# Does privacy concern influence e-shoppers choices for online retailers? An application of the Integrated Latent Variable Choice Model

Juan Palacios

CeDInt

Technical University of Madrid
Spain
juanfrancisco.palacios@upm.es

Dimitris Potoglou

School of Planning and Geography
Cardiff University
United Kingdom
potogloud@cardiff.ac.uk

Corresponding author

Claudio Fejoo

CeDInt

Technical University of Madrid
Spain
Claudio.feijoo@upm.es

Suggested Citation:

Palacios, J., Potoglou, D., and Fejoo, C. "Does privacy concern influence e-shoppers choices for online retailers? An application of the Integrated Latent Variable Choice Model" International Choice Modelling Conference 2015, 10-13 May, Austin, Texas, USA

**Abstract**

The collection, management and sharing of customers' personal data represents a new set of opportunities for online businesses. The provision of personal data by consumers is a source of benefits and threats at the same time. Online retailers and service providers often seek ways to gain consumers' confidence and minimise privacy concerns in online transactions. This paper aims to establish a better understanding of the role of attitudes in respondents' willingness to provide personal information in online transactions. We employ a stated preference discrete choice experiment to collect respondents' choices across online retailers and a conventional store under different levels of personal-information requirements. An opt-out option is also available in each of the choice exercises. Personal information in the experiment is presented across three dimensions (attributes): amount/type of information collected, duration of storage and the likelihood of this information being shared with third parties. These dimensions are introduced in order to be able to capture specific threats involved in online transactions according to consumer perceptions. The choice data are complemented with a set of attitudinal indicators (psychometric scales) corresponding to latent constructs such as privacy concern, general caution and technical protection. The data comes from 502 participants reflecting a nationally representative sample of the online-user population in the UK. We report on Integrated Latent Variable models which test the influence of these latent constructs in the respondents' decision to engage in an online transaction. Model estimation results show that the higher an individual's concern, general caution and technical protection the less likely this individual is to purchase a product online. In a joint model, the influence of privacy concern is found to outweigh the influence of general caution and technical protection.

## 1. Introduction

The increasing uptake of broadband, a more technologically aware population, and improved digital infrastructure have facilitated an impressive rise of global e-commerce services (OECD, 2012). According to the Centre for Retail Research (2015), a higher-education

research institute, online sales in Western Europe are expected to grow by 18.4%[1] in 2015 and to reach £185.44bn (€219.44bn) in 2016.

Purchase data and personal information collected from e-consumers are nowadays highly valuable assets for e-commerce stakeholders. Using data mining tools enables online retailers to uncover purchase patterns, which help them tailor products, identify consumer segments and thus better predict consumer behaviour. Advertisement revenues are also at the heart of e-commerce business models. E-commerce platforms sell personal information about their visitors to advertisers, who are willing to pay a premium for increasing the efficiency of their adds by targeting them to individuals with certain characteristics or behaviours (Godel, Litchfield, & Mantovani, 2012). In fact, personal information has become a critical input for e-commerce. Companies without access to consumer data generally face significant barriers, for example, to enter in certain industries, and thus are at competitive disadvantage against those firms that hold this asset (Acquisti, 2010).

Normally, information on consumers' online activity such as the number of visits to the e-retailer's site or what products they have bought are monitored, collected, traded and retained by online companies. In this context, the provision of personal information to online retailers represents a source of potential benefits and threats for costumers. When accessing e-commerce services, e-shoppers benefit from lower prices and personalised consumer experiences, such as purchase recommendations. On the other hand, the access to such services requires the provision of personal data by Internet users. Those individuals who disclose their personal information to retailers may face threats in the form of, for example, adverse price discrimination or invasive advertising (Goldfarb & Tucker, 2011).

As a result, privacy concerns in Europe have remained high and are reported to be one of the main factors inhibiting e-commerce (EC, 2011). Individuals usually report high levels of (information) privacy concerns in public opinion polls, however, their actions in real life scenarios are inconsistent with their stated concerns (Spiekermann, Grossklags, & Berendt, 2001). The literature refers to this phenomenon as the 'Privacy Paradox' – that is the situation where although consumers are aware of the potential threats to their privacy, in practice they continue using online services, e-commerce platforms and purchase products even when they know their privacy may be under threat (Monteleone, 2013). For example, while people generally report high levels of disruption when asked about tracking technology or information requirements in web forms, their behaviour is generally far from protective; being so unlikely that internet users opt-out online transactions, once they are in the subscription process or check out. Dinev & Hart (2006) argue that this phenomenon may be occurring either because consumers' behaviour actually 'reflects lower privacy concerns than polls and research would suggest or other factors mitigate privacy concerns'.

A large body of work has focused on understanding the benefits and drawbacks (trade-offs) of personal data being disclosed or protected by the two sides of the market. Information privacy is commonly defined as “the desire of individuals to control or have some influence over data about themselves” (Bélanger & Crossler, 2011). In their decision to provide information to retailers, customers weigh out perceived risk and benefits of disclosing their personal data to retailers performing a cost-benefit analysis (Dinev & Hart, 2006). Smith et al. (1996) identify four central dimensions of individuals’ concerns about organizational information privacy practices: collection of personal information, handling errors, unauthorized secondary use and improper access. Following this rationale, experimental studies show that respondents tend to search for online retailers, which require the minimum

[1] Prediction based on data from in the UK, Germany, France Sweden, The Netherlands. Italy Poland and Spain.

amount of personal information (Hui, Teo, & Lee, 2007) and those committed to limit future use and access to/use of information by third parties (Hann, Hui, Lee, & Png, 2007). Smith et al. (2011) note the existence of heterogeneity in preferences, which can be determined by observable (e.g. gender, age), and especially by latent psychological characteristics of consumers such as attitudes and perceptions.

In a disjoint, but extensive stream of research, a number of studies have used self-reported scales to examine the role of information-privacy concern as a central construct in shaping consumer behaviour (Smith et al., 2011). Information privacy concerns are measured through psychometric instruments (scales, indicators) based on respondents' opinions about privacy-related issues or past behaviour (for a collection of psychometric measures, see Preibusch (2013)). In an excellent review of the articles by Smith et al. (2011) and Bélanger & Crossler (2011), Pavlou (2011) notes that there is a 'general consensus among researchers that information privacy concerns correspond to a person's willingness to render personal information, transaction activity and government regulation'. Yet, with the exception of Dinev & Hart (2006), there have been very few studies on how consumers' willingness to disclose their personal information in e-commerce relates to latent constructs (e.g. attitudes) and how these constructs are influenced through observed individual characteristics (e.g. age).

This paper is aimed at better understanding of the decision of internet users to engage in an online purchase, where the provision of personal information is needed to complete the transaction. More precisely, we explore: (a) how consumers react to different levels of information requirements based on the type/load of information they have to disclose, for how long this information may be retained by retailers and whether their personal information may be shared with third parties; and (b) how privacy related attitudes in the form of, for example, privacy concerns influence e-shoppers choices for online retailers. We argue that a person's willingness to render their personal information is dependent upon their preferences against personal data requirements and latent constructs (attitudes) such as privacy concerns, which simultaneously influence these preferences. The objective of this paper is therefore to test the following hypothesis: 'Higher levels of an individual's privacy concern, caution and technical protection are related to a lower level of willingness to engage in online transactions and thus provide personal information when purchasing products from retailers online'.

Our study is based on data collected via a Discrete Choice Experiment (DCE) – an advanced, survey-based preference elicitation instrument, which simulates an e-commerce scenario of online retailer choice with varying levels of information collection, handling and storage. Using the respondents' preference data for online retailers along with a set of psychometric scales relevant to privacy concerns we develop a model known as Integrated Choice and Latent Variable (ICLV) that allows to make use of information provided by psychometric indicators measuring respondents' attitudes toward privacy and data collection (Daly, Hess, Patruni, Potoglou, & Rohr, 2012). This modelling framework also provides us with a better understanding of what drives e-shopper choices of retailers. Further, ICLV allows for exploring the link between such choices and self-reported concerns, gauging to what extent information concerns drive individuals to opt-out from their e-commerce transactions.

