mean can result in higher profits to sellers and greater total welfare. This strategy also reduces dead-weight loss, when the marginal cost is very low. Intellectual Property Rights (IPR) promote creation of content, but limitations imposed by IPR increase the dead-weight loss and reduce social welfare. However, on the whole, retail e-commerce is likely to increase social welfare through lower prices, more choices, lower “fit costs,” and first-order increase in welfare from product offerings. Even price discrimination may increase social welfare by increasing the number of buyers.

2.2.2 Social Impacts of the Internet

Shah, Kwak, and Holbert (2001) investigate the relationship between Internet use and the individual-level production of social capital. They analyze the predictive power of new media use using 1999 DDB Life Style Study data. They find that the informational uses of the Internet are positively related to differences in the production of social capital; that socio-recreational uses are negatively related; and that social capital production is related to use among generation X.

2.2.3 Behavioural Impacts of the Internet

Stevenson (2003) studies behavioural impacts of the Internet in the context of job search by unemployed and employed persons who wish to change jobs. Stevenson uses the state average ownership rates of household appliances in 1960 as instrument to explain the adoption pattern in various states in the U.S. and finds that, in states that adopted the Internet rapidly, the unemployed have increased their job searches. The research also finds that the employed also seem to have increased their job search for changing jobs.

2.2.4 Geography and Impacts of the Internet

Pohjola (2002) reviewed the literature and analyzed WITSA and WBDI data to study the New Economy, characterized by the Internet and globalization, with respect to its various facts, its impacts, and public policy implications. He observes that people in rich countries, having the required infrastructure and skills, are in a much better position to benefit from the Internet than people in poor countries are.

2.3 Diffusion of the Internet and Applications

2.3.1 Diffusion of Technologies

Griliches (1957), in his seminal study of diffusion of innovations, analyzed the factors responsible for cross-sectional differences in the rates of use of hybrid seed corn in the U.S. He fit logistic growth functions to data and interpreted the differences among areas as differences in the estimates of origins, slopes, and ceilings (which in turn are explained by different economic factors that are characteristic of areas). Through this study, he showed that the process of innovation, the process of adapting and distributing a particular invention, and the rate at which it is accepted by entrepreneurs are amenable to economic analysis.

Saloner and Shepard (1995) studied time until adoption of technologies with network effects in the context of banks' adoption of ATMs. They first develop a theoretical framework for the relationship between network size and (banks') propensity to adopt (ATMs). They then use different standard duration models (Weibull, log-logistic, and non-parametric Cox partial likelihood estimator) for empirical tests. They use data from FDIC and from Hannan and McDowell (1987) surveys. They find that the number of branches (ATM locations) increases the propensity to adopt early, consistent with the presence of a "network effect."

Banerjee (1992) analyzes sequential decision making in which each decision maker looks at the decisions made by previous decision makers in taking his or her own decision ("herd behaviour") through a game-theoretic model with Bayesian-Nash equilibrium. He theorizes that people will do what others are doing rather than use their own information and that the resulting equilibrium is inefficient. We may interpret from this result, in terms of adoption by households, that the situation of each household is different and known only to the household; and that households will adopt the Internet whatever their situations may be, if adoption is high in their neighbourhood, resulting in inefficient equilibrium[5].

Geroski (2000), in a survey of technology diffusion models, reviews the common S-curve generating models, epidemic and probit, and argues for two alternate models, competition and information cascades. He identifies new implications for public policy based on the alternative models. Hall (2004), in her study of determinants of diffusion, and modeling strategies, provides a historical and comparative review of determinants and a non-technical review of modeling strategies. She advocates the real options approach as a promising modeling avenue. The real options approach would naturally yield a hazard rate/waiting time model and an S-shaped cumulative distribution, while explicitly incorporating uncertainties of decisions.

2.3.2 Determinants of Internet Use

2.3.2.1 Socio-Economic Determinants of Internet Use and the Digital Divide

Venkatesh and Brown (2001) identify drivers of technology adoption in homes through a nationwide, two-wave, longitudinal investigation of factors driving personal computer adoption in U.S. homes. Their research found that adopters of technology were driven by utilitarian, hedonic, and social outcomes whereas non-adopters were influenced by rapid changes in technology and the fear of obsolescence. They also found an asymmetrical relationship between intent and behaviour in connection with adoption decisions.

Kiiski and Pohjola (2002)'s research looked for factors that determine the diffusion of the Internet across countries. Using the Gompertz model of technology diffusion and data on hosts per capita for 1995-2000, they found that, for a sample of OECD countries, the per capita gross domestic product and Internet access cost best explain the growth in computer hosts per capita; however, investment in education is not a significant predictor. For a larger sample of industrial and developing countries, education also becomes a significant factor.

[5] The resulting equilibrium in such cases is inefficient because it reduces welfare.

Sexton et al. (2002) analyzed a wide range of variables to identify accurate predictors of Internet and e-commerce use among individuals. With the aid of survey research and a neural network model, they found that a person's gender, overall computer usage, job-related use of the Internet, and access to the Internet from home are important influences of the use of the Internet and e-commerce.

Mills and Whitacre (2003) studied the gap between home Internet use in metropolitan areas and non-metropolitan areas in the U.S. They used data from the 2001 U.S. Current Population Survey and modeled the household Internet adoption decision using a logit estimation approach. Decomposing the estimates for metropolitan and non-metropolitan areas, they found that differences in household attributes, especially education and income, account for 63% of the digital divide between metropolitan and non-metropolitan areas.

Hoffman and Novak (1999) surveyed the literature and used data from various sources to study the digital divide among sections of American society. They found that the digital divide between Whites and African Americans is growing and that differences in the levels of education, income and wealth, and gender also contribute to the digital divide[6].

2.3.2.2 Behavioural Determinants of Internet Use

Kraut et al. (1998) researched the relative importance of alternative uses of the Internet in households, especially communication and information-related uses. In the research, they treated the use of email as related to interpersonal communication and the use of World-Wide Web (WWW) as related to information acquisition and entertainment. They analyzed data from a longitudinal survey of 229 individuals in their first year of using the Internet. The research showed interpersonal communication to be a stronger driver of Internet use than are information and entertainment applications.

[6] In our research, we do not study household e-buying from the racial perspective.

Goolsbee and Klenow (2002) researched the importance of local spillover (such as network externalities and learning from others) in the diffusion of home computers. Using a linear probability model and instrument variables estimation, they analyzed data on 110,000 U.S. households in 1997 and found that people are more likely to buy their first home computer in areas where a high fraction of households already own computers or when a large share of their family and friends own computers; that the spillovers appear to come from experienced and intensive computer users; that the spillovers are not associated with use of any particular software, but seem to be tied to the use of email and the Internet.

Shih and Fang(2004) attempt to identify behavioural factors that can predict Taiwanese bank customers' intention to adopt banking. They analyze data collected from 425 respondents, using models of the Theory of Planned Behaviour (TPB) and the Theory of Reasoned Action (TRA). They use structural equation modeling for the empirical analysis. They find that attitudes towards Internet banking influence adoption significantly, whereas subjective norms (that is, what individuals think others would expect or do) do not. Intention to adopt is in turn a significant determinant of actual adoption. The authors reach this conclusion from both models, TPB and TRA. In our opinion, these results suggest that income, education, and age levels might be among factors influencing household online buying. Households with lower incomes, lower levels of education, and aged people may not have a favourable attitude towards online banking or buying.

Cheong and Park (2005) researched factors that promote the use of the Mobile Internet (M-Internet). They developed hypotheses from the literature on the Technology Acceptance Model (TAM) and empirically tested them using survey data obtained from 1279 respondents in Korea. They found that individuals' attitudes towards the service, their perception of M-Internet's "playfulness,"[7] and its usefulness are significant predictors of their intention to adopt the M-Internet.

[7] Cheong and Park (2005) use the phrase "perceived playfulness" to mean how entertaining the Mobile Internet is.

2.3.2.3 Determinants of Internet Use and its Applications: Organizations and Individuals

In this section, we review some literature on the determinants of Internet use and its applications in organizations. The objective is to compare these determinants to those suggested by the literature on individuals' adoption of the Internet.

Chwelos, Benbasat, and Dexter (2001) surveyed the research and diverse sets of models of EDI adoption by organizations, developed a compact model by synthesizing these models, collected survey data, and conducted the first empirical test of their model. The model hypothesizes that *external pressure*, *readiness*, and *perceived benefits* are the three fundamental determinants of EDI adoption by organizations. In developing their model, the authors hypothesized that H1: *Higher perceived benefits will lead to greater intent to adopt EDI*; H2: *Higher external pressure will lead to greater intent to adopt EDI;* and H3: *Higher readiness will lead to greater intent to adopt EDI*. The model is tested by structural equation modeling, with data collected through a survey of a sample of senior purchasing managers (N = 268) throughout Canada.

The authors used the Partial Least Squares (PLS) statistical analysis technique. *Intent to adopt* was the dependent variable. The findings support the hypotheses. The independent constructs - *Perceived benefits*, *readiness,* and *external pressure* - all positively relate to the intent to adopt EDI at the $p < 0.001$ level. Approximately 32% of the variance in *intent to adopt* is accounted for by the three constructs ($R^2 = 0.318$). The path coefficients range from 0.11 to 0.37, with the two paths from *readiness* and *external pressure* to *intent to adopt* above the suggested minimum of 0.2. These imply that the fit of the model is significant and that the two constructs, *external pressure* and *readiness,* are more important than *perceived benefits*, but that all three constructs are significant determinants of EDI adoption.

Forman (2005) researched Internet adoption decisions in organizations (joint decisions of basic access and applications such as e-commerce). The research built on hypotheses from prior results and used discrete choice analysis (nested logit model) to test the hypotheses with data from Harte Hanks Computer Intelligence Technology Database for the years 1996-98. It was found that the two technologies, access and applications, diffused at different rates (the latter being much slower); that prior investments in Information Technologies (IT) affect adoption decisions (some may affect negatively by acting as short-term substitutes, or through a “lock-in” effect); and that geographic dispersion of employees, organizational size, and external pressure increase the likelihood of adoption.

2.3.2.4 Taxes, and Internet and E-Commerce Adoption

Consumers find incentives to engage in e-commerce in other, albeit non-economic, areas. Customers incur effectively less sales tax when they buy online. This section deals with the literature on e-commerce and taxes. The *Economist* (January, 2000) carried a survey on globalization and tax. Matthew Bishop investigated if consumers will pay less tax in the future, as globalization is accelerated by the Internet. Globalization means the gradual fusing of national markets into a single world market. The Internet reduces the relevance of being in a particular location and enhances the pace of globalization, making it more difficult for governments to collect taxes. The Internet also makes it relatively easy to break the law and evade taxes.

Current tax systems rely on knowing where a particular economic activity is located. In the U.S., companies without sufficient “nexus” in a state are not legally obliged to collect sales tax. Online buyers are required to pay a “use tax” equal to the sales tax, but they hardly ever do. Online sellers such as Amazon.com have not bothered to remind their customers about the use tax. Less affluent members of society without access to the Internet cannot take advantage of Internet buying. Even traditional retail companies have attempted to spin off their Internet sales divisions into separate companies to take advantage of the Internet. Traditional retailers may convert their check-out counters to Internet terminals where the customers can “order” online what they picked from the shelves, if proposals to make Internet selling tax free take effect. Taxing Internet sales will become much less acceptable if consumers become accustomed to tax-free Internet buying. Taxing individuals is also becoming more difficult. For example, capital, if not people themselves, moves to tax havens easily in the Internet-influenced globalized world, thereby decreasing tax revenues from interest income and capital gains.

Cline and Neubig (1999) investigate if U.S. state and local tax revenues have eroded with the growth of e-commerce. Interstate sales are not subject to sales or use-tax collection by companies without nexus, as per U.S. Supreme court rulings. Most services and intangibles are not subject to such tax anyway. Some goods, such as groceries and prescription drugs, are also exempt in many states. Taxable sales through the Internet are subject to use tax payable by buyers, but governments do not enforce the collection of this tax. As of 1998, B2C Internet sales were only $20 billion or 0.3% of total consumer spending. The estimated revenue loss is only $170 million or 0.1% of total state and local sales and use tax collections. About 80% of e-commerce is Business to Business (B2B), which is generally tax exempt or effectively subject to use-tax payments. About 63% of current e-commerce B2C sales relate to intangible services (e.g., travel and financial services) or exempt goods such as groceries that are not generally subject to state or local taxes. About 60% of e-commerce sales are those that would otherwise have evaded tax through telephonic sales or mail order sales and hence do not constitute fresh revenue loss. Thus, it is estimated that only about 13% of all or about one third of taxable e-commerce sales have tax collection issues.

Goolsbee (2000) tries to determine if taxes influence the Internet buying decisions of individuals. With controls for various buyer characteristics, the results suggest that online users in high sales-tax locations are more likely to buy over the Internet[8]. The author estimates that the number of online buyers would drop by as much as 24% if sales taxes were applied to online purchases. The tax sensitivity of sales has always been significant, with elasticity as high as 5 or 6, independent of the Internet, as is observed from the buying behaviour of people residing near state borders. Internet commerce has similar

[8] A buyer may not pay tax in an online purchase; if he or she goes instead to a local store to buy the same good, he or she will pay sales tax.

characteristics as traditional commerce in cross-border locations. An earlier survey conducted nationally for Forrester Research, Massachusetts involved 110,000 U.S. households. The author uses the data pertaining to about 25,000 of these that had online access.

A probit model estimates the probability that an individual with online access will buy over the Internet, as a function of the sales tax rate and various controls such as income, education, age, ethnicity, child/adult, marital status, gender, access to computer, running one's own business or not, etc. With regards to the sales tax, the more likely an individual is to buy over the Internet, the greater the ratio $P_s(1+t)/P_i$ is, where P_s is the store price, t is the sales tax and P_i is the price over the Internet. It is assumed that the relative price, P_s/P_i, is constant across locations. The dependent variable becomes a function of (1+t) and other control variables. The sign of the coefficient of the variable (1+t) is anticipated to be significantly positive.

The results suggest that the mean probability of buying over the Internet, conditional on having Internet access, is 20.3%. The sales tax is found to have significant impact on the decision to buy online. Raising the tax by 0.01 increases the mean probability by 0.005, implying that the elasticity of online buying with respect to tax price (1+t) is 2.3. All the control variables, except ethnicity, are also generally significant. The mean frequency of use is found to be 16.7 days per month. It is also found that higher sales taxes do not make an individual more likely to obtain online access or own a computer. It is found also that taxes do not influence buying decisions on products that proxy for technological sophistication.

2.4 Summary of Literature

The impacts of the Internet are primarily economic and involve trade. For example, the Internet impacts the marketplace through reduced buyer search costs (Bakos, 1997). Buyers gain also from rent redistribution through lower prices, in addition to the social benefits of lower search costs and better matching of buyers and sellers. The earliest effect of the Internet on markets is incremental and is basically cost reduction. In due course, drastic shifts arise: through e-commerce, we can restructure business (Borenstein and Saloner, 2001). The Internet creates value by greatly reducing the cost of transferring information. If the product is itself information, the possibilities for value creation are tremendous. Value is created either on the cost side or on the demand side. On the demand side, improved matching of supplier and buyer occurs through improved information, better access to goods, and ease of customization. Still, issues of fit, feel, quality, and real appearance of goods remain. Automated e-commerce incorporates personalization technologies that track an online buyer, identify his or her preferences, recommend products, and permit customization (Vulkan, 2003). This ability of sellers can blunt the buyer's price comparison ability, thereby enabling the seller to engage in "dynamic pricing" and charge a different and usually higher price. However, the buyer also has software to help him or her. In effect, the individual consumer becomes a separate market of his or her own that many firms compete to capture. The resulting dynamics are complex, so how value is distributed is difficult to predict. Brynjolfsson and Smith (2000) find that trust and brand loyalty are important in Internet trade and that these cause heterogeneity among online retailers. In fact, the Internet seems to strengthen the importance of trust and loyalty. Price is generally lower in Internet trade, suggesting that the Internet lowers costs. Apart from the reason of retailer heterogeneity, Internet retailers may charge higher prices to uninformed customers. It is unclear whether the price dispersion caused by heterogeneities arising from information asymmetry and customer loyalty will hold in equilibrium. Thus, whether Internet prices are lower than physical market prices is an empirical question, rather than a theoretical one. Retail e-commerce, on the whole, is likely to increase social welfare through lower prices, more choices, lower "fit costs," first-order increase in welfare from product offerings, etc. Even price discrimination may increase social welfare by increasing the number of buyers (Bakos, 2001).

Because of numerous positive impacts, the Internet has diffused widely and rapidly. The process of innovation, the process of adapting and distributing a particular invention, and the rate at which it is accepted by entrepreneurs are amenable to economic analysis

(Griliches, 1957). The diffusion of the Internet as an innovation is thus amenable to analysis and there has been considerable research on the topic. The Internet is a network. In a study of the adoption of ATMs by banks, Saloner and Shepard (1995) find that the number of branches (ATM locations) increases the propensity to adopt early, consistent with the presence of "network effect." Thus, in the case of the Internet, when the number of nodes (that is, computers connected to the Internet) increases, further adoption of the Internet would gain even greater momentum. Similarly, "herd behaviour" (Banerjee, 1992) implies that further adoption of the Internet will quicken when more and more individuals adopt the technology. The diffusion of technologies has been primarily modeled with S-Curve, first proposed by Griliches (1957). Alternate models have also been advocated by researchers (e.g., Geroski, 2000; Hall, 2004) to explain diffusion of technologies. The estimation of parameters of S-Curve or similar models leads us to the study of fundamental factors that drive adoption of technologies by users. As a result, there has been considerable research on the determinants of Internet use.

The determinants of Internet use have been studied from socio-economic, behavioural, and other points of view. Venkatesh and Brown (2001) found that adopters of personal computers in homes were driven by utilitarian, hedonic, and social outcomes but that non-adopters were influenced by rapid changes in technology and the fear of obsolescence. Kiiski and Pohjola (2002) found that the per capita gross domestic product and Internet access cost explain best the growth in computer hosts per capita in OECD countries, and these factors and investment in education explained the same in a broader sample of countries. Sexton et al. (2002) found that a person's gender, overall computer usage, job-related use of the Internet, and access to the Internet from home are important influences of the use of the Internet and e-commerce. Mills and Whitacre (2003) found that differences in household attributes, especially education and income, account for 63% of the digital divide between metropolitan and non-metropolitan areas. Hoffman and Novak (1999) found that the digital divide between Whites and African Americans is growing and that differences in the levels of education, income and wealth, and gender also contribute to the digital divide. Kraut et al. (1998) showed interpersonal communication to be a stronger driver of Internet use than are information and entertainment applications. Goolsbee and Klenow (2002) found that people are more likely to buy their first home computer in areas where a high fraction of households already own computers or when a large share of their family and friends own computers, and that computer adoption seems to be tied to the use of email and the Internet . Shih and Fang (2004) find that individuals' attitudes towards online banking influences significantly their intention to adopt it, while subjective norms (that is, what individuals think others would expect or do) do not. Intention to adopt is in turn a significant determinant of actual adoption. Cheong and Park (2005) found that individuals' attitudes towards the service, their perception of the M-Internet's playfulness, and its usefulness are significant predictors of their intention to adopt the M-Internet.

Chwelos, Benbasat, and Dexter (2001) found *external pressure* faced by firms and their *readiness* to adopt technologies to be more important than the *perceived benefits* of technologies. Forman (2005) found that access and applications diffused at different rates (the latter being much slower); that prior investments in information technologies (IT) affect adoption decisions; and that external pressure, among other factors, increases the likelihood of adoption.

Finally, researchers have also explored whether taxes are a determinant of Internet adoption decisions by individuals. The *Economist* (January, 2000) highlighted that globalization driven by the Internet has led to loss of tax revenues to governments. Online buyers are supposed to pay “use tax” in place of “sales tax,” but they rarely do. Borenstein and Saloner (2001) state that online buyers enjoy effectively lower taxes. Cline and Neubig (1999) argue that tax-loss due to online trade is not as large a loss of tax revenue as it is generally believed to be. Goolsbee (2000), however, found that online buyers are sensitive to these tax issues but still will not buy computers or subscribe to Internet services simply because of the effective lower taxes in online buying. That is, taxes are not a significant determinant of Internet use.

Chapter 3

Model

3.1 Introduction

Data collected by Statistics Canada through its Household Internet Use Survey (HIUS) from 1997 - 2003 show that the level of online or electronic buying (e-buying) activity by Canadian households steadily increased over the period. Obviously, not all consumers adopted e-buying equally when presented with the technological opportunity of e-buying. Although the maturity of the electronic marketplace and the state of technology affect how conducive it is for consumers to adopt e-buying, individual consumer-specific factors such as age, education, and income probably play a role in e-buying adoption decisions. The individual-specific factors may explain why some consumers e-buy but others do not. To explore factors that influence e-buying adoption by the public, we chose as our research question, "how does e-buying intensity vary across age groups, gender, education levels, income levels, and the nature of goods?" We first develop a simple theoretical model[9] that leads us to some hypotheses concerning the above question.

For our theoretical model, we start with the premise that individuals maximize their utility. An individual consumer makes a purchase when he gains positive utility from doing so. Utility arises from the good's specific nature (the quality parameter). But utility from the same good varies from consumer to consumer because of individual-specific factors (the taste parameter). Costs associated with a purchase, such as the price of the good and search costs, reduce the utility of purchase. If we were to consider the possibility of e-buying a good to the possibility of conventionally buying the good from a local store, we may associate different utilities to the consumer. We may then say that a consumer will e-buy rather than conventionally buy if the utility of e-buying is greater. The overall demand for a good will depend on population and the distribution of the taste parameter. Our theoretical model will thus have a quality parameter, taste parameter, and

[9] The model is adapted from Tirole (1989), chapter 2.

costs (such as price and search cost), among others, as variables that explain e-buying demand.

The HIUS data, which we wish to use for the empirical tests, have variables such as individual-specific (or household-specific) details such as age group, gender, and education level of the heads of the households, household income levels, type of Internet connection, whether the households ordered any good online in the past twelve months, the type of goods ordered, and if so, the number of orders placed and their dollar value. We can group goods ordered into two broad types: homogeneous and heterogeneous. Thus, in the HIUS datasets we have data pertaining to the quality parameter for a good, the taste parameter from the individual or household-specific factors, and search costs from the type of Internet connection. With this background, we develop hypotheses to explain overall e-buying adoption by households.

3.2 Theoretical Model

3.2.1 Assumptions

We develop our simple model to explain, with certain assumptions, total e-buying demand of a specific good from a population of individual consumers. There are *N* households in a population. Each household represents one consumer. A consumer buys one unit of the good. He or she has a choice of buying it either through the Internet (e-buying) or in a local store (non-ebuying)[10].

A good's quality parameter is defined as *s*. Through this parameter *s*, we differentiate a homogeneous good from a heterogeneous one. When a consumer buys a good in a store, he or she can physically inspect the good, feel it, smell it, and determine its visual appeal.

[10] A consumer who e-buys may do so either through his or her home connection or from outside locations such as a public library, workplace, university, or café, though most Canadian households access the Internet through their home connections. For example, out of the total sample of 23113 Canadian households of the HIUS 2003, 11868 had some kind of connection and 14159 used the Internet in a typical month from any location.

When e-buying the good, the customer cannot physically inspect it. He or she can ascertain online the quality to some extent of the good, subject to the limitations of online verifying. Thus, we assume that the quality parameter of a good bought online, designated s_e, cannot exceed the quality parameter of a good bought in a local store, designated s_n. However, the difference between s_e and s_n can be practically zero in the case of homogeneous goods such as a book or a music CD.

A buyer is almost as sure of the quality of a homogeneous good when he or she e-buys it as he or she would be when conventionally buying it in a local store. Hence we have $s_e \sim s_n$ for homogeneous goods. In the case of heterogeneous goods such as flowers, beauty products, clothing, jewellery, and real estate, a buyer is likely to be less sure of the quality of the good when he or she e-buys it than he or she would be when buying it through conventional means. Hence we assume that $s_e < s_n$ for heterogeneous goods. Thus we have assumed that the quality of heterogeneous goods can be ascertained only by physical contact whereas the quality of homogeneous goods can be verified indirectly through the Internet[11]. The parameter s captures this aspect and is thus a measure of how confident a consumer feels about the quality of the good in a particular purchase. The quality parameter s in turn may be a function of, apart from the good's degree of homogeneity, several factors such as seller's reputation, whether the good is branded or not, the buyer's risk attitude, and the transaction value. We are, however, not concerned with these other factors in our simple model.

For a given value of the quality parameter s from a purchase of a good, two consumers will derive different amounts of utility, due to differences in individual tastes. We capture the taste differences through a taste parameter designated as θ. Consequently, a consumer's taste parameter is θ, which varies by consumer. Individual taste differences may arise due to individual-specific factors such as age, gender, education, and income. The parameter θ captures these differences. The utility derived by a consumer from a purchase is θs_e, the product of the taste parameter and the quality parameter. The taste

[11] Homogeneous goods (for example, books or music CDs) do not call for rigorous inspection as heterogeneous goods do.

parameter θ may be distributed in a population of consumers in a certain way. The distribution of θ is $F(\theta)$, the cumulative distribution function (CDF). At this stage we do not make any assumption about the type of distribution $F(\theta)$. To elaborate on the taste parameter, if we consider an online purchase of a costly pen from the viewpoints of two consumers, this purchase may have different utilities to the two consumers, so one may buy it but the other may not. The purchase will likely be viewed equally by the two with respect to its quality parameter, but one consumer may not be able to afford it and hence he or she may not have developed a taste for such costly goods (income factor). Similarly, a particular overseas travel package may be more appealing to a woman than to a man (gender factor). A trip to a theme park may be more exciting to a child than to his or her parents (age factor). Other factors may contribute to differences in individual tastes for a good.