The remainder of the paper is organised as follows. Section 2 presents an overview of the experimental design and the psychometric indicators used. Section 3 describes the analytical approach based on ICLV. Finally, the paper closes with the discussion of results.

# 2. Data and Methods

## 2.1 Data

The data come from a UK-representative sample of online users collected through an Internet Panel in August 2012. Sample quotas are specified in order for the sample to match the profile of the 2011 (Q4) Internet-user population in the UK (ONS, 2011). The total sample size includes 502 respondents (see, Table 1).

Participants' ages range from 18 to 82 years old (with mean years of age 43). The sample includes 248 male (49.6%) and 269 female (52.0%) respondents. More than half of the participants in the sample work full or part time, 7% of individuals are students, and the rest of participants are unemployed or retired. Most participants report an individual income below £50,000. Most respondents are frequent internet users; 50% of respondents in the sample spend more than 14 hours per week and the majority are regular e-shoppers. Sixty-six percent of the sample search for products at least once a month and 59% report that they buy products online monthly.

**Table 1. Sample characteristics vs. the 2011 UK online user population (n=502)**

| Variable | Sample (%) | Internet users in UK (2011 Q4, %) | Variable | Sample (%) | Internet users in UK (2011 Q4, %) |
|---|---|---|---|---|---|
| Gender (female) | 52.0 | 49.6 | *Region* | | |
| *Age group* | | | East of England | 10.1 | 7.2 |
| 18-24 | 13.9 | 17.1 | East Midlands | 7.2 | 9.5 |
| 25-34 | 21.5 | 19.6 | London | 12.8 | 13.3 |
| 35-44 | 19.3 | 19.5 | North East | 3.7 | 4.0 |
| 45-54 | 18.4 | 18.8 | North West | 11.6 | 11.0 |
| 55-64 | 15.9 | 14.0 | Northern Ireland | 2.3 | 2.5 |
| 65-74 | 7.9 | 7.9 | Scotland | 8.5 | 8.3 |
| 75 and over | 3.1 | 3.2 | South East | 13.7 | 14.1 |
| | | | South West | 9.3 | 8.7 |
| *Annual individual income* | | | Wales | 4.5 | 4.7 |
| Less than £10,399 | 27.8 | 20.9 | West Midlands | 8.3 | 8.3 |
| £10,400 - £15,599 | 14.1 | 15.2 | Yorkshire / Humberside | 8.1 | 8.4 |
| £15,600 - £20,799 | 12.6 | 15.9 | | | |
| £20,800 - £25,999 | 9.3 | 12.9 | *Occupational status* | | |
| £26,000 - £31,199 | 6.6 | 10.4 | Working full time | 41.0 | |
| £31,200 - £36,399 | 6.6 | 7.3 | Working part time | 17.2 | |
| £36,400 - £41,599 | 4.1 | 4.6 | Student | 7.2 | |
| £41,600 - £46,799 | 2.5 | 3.8 | Retired | 16.1 | |
| £46,800 - £51,999 | 2.7 | 2.7 | Not in paid work because of long term illness or disability | 7.0 | |
| £52,000 - £77,999 | 2.9 | 4.1 | Seeking work | 5.8 | |
| £78,000 - £103,999 | 1.2 | 1.8 | Other | 5.8 | |
| £104,000 or higher | 0.0 | 0.3 | | | |
| Not reported | 9.7 | 20.9 | | | |

## 2.2 Choice of online retailers involving varying aspects of information privacy

Respondents' choice for online retailers is captured through a stated preference discrete choice experiment (DCE), a core component of the online survey. This method is a multi-attribute survey-based approach for eliciting consumer's preferences and has been applied in a wide range of fields such as health economics, transport or marketing (Louviere, Hensher, & Swait, 2000). In the survey, respondents are firstly asked to recall an online

purchase they have recently completed[2]. This strategy of linking the experiment to a real-life experience helps respondents to frame the choice scenario (Hess & Rose, 2009). As shown in Figure 1, respondents are then offered the choice to buy the product from three different online retailers, a Conventional Store and an Opt-out option. Online retailers are described by the setting under which they collect and handle customer information, namely (see, also Table 2):

- The information online retailers collect in a given transaction (e.g. email, purchase history or browsing history);
- The time period they may store customer data such as purchases; and
- Whether they share the information with third parties.

Retailers are also presented with a cost attribute which is defined as either an extra-payment to be paid by participants to acquire the product or discount off the market price of the product. Also, the additional-services attribute reflects potential benefits offered by online retailers in exchange of the personal information provided by respondents. These attributes are combined such that participants face trade-offs between restricting the conditions under which retailers can collect and handle their information, and the access to benefits such as discount or additional services (e.g. priority shipping, etc.).

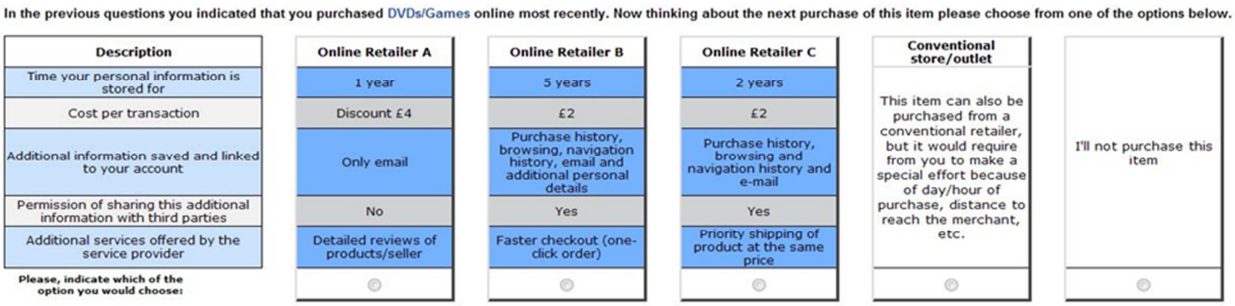


**Figure 1. An example choice exercise**

In the context of e-commerce, geographic proximity to conventional stores is a determinant aspect to complete e-commerce transactions (Dinev & Hart, 2006). An individual might decide that the need for a highly desired product that is not in close proximity overrides privacy risk and concerns. In our experiment, we explore this dimension by including a conventional-retailer alternative which is described by one attribute reflecting the ease of access to a physical store (conventional store) with the desired product in stock. Finally, an opt-out option allows participants not to buy the product under any of the conditions offered in the choice exercise. The inclusion of the Opt-out option does not force respondents to choose an undesired alternative and prevents not-participation (Kjaer, 2005).

The combinations of the attribute levels of each attribute are combined to form the Online Retailer and Conventional Store alternatives using a D-efficient experimental design with zero priors (Rose & Bliemer, 2009). The final design contains 60 rows (scenarios) which are divided into 12 blocks so that each respondent is presented with a set of five choice tasks. The experimental design matrix is generated using the software Ngene (ChoiceMetrics, 2014) The 60 scenarios are further divided into 12 orthogonal blocks so that each respondent is presented with a set of five scenarios corresponding to one block. Each

[2] The choice experiment is introduced to the subjects by the following statement *"In the previous questions you indicated that you purchased <PRODUCT> online most recently. Now thinking about the next purchase of this item please, choose from one of the options below."*

orthogonal block is randomly assigned to respondents to ensure that any heterogeneity retrieved in the parameter estimates is not affected by the attribute levels presented to respondents (Arentze, Borgers, Timmermans, & DelMistro, 2003).

In this study, we use responses from 502 individuals, which resulted into 2,510 observations.

**Table 2. Attributes and levels in the choice of online-retailer**

| Attribute | Levels |
|---|---|
| *Premium/discount against security costs* | (1) Discount £4.00<br>(2) Discount £2.00<br>(3) No charge<br>(4) Premium £2.00<br>(5) Premium £4.00 |
| *Additional information saved and linked to your account* | (1) Only email [Reference Level]<br>(2) Purchase history and email<br>(3) Purchase history, browsing and navigation history and email<br>(4) Purchase history, browsing, navigation history, email and additional personal details |
| *Permission of sharing personal information with third parties* | (1) Yes<br>(2) No [Reference Level] |
| *Time your personal information is stored for* | (1) 1 year [Reference Level]<br>(2) 2 years<br>(3) 5 years<br>(4) Without an explicit temporal limit |
| *Availability of product or service at a conventional store/outlet*<br>(Only available in the Conventional store/outlet alternative) | (1) This item can also be easily purchased in your neighbourhood at a conventional retailer<br>(2) This item can also be purchased from a conventional retailer, but it would require from you to make a special effort because of day/hour of purchase, distance to reach the merchant, etc.)<br>(3) This item is not available to purchase from a conventional retailer in your neighbourhood [Reference Level] |
| *Additional services offered by the product provider at the same price* | (1) None [Reference Level]<br>(2) Faster checkout (one-click order)<br>(3) Detailed reviews of products/seller<br>(4) Priority shipping of product |

## 2.3 Privacy and information protection psychometric scales

As part of the survey questionnaire, respondents are also asked to indicate their strength of preference to a series of statements relating to attitudes toward privacy and information protection. As shown in Table 4, these statements o three sets of attitudinal scales previously validated by other studies (Buchanan, Paine, Joinson, & Reips, 2007; Robinson, Potoglou, Kim, Burge, & Warnes, 2010)).

The explanatory factor analysis used is a principal-axis factor analysis with promax rotation (Tabachnick & Fidell, 2007). Eigenvalue scores greater than 0.9 and the percentage of variance explained being smaller than 5% are the criteria used for selecting the number of components to extract from the analysis obtaining a three factor solution. Table 3 shows the rotated component matrix arising from the three factor solution. We obtained a similar structure as Buchanan et al. (2007) and Daly et al. (2012), where each of the significant instruments loads in the expected factor. Thus, the results from the explanatory factor analysis provide evidence on the existence of three latent variables capturing respondents': (a) General Caution (alpha = 0.877), (b) Technical Protection (alpha = 0.804) and (c) Privacy Concern (alpha = 0.611).

**Table 3. Rotated factor loadings against psychometric indicators**

| Psychometric Indicator | Factor1 | Factor2 | Factor3 |
|---|---|---|---|
| Do you shred/burn your personal documents when you are disposing of them? | 0.244 | 0.135 | 0.317 |
| Do you hide your bank card PIN number when using cash machines/making purchases? | 0.242 | 0.193 | 0.248 |
| Do you only register for websites that have a privacy policy? | **0.625** | | 0.118 |
| Do you read a website's privacy policy before you register your information? | **0.883** | | |
| Do you look for a privacy certification on a website before you register your information? | **0.784** | | |
| Do you read license agreements fully before you agree to them? | **0.812** | | |
| Do you watch for ways to control what people send you online (such as check boxes that allow you to opt-in or opt-out of certain offers)? | | 0.383 | 0.298 |
| Do you remove cookies? | 0.103 | **0.707** | -0.137 |
| Do you use a pop up window blocker? | -0.213 | **0.656** | 0.107 |
| Do you check your computer for spy ware? | | **0.646** | 0.107 |
| Do you clear your browser history regularly? | 0.103 | **0.751** | -0.185 |
| Do you block messages/emails from someone you do not want to hear from? | | **0.612** | |
| Protecting the privacy of my personal information is… | | -0.123 | **0.680** |
| Taking action against important security risks (e.g. international terrorism, organised crime) is… | | | **0.589** |
| Defending current liberties and human rights is… | | | **0.602** |
| Technology has almost got out of control | | 0.143 | -0.167 |
| Government can generally be trusted to look after our interests | | -0.114 | |
| The way one votes has no effect on what the Government does | 0.102 | -0.118 | |
| In general business helps us more than it harms us | | -0.111 | |

**Table 4. Psychometric scales for privacy concern, general and technical protection**

| Latent Construct | Description | Never (1) | Rarely (2) | Sometimes (3) | Very often (4) | Always (5) |
|---|---|---|---|---|---|---|
| **General Caution** Source: (Buchanan et al. 2007) | Do you shred/burn your personal documents when you are disposing of them? | 5.2% | 5.6% | 17.0% | 21.1% | 51.1% |
| | Do you hide your bank card PIN number when using cash machines/making purchases? | 1.6% | 3.5% | 13.8% | 16.83% | 65.4% |
| | Do you only register for websites that have a privacy policy? | 4.6% | 9.1% | 31.1% | 27.7% | 27.5% |
| | Do you read a website's privacy policy before you register your information? | 11.4% | 20.3% | 32.5% | 18.0% | 17.8% |
| | Do you look for a privacy certification on a website before you register your information? | 12.2% | 18.2% | 27.1% | 21.5% | 21.1% |
| | Do you read license agreements fully before you agree to them? | 12.4% | 23.2% | 29.6% | 19.3% | 15.5% |
| **Technical Protection** Source: (Buchanan et al. 2007) | Do you watch for ways to control what people send you online (such as check boxes that allow you to opt-in or opt-out of certain offers)? | 4.3% | 5.6% | 17.8% | 28.6% | 43.7% |
| | Do you remove cookies? | 7.5% | 13.5% | 35.2% | 24.4% | 19.3% |
| | Do you use a pop up window blocker? | 4.1% | 4.6% | 18.4% | 26.7% | 46.2% |
| | Do you check your computer for spy ware? | 4.3% | 4.3% | 17.4% | 27.5% | 46.6% |
| | Do you clear your browser history regularly? | 5.8% | 11.0% | 26.3% | 22.0% | 34.8% |
| | Do you block messages/emails from someone you do not want to hear from? | 6.0% | 9.1% | 21.2% | 20.7% | 41.0% |
| | | **Not at all important (1)** | **Not very important (2)** | **Neither (3)** | **Somewhat important (4)** | **Very important (5)** |
| **Privacy Concern** Source: (Daly et al. 2012) | Protecting the privacy of my personal information is… | 0.6% | 1.9% | 1.6% | 13.3% | 82.6% |
| | Taking action against important security risk (e.g. international terrorism, organised crime) is… | 1.0% | 2.7% | 2.1% | 15.9% | 78.3% |
| | Defending current liberties and human rights is… | 1.0% | 6.2% | 2.5% | 24.6% | 65.8% |

The first two factors reflect two latent variables which measure individuals attitudes toward privacy protection in online contexts based on privacy-protective measures. These sets of scales were derived and validated by Buchanan et al. (2007). The latent variable *General Caution* reflects general concern with regard to protection of privacy in different ways. Finally,

the latent variable *Technical Protection* comprised a series of questions specific to the use of technology for the protection of information privacy (Buchanan et al., 2007). The final set of statements asked respondents to indicate their strength of preference on a five-point Likert scale against an adapted version of the *Privacy Concern Index* (Kumaraguru & Cranor, 2005) consisting of a series of questions about their attitudes towards privacy, security and liberty.