A particular benefit of e-buying is the ease with which information about goods and sellers can be found by buyers. Information is often just a few clicks away. Electronic commerce has cost advantages over conventional commerce (Bakos, 1997, 2001; Borenstein and Saloner, 2001; Vulkan, 2003). A major source of cost reduction is reduced search cost. The lower search cost of e-buying enhances the utility derived from e-buying. We factor this aspect into our model through the effort parameter for information search a. The greater the cost of information search, the greater the value of a. We designate the effort parameter for information search of e-buying as a_e and that of conventional buying as a_n.

The Price of the good, which we designate as p, considerably reduces the utility of purchase to most consumers and hence influences the demand for the good. Price p_n is the local price. Some empirical evidence suggests that online prices, which we designate as p_e, are lower than conventional market prices (e.g., Brynjolfsson and Smith, 2000). However, conflicting evidence suggests that online prices sometimes may be higher (e.g., Brynjolfsson and Smith, 2000) and that sellers may engage in dynamic and discriminatory pricing (e.g., Vulkan, 2003). Consistent with the literature, we designate prices of online and conventional purchases differently, and these prices may or may not be equal. Additional transportation costs and effectively lower taxes are associated with online buying. The geographical market reach of electronic commerce may be considerably enhanced for tangible goods (Borenstein and Saloner, 2001) because greater transportation costs can be offset by other cost savings and a distant market reached. Since tax collection is difficult to enforce in electronic commerce, online buyers pay effectively lower sales taxes (e.g., *Economist*, January 2000). These issues of transportation costs and tax affect the effective price of online purchases; hence, our variable for the price of e-buying, p_e, includes transportation costs and excludes tax (in some cases). Although the online price includes additional transportation costs, it can still be competitive due to a larger number of suppliers (resulting from the extended geographic market reach of electronic commerce, which is consistent with the evidence for lower online prices cited above).

3.2.2 Model

Let U_e and U_n be the utilities derived by a consumer from an online purchase and a conventional purchase (in a store) of a certain good. Then we have

(301) $U_e = \theta\, s_e - p_e - a_e$

(302) $U_n = \theta\, s_n - p_n - a_n$

As described earlier, $\theta\, s$ is the positive utility from purchase; price p and search cost, characterized by the effort parameter of information search a, reduce the utility of purchase. Hence we have the above expressions (301) and (302) for the utilities of purchase U. The consumer will e-buy instead of buying in a store if

(303) $U_e > U_n$, or

(304) $U_e - U_n = \theta\,(s_e - s_n) - (p_e - p_n) - (a_e - a_n) > 0$

We can redefine parameters as the difference between e-buying and non-ebuying; then we have

(305) $U = \theta s - p - a > 0$, or

(306) $\theta > [(p + a) / s]$

The higher the price advantage and the search cost advantage of e-buying, the more negative the numerator of the right-hand side of the above inequality[12]. The greater the certainty of the quality of the good bought online, the less negative the denominator. Thus, the above inequality (306) implies that even a consumer with a low taste for a good will e-buy if the online price and search cost are sufficiently low and the quality of the good bought online is sufficiently certain.

There are N households (i.e., consumers) in the population, and consumers whose utility is greater than 0 or whose taste parameter is $\theta > [(p + a)/s]$ will demand the good. Thus, the demand function is

(307) $D(p) = N[1 - F(\theta)]$, or

(308) $D(p) = N[1 - F((p + a)/s)]$

where $D(p)$ is the demand for the good, $F(\theta)$ is the distribution of the taste parameter θ in the population, s is the quality parameter, and the other variables are as explained earlier. $F(\theta)$ represents the area under the distribution curve for values of taste between -8 and the taste parameter θ. This area is the proportion of the population that represents consumers with taste for the good below the taste parameter θ. These consumers will not e-buy the good. The area under the curve $F(\theta)$ lying on the right side of the vertical line $x = \theta$ represents the proportion of the population of consumers with taste for the good above the taste parameter θ. These consumers will e-buy the good. Thus, we find from (307), the demand function $D(p)$, that the higher the taste parameter θ, the lower the demand for the good. Alternately, from (308), we find that the higher the price and cost

[12] The numerator ($p + a$) represents the disadvantage of e-buying with respect to price and search cost; since e-buying is generally advantageous, ($p + a$) would often be negative. The numerator can also be interpreted more generally as a measure of the total cost advantage of e-buying, rather than as only a measure of price and search-cost advantages.

advantages of e-buying and lesser the uncertainty about the quality of the good bought online, the higher the demand for e-buying.

3.3 Empirical Model

3.3.1 The Implications of the Theoretical Model

The demand function $D(p)$, as expressed in (308), may be interpreted in terms of the variables for which data are available in the HIUS dataset (namely age, gender, education, and income), so we can use the HIUS dataset for empirical tests of the above model. Equation (308) essentially relates demand with the individual-specific taste parameter θ, which in turn is related to the expression $[(p + a) / s]$. The price p and search-cost advantage a affect individuals differently since their income levels are different. Some are very sensitive to price and costs, while others are not at all. The search-cost advantage a may be high for consumers with high-speed Internet connections at home. It may be high for those who are young since young people are probably more tech-savvy and adept at the use of computers and the Internet. It can also be high for consumers with higher levels of education. The quality parameter (gap) s may affect old consumers, less educated consumers, and low-earning consumers more than it affects others. Consequently, individual factors such as age, gender, education, and income may be reasonably considered as representing (or as proxies for) the individual-specific taste parameter.

3.3.2 Literature and Proposed Theoretical Model

At this point, it may be relevant to refer to the literature for consistency with the above reasoning. If we find consistency, we would have additional justification for the above reasoning and for hypotheses we develop subsequently for empirical tests. Kiiski and Pohjola (2002) find that per capita GDP and Internet access cost explain best the growth in computer hosts per capita in a sample of countries and that investment in education also is significant in a larger sample of countries. That is, income and education promote use of the Internet. Sexton et al. (2002) find that gender, overall computer use, job-related use, and home access are important influences of Internet use and e-commerce. This finding suggests that age, gender, education, and income influence Internet use. Mills and Whitacre (2003) find that differences in household attributes, especially education and income, account for 63% of the digital divide between metropolitan and non-metropolitan areas in the U.S. Brynjolfsson and Smith (2000) find that trust and brand loyalty are important in online trades. This finding suggests that age probably influences Internet use since older people are likely to be more conscious of online security issues and hence more circumspect about online transactions.

Borenstein and Saloner (2001) state that most customers value trustworthiness, after-sales support, etc. in online transactions, suggesting that age may be an important factor in Internet use. Pohjola (2002) observes that people in rich countries, having the required infrastructure and skills, are in a much better position to benefit from the Internet than people in poor countries are. From this observation, we can infer that income and education probably affect Internet use. Hoffman and Novak (1999) observe that education, income/wealth, and gender contribute to the digital divide among races in the U.S. Shah, Kwak, and Holbert (2001) find that informational uses of the Internet are positively related to differences in production of social capital, which in turn is related to Internet use among generation X. That is, age is likely a factor influencing use of the Internet. Kraut et al. (1998) find that interpersonal communication influences use of the Internet. Anecdotal evidence suggests that online chatting is popular with very young people, suggesting that age may influence Internet use. Cheong and Park (2005) find empirically that perceived playfulness is one of the significant predictors of intention to adopt the M-Internet. This finding suggests that age probably influences Internet use. Venkatesh and Brown (2001) find that non-adopters of personal computers in U.S. homes are influenced by rapid changes in technology and the fear of obsolescence. Age, education, and income are probably reflected in the behaviour and attitudes of non-adopters of technology. Consistent with the literature, it is reasonable to consider that consumer-specific factors such as age, gender, education, and income influence use of the Internet.

3.3.3 Development of Hypotheses

Consistent with our reasoning from our theoretical model and the literature, we develop the following hypotheses for empirical tests:

H1: The higher the education level of head of household[13], the higher the level of e-buying by household
H2: The higher the level of household income, the higher the level of e-buying[14]
H3: The level of e-buying is less for households with a female as the head of household
H4: The older the head of household, the less the level of e-buying by household
H5: The level of e-buying is greater for homogeneous goods than for heterogeneous goods[15]

While Hypotheses *H1* to *H4* arise from our discussions in the earlier sections, we test Hypothesis *H5* based on the assumption of our theoretical model that the quality parameter difference s is greater (that is, more negative) for heterogeneous goods. The greater this difference, the lesser is the demand as per our model of the demand function $D(p)$.

3.3.4 Model

The following is the equation that we use for the empirical tests with three different sets of data: (1) e-buy ordered, but not paid for online, (2) e-buy ordered and paid for online, and (3) e-buy ordered, both paid for online and not paid for online. The tests were conducted for full samples[16] and then for partial samples for which there was Internet

---

[13] The survey data pertain to heads of households.
[14] Since, normally, the higher the income, the greater the consumption (and hence the higher the level of buying in general), our conclusions do not address whether if particularly high income consumers e-buy more than others do.
[15] Since we do not compare e-buying of homogeneous goods (relative to heterogeneous goods) to overall buying of homogeneous goods (relative to heterogeneous goods), our conclusions do not address whether households particularly e-buy one type of good more than they e-buy another type.
[16] For some samples the type of good ordered (homogeneous or heterogeneous) was suppressed in the public micro dataset. Such samples constitute less than 5% of the total set of samples and were excluded from estimations.

connection in the household[17]. The estimations are made for two types of regressions: Tobit and Poisson (epoisson). While both types were used for the estimations involving 'number of separate orders' as the dependent variables, only the Tobit type was used for the estimations involving 'estimated total cost (of purchases)' as the dependent variable.

3.3.4.1 Tobit Model

(309) $y = a + \sum_i b_i x_i + \mu$

where

y is (1) number of separate orders ordered online in the last 12 months, = 0

(2) estimated total cost, in Canadian dollars, of online purchases made in the last 12 months, = 0

i = 1, 2, …, 13

x_1 = 1 if a homogeneous good was ordered; 0 if not

x_2 = 1 if a heterogeneous good was ordered; 0 if not

x_3 = 1 if the age of the head of the household sample is of age group 2 (35 – 54 years); 0 if not

x_4 = 1 if the age of the head of the household sample is of age group 3 (55 – 64 years); 0 if not

x_5 = 1 if the age of the head of the household sample is of age group 4 (65 +); 0 if not

x_6 = 1 if the head of the household sample is female; 0 if not

x_7 = 1 if the education level of the head of the household sample is high school or some college; 0 if not

x_8 = 1 if the education level of the head of the household sample is university degree; 0 if not

[17] It may be noted that a household that has no Internet connection at home may still e-buy from a location outside home; this can be seen from, for example, the HIUS 2003 data.

x_9 = 1 if the income level of the household sample is in the quartile 2; 0 if not[18]
x_{10} = 1 if the income level of the household sample is in the quartile 3; 0 if not
x_{11} = 1 if the income level of the household sample is in the quartile 4; 0 if not
x_{12} = 1 if the type of the connection of the household sample is a type other than a telephone line; 0 if not
x_{13} = 1 if the household sample has no connection; 0 if not
μ = error term, normally distributed

3.3.4.2 Poisson Regression Model

(310) $\Pr(Y_t) = \dfrac{l^{Y_t} e^{-l_t}}{Y_t!}$ for $Y_t = 0,1,2,\ldots$and

(311) $?_t = e^{X_t' b}$

where Y_t is *y* as in the Tobit model. The exponent of the right-hand side of (311) is akin to the right-hand side of (309). $?_t$ is the parameter of the Poisson probability distribution that is influenced by the explanatory variables X_t, all of which are the same as in the Tobit model. This is a non-normal model, unlike the Tobit model since errors are not assumed to be normally distributed. We used this model for the empirical tests of the equations for the dependent variables 'number of separate orders' only, not for those involving the dependent variables 'estimated total cost (of purchases).'

[18] The income ranges for the quartiles increased over the years; the ranges for 2003 (1999) were for quartile 1: < $ 24,001 (< $ 20001), quartile 2: $ 24,001 - $ 43,999 ($ 20,001 - $ 35,999), quartile 3: $ 44,000 - $ 69,999 ($ 36,000 - $ 59,999), and quartile 4: $ 70,000 plus ($ 60,000 plus).