## 3. Analytical Approach: The Integrated Latent Variable Choice Model

The analytical framework in this paper is based on the Integrated Choice and Latent Variable (ICLV) model (Bolduc, Ben-Akiva, Walker, & Michaud, 2005; Walker & Ben-Akiva, 2002). This model structure involves the simultaneous estimation of a choice model and attitudinal - structural equation (SEM) model. The ICLV model assumes that both choice and responses to attitudinal indicators are influenced by latent constructs (e.g. privacy concerns) modelling together both outcomes to give more insights into the process that motivate respondents' choices (Daly et al., 2012). ICLV updates the standard choice models based on characteristics of the alternatives and socio-economic characteristics of the respondents by incorporating latent variables describing attitudes and perceptions of individuals. At the same time, these attitudinal variables are used to explain respondents' answers to attitudinal indicators used to quantify the psychological characteristics of individuals. The ICLV model can be then used to explicitly explore the role of the latent constructs in the context of e-commerce transactions – i.e., the willingness of individuals to render their personal information online and while they engage in an online transaction.

In this study, the choice model is employed to explain how respondents make choices between the three online retailers, the conventional-store and the opt-out options given the different attributes describing each option and respondents' latent (unobserved) attitudinal constructs. In the specification of the utilities of the choice model, therefore, we incorporate the three latent variables (latent constructs) identified in Table 4. These factors cannot be directly observed, but can be inferred by respondents' socio-economic characteristics and their responses to the corresponding psychometric scales for privacy concern, general and technical protection, respectively (see, Latent Variable Model section below).

To incorporate attitudinal concerns in a discrete choice model without introducing the risk of measurement errors and endogeneity bias, the ICLV approach treats the answer to attitudinal instruments as dependent rather than explanatory variables (for a comprehensive overview of the methodological framework, see Daly et al. (2012)). Thus, the ICLV model requires the specification of two components and the associated set of structural equations: (1) a Latent Variable Model (LVM) describing respondents' responses to psychometric indicators, and (2) a choice model describing the choices by individuals in the choice tasks forming the DCE (see Figure 2).

On the right-hand side of the graph, there is a set of psychometric indicators capturing respondents' attitudes. These indicators are explained using the three latent variables General Caution, Technical Protection and Concern, which are in turn a function of socio-demographic characteristics of individuals (see top-left part in Figure 2). These latent variables are then used as predictors, along with the other attributes in the choice model, in order to explain the choices of respondents across the Online retailer, Conventional Store and Opt-out alternatives.

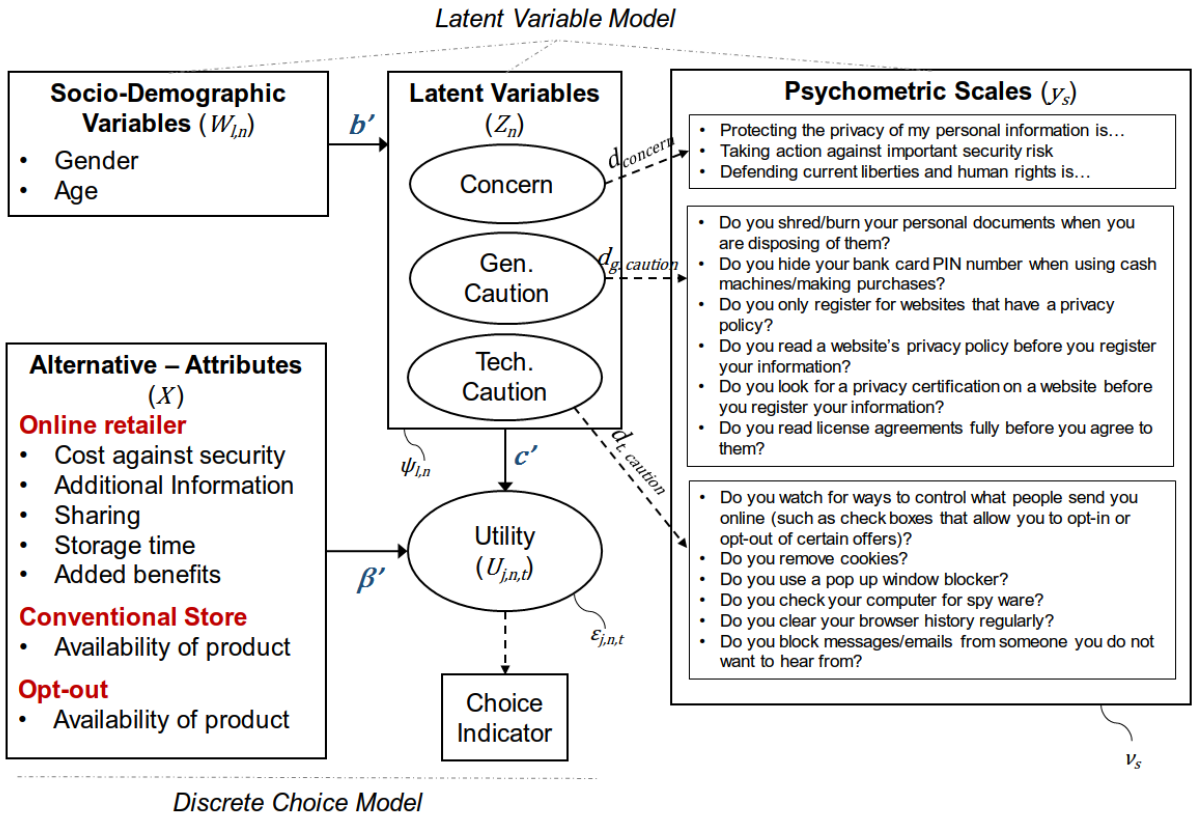


**Figure 2. The Integrated Choice and Latent Variable Modelling approach (adapted from: Walker & Ben-Akiva (2002); Bolduc et al. (2005).**

*Choice Model*

Five utility functions are specified in the choice model: one for each of the three Online Retailers, one for the Conventional Store (CS) and one the opt-out (OO) alternative.

The utility of individual $n$ when considering the online-retailer $j$ in the choice task $t$ is described as follows:

$$U_{j,n,t} = \beta_1 Cost_{j,n,t} + \beta_2' Add.Information_{j,n,t} + \beta_3 Sharing_{j,n,t} + \beta_4' StorageTime_{j,n,t} + \beta_5' Add.Services_{j,n,t} + \varepsilon_{j,n,t} \qquad \forall\, 1 \le j \le 3 \tag{1}$$

Where $Cost, Add.Information, Sharing, StoreageTime, Add.Services$ correspond to the attributes describing the online retailer $j$ and $\beta s$ are unknown coefficients to be estimated. Specifically, $\beta_1$is the coefficient representing the influence of variation in the cost attribute (premium or discount) in selecting an online retailer.

All privacy-related attributes (i.e., except the cost attribute) are dummy coded. For each attribute, we set a "base level" (see, Table 2), which represents maximum level of information protection by retailers, thus implying minimum levels of: information requirements, length of storage and restricted access to such information by third parties. The remaining attribute levels describe retailers with higher information requirements in terms relative to the corresponding base level - i.e., collecting additional details of respondents, having the right to share the information with third parties or storing respondents' information for longer periods of time.

$\beta_2'$- $\beta_5'$ are vectors of coefficients capturing the effect of the associated attribute level (against its base level) on the likelihood of choosing a specific online retailer. These parameters capture the effect of relaxing the restrictions to retailers on data collection and data handling. For instance, the coefficient $\beta_3$ captures the effect of unrestricted access to respondents' personal information on the odds of choosing a retailer, relative to the base level, in which the retailer guarantees that the information will not be shared with third parties.

The random error $\varepsilon_{j,n,t}$ is i.i.d. Type I Extreme Value (Gumbel) distributed and captures the effect of unobservable attributes, unobserved taste variations and measurement or specification errors (M. E. Ben-Akiva & Lerman, 1985).

The utilities associated with the conventional store and opt-out alternatives are as follows:

$$U_{CS,n,t} = ASC_{CS} + \beta_6 Spec.Effort_{CS,n,t} + c'Z_n + \varepsilon_{CS,n,t} \quad (2)$$

$$U_{OO,n,t} = ASC_{OO} + c'Z_n + \varepsilon_{OO,n,t} \quad (3)$$

where $ASC_{CS}$ and $ASC_{OO}$ are the alternative specific constants and capture the average effect of unobserved factors in the utility of the conventional retailer and opt-out options, respectively, and relative to online retailers (Train 2003, p.20).

The conventional-store utility function includes the attribute Spec.Effort, which describes the availability of the product at a conventional store. This variable takes the value of zero when the product can be easily bound at a nearby store and takes the value of one when special effort is required by respondents to purchase it.

$Z_n$ reflects the latent variables considered in this study, namely Privacy Concern, General Caution and Technical Protection and the coefficients $c'$ to be estimated indicate the marginal probability of selecting the Conventional Store alternative or the Opt-out relative to the probability of selecting an online retailer option.

The probability for the sequence of $T$ choices $k = \{k_1, \dots, k_T\}$ made by respondent $n$, conditional on $Z_n$ is given by:

$$P(k|Z_n, \xi_n) = \prod_T \left( \frac{\exp(\beta' X_{k,n,t} + c_k' Z_n}{\sum_j \exp(\beta' X_{j,n,t} + c_j' Z_n)} \right) \quad (4)$$

Where $X_{j,n,t}$ corresponds to the vector of attributes describing alternative $j$, and $\beta'$ is the vector of unknown coefficients to be estimated.

*Latent Variable Model*

As mentioned in the previous section, the analysis focuses on three latent variables: (a) General Caution, $z_{Gen\_caution,n}$, (b) Technical Protection, $z_{Tech\_protection,n}$ and (c) Privacy Concern, $z_{Concern,n}$. These variables are measured by the set of psychometric scales shown in Table 4. The responses to psychometric scales are modelled via 15 equations known as *measurement equations,* which are specified as follows:

$$y_s = \delta_s + d_s Z + v_s \quad (5)$$

where $y_s$ is the observed response to the $s$th psychometric scale (out of S=15). The effect of the latent variables on the value of the scale is given by the vector of parameters $d_s$ which are specific to a given attitudinal indicator. Finally, $v_s$ is the random component of the response to the psychometric indicator.

The three latent variables are assumed to be determined by a series of linear 'structural' relationships:

$$Z_{l,n} = b_l W_{l,n} + \sigma_l \psi_{l,n} \quad \text{For } l = \text{General Concern, General Caution, Technical Protection} \tag{6}$$

where $b_l W_{l,n}$ represents the deterministic part of $Z_{l,n}$, with $W_{l,n}$ corresponds to a set of socioeconomic variables (e.g. age and gender) and $\sigma_l \psi_{l,n}$ is a normally-distributed error term - with mean zero and standard deviation $\sigma_l$.

Following Daly et al. (2012), we take into consideration the ordinal character of the *S* scales included in our survey by using an ordered logit model specification. Thus, the probability that individual $n$ gives the observed response $q$ in the psychometric indicator $s$ is described by:

$$Pr\{y_{s,n} = q | Z_n\} = \Lambda(\mu_{y_{s,n}} - d_s Z_n) - \Lambda(\mu_{y_s - 1} - d_s Z_n) \tag{7}$$

The likelihood of the series of answers of respondent *n* to the psychometric indicators can be described as follows:

$$Pr\{y_n = q | Z_n\} = \prod_s \left(\Lambda(\mu_{y_{s,n}} - d_s Z_n) - \Lambda(\mu_{y_s - 1} - d_s Z_n)\right) \tag{8}$$

Where $y_{s,n}$ is the observed response to the $s$th attitudinal indicator by respondent *n*; $Z_n$ is a vector of latent variables; and $d_s$ is the vector of parameters describing the impact of each latent variable on the value of the $s$th indicator, which takes the value of zero when a latent factor does not have any impact on a specific indicator. We follow Ben-Akiva et al. (1999) for the normalisation of the scale for the measurement equations, which consists of normalising one of the parameters associated with each of the latent variables $d_s$ to one.

As part of the ICLV specification shown in Figure 2, the utility functions in the choice model component incorporate a vector of latent variables $Z_n$ as an additional predictor for the choices made by the respondents. Where $c'$ is the vector of coefficients capturing the effect of the latent variables on the utility of selecting alternative $j$ – specifically here the probability of selecting the conventional store and opt-out options. In this study, we normalise the value of $c_1, c_2$ and $c_3$ to zero such that $c_{CS}$ and $c_{OO}$ correspond to the effect of variation on the level of privacy concern on the probability of opting-out an e-commerce transaction.

The likelihood function of the ICLV model is the product of the likelihood for the measurement model (Eq. 8) and the likelihood of the series of choices (Eq. 4). For details on the estimation of maximum likelihood estimation of the ICLV model, see Daly et al. (2012). The estimated parameters are estimated by maximising the simulated likelihood function given by the product of the likelihoods of each observation – i.e. the product of the series of five choice tasks and the set of responses to the S attitudinal indicators. As a base model, we also estimated a simple Multinomial Logit (MNL) model without latent variables.

# 4. Results

This section discusses the results of the estimated models based on the collected stated choice, individual characteristics and psychometric-scale response data.

Tables 6, 7 and 8 present the estimated parameters of four ICLV models each split into its corresponding model components, namely: the choice, structural equation and the measurement equations models. The first three ICLV models introduce one latent variable at a time and thus are named as ICLV Concern when the latent variable (Privacy) Concern is introduced, ICLV General Caution and ICLV Technical Protection when the latent variables General Caution and Technical Protection are introduced in the choice model, respectively. Also, Table 6 presents the estimated parameters of a multinomial logit model (MNL), a model specification which only includes the estimates of the attributes describing online retailers, the conventional store and the opt-out alternatives and does not include any latent variables.

Table 5 presents a summary of each model in terms of: the number of respondents, the number of observations, the number of Halton draws used in the simulations (where applicable), the number of parameters in the discrete choice model (DCM) component and the ICLV overall, and the values of the corresponding likelihood functions. All models reported in the following sections are estimated in Python-Biogeme (Bierlaire, 2003; Bierlaire & Fetiarison, 2009) and the estimates were confirmed by re-estimating the models in Ox (Doornik, 2001).

As shown in Table 5, the log-likelihood for the choice component of the model is substantially improved with the inclusion of the latent variables and the corresponding latent-variable model components. This is particularly the case for the ICLV Concern and ICLV General Caution whereas the inclusion of all variables in the Joint ICLV only marginally improves the model estimation when compared to the ICLV concern.

**Table 5. Models statistics**

| | MNL | ICLV Concern | ICLV General Caution | ICLV Technical Protection | Joint ICLV |
|---|---|---|---|---|---|
| Number of individuals | 502 | 502 | 502 | 502 | 502 |
| Number of observations | 2,510 | 2,510 | 2,510 | 2,510 | 2,510 |
| Number of Halton draws | - | 1,000 | 1,000 | 1,000 | 1,000 |
| Log-likelihood (DCM) | -3,497.16 | -3,177.28 | -3,060.65 | -3,486.45 | -3,177.76 |
| Number of parameters (DCM) | 13 | 14 | 14 | 14 | 16 |
| Number of parameters (ICLV) | - | 31 | 36 | 41 | 82 |
| Log-likelihood (ICLV) | - | -4265.7 | -5974.9 | -7010.3 | -10764.7 |
| AIC (DCM) | 7020.3 | 6382.6 | 6149.3 | 7000.9 | 6387.5 |
| BIC(DCM) | 7038.5 | 6402.2 | 6168.9 | 7020.5 | 6409.9 |

Each of the ICLV models specifically tests the hypothesis that the higher an individual's concern, general caution and technical protection the more likely these consumers are to opt-in in an online transaction. All else being equal, the estimated coefficients of Concern, General Caution and Technical Protection each in the corresponding ICLV Concern, ICLV General Caution and ICLV Technical Protection models, satisfy the above hypothesis. For example, the estimates in the choice-model component of the ICLV Concern model show that the higher the (Privacy) Concern of an individual the more likely these consumers are to purchase a product from a conventional store or opt-out from the choice task, as a matter of protest vote, for example.