Chapter 4
Data and Method

4.1 Introduction

Our research uses the Household Internet Use Survey (HIUS) datasets of Statistics Canada for empirical tests of the hypotheses. We need e-commerce data pertaining to the level of e-buying activity by households and types of goods ordered by households; demographic data such as age, gender, and education of individual consumers representing households; and technical data such as Internet connection type to carry out the necessary empirical tests. All these relevant data are contained in the HIUS data. The purpose of our empirical model is to test our hypotheses. Since one of our dependent variables, namely the 'number of separate orders' ordered, is count-type we use Tobit and Poisson regression techniques for estimations involving this dependent variable. The other dependent variable, namely the 'estimated total cost (of purchases)', is positive or zero but continuous (and not count-type), so we use only the Tobit regression technique for estimations involving this dependent variable.

4.2 Data

We empirically test our hypotheses using detailed data on the Internet activities of Canadian households, collected by the Science, Innovation, and Electronic Information Division of Statistics Canada for 1999 – 2003[19] through annual surveys known as the Household Internet Use Survey (HIUS)[20]. This survey reports on Canadians using the Internet and measures the extent of their use, location of use, frequency of use, and their reasons for using or not using the Internet. The HIUS has been conducted since1997 and has evolved to capture increasingly more detail. In 1999, data on electronic commerce (e-commerce) from home were provided. The 2003 survey examined Canadian households' access to the Internet at home, in the workplace, and other locations such as public

[19] The data of 2002 were not used in the regressions due to issues (the variable "quartile" had values 0 to 9, instead of 1 to 4)
[20] Most of the content of this section is extracted from the HIUS 2003 User Guide published by Statistics Canada.

libraries, schools / universities, and Internet cafés. The collected data reveal relationships between Internet use and household income, location of use, and demographic factors such as age and education. The detailed questions dealing with household e-commerce that were introduced in 1999 was repeated each year thereafter until 2003.

The objectives of the HIUS survey are, among others, to gain a better understanding of how Canadian households use the Internet, identify the types of Internet services used at home, find reasons for non-usage of the Internet, determine what factors would induce households to start using the Internet, understand the impact of the Internet on purchases of goods and services, etc. In assessing the use of the Internet, Statistics Canada has measured the accessibility of the Internet from different locations as well as the frequency and intensity of use from home.

The HIUS survey datasets published by Statistics Canada contain data directly collected from the HIUS as well as data derived from another source: the Labour Force Survey (LFS). Demographic and employment data collected through the LFS were added to the HIUS household data. The LFS and HIUS data were collected from the same households, but not all households surveyed for the LFS were surveyed for the HIUS. For example, the total number of households surveyed for the HIUS 2003 is 23,113 while that for the LFS is 34,674. The data were collected through computer-assisted telephone surveys.

An important aspect of the HIUS to be noted for our empirical tests is the need to use population weights (provided through the variable WTHP). The weight applied to each observation is the total number of Canadian households represented by that sample. The regressions thus represent population estimates for the total number of households in Canada.

The data are available through the Tri-University Data Resources (TDR) website or through the new Nesstar website. The old website is no longer updated, but the full HIUS datasets of years 1997 to 2003 are available under the data group “Communications.”

Either a full set or a sub-sample of observations can be downloaded, based on categories such as province, gender, etc. Furthermore, all variables (columns that pertain to questions in the questionnaire) or a subset of variables can be downloaded for each observation.

4.3 Method

4.3.1 Tobit Regression

If we were to use Ordinary Least Squares (OLS) for estimation of our empirical models, we may obtain coefficient estimates that are unrealistic to our Data Generating Process (DGP). Our DGP has dependent variables taking values of zero or positive real numbers, representing a limited dependent variable. The number of separate orders placed in a period cannot be negative. Furthermore, the total dollar value of transactions cannot be negative. Since the OLS estimate can return negative dependent variable values for a given set of coefficient estimates and a set of values for explanatory variables, its coefficients estimates are biased and inconsistent. For example, our initial tests used OLS and we obtained counter-intuitive results.

Tobit regression is suitable to our requirement because it is specifically designed to account for a truncated sample. It is a special type of probit model, propounded by economist Tobin (Tobin's probit). It provides Maximum Likelihood (ML) estimates. The structure of the Tobit model is as follows:

$$Y_t^* = X'_t ß + e_t$$
$$Y_t = 0 \qquad \text{if } Y_t^* = 0$$
$$Y_t = Y_t^* \qquad \text{if } Y_t^* > 0$$

A limit value other than zero can be specified, which is however unnecessary in our case; our limits for the number of orders and the dollar value of transactions are zero. In the above structure of the Tobit model, Y_t^* is a latent variable, not the dependent variable. Y_t

is the dependent variable. The model returns estimates for coefficients of the explanatory variables such that the dependent variable's estimate is always zero or positive, never negative. Since the Tobit model is an extension of the probit model, its error term e is normally distributed as is the error term of the probit model. The dependent variable's estimate is a continuous value. Tobit models can be executed in SHAZAM using the command

TOBIT *depvar indeps /options*

where *depvar* is the dependent variable, *indeps* is a list of independent variables, and *options* is a list of desired options. The options WEIGHT= and NONORM[21] applicable to the OLS procedure are also applicable to Tobit models. In our case, we have to use the weights of individual samples of the HIUS survey to generalize to the entire population. The option NONORM ensures that the weighting is not normalized so that weighting is for the purpose of replicating the samples. The option REPLICATE, which is applicable to OLS, is not valid for Tobit and is in fact unnecessary. The option NONORM is sufficient to ensure replication.

The Tobit model is also known as a "censored regression model." The censored model is applicable in two-stage decision situations. In our case, the first-stage decision for a household is whether to e-buy or not. If it decides to e-buy, it proceeds to the second-stage decision of how much to e-buy. Thus the Tobit model seems appropriate to our problem.

4.3.2 Poisson Regression

In our case, the dependent variable the 'number of separate orders' ordered in the last twelve months by a household, can take only count values, that is values 0,1,2,…, and so on. This dependent variable cannot take negative or fractional values. It takes only

[21] With this option, SHAZAM does not normalize weights, since we weight in order to replicate the grouped survey data, not to correct for heteroskedasticity.

discrete positive values. Similar applications include the number of patents received by a firm in a year, the number of customers arriving at a bank teller every five minutes, etc. In such cases, the assumption of normally distributed error terms may be inappropriate[22]. The probability distribution specifically suited for count data is a Poisson probability distribution. Poisson regression does away with the assumption of normally distributed errors. Hence it is a non-normal model; however, the mean of the distribution is equal to the variance. We use Poisson regression technique for the estimations involving the above dependent variable, in addition to Tobit technique.

The Poisson regression model has the following structure:

$\Pr(Y_t) = \dfrac{l^{Y_t} e^{-l_t}}{Y_t!}$ for $Y_t = 0,1,2,\ldots$and

$?_t = e^{X_t^{'} b}$

The above model is called EPOISSON in SHAZAM and it is executed using the Maximum Likelihood Estimate (MLE) command as follows:

MLE *depvar indeps /*TYPE=EPOISSON

If we use TYPE=POISSON, a different kind of Poisson regression in which

$?_t = X'_t$ß

is used. In this case, the log-likelihood function is not defined for values of X'_tß < 0. Hence, especially when we have negative X'_tß, TYPE=EPOISSON is preferable to TYPE=POISSON. In our case, we have a large number of negative X'_tß in almost all the estimations, so we use TYPE=EPOISSON[23]. Poisson regression has the OLS option

[22] SHAZAM User's Reference Manual, Version 10, p239.
[23] X'_tß (or simply XBs) are products of the observed values of independent variables and their estimated coefficients. Thus an XB is the portion of the estimate of dependent variable that is explained by the independent variable X.

WEIGHT= and NONORM, which we use in our estimations since our data are grouped. Poisson regression does not have the REPLICATE option as OLS does, but the NONORM option ensures replication of samples.

Chapter 5

Results

5.1 Introduction

This chapter summarizes in tabular form the estimation results with respect to individual hypotheses. These tables are in non-standard format, but help analyze the results against the hypotheses. Before reading these summary tables, please review the description of the variables, the descriptive statistics for the variables, and the regression results that are appended in standard format. Any notable points of the estimations are mentioned in the chapter. After the presentation of results, we analyze the same to draw conclusions about our hypotheses. We finally list the issues we have with our theoretical model and the issues we faced in the empirical tests.

5.2 Results

5.2.1 Hypothesis I

H1: The higher the education level of head of household, the higher the level of e-buying by household

Table 5.2.1: Results of Hypothesis I

Equation	Sample	Type of Regression	EDU 2[24]	EDU 3	Result
1[25]	Full	Tobit[26]	0.637 ***	0.808 ***	Not rejected[27]

[24] EDU 1: Less than high school; EDU 2: High school or some college; and EDU 3: University degree. The equation skips EDU 1 and keeps the other two binary variables out of the three. Significance levels are: *** 1%, ** 5%, and * 10%

[25] The equations are distinguished mainly with respect to the dependent variable as follows: Eq. 1 - number of separate orders, ordered online but not paid for online; Eq. 2 – estimated total cost of purchases, ordered online but not paid for online; Eq. 3 – number of separate orders, ordered online and paid for online; Eq. 4 – estimated total cost of purchases, ordered online and paid for online; Eq. 5 – number of separate orders, ordered online (whether paid for online or not); and Eq. 6 – estimated total cost of purchases, ordered online (whether paid for online or not).

2	Full	Tobit	177.3 ***	294.1 ***	Not rejected
3	Full	Tobit	1.326 ***	1.76 ***	Not rejected
4	Full	Tobit	141.9 ***	263.3 ***	Not rejected
5	Full	Tobit	1.858 ***	2.407 ***	Not rejected
6	Full	Tobit	231.3 ***	391.9 ***	Not rejected
1	Sub	Tobit	0.153 ***	0.168 ***	Not rejected
2	Sub	Tobit	74.12 ***	164.9 ***	Not rejected
3	Sub	Tobit	1.336 ***	1.823 ***	Not rejected
4	Sub	Tobit	99.72 ***	222.8 ***	Not rejected
5	Sub	Tobit	1.112 ***	1.504 ***	Not rejected
6	Sub	Tobit	112.5 ***	248.6 ***	Not rejected
1	Full	Poisson	0.091 ***	0.141 ***	Not rejected
3	Full	Poisson	0.205 ***	0.243 ***	Not rejected
5	Full	Poisson	0.238 ***	0.302 ***	Not rejected
1	Sub	Poisson	0.011 insig	0.040 insig	Reject
3	Sub	Poisson	0.183 ***	0.254 ***	Not rejected
5	Sub	Poisson	0.138 ***	0.200 ***	Not rejected

The higher the education level, the higher the e-buying with e-pay, and the higher the e-buying with or without e-pay (that is, overall e-buying). In the case of e-buying without e-paying (equation 1), such a relationship is rejected by Poisson regression on the sub-samples. Those with higher levels of education may be willing to e-pay. The evidence, as a whole, is consistent with *H1*.

[26] For Tobit regressions, we present regression coefficients, not normalized coefficients.
[27] If the coefficients of EDU3 and EDU2 are significant, and that of EDU3 is higher than that of EDU2, the hypothesis is not rejected; otherwise, it is rejected.

5.2.2 Hypothesis II

H2: The higher the level of household income, the higher the level of e-buying

Table 5.2.2: Results of Hypothesis II

Equa-tion	Sample	Type of Regression	QUART2[28]	QUART3	QUART4	Result
1	Full	Tobit	0.389 ***	1.095 ***	0.588 ***	Reject[29]
2	Full	Tobit	30.96 ***	117.1 ***	338.6 ***	Not rejected
3	Full	Tobit	0.686 ***	1.126 ***	1.03 ***	Reject
4	Full	Tobit	97.23 ***	178.5 ***	331.7 ***	Not rejected
5	Full	Tobit	0.918 ***	1.773 ***	1.745 ***	Reject
6	Full	Tobit	113.7 ***	236.3 ***	478.2 ***	Not rejected
1	Sub	Tobit	0.095 ***	0.709 ***	0.321 ***	Reject
2	Sub	Tobit	-19.6 ***	33.19 ***	283.2 ***	Reject
3	Sub	Tobit	0.433 ***	1.086 ***	0.917 ***	Reject
4	Sub	Tobit	78.23 ***	181.7 ***	337.7 ***	Not rejected
5	Sub	Tobit	0.424 ***	1.221 ***	1.161 ***	Reject
6	Sub	Tobit	60.04 ***	164 ***	409.6 ***	Not rejected
1	Full	Poisson	0.16 ***	0.268 ***	0.102 ***	Reject
3	Full	Poisson	0.169 ***	0.176 ***	0.237 ***	Not rejected
5	Full	Poisson	0.245 ***	0.287 ***	0.295 ***	Not rejected
1	Sub	Poisson	0.051 *	0.148 ***	0.008 insig	Reject
3	Sub	Poisson	0.124 ***	0.155 ***	0.200 ***	Not rejected
5	Sub	Poisson	0.123 ***	0.165 ***	0.179 ***	Not rejected

From the above results, we can observe the following with regard to Hypothesis *H2*:

[28] QUART1: = $ 24,000; QUART2: $ 24,001 - $ 43,999; QUART3: $ 44,000 - $ 69,999; and QUART4: $ 70,000 +

[29] If the coefficients of QUART2, QUART3, and QUART4 are significant and progressively higher in that order, the hypothesis is not rejected; otherwise, it is rejected.