The Joint ICLV model tests the above hypothesis when all three latent variables are introduced in the Conventional Store and Opt-out alternatives. In the Joint model, it is shown that the effect of the Concern latent variable remains almost unaffected and is strongly significant (t-ratio = 4.43) whereas influence of the General Caution and Technical Protection diminishes. This is an indication that the latent variable Concern outweighs the other two latent variables, General Caution and Technical Protection. It should be noted that the sign of the coefficient of Technical Protection becomes negative, though remains insignificant. The change of the coefficient possibly implies that after controlling for an individual's level of (Privacy) Concern and General Caution as observed through perception and general practice when using the Internet, it is shown that for a (non-statistically significant) proportion in the sample, those who seek technical protection are more likely to engage in an online transaction than purchasing a product online or opt-out from the choice exercise.

*Choice Model Component*

The parameter estimates in the MNL model in Table 6 reveal the effect of each attribute describing the online retailers and the Conventional Store alternatives without the specification of any latent attitude variables in the utilities. The effect of premium/discount (*cost per transaction)* included in the observed part of the utility of the online-retailer is specified with a linear term. As expected, the parameter of the premium/discount has negative sign implying that the higher the premium the less likely a consumer is to choose an online retailer presenting that premium.

The coefficients in the attribute levels of the a*dditional information saved and linked to your account* attribute were dummy coded. 'Only e-mail' corresponds to the minimum information required to complete the purchase, and is set as the reference level. Again, the negative sign of the coefficients indicates that consumers are more likely to choose retailers that collect the minimum amount of personal information (i.e., only email) to complete the purchase. Therefore, consumers are less likely to select retailers that, besides email address, record and store the 'purchase history' or the 'purchase history and browsing history' of their customers.

Preferences in favour of restricting access to consumer information beyond the online retailer are also reflected in the negative effect of *sharing information with third parties* attribute. At the reference level, retailers compromise to restrict the access of any recorded information within the retailer and not share that information with any other public or private institution. While this strategy reduces the potential profitable strategies by companies, it definitely increases the odds of being selected by individuals. It is worth noting that the absolute value of the coefficient describing the effect of unbounded access to personal information is the highest among the privacy-related (dummy coded) coefficients.

To complete the results of privacy-related attributes, we present the effect of increasing the timeframe within which retailers store respondents' information. Setting "one year" as the reference level, the estimated coefficients show that as retailers increase the timeframe during which the collected information is stored, the likelihood of a retailer being selected decreases. Preliminary estimations showed no statistical difference storing personal information between one and two years; hence the reference level in Table 6 is set at 1-2 years. Further, the significantly negative coefficient associated with "without an explicit temporal limit" indicates that participants tend to avoid retailers who do not compromise to remove consumers' information after a specified period of time. However, the (negative) effect between a 5-year storage limit and a 'without an explicit temporal limit' are valued (statistically) equally negative – i.e., there is not significant difference between the 5-year and 'no explicit temporal limit' coefficients.

Online retailers may persuade individuals to select them by reducing the price of the product, protecting their personal data, and offering additional services. To capture the latter, we included three different additional services that online retailers may offer free of charge. We set the reference level as the situation in which retailers do not offer any additional service. As shown in Table 6, results indicate that – all else being equal, additional services free of charge effectively increase the likelihood of an online retailer being selected by respondents. These services are equally valued as there are no significant differences in the coefficients the free services on offer, namely *faster check out*, *detailed reviews* of the product or the seller and *priority shipping* of the purchased product.

Finally, the utility of the fourth and fifth alternatives are given by two alternative specific constants (ASC). The negative coefficients of the ASC for Conventional Store and Opt-out options show that, all else being equal, respondents are less likely to buy products from a conventional retailer or opt out from the choice context presented to them when compared to the reference levels describing an online retailer. As shown in Table 2, the reference levels describing an online retailer include the maximum level of information protection that is collecting the minimum amount of information, restricting third parties from access to such information and storing customers' information just for one-two years and offer no additional services.

The above patterns of attribute-level coefficients remain unchanged across the DCM parts of the ICLV models reported in Table 6.

*Latent Variable Model Component*

As mentioned at the beginning of this section, we explore the impact of the latent constructs (Concern, General Caution, Technical Protection) on respondents' choices of retailers by adding the three latent variables as regressors in the utilities for Conventional Store and Opt-out options in the choice-model component of the ICLV both separately and in a joint model. The choice- and latent-variable model components are estimated simultaneously thus achieving consistent and efficient parameter estimates (Daly et al. 2012). The repeated-response (panel) nature of the data is also taken into considerations across all models.

The parameter estimates of the latent variable are better understood in conjunction with the structural and measurement equations in Tables 7 and 8. Table 7 shows the estimation results for the structural equations. The estimated coefficients indicate that respondents' age and gender are positively associated with each of the latent constructs, except in the last latent construct, Technical Protection, in which it is shown that women are less likely to use available technology to protect the privacy of their information when they are online. Other socio-economic effects tested included respondents' familiarity with the Internet measured as the 'number of hours spent on the Internet' and social class, a composite index of education and occupation status, but these were not statistically significant.

**Table 6. Parameter estimates in the choice model**

| | MNL | | ICLV Concern | | ICLV General Caution | | ICLV Technical Protection | | Joint ICLV | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Est. | t-ratio | Est. | t-ratio | Est. | t-ratio | Est. | t-ratio | Est. | t-ratio |
| **Online retailer Alternatives** | | | | | | | | | | |
| *Cost per transaction (in £)* | -0.128 | -14.49 | -0.130 | -14.58 | -0.129 | -14.55 | -0.128 | -14.51 | -0.130 | -14.58 |
| *Additional information saved and linked to your account* | | | | | | | | | | |
| Only mail | Reference level | | | | | | | | | |
| Purchase history and email | -0.165 | -2.42 | -0.182 | -2.60 | -0.162 | -2.37 | -0.165 | -2.41 | -0.181 | -2.59 |
| Purchase history, browsing, navigation history and email | -0.470 | -6.49 | -0.479 | -6.48 | -0.470 | -6.48 | -0.470 | -6.48 | -0.478 | -6.47 |
| Purchase history, browsing, navigation history, additional personal details and email | -0.403 | -5.58 | -0.413 | -5.56 | -0.396 | -5.48 | -0.402 | -5.56 | -0.411 | -5.54 |
| *Information is shared with third parties* | -0.704 | -13.66 | -0.722 | -13.79 | -0.709 | -13.74 | -0.705 | -13.67 | -0.723 | -13.79 |
| *Time your personal information is stored for* | | | | | | | | | | |
| 1-2 years | Reference level | | | | | | | | | |
| 5 years | -0.311 | -4.83 | -0.352 | -5.34 | -0.313 | -4.86 | -0.315 | -4.90 | -0.353 | -5.36 |
| Without an explicit temporal limit | -0.340 | -5.27 | -0.362 | -5.51 | -0.344 | -5.32 | -0.342 | -5.30 | -0.363 | -5.53 |
| *Additional services free of charge* | | | | | | | | | | |
| None | Reference level | | | | | | | | | |
| Faster checkout (one-click order) | 0.369 | 4.92 | 0.385 | 5.06 | 0.369 | 4.91 | 0.369 | 4.93 | 0.385 | 5.06 |
| Detailed reviews of products/seller | 0.438 | 5.82 | 0.435 | 5.69 | 0.440 | 5.85 | 0.438 | 5.83 | 0.435 | 5.70 |
| Priority shipping of product | 0.441 | 5.84 | 0.469 | 6.09 | 0.442 | 5.84 | 0.442 | 5.85 | 0.469 | 6.10 |
| **Conventional Store Alternative** | | | | | | | | | | |
| Alternative Spec. Constant - Conventional Store | -0.481 | -4.31 | -3.07 | -5.89 | -0.828 | -5.90 | -0.516 | -4.38 | -3.11 | -5.92 |
| This item can also be purchased in your neighbourhood at a conventional retailer | Reference Level | | | | | | | | | |
| This item can also be purchased from a conventional retailer, but it would require from you to make a special effort (because of opening hours, distance to reach the merchant, etc.) | -0.637 | -4.51 | -0.835 | -5.08 | -0.672 | -4.72 | -0.654 | -4.62 | -0.834 | -5.09 |
| **Opt-Out Alternative** | | | | | | | | | | |
| Alternative Spec. Constant – Opt-Out | -0.869 | -9.64 | -3.66 | -7.06 | -1.24 | -9.84 | -0.913 | -9.29 | -3.70 | -7.08 |
| **Latent Variables (Conv. Store & Opt-Out alt.)** | | | | | | | | | | |
| Concern | | | 3.17 | 4.31 | | - | | | 2.97 | 4.43 |
| General Caution | | - | | | 0.217 | 6.05 | | - | 0.0951 | 0.99 |
| Technical Protection | | | | - | | - | 0.095 | 2.87 | -0.119 | -1.31 |

**Table 7. Parameter estimates in the structural equation model component of ICLV**

| | ICLV Concern | | ICLV General Caution | | ICLV Technical Protection | | Joint ICLV | |
|---|---|---|---|---|---|---|---|---|
| | Est. | t-ratio | Est. | t-ratio | Est. | t-ratio | Est. | t-ratio |
| *Concern* | | | | | | | | |
| Age | 0.008 | 2.19 | | | | | 0.00852 | 2.27 |
| Gender (female) | 0.348 | 2.89 | | - | | - | 0.339 | 2.78 |
| Standard deviation | 0.864 | 4.55 | | | | | 0.926 | 4.82 |
| *General Caution* | | | | | | | | |
| Age | | | 0.029 | 4.80 | | | 0.0268 | 4.43 |
| Gender (female) | | - | 0.464 | 2.52 | | - | 0.362 | 2.03 |
| Standard deviation | | | 1.85 | 12.05 | | | 1.87 | 11.97 |
| *Technical Protection* | | | | | | | | |
| Age | | | | | 0.014 | 1.98 | 0.00972 | 1.43 |
| Gender (female) | | - | | - | -0.462 | -2.06 | -0.670 | -3.04 |
| Standard deviation | | | | | 2.15 | 11.03 | 2.13 | 11.05 |

Table 8 shows that the latent variable Concern is positively associated with the privacy, security and liberty indicators. The highest rating is attributed to the privacy indicator when compared with the security and liberty indicators. Similarly for the other two latent variables, General Caution, Technical Protection, these are positively associated with all their indicators. Individuals with high levels of 'Concern' are those who attach high relevance to the indicators 'security', 'liberty' and especially 'privacy'. High values of the latent variables 'General Caution' and 'Technical Protection' are both associated with those individuals who more regularly implement measures to protect the privacy of their personal information in online scenarios. Therefore, an increased value of any of the abovementioned latent variables implies a positive association with all their corresponding indicators indicating 'strong agreement' or 'always' in the psychometric scales presented in Table 4.

**Table 8. Parameter estimates in the measurement model component of ICLV**

| | ICLV Concern | | ICLV General Caution | | ICLV Technical Protection | | Joint ICLV | |
|---|---|---|---|---|---|---|---|---|
| | Est. | t-ratio | Est. | t-ratio | Est. | t-ratio | Est. | t-ratio |
| ***Concern indicators (d parameters)*** | | | | | | | | |
| Privacy (Indicator 10) | 1 | - | | | | | 1 | - |
| Security (Indicator 11) | 0.613 | 3.21 | | - | | - | 0.607 | 3.39 |
| Liberty (Indicator 12) | 0.543 | 3.36 | | | | | 0.544 | 3.56 |
| ***General Caution indicators (d parameters)*** | | | | | | | | |
| Register only if privacy policy( Indicator 1) | | | 1 | - | | | 1 | - |
| Read privacy policy (Indicator 2) | | - | 2.770 | 6.30 | | - | 2.650 | 6.63 |
| Privacy certification (Indicator 3) | | | 1.400 | 9.78 | | | 1.450 | 9.64 |
| License agreements (Indicator 4) | | | 1.490 | 9.05 | | | 1.500 | 8.95 |
| ***Technical Protection indicators (d parameters)*** | | | | | | | | |
| Cookies (Indicator 5) | | | | | 1 | - | 1 | - |
| Pop-up blocker (Indicator 6) | | | | | 0.712 | 7.48 | 0.718 | 7.54 |
| Spy ware (Indicator 7) | | - | | - | 1.180 | 7.18 | 1.190 | 7.22 |
| Clear browsing history (Indicator 8) | | | | | 1.230 | 8.50 | 1.240 | 8.58 |
| Block messages (Indicator 9) | | | | | 0.756 | 7.82 | 0.747 | 7.88 |
| **Constants (δs)** | | | | | | | | |
| Privacy | 6.110 | 6.02 | | | | | 6.17 | 6.07 |
| Security | 4.950 | 8.43 | | - | | | 4.96 | 8.45 |
| Liberty | 4.950 | 8.46 | | | | | 4.97 | 8.48 |
| Register only if privacy policy | | | 2.950 | 7.73 | | - | 3.10 | 8.23 |
| Read privacy policy | | - | 2.450 | 2.83 | | | 2.80 | 3.44 |
| Privacy certification | | | 1.590 | 3.59 | | | 1.84 | 4.08 |