1. For the full samples and the sub-samples, the Poisson regression results do not reject the hypothesis, in the case of number of separate orders ordered, for (a) e-buy with e-pay and (b) e-buy with or without e-pay; but they reject the hypothesis for e-buy without e-pay.
2. For the full samples, the Tobit regression results do not reject the hypothesis, in the case of the estimated total cost (of purchases).

From the above, we can conclude that, across the entire population of households, the higher the income level, the higher the level of e-buying (in terms of dollar value).

5.2.3 Hypothesis III

H3: The level of e-buying is less for households with a female as the head of household

Table 5.2.3: Results of Hypothesis III

Equation	Sample	Type of Regression	FEMALE	Result
1	Full	Tobit	-0.702 ***	Not rejected[30]
2	Full	Tobit	-59.31 ***	Not rejected
3	Full	Tobit	-0.751 ***	Not rejected
4	Full	Tobit	-92.27 ***	Not rejected
5	Full	Tobit	-0.894 ***	Not rejected
6	Full	Tobit	-90.29 ***	Not rejected
1	Sub	Tobit	-0.733 ***	Not rejected
2	Sub	Tobit	-69.85 ***	Not rejected
3	Sub	Tobit	-1.245 ***	Not rejected
4	Sub	Tobit	-128.1 ***	Not rejected
5	Sub	Tobit	-1.243 ***	Not rejected

[30] If the coefficient of FEMALE is significant and negative, the hypothesis is not rejected; otherwise, it is rejected.

6	Sub	Tobit	-122.6 ***	Not rejected
1	Full	Poisson	-0.331 ***	Not rejected
3	Full	Poisson	-0.175 ***	Not rejected
5	Full	Poisson	-0.225 ***	Not rejected
1	Sub	Poisson	-0.35 ***	Not rejected
3	Sub	Poisson	-0.279 ***	Not rejected
5	Sub	Poisson	-0.275 ***	Not rejected

Both the Poisson regression results and the Tobit regression results do not reject Hypothesis *H3*.

5.2.4 Hypothesis IV

H4: The older the head of household, the less the level of e-buying by household

Table 5.2.4: Results of Hypothesis IV

Equation	Sample	Type of Regression	AGE 2[31]	AGE 3	AGE 4	Result
1	Full	Tobit	-0.145 ***	0.062 ***	-1.288 ***	Reject[32]
2	Full	Tobit	-25.43 ***	1.77 insig	-90.26 ***	Reject
3	Full	Tobit	-0.725 ***	-0.424 ***	-1.608 ***	Reject
4	Full	Tobit	-90.54 ***	-53.71 ***	-160.2 ***	Reject
5	Full	Tobit	-0.754 ***	-0.426 ***	-2.527 ***	Reject
6	Full	Tobit	-86.91 ***	-65.87 ***	-251.3 ***	Reject
1	Sub	Tobit	-0.039 ***	0.144 ***	-0.497 ***	Reject
2	Sub	Tobit	4.259 insig	42.86 ***	126.2 ***	Reject

[31] AGE 1: < 35 years; AGE 2: 35 – 54 years; AGE 3: 55 – 64 years; and AGE 4: 65 + years
[32] If AGE2, AGE3, and AGE4 are significant and progressively less in that order, the hypothesis is not rejected; otherwise, it is rejected.

3	Sub	Tobit	-0.623 ***	-0.065 ***	-1.0 ***	Reject
4	Sub	Tobit	-63.44 ***	-20.11 ***	-48.67 ***	Reject
5	Sub	Tobit	-0.497 ***	0.109 ***	-0.961 ***	Reject
6	Sub	Tobit	-40.70 ***	24.70 ***	9.496 ***	Reject
1	Full	Poisson	-0.150 ***	-0.123 ***	-0.492 ***	Reject
3	Full	Poisson	-0.103 ***	-0.087 ***	-0.24 ***	Reject
5	Full	Poisson	-0.103 ***	-0.074 ***	-0.335 ***	Reject
1	Sub	Poisson	-0.128 ***	-0.088 ***	-0.283 ***	Reject
3	Sub	Poisson	-0.090 ***	-0.045 ***	-0.141 ***	Reject
5	Sub	Poisson	-0.086 ***	-0.032 ***	-0.165 ***	Reject

We may note from the above results that Hypothesis *H4* is rejected. We may hence conclude that we do not find empirical evidence for the hypothesis that the older the head of household, the less the level of e-buying by household.

5.2.5 Hypothesis V

H5: The level of e-buying is greater for homogeneous goods than for heterogeneous goods[33]

[33] We categorize different kinds of goods noted in the HIUS 2003 survey as either “homogeneous” or “heterogeneous” as follows: homogeneous goods – computer software, computer hardware, Music (CDs, tapes, and MP3), books and magazines, videos and DVDs, consumer electronics, and travel arrangements; heterogeneous goods – other entertainment products, “food, condiments, and beverages,” “health, beauty, and medical,” clothing and jewelry, house wares, automotive, flowers and gifts, sports equipment, toys and games, real estate, crafts and hobbies, other household related items, “other, , and renovations,” and “other, specify.”

Table 5.2.5: Results of Hypothesis V

Equation	Sample	Type of Regression	HOMOGENEOUS[34]	HETEROGENEOUS	Result
1	Full	Tobit	19.72 ***	20.56 ***	Reject[35]
2	Full	Tobit	4121 ***	4059 ***	Not rejected
3	Full	Tobit	23.9 ***	21.82 ***	Not rejected
4	Full	Tobit	3408 ***	2931 ***	Not rejected
5	Full	Tobit	22.31 ***	21.14 ***	Not rejected
6	Full	Tobit	3305 ***	2939 ***	Not rejected
1	Sub	Tobit	19.41 ***	20.25 ***	Reject
2	Sub	Tobit	4092 ***	4027 ***	Not rejected
3	Sub	Tobit	22.73 ***	20.93 ***	Not rejected
4	Sub	Tobit	3252 ***	2791 ***	Not rejected
5	Sub	Tobit	21.63 ***	20.37 ***	Not rejected
6	Sub	Tobit	3225 ***	2844 ***	Not rejected
1	Full	Poisson	1.866 ***	3.033 ***	Reject
3	Full	Poisson	2.011 ***	2.103 ***	Reject
5	Full	Poisson	1.596 ***	1.954 ***	Reject
1	Sub	Poisson	1.834 ***	2.906 ***	Reject
3	Sub	Poisson	1.74 ***	1.916 ***	Reject

[34] We include both the "homogeneous" and "heterogeneous" variables, since households may order both types of goods, so these two variables are not mutually exclusive. The actual names of the variables in the regressions are "homobuy" and "heterobuy" for equations 1 and 2, "homopay" and "heteropay" for 3 and 4, and "homogood" and "heterogood" for 5 and 6, respectively.

[35] If the coefficients of the variables HOMOGENEOUS and HETEROGENEOUS are significant and that of the former is more than that of the latter, the hypothesis is not rejected; otherwise, it is rejected.

5	Sub	Poisson	1.547 ***	1.861 ***	Reject

From the above results, we may note the following with regard to Hypothesis *H5*:

1. For the full samples, which consist of almost the entire set of observations of the surveys, and the sub-samples, which consist of only the households with an Internet connection, the Tobit results do not reject the hypothesis, in the case of estimated total cost of orders in dollars.
2. For the full samples and the sub-samples, the Poisson regression results reject Hypothesis *H5* and the Tobit results are mixed, in the case of number of separate orders ordered online.

Thus from the above empirical results, we find evidence for Hypothesis *H5* that the level of e-buying is greater for homogeneous goods than for heterogeneous goods (in terms of dollar value).

It is possible that homogeneous goods are ordered for larger dollar value with fewer orders compared to heterogeneous goods. However, we would advise caution in concluding so, since the results of the Tobit models and the Poisson models do not exactly agree.

5.3 Analysis

We can summarize the findings of the empirical tests as follows:

H1: The higher the education level of head of household, the greater is e-buying with e-paying, and the greater is overall e-buying. In the case of e-buying without e-paying (equation 1), such a relationship is rejected. Those with higher levels of education may be willing to e-pay; they may have better knowledge of latest developments in online security or they may know better how to safeguard themselves online. Another aspect to note is that a higher income level is normally associated with a higher education level, so income may be the real driver of e-buying, rather than education. However, we have controlled for income level in the regressions. Hence we conclude that the evidence, as a whole, is consistent with *H1* that the higher the education level of head of household, the greater the e-buying by household irrespective of income level.

H2: From the empirical tests we conclude that, across the entire population of households, the higher the level of household income, the higher the level of e-buying (in terms of dollar value). We may note that it is normal for consumption and buying to increase as household income increases. Our results may simply reflect this fact rather than imply that e-buying as a proportion of overall buying is higher for higher-income households.

H3: Both the Poisson regression results and the Tobit regression results do not reject Hypothesis *H3* that the level of e-buying is less for households with a female as the household head. It may be noted that a single-parent household is likely to have a female as its head. The single-parent household income level is most likely low, leading to less e-buying. Since we have controlled for income in the regressions, we may conclude that households with a female as the head of household e-buy less, irrespective of income level.

H4: We do not find any unambiguous empirical evidence for the hypothesis that the older the head of household, the less the level of e-buying by household. Hence we reject Hypothesis *H4*.

H5: We found that the Tobit results do not reject the hypothesis, in the case of estimated total cost of orders in dollars. We also found that the Poisson regression results reject Hypothesis *H5* and the Tobit results are mixed, in the case of number of separate orders ordered online. Thus from the above empirical results, we find evidence for Hypothesis *H5* that the level of e-buying is greater for homogeneous goods than for heterogeneous goods (in terms of dollar value).

It is possible that homogeneous goods are ordered for larger dollar value with fewer orders compared to heterogeneous goods, since goods under the category "heterogeneous" include frequent-purchase items such as food and clothing. However, we would advise caution in concluding so, since the results of the Tobit models and the Poisson models do not agree. Moreover, our classification of different goods into homogeneous goods and heterogeneous goods may be imprecise.

Our theoretical model suggests that homogeneous goods may be more readily bought online. The quality parameter of homogeneous goods is almost the same for e-buying as for non-ebuying ($s_e \sim s_n$) whereas that of heterogeneous goods is lower for e-buying ($s_e < s_n$). Though we find evidence that homogeneous goods are ordered more than the heterogeneous goods (in terms of dollar value), we need to compare online buying of homogeneous goods (relative to heterogeneous goods) to conventional or overall buying of the same (relative to heterogeneous goods), in order to test this implication of the theoretical model.

5.4 Issues

5.4.1 Issues in the Theoretical Model

The key aspect of our theoretical model is the use of the variable ?, the taste parameter of an individual consumer. Empirical data for many individual-specific factors such as the demographic factors age, gender, education, and income are available from the Labour Force Survey (LFS), the Household Internet Use Survey (HIUS), and other such statistics collected by Statistics Canada and other sources. We build our thesis on the premise that the taste parameter ? of our model is probably determined by a combination of such factors. Our theoretical model does not suggest what factors determine the taste parameter; it does not spell out any specific relationship of the taste parameter ? to the individual-specific factors; and it does not express the parameter in terms of these factors. In view of this shortcoming, empirically testing the model using available statistics is difficult. We overcome this shortcoming by making use of the literature in conjunction with the theoretical model to develop hypotheses for empirical tests. More precise or more pertinent hypotheses would be possible if we express the taste parameter ? in terms of its determinants.

Another issue with the theoretical model is that we do not suggest specific distributions for *F*(?), which is the distribution of the taste parameter ?. Research into the determinants of the taste parameter might help determine possible distributions. Instead of directly considering the distribution of the taste parameter, we look at the nature of the available variables for the demand function *D*(*p*) of the theoretical model. The variables "number of separate orders" and "estimated total cost" of e-buying, for which data are readily available in the HIUS 2003 dataset, are always positive or zero. Hence we chose a Tobit regression for empirical tests. Since "number of separate orders" is essentially a count-type variable, taking discrete values 0,1,2,…, normality assumption for the error terms of the regression models may be inappropriate, so we also chose Poisson regression models for tests involving this variable.

5.4.2 Issues in the Empirical Tests

The empirical tests often returned inconclusive results in the case of e-buying without e-paying, while we often had intuitive results for e-buying with e-paying as well as for all e-buying (with and without e-paying). One particular point to note is that for about 1% of the total sample size of 23113 for 2003, the data as to whether a homogeneous good was ordered or a heterogeneous good was ordered by an e-buying household are suppressed in the public use micro-data file of the HIUS 2003. Although only 6 observations had this data suppressed in the case of e-buying with e-paying, 228 observations had this data suppressed in the case of e-buying without e-paying. We had to exclude these observations from our empirical tests. Similar instances occurred for other years (though less severely). Though only a few hundred samples were involved, in the empirical tests of only e-buying without e-paying, the reduction in the number of influential samples would have been much more, proportionally. This reduction probably led to biased estimates for the coefficients. If we had had a complete set of observations for empirical tests of e-buying without e-paying, we might have obtained intuitive results.

Another point to note from the empirical results is that we often obtained inconclusive results for tests that had “number of separate orders” as the dependent variable, but we often obtained clear and intuitive results for tests that had “estimated total cost” as the dependent variable. We suggest that the respondents to the HIUS surveys might be less precise with respect to their responses about the number of online orders they placed relative to how much they paid. The amount of money spent e-buying was easier to retain than the number of times they ordered online.[36]

Another issue with the HIUS 2003 public micro-data set is that the question pertaining to a high-speed connection is suppressed. Data from other questions concerning telephone line connection and other types of connections do not lend themselves to deriving data about the speed of connection, since such connections can be either high-speed or 56 Kbps (56 kilobits per second is the standard for dial-up modems). Connection speed can be associated with search cost (among other things, such as a high-intensity Internet user), but we are not able to test whether households with a high-speed connection e-buy more than those without one.

[36] It may be noted that a survey respondent quickly responds from memory during a conversation with a telephone surveyor.

Chapter 6
Conclusion

6.1 Approach

Our approach and results need to be discussed in the context of our research objective, which was to explore factors that influence e-buying adoption by the Canadian public. To address the objective, we first built a simple theoretical model in Chapter 2 by adapting from Tirole (1989). The model assumed a single good, one unit of which is bought by every consumer (household) in a population. Consumers can either e-buy or buy conventionally from a local store. A good's quality is captured by a quality parameter, and consumer taste is captured by a taste parameter. These parameters are themselves determined by many other factors. The taste parameter is influenced by individual-specific factors. The taste parameter is distributed in a certain way in a population. Since search, transportation costs, and market reach are different for e-buying and conventional buying, final prices are expected to be different. Individuals maximize their utilities, so a consumer buys a good if he or she derives utility from the purchase; he e-buys if he derives greater utility from e-buying than from conventional buying. There are N consumers in the population; the demand for the good is derived in terms of N, price and cost, the quality parameter, and the distribution function of the taste parameter. We then differentiated possible goods as homogeneous or heterogeneous. The model implies that the distribution of the demand of the good is essentially that of the taste parameter in the population of consumers (or households), which in turn is influenced by price, search cost, and the quality parameter.

We then discussed the associations among price, search cost, and the quality parameter and individual-specific demographic factors such as age, gender, education, and income, which are central to our research question. Our premise is that the taste parameter of a consumer is influenced by such individual-specific factors. That is, the demand for the good is affected ultimately by individual-specific factors that shape individuals' tastes.

We did not, however, derive the taste parameter (and hence the demand function) directly in terms of demographic factors.

We then extracted the results of the literature that pertain to demographics and other determinants of technology use by households or individual consumers. The literature helped us to identify potential factors influencing the taste parameter of individuals. Reference to the literature was necessary in the absence of direct derivation of our theoretical model from the demographic variables.

Although the literature provides evidence that education and income influence household adoption of computers and the Internet, past research also weakly suggests that gender and age may also affect technology adoption by individuals. Based on these observations and from the implications of our theoretical model, we developed hypotheses, associating households' e-buying activities and demographic variables, for empirical tests. Especially based on our theoretical model, we developed a hypothesis associating the level of e-buying activity with the type of good. We did not test any hypothesis concerning the type of connection a household has and its e-buying intensity (though it formed a part of our research question), since we do not have data pertaining to connection type in the HIUS 2003 public use micro-data set.

We tested our hypotheses empirically using the Household Internet Use Survey data collected by Statistics Canada pertaining to the years 1999, 2000, 2001, and 2003. The dependent variables of our empirical models, namely "number of separate orders (ordered online)" and "estimated total cost (of online purchases)," are limited dependent variables. Since they are positive or zero, we used Tobit regression suitable for censored models. Since they are also count-type data, taking discrete values 0, 1, 2, …, the normality assumption of error terms may be inappropriate. Hence we assumed a Poisson probability distribution for the dependent variables and carried out the empirical tests using Maximum Likelihood Estimation with Poisson type errors. For Poisson type regression, we adopted the EPOISSON regression, in which the Poisson parameter ? is itself exponentially influenced by the explanatory variables, instead of the POISSON regression, in which ? is linearly influenced. We did so because we had many negative XBs in the estimations; in such cases, the POISSON regression is inappropriate since its log-likelihood function is not defined for negative XBs. In other words, we assumed at the empirical-test stage that the taste parameter ? is distributed in a population as per Poisson probability. We did not make any specific assumption about the distribution of ? while we developed our theoretical model. We found empirical evidence consistent with three (*H1*, *H2*, and *H3*) of our five hypotheses.

6.2 Results Revisited

The empirical test results are consistent with our hypotheses that (1) the higher the education level of head of household, the higher the level of e-buying by household, (2) the higher the level of household income, the higher the level of e-buying, (3) the level of e-buying is less for households with a female as the head of household, and (4) the level e-buying is greater for homogeneous goods than for heterogeneous goods; and the results reject the hypothesis that the older the head of household, the less the level of e-buying by household.

The above results are irrespective of income level, since we have controlled the estimations for income level. The result that "the higher the income, the higher the e-buying" does not necessarily mean that high-income households e-buy more. This result may simply reflect that consumption (and hence overall buying) is more for households with higher incomes.

6.3 Limitations of the Study

6.3.1 Limitations of the Theoretical Model

Our theoretical model does not determine what factors determine the taste parameter; it does not spell out any specific relationship of the taste parameter ? to the individual-specific factors; and it does not express the parameter in terms of these factors. More precise or more pertinent hypotheses would be possible if we express the taste parameter ? in terms of its determinants.

Another limitation of the theoretical model is that it does not suggest specific distributions for *F*(?) - the distribution of the taste parameter ?. However, the model allows for a range of distributions to be considered, which we did in our empirical tests. If the true distribution is suggested by the theoretical model, hypotheses development and empirical tests would be easier. It would prevent mistakes in method selection for empirical tests and hence help us to arrive at right conclusions.

The above two limitations may be linked, since identification of the determinants of the taste parameter may help determine the true distribution of the parameter. Knowledge of the distribution will help firms in their online marketing. The theoretical model is limited also in its treatment of the quality parameter *s*. Further insights about the quality parameter will help firms find ways to improve online sales of goods, especially heterogeneous goods.

6.3.2 Limitations of the Empirical Tests

Limitations of the empirical tests mostly pertain to limitations with data. Some data of the HIUS public use micro-data sets are suppressed reducing the utility of other data available. The estimations are hampered seriously because of this limitation, and the coefficient estimates obtained in the tests may be biased to that extent. We list below some instances of suppressed data affecting our empirical tests.

For about 1% of the total sample size of 23113 for 2003, the data as to whether a homogeneous good was ordered or a heterogeneous good was ordered by an e-buying household are suppressed in the public use micro-data file of the HIUS 2003. Similar suppression was found with the other years. This limitation, mainly in the case of e-buying without e-paying, forced us to exclude these observations from our empirical tests; this exclusion probably led to biased estimates for the coefficients.

Data as to whether a household's Internet connection is a high-speed connection are suppressed in the HIUS public use micro-data. This suppression limits us from testing if households with high-speed connections e-buy more than those without. Some key pieces of data are not directly collected in surveys; some are not derived and reported in the public use micro-data set. Users of data have to derive such data, leading perhaps to mistakes and hence biased estimates. For example, in our case, we had to infer which observations (sample households) had any type of Internet connection from the data as to whether a household used the Internet at home in a typical month. Our inference could be an under-estimate.

6.4 Conclusions and Implications for Future Research

The research has presented a simple model of a demand function for e-buying in terms of a taste parameter distributed in some way in a population of consumers. It also expresses the taste parameter in terms of price and cost advantages of e-buying, and the quality parameter differential of the good between e-buying and conventional buying. It facilitates analysis of e-buying demand with respect to the type of good (homogeneous or heterogeneous). It allows for a range of distributions to be assumed for the taste parameter. However it does not pinpoint the individual-specific determinants of the taste parameter; this fact limits our ability to conduct empirical tests.

The empirical tests of the hypotheses, developed with the aid of the literature in conjunction with the theoretical model, show that the level of e-buying is greater for households with higher levels of income and education, and less for households with females as their heads. The methods used for empirical tests suggest that the taste parameter ? may be distributed in a population as per Poisson probability distribution.

Further research may investigate the determinants of the taste parameter and the quality parameter. Identification of the determinants of the taste parameter may help pinpoint its true distribution. Such knowledge would help firms plan their online sales and marketing of goods, promote the use of the Internet, and promote e-buying by households; greater use of the Internet and greater levels of e-buying will increase overall welfare.

This research focussed on the determinants of e-buying by households; future research may investigate diffusion path of the use of the Internet and e-buying by Canadian households; estimate the parameters of origin, slope, and ceiling of the sigmoid curve; and identify the determinants of these parameters. Such knowledge may help public authorities and other interested parties plan for the accelerated diffusion of similar technologies in the future, in Canada and elsewhere.

REFERENCES

Bakos, Y. 2001. "The Emerging Landscape for Retail E-Commerce." *The Journal of Economic Perspectives*, Vol. 15, No. 1 (Winter, 2001), pp. 69-80.

Bakos, Y. 1997. "Reducing Buyer Search Costs: Implications for Electronic Marketplaces." *Management Science*, Vol. 43, No. 12 Frontier Research on Information Systems and Economics (Dec., 1997), pp. 1676 – 1692.

Banerjee, A. V. 1992. "A Simple Model of Herd Behavior." *The Quarterly Journal of Economics,* Vol. 107, No. 3. (Aug., 1992), pp. 797 - 817.

Borenstein S. and Saloner G. 2001. "Economics and E-commerce." *The Journal of Economic Perspectives*, Vol. 15, No. 1 (Winter, 2001), pp. 3-12.

Brynjolfsson, E. and Smith, Michael D. 1999. "Frictionless Commerce? An Exploratory Analysis of Internet Pricing Behavior." Retrieved August 10, 2005, from *http://ecommerce.mit.edu/papers/friction*

Cheong, J. H., and Park, M-C. "Mobile internet acceptance in Korea." *Internet Research,* Vol. 15 No. 2, 2005, pp. 125 – 140.

Chwelos, P., Benbasat, I. and Dexter, A.S. 2001. "Research Report: Empirical test of an EDI adoption model," *Information Systems Research* 12(3), 304-321.

Cline, R.J., and T.S. Neubig. 1999. "The sky is not falling: Why state and local revenues were not significantly impacted by the internet in 1998," Ernst & Young Economic Consulting and Quantitative Analysis discussion paper, June, 18, 1999.

Economist, The. 2000. "Survey on globalization and tax," January 29, 2000, 3-22.

Forman, C. 2005. “The Corporate Digital Divide: Determinants of Internet Adoption.” *Management Science*, Vol. 51, No. 4, April 2005, pp. 641 – 654.

Geroski, P. A. 2000. “Models of technology diffusion.” *Research Policy*, 29 (2000) 603 – 625.

Goolsbee, A. 2000. “In a world without borders: The impact of taxes on Internet Commerce,” *Quarterly Journal of Economics*, 115(2), 561-76.

Goolsbee, A. and Klenow, P. J. 2002. “Evidence on Learning and Network Externalities in the Diffusion of Home Computers.” *Journal of Law and Economics,* vol. XLV (October 2002), pp. 317 – 343.

Griliches, Z. 1957. “Hybrid Corn: An Exploration in the Economics of Technological Change.” *Econometrica,* Vol. 25, No. 4 (Oct., 1957), pp. 501 - 522.

Gujrati, D. N. 2003. “Qualitative Response Regression Models” in *Basic Econometrics.* McGraw-Hill/Irwin, New York, NY.

Hall, B. H. 2004. “Innovation and Diffusion.” Retrieved August 13, 2006, from *http://www.nber.org/papers/w10212*

Hoffman, D. L. and Novak, T. P. 1999. “The Growing Digital Divide: Implications for an Open Research Agenda.” Retrieved August 13, 2006, from *http://199.239.233.76/downloadable_assets/digitaldivide_openresearch.pdf*

Household Internet Use Survey (HIUS) data released by Statistics Canada (2004, September 21). Retrieved August, 2006, from the TDR Tri-University Data Resources public use micro-data sets maintained by the University of Guelph, Guelph, Canada, under “Communication.”