| | ICLV Concern | | ICLV General Caution | | ICLV Technical Protection | | Joint ICLV | |
|---|---|---|---|---|---|---|---|---|
| | Est. | t-ratio | Est. | t-ratio | Est. | t-ratio | Est. | t-ratio |
| License agreements | | | 1.600 | 3.41 | | | 1.83 | 3.95 |
| Cookies | | | | | 3.710 | 8.39 | 4.05 | 9.04 |
| Pop-up blocker | | | | | 3.960 | 10.48 | 4.22 | 11.03 |
| Spy ware | | | | - | 4.900 | 8.87 | 5.34 | 9.25 |
| Clear browsing history | | | | | 4.570 | 8.06 | 5.03 | 8.58 |
| Block messages | | | | | 3.520 | 9.95 | 3.77 | 10.61 |
| **Thresholds (μy,s,n )** | | | | | | | | |
| **Concern** | | | | | | | | |
| *Privacy Thresholds* | | | | | | | | |
| Threshold 1 | 0 | - | | | | | 0 | - |
| Threshold 2 | 2.350 | 2.47 | | | | - | 2.350 | 2.46 |
| Threshold 3 | 2.919 | 2.99 | | | | | 2.926 | 2.99 |
| Threshold 4 | 4.819 | 4.79 | | | | | 4.856 | 4.81 |
| *Security Thresholds* | | | | | | | | |
| Threshold 1 | 0 | - | | | | | 0 | - |
| Threshold 2 | 1.720 | 3.27 | | | | - | 1.720 | 3.27 |
| Threshold 3 | 2.240 | 4.07 | | | | | 2.242 | 4.07 |
| Threshold 4 | 3.900 | 6.70 | | | | | 3.902 | 6.71 |
| *Liberty Thresholds* | | | | | | | | |
| Threshold 1 | 0 | - | | | | | 0 | - |
| Threshold 2 | 2.480 | 4.47 | | | | - | 2.490 | 4.48 |
| Threshold 3 | 2.826 | 5.01 | | | | | 2.838 | 5.02 |
| Threshold 4 | 4.556 | 7.81 | | | | | 4.578 | 7.83 |
| **General Caution** | | | | | | | | |
| *Register only if privacy policy* | | | | | | | | |
| Threshold 1 | | | 0 | - | | | 0 | - |
| Threshold 2 | | - | 1.660 | 7.00 | | - | 1.650 | 6.99 |
| Threshold 3 | | | 4.160 | 13.70 | | | 4.160 | 13.70 |
| Threshold 4 | | | 6.030 | 17.80 | | | 6.030 | 17.81 |
| *Read privacy policy* | | | | | | | | |
| Threshold 1 | | | 0 | - | | | 0 | - |
| Threshold 2 | | - | 3.950 | 7.29 | | - | 3.810 | 7.68 |
| Threshold 3 | | | 8.580 | 10.58 | | | 8.270 | 11.22 |
| Threshold 4 | | | 11.790 | 12.66 | | | 11.360 | 13.40 |
| *Privacy certification* | | | | | | | | |
| Threshold 1 | | | 0 | - | | | 0 | - |
| Threshold 2 | | - | 2.100 | 9.95 | | - | 2.140 | 9.82 |
| Threshold 3 | | | 4.320 | 15.29 | | | 4.420 | 15.04 |
| Threshold 4 | | | 6.330 | 18.68 | | | 6.490 | 18.43 |
| *License agreements* | | | | | | | | |
| Threshold 1 | | | 0 | - | | | 0 | - |
| Threshold 2 | | - | 2.620 | 11.10 | | - | 2.640 | 11.05 |
| Threshold 3 | | | 5.130 | 16.41 | | | 5.160 | 16.32 |
| Threshold 4 | | | 7.360 | 19.37 | | | 7.400 | 19.27 |
| **Technical Protection** | | | | | | | | |
| *Cookies* | | | | | | | | |
| Threshold 1 | | | | | 0 | - | 0 | - |
| Threshold 2 | | - | | - | 1.770 | 8.590 | 1.780 | 8.56 |
| Threshold 3 | | | | | 4.340 | 15.190 | 4.340 | 15.19 |
| Threshold 4 | | | | | 6.470 | 18.73 | 6.460 | 18.76 |
| *Pop-up blocker* | | | | | | | | |
| Threshold 1 | | | | | 0 | - | 0 | - |
| Threshold 2 | | - | | - | 1.090 | 5.050 | 1.090 | 5.05 |
| Threshold 3 | | | | | 2.850 | 10.23 | 2.860 | 10.22 |
| Threshold 4 | | | | | 4.430 | 14.43 | 4.440 | 14.41 |
| *Spy ware* | | | | | | | | |
| Threshold 1 | | | | | 0 | - | 0 | - |
| Threshold 2 | | - | | - | 1.160 | 4.730 | 1.170 | 4.74 |
| Threshold 3 | | | | | 3.320 | 9.780 | 3.340 | 9.76 |
| Threshold 4 | | | | | 5.510 | 14.020 | 5.540 | 14.01 |
| *Clear browsing history* | | | | | | | | |
| Threshold 1 | | - | | - | 0 | - | 0 | - |
| Threshold 2 | | | | | 1.970 | 7.490 | 2.000 | 7.46 |

| | ICLV Concern | | ICLV General Caution | | ICLV Technical Protection | | Joint ICLV | |
|---|---|---|---|---|---|---|---|---|
| | Est. | t-ratio | Est. | t-ratio | Est. | t-ratio | Est. | t-ratio |
| Threshold 3 | | | | | 4.310 | 12.640 | 4.340 | 12.61 |
| Threshold 4 | | | | | 6.190 | 15.900 | 6.210 | 15.87 |
| *Block messages* | | | | | | | | |
| Threshold 1 | | | | | 0 | - | 0 | - |
| Threshold 2 | - | | - | | 1.360 | 7.080 | 1.350 | 7.07 |
| Threshold 3 | | | | | 3.090 | 12.620 | 3.070 | 12.61 |
| Threshold 4 | | | | | 4.340 | 16.020 | 4.310 | 16.01 |

# 5. Discussion and Conclusion

We report on respondents' preferences from a nationally representative sample of the Online User population in the UK. Therefore, this paper addresses previous calls for research on personal information to focus on representative and particularly non-student samples (Belanger & Crossler, 2011, pp.1026). Another critique on previous studies is that they are primarily US-based (Belanger & Crossler, 2011, pp.1019). Our study draws findings from a nationally representative sample of online users in the UK, ensuring the representativeness of our results for the British population of internet users and not specific to the nature of the subjects studied (e.g. university students).The British population is one of the most active in terms of internet adoption with 76% of adults (38 million) accessing the Internet on a daily basis; purchasing goods or services online is of the most popular activities (Office for National Statistics, 2014). Thus, this paper widens the existing empirical evidence departing from narrow socio-geographic based samples generally found in the literature (Bélanger & Crossler, 2011)[3] by gathering nationally representative data.

This study contributes to a growing body of literature that has used different experimental and laboratory approaches to study e-shopper behaviour and their willingness to provide their personal information in online transactions (see, for example Tsai et al. 2010; Acquisti et al. 2013; Jentzsch et al. 2012; Beresford et al. 2010; Hui et al. 2007; Hann et al. 2007). Our paper illustrates how data from advanced preference elicitation methods, such as stated preference discrete choice experiments, and discrete choice models used for their analysis can be employed to obtain a better understanding of e-shoppers' choices for retailers that involve varying levels of personal information. Among the alternatives on offer, there is an option to purchase the same product from a conventional store or opt-out from the hypothetical choice exercise. In particular, we show how discrete choice experiments can be used to explore the relevance of protecting the privacy of their personal information in e-commerce transactions. Respondents face trade-offs between restricting the conditions under which retailers can collect and handle their information, and the access to benefits such as discount or additional services (e.g. priority shipping, etc.). Respondents are more likely to buy products from retailers that only require consumers' email, store their personal information for one year and do not share that information with third parties. On the other hand, they are also against paying premiums in order to introduce control of their personal data.

While individuals' concern with regard to their privacy have been included in several models attempting to explain willingness to provide information in online transactions (Dinev & Hart 2006) or willingness to accept for providing personal data (Motiwalla, Li, & Liu, 2014); this is the first time (to the knowledge of the authors) that it is explicitly incorporated in a model describing choices for online retailers and how these are simultaneously influenced by

[3] For an existing review of empirical studies in the literature , see among others Bélanger & Crossler (2011); Smith et al. (2011); or Pavlou (2011).

respondents'/consumers' attitudes such as privacy concern, general caution and technical protection.

Our study shows that individuals' willingness to engage in online transactions and disclose their personal information are dependent on age and their level of concern, general caution and technical protection. The results of our analysis show that latent attitudinal constructs such as Privacy Concern, General Caution and Technical Protection do play a significant role in their engagement with e-commerce services, such purchasing a good. In particular, the higher the levels of an individual's Privacy Concern and General Caution the more likely the individual is to purchase the product from a conventional store or opt-out from purchasing that product. Furthermore, these latent constructs positively influence responses to a set of psychometric scales related to privacy, security and liberty concerns as well as perceptions and practises reflecting general caution and technical protection. These latent constructs vary by age and gender; older individuals and women are generally more concerned and cautious and older individuals are more likely seek for technical protection. On the other hand, women are less likely to enhance their technical protection when using the Internet. Finally, in preliminary estimations, we find that frequency of using the Internet for private use and social status do not influence the abovementioned latent constructs. By specifying the latent attitudinal constructs in the discrete choice models we do find that the explanatory power of these models does improve substantially