Kiiski, S. and Pohjola, M. 2002. “Cross-country diffusion of the Internet.” *Information Economics and Policy*, 14 (2002) 297 – 310.

Kraut, R. et al. 1998. “Communication and Information: Alternative Uses of the Internet in Households.” Retrieved August 13, 2006, from http://portal.acm.org/citation.cfm?doid=274644.274695

Michalak, W. and Jones, K. 2003. “Canadian E-Commerce.” *International Journal of Retail and Distribution Management,* Volume 31. Number 1. 2003. pp. 5-15.

Mills, B. F. and Whitacre, B. E. 2003. “Understanding the Non-Metropolitan – Metropolitan Digital Divide.” *Growth and Change,* Vol. 34. No. 2 (Spring 2003). pp. 219 - 243.

Pohjola, M., 2002. “The New Economy: facts, impacts, and policies.” *Information Economics and Policy,* 14 (2002) 133 - 144.

Saloner, G. and Shepard, A. 1995. “Adoption of technologies with network effects: an empirical examination of the adoption of automated teller machines.” *RAND Journal of Economics*, Vol. 26. No. 3. Autumn 1995. pp. 479 - 501.

Sexton, R. S., Johnson, R. A. and Hignite, M. A., 2002. “Predicting Internet / e-commerce use.” *Internet Research: Electronic Networking Applications and Policy*, Volume 12 . Number 5 . 2002 . pp. 402 – 410.

Shah, D. V., Kwak, N. and Lance Holbert, R. 2001. “ ‘Connecting’ and ‘Disconnecting’ With Civic Life: Patterns of Internet Use and the Production of Social Capital.” *Political Communication,*18: 141 – 162, 2001

Shih, Y. Y. and Fang, K. 2004. “The use of a decomposed theory of planned behavior to study Internet banking in Taiwan.” *Internet Research,* Volume 14 . Number 3 . 2004 . pp. 213 – 223.

Stevenson, B. 2003. “The Internet, Job Search, and Worker Mobility.” Retrieved August 13, 2006, from http://www.gsb.stanford.edu/facseminars/events/political_economy/pdfs/Stevenson_Internet_Job_Search.pdf

Tirole, J. 1989. “The Theory of Industrial Organization.” MIT Press. 1988. Chapter 2.

Venkatesh, V. and Brown, S. A. 2001. “A Longitudinal Investigation of Personal Computers in Homes: Adoption Determinants and Emerging Challenges.” *MIS Quarterly;* March 2001; 25, 1; ABI / INFORM Global pg. 71.

Vulkan, N. 2003. “Automated E-Commerce” in *The Economics of E-Commerce: A Strategic Guide to Understanding and Designing the Online Marketplace.* Princeton University Press, Princeton, N J.

APPENDICES

Table A 1: Description of Variables of Empirical Models

Variable	Description
Y99	Year control - 1 if 1999; 0, otherwise
Y00	Year control - 1 if 2000; 0, otherwise
Y01	Year control - 1 if 2001; 0, otherwise
Y03[37]	Year control - 1 if 2003; 0, otherwise
WTHP	Weight for each observation (for replication)
Nosobuy	Number of separate orders ordered online in the pervious 12 months without online payment
Nosopay	Number of separate orders ordered online in the pervious 12 months with online payment
Nosoall	Number of separate orders ordered online in the pervious 12 months with or without online payment
Etcbuy	Expected total cost of online purchases in the previous 12 months without online payment (C$)
Etcpay	Expected total cost of online purchases in the previous 12 months with online payment (C$)
Etcall	Expected total cost of online purchases in the previous 12 months with or without online payment (C$)
Homobuy	1 if a homogeneous good was ordered online in the pervious 12 months without online payment; 0, otherwise
Homopay	1 if a homogeneous good was ordered online in the pervious 12 months with online payment; 0, otherwise
Homogood	1 if a homogeneous good was ordered online in the pervious 12 months with or without online payment; 0, otherwise
Heterobuy	1 if a heterogeneous good was ordered online in the pervious 12 months without online payment; 0, otherwise
Heteropay	1 if a heterogeneous good was ordered online in the pervious 12 months with online payment; 0, otherwise
Heterogood	1 if a heterogeneous good was ordered online in the pervious 12 months with or without online payment; 0, otherwise
Age1	1 if the age of head of household is < 35; 0, otherwise
Age2	1 if the age of head of household is between 35 and 54; 0, otherwise
Age3	1 if the age of head of household is between 55 and 64; 0, otherwise
Age4	1 if the age of head of household is 65 plus; 0, otherwise
Male	1 if head of household is a male; 0, otherwise
Female	1 if head of household is a female; 0, otherwise
Edu1	1 if the education level of head of household is less than high-school; 0, otherwise
Edu2	1 if the education level of head of household is high-school or some college; 0, otherwise
Edu3	1 if the education level of head of household is university degree; 0, otherwise

[37] The data of 2002 were not used in the regressions due to issues (the variable “quartile” had values 0 to 9, in stead of 1 to 4)

Quart1	1 if the income level of household is in the first (lowest) quartile; 0, otherwise
Quart2	1 if the income level of household is in the second quartile; 0, if otherwise
Quart3	1 if the income level of household is in the third quartile; 0, if otherwise
Quart4	1 if the income level of household is in the fourth quartile; 0, if otherwise
I_tel	1 if the internet connection of household is through telephone cable; 0, otherwise
I_otherconnect	1 if the internet connection of household is some other type; 0, otherwise
I_noconnect	1 if household has no internet connection; 0, otherwise

Table A 2.1: Descriptive Statistics, Equations 1 & 2, Full Sample[38] Regressions

NAME	N	MEAN	ST. DEV	VARIANCE	MINIMUM	MAXIMUM
Y99	126938	0.28547	0.45164	0.20398	0.0000	1.0000
Y00	126938	0.26589	0.44181	0.19520	0.0000	1.0000
Y01	126938	0.26835	0.44310	0.19634	0.0000	1.0000
Y03	126938	0.18028	0.38443	0.14778	0.0000	1.0000
WTHP	126938	374.86	334.24	0.11172E+06	41.124	4186.9
NOSOBUY	126938	0.20905	2.1062	4.4362	0.0000	150.00
ETCBUY	126938	26.659	407.32	0.16591E+06	0.0000	50000.
HOMOBUY	126938	0.25272E-01	0.15695	0.24634E-01	0.0000	1.0000
HETEROBU	126938	0.26304E-01	0.16004	0.25612E-01	0.0000	1.0000
AGE1	126938	0.18390	0.38740	0.15008	0.0000	1.0000
AGE2	126938	0.44246	0.49668	0.24669	0.0000	1.0000
AGE3	126938	0.15113	0.35818	0.12829	0.0000	1.0000
AGE4	126938	0.22251	0.41593	0.17300	0.0000	1.0000
MALE	126938	0.75707	0.42885	0.18392	0.0000	1.0000
FEMALE	126938	0.24293	0.42885	0.18392	0.0000	1.0000
EDU1	126938	0.28780	0.45274	0.20497	0.0000	1.0000
EDU2	126938	0.56191	0.49615	0.24617	0.0000	1.0000
EDU3	126938	0.15029	0.35735	0.12770	0.0000	1.0000
QUART1	126938	0.26704	0.44242	0.19573	0.0000	1.0000
QUART2	126938	0.26150	0.43945	0.19312	0.0000	1.0000
QUART3	126938	0.24685	0.43118	0.18592	0.0000	1.0000
QUART4	126938	0.22461	0.41732	0.17416	0.0000	1.0000
I_TEL	126938	0.29169	0.45454	0.20661	0.0000	1.0000
I_OTHERC	126938	0.92352E-01	0.28952	0.83824E-01	0.0000	1.0000
I_NOCONN	126938	0.61460	0.48669	0.23687	0.0000	1.0000

[38] Full samples, however, exclude observations for which the type of good ordered (homogeneous or heterogeneous) is unknown. Such exclusions amounted to 5% or less of total sample sizes. The estimation sample size N is different for different equations.

Table A 2.2: Descriptive Statistics, Equations 3 & 4, Full Sample Regressions

NAME	N	MEAN	ST. DEV	VARIANCE	MINIMUM	MAXIMUM
WTHP	127184	375.11	334.47	0.11187E+06	41.124	4186.9
NOSOPAY	127184	0.61393	4.3797	19.182	0.0000	500.00
ETCPAY	127184	79.353	664.03	0.44093E+06	0.0000	0.10000E+06
HOMOPAY	127184	0.71715E-01	0.25802	0.66572E-01	0.0000	1.0000
HETEROPA	127184	0.62162E-01	0.24145	0.58298E-01	0.0000	1.0000
AGE1	127184	0.18398	0.38747	0.15013	0.0000	1.0000
AGE2	127184	0.44281	0.49672	0.24673	0.0000	1.0000
AGE3	127184	0.15108	0.35813	0.12826	0.0000	1.0000
AGE4	127184	0.22213	0.41568	0.17279	0.0000	1.0000
MALE	127184	0.75723	0.42876	0.18383	0.0000	1.0000
FEMALE	127184	0.24277	0.42876	0.18383	0.0000	1.0000
EDU1	127184	0.28743	0.45256	0.20481	0.0000	1.0000
EDU2	127184	0.56200	0.49614	0.24616	0.0000	1.0000
EDU3	127184	0.15057	0.35763	0.12790	0.0000	1.0000
QUART1	127184	0.26669	0.44223	0.19557	0.0000	1.0000
QUART2	127184	0.26122	0.43930	0.19299	0.0000	1.0000
QUART3	127184	0.24702	0.43128	0.18600	0.0000	1.0000
QUART4	127184	0.22507	0.41763	0.17441	0.0000	1.0000
I_TEL	127184	0.29232	0.45483	0.20687	0.0000	1.0000
I_OTHERC	127184	0.92842E-01	0.29021	0.84223E-01	0.0000	1.0000
I_NOCONN	127184	0.61348	0.48695	0.23712	0.0000	1.0000

Table A 2.3: Descriptive Statistics, Equations 5 & 6, Full Sample Regressions

NAME	N	MEAN	ST. DEV	VARIANCE	MINIMUM	MAXIMUM
WTHP	126979	375.01	334.34	0.11179E+06	41.124	4186.9
NOSOBUY	126979	0.21425	2.1256	4.5181	0.0000	150.00
NOSOPAY	126979	0.61627	4.3849	19.227	0.0000	500.00
NOSOALL	126979	0.83052	5.1378	26.397	0.0000	550.00
ETCBUY	126979	27.465	410.11	0.16819E+06	0.0000	50000.
ETCPAY	126979	79.586	664.62	0.44172E+06	0.0000	0.10000E+06
ETCALL	126979	107.05	805.41	0.64868E+06	0.0000	0.10000E+06
HOMOGOOD	126979	0.88668E-01	0.28427	0.80807E-01	0.0000	1.0000
HETEROGO	126979	0.80950E-01	0.27285	0.74445E-01	0.0000	2.0000
AGE1	126979	0.18394	0.38743	0.15010	0.0000	1.0000
AGE2	126979	0.44254	0.49669	0.24670	0.0000	1.0000
AGE3	126979	0.15110	0.35814	0.12827	0.0000	1.0000
AGE4	126979	0.22243	0.41588	0.17296	0.0000	1.0000
MALE	126979	0.75709	0.42884	0.18390	0.0000	1.0000
FEMALE	126979	0.24291	0.42884	0.18390	0.0000	1.0000

EDU1	126979	0.28772	0.45270	0.20494	0.0000	1.0000
EDU2	126979	0.56195	0.49615	0.24616	0.0000	1.0000
EDU3	126979	0.15033	0.35740	0.12773	0.0000	1.0000
QUART1	126979	0.26696	0.44237	0.19569	0.0000	1.0000
QUART2	126979	0.26135	0.43937	0.19305	0.0000	1.0000
QUART3	126979	0.24690	0.43121	0.18594	0.0000	1.0000
QUART4	126979	0.22479	0.41745	0.17426	0.0000	1.0000
I_TEL	126979	0.29176	0.45457	0.20664	0.0000	1.0000
I_OTHERC	126979	0.92456E-01	0.28967	0.83909E-01	0.0000	1.0000
I_NOCONN	126979	0.61442	0.48673	0.23691	0.0000	1.0000

Table A 3: Results, Full Sample Tobit Regression

Dependent variable: Eq. 1: nosobuy; Eq. 2: etcbuy; Eq. 3: nosopay; Eq. 4: etcpay; Eq. 5: nosoall; and Eq. 6: etcall

Eq.	Homo	Hetero	Age2	Age3	Age4	Female	Edu2	Edu3	Quart2	Quart3	Quart4	I_otherconnect	I_noconnect	Y00	Y01	Y03	constant
1	19.72 ***	20.56 ***	-0.145 ***	0.062 ***	-1.288 ***	-0.702 ***	0.637 ***	0.808 ***	0.389 ***	1.095 ***	0.588 ***	-0.619 ***	-4.572 ***	2.808 ***	3.107 ***	3.199 ***	-23.57 ***
2	4121 ***	4059 ***	-25.43 ***	1.77 insig	-90.26 ***	-59.31 ***	177.3 ***	294.1 ***	30.96 ***	117.1 ***	338.6 ***	-34.23 ***	-1010 ***	406.2 ***	632.9 ***	534.2 ***	-5258 ***
3	23.9 ***	21.82 ***	-0.725 ***	-0.424 ***	-1.608 ***	-0.751 ***	1.326 ***	1.76 ***	0.686 ***	1.126 ***	1.03 ***	0.264 ***	-5.517 ***	-0.372 ***	0.921 ***	1.169 ***	-27.13 ***
4	3408 ***	2931 ***	-90.54 ***	-53.71 ***	-160.2 ***	-92.27 ***	141.9 ***	263.3 ***	97.23 ***	178.5 ***	331.7 ***	51.1 ***	-750.8 ***	-72.11 ***	112.8 ***	243 ***	-3922 ***
5	22.31 ***	21.14 ***	-0.754 ***	-0.426 ***	-2.527 ***	-0.894 ***	1.858 ***	2.407 ***	0.918 ***	1.773 ***	1.745 ***	-0.039 ***	-8.99 ***	2.762 ***	4.138 ***	4.429 ***	-29.39 ***
6	3305 ***	2939 ***	-86.91 ***	-65.87 ***	-251.3 ***	-90.29 ***	231.3 ***	391.9 ***	113.7 ***	236.3 ***	478.2 ***	28.58 ***	-1308 ***	363.9 ***	630.4 ***	708.4 ***	-4467 ***

Significance level: *** 1%, ** 5%, and * 10%

Table A 4: Results, Full Sample Poisson Regression

Dependent variable: Eq. 1: nosobuy; Eq. 3: nosopay; and Eq. 5: nosoall;

Eq.	Homo	Hetero	Age2	Age3	Age4	Female	Edu2	Edu3	Quart2	Quart3	Quart4	I_otherconnect	I_noconnect	Y00	Y01	Y03	constant
1	1.866 ***	3.033 ***	-0.150 ***	-0.123 ***	-0.492 ***	-0.331 ***	0.091 ***	0.141 ***	0.16 ***	0.268 ***	0.102 ***	-0.131 ***	-1.840 ***	0.676 ***	0.715 ***	0.726 ***	-2.778 ***
3	2.011 ***	2.103 ***	-0.103 ***	-0.087 ***	-0.24 ***	-0.175 ***	0.205 ***	0.243 ***	0.169 ***	0.176 ***	0.237 ***	0.114 ***	-0.98 ***	0.046 ***	0.170 ***	0.270 ***	-1.908 ***
5	1.596 ***	1.954 ***	-0.103 ***	-0.074 ***	-0.335 ***	-0.225 ***	0.238 ***	0.302 ***	0.245 ***	0.287 ***	0.295 ***	0.059 ***	-1.807 ***	0.487 ***	0.580 ***	0.666 ***	-1.844 ***

Significance level: *** 1%, ** 5%, and * 10%

Table A 5.1: Descriptive Statistics, Equations 1 & 2, Sub-Sample[39] Regressions

NAME	N	MEAN	ST. DEV	VARIANCE	MINIMUM	MAXIMUM
WTHP	48922	418.05	363.30	0.13199E+06	41.124	4186.9
NOSOBUY	48922	0.51527	3.3129	10.975	0.0000	150.00
ETCBUY	48922	66.165	649.17	0.42142E+06	0.0000	50000.
HOMOBUY	48922	0.62283E-01	0.24167	0.58405E-01	0.0000	1.0000
HETEROBU	48922	0.64163E-01	0.24505	0.60048E-01	0.0000	1.0000
AGE1	48922	0.20911	0.40668	0.16539	0.0000	1.0000
AGE2	48922	0.58095	0.49341	0.24345	0.0000	1.0000
AGE3	48922	0.13759	0.34447	0.11866	0.0000	1.0000
AGE4	48922	0.72360E-01	0.25909	0.67125E-01	0.0000	1.0000
MALE	48922	0.85358	0.35353	0.12498	0.0000	1.0000
FEMALE	48922	0.14642	0.35353	0.12498	0.0000	1.0000
EDU1	48922	0.11968	0.32459	0.10536	0.0000	1.0000
EDU2	48922	0.62377	0.48444	0.23469	0.0000	1.0000
EDU3	48922	0.25655	0.43673	0.19074	0.0000	1.0000
QUART1	48922	0.11042	0.31342	0.98230E-01	0.0000	1.0000
QUART2	48922	0.20502	0.40372	0.16299	0.0000	1.0000
QUART3	48922	0.30095	0.45868	0.21038	0.0000	1.0000
QUART4	48922	0.38361	0.48627	0.23646	0.0000	1.0000
I_TEL	48922	0.75684	0.42900	0.18404	0.0000	1.0000
I_OTHERC	48922	0.23963	0.42686	0.18221	0.0000	1.0000

Table A 5.2: Descriptive Statistics, Equations 3 & 4, Sub-Sample Regressions

NAME	N	MEAN	ST. DEV	VARIANCE	MINIMUM	MAXIMUM
WTHP	48986	412.28	363.75	0.13231E+06	41.124	4186.9
NOSOPAY	48986	1.4382	6.6700	44.489	0.0000	500.00
ETCPAY	48986	186.96	1015.3	0.10309E+07	0.0000	0.10000E+06
HOMOPAY	48986	0.16013	0.36673	0.13449	0.0000	1.0000
HETEROPA	48986	0.14086	0.34788	0.12102	0.0000	1.0000
AGE1	48986	0.20900	0.40660	0.16532	0.0000	1.0000
AGE2	48986	0.58141	0.49333	0.24338	0.0000	1.0000
AGE3	48986	0.13753	0.34441	0.11862	0.0000	1.0000
AGE4	48986	0.72061E-01	0.25859	0.66870E-01	0.0000	1.0000
MALE	48986	0.85343	0.35368	0.12509	0.0000	1.0000

[39] Sub-samples exclude both the observations for which the type of good ordered is unknown and the observations for which there is no Internet connection in the household. The sample size N is different for different equations.

FEMALE	48986	0.14657	0.35368	0.12509	0.0000	1.0000
EDU1	48986	0.11948	0.32436	0.10521	0.0000	1.0000
EDU2	48986	0.62381	0.48443	0.23468	0.0000	1.0000
EDU3	48986	0.25671	0.43682	0.19081	0.0000	1.0000
QUART1	48986	0.11013	0.31306	0.98006E-01	0.0000	1.0000
QUART2	48986	0.20492	0.40364	0.16293	0.0000	1.0000
QUART3	48986	0.30096	0.45868	0.21039	0.0000	1.0000
QUART4	48986	0.38399	0.48636	0.23655	0.0000	1.0000
I_TEL	48986	0.75895	0.42772	0.18295	0.0000	1.0000
I_OTHERC	48986	0.24105	0.42772	0.18295	0.0000	1.0000

Table A 5.3: Descriptive Statistics, Equations 5 & 6, Sub-Sample Regressions

NAME	N	MEAN	ST. DEV	VARIANCE	MINIMUM	MAXIMUM
WTHP	48960	418.38	363.48	0.13212E+06	41.124	4186.9
NOSOBUY	48960	0.52665	3.3331	11.109	0.0000	150.00
NOSOPAY	48960	1.5443	6.7875	46.070	0.0000	500.00
NOSOALL	48960	2.0709	7.9360	62.979	0.0000	550.00
ETCBUY	48960	68.101	653.16	0.42662E+06	0.0000	50000.
ETCPAY	48960	198.70	1040.5	0.10826E+07	0.0000	0.10000E+06
ETCALL	48960	266.80	1261.3	0.15910E+07	0.0000	0.10000E+06
HOMOGOOD	48960	0.22010	0.41432	0.17166	0.0000	1.0000
HETEROGO	48960	0.19898	0.39939	0.15951	0.0000	2.0000
AGE1	48960	0.20917	0.40672	0.16542	0.0000	1.0000
AGE2	48960	0.58103	0.49340	0.24344	0.0000	1.0000
AGE3	48960	0.13752	0.34440	0.11861	0.0000	1.0000
AGE4	48960	0.72283E-01	0.25896	0.67060E-01	0.0000	1.0000
MALE	48960	0.85359	0.35352	0.12497	0.0000	1.0000
FEMALE	48960	0.14641	0.35352	0.12497	0.0000	1.0000
EDU1	48960	0.11961	0.32451	0.10530	0.0000	1.0000
EDU2	48960	0.62382	0.48443	0.23467	0.0000	1.0000
EDU3	48960	0.25658	0.43675	0.19075	0.0000	1.0000
QUART1	48960	0.11033	0.31331	0.98163E-01	0.0000	1.0000
QUART2	48960	0.20468	0.40347	0.16279	0.0000	1.0000
QUART3	48960	0.30100	0.45870	0.21040	0.0000	1.0000
QUART4	48960	0.38399	0.48636	0.23655	0.0000	1.0000
I_TEL	48960	0.75668	0.42909	0.18412	0.0000	1.0000
I_OTHERC	48960	0.23979	0.42696	0.18229	0.0000	1.0000

Table A 6: Results, Sub-Sample Tobit Regression

Dependent variable: Eq. 1: nosobuy; Eq. 2: etcbuy; Eq. 3: nosopay; Eq. 4: etcpay; Eq. 5: nosoall; and Eq. 6: etcall

Eq.	Homo	Hetero	Age2	Age3	Age4	Female	Edu2	Edu3	Quart2	Quart3	Quart4	I_otherc onnect	Y00	Y01	Y03	constant
1	19.41 ***	20.25 ***	-0.039 ***	0.144 ***	-0.497 ***	-0.733 ***	0.153 ***	0.168 ***	0.095 ***	0.709 ***	0.321 ***	-0.484 ***	2.756 ***	2.606 ***	2.589 ***	-22.47 ***
2	4092 ***	4027 ***	4.259 insig	42.86 ***	126.2 ***	-69.85 ***	74.12 ***	164.9 ***	-19.6 ***	33.19 ***	283.2 ***	-6.073 ***	387 ***	524.9 ***	394.5 ***	-5075 ***
3	22.73 ***	20.93 ***	-0.623 ***	-0.065 ***	-1.0 ***	-1.245 ***	1.336 ***	1.823 ***	0.433 ***	1.086 ***	0.917 ***	0.247 ***	5.468 ***	6.26 ***	6.446 ***	-31.54 ***
4	3252 ***	2791 ***	-63.44 ***	-20.11 ***	-48.67 ***	-128.1 ***	99.72 ***	222.8 ***	78.23 ***	181.7 ***	337.7 ***	48.98 ***	703.8 ***	813.5 ***	939.9 ***	-4479 ***
5	21.63 ***	20.37 ***	-0.497 ***	0.109 ***	-0.961 ***	-1.243 ***	1.112 ***	1.504 ***	0.424 ***	1.221 ***	1.161 ***	0.284 ***	2.62 ***	3.193 ***	3.358 ***	-27.16 ***
6	3225 ***	2844 ***	-40.70 ***	24.70 ***	9.496 ***	-122.6 ***	112.5 ***	248.6 ***	60.04 ***	164 ***	409.6 ***	76.11 ***	341.2 ***	492.5 ***	553.2 ***	-4185 ***

Significance level: *** 1%, ** 5%, and * 10%

Table A 7: Results, Sub-Sample Poisson Regression

Dependent variable: Eq. 1: nosobuy; Eq. 3: nosopay; and Eq. 5: nosoall;

Eq.	Homo	Hetero	Age2	Age3	Age4	Female	Edu2	Edu3	Quart2	Quart3	Quart4	I_otherconnect	Y00	Y01	Y03	constant
1	1.834 ***	2.906 ***	-0.128 ***	-0.088 ***	-0.283 ***	-0.35 ***	0.011 insig	0.040 insig	0.051 *	0.148 ***	0.008 insig	-0.100 ***	0.631 ***	0.572 ***	0.584 ***	-2.424 ***
3	1.74 ***	1.916 ***	-0.090 ***	-0.045 ***	-0.141 ***	-0.279 ***	0.183 ***	0.254 ***	0.124 ***	0.155 ***	0.200 ***	0.124 ***	1.154 ***	1.223 ***	1.322 ***	-2.572 ***
5	1.547 ***	1.861 ***	-0.086 ***	-0.032 ***	-0.165 ***	-0.275 ***	0.138 ***	0.200 ***	0.123 ***	0.165 ***	0.179 ***	0.091 ***	0.444 ***	0.470 ***	0.548 ***	-1.472 ***

Significance level: *** 1%, ** 5%, and * 10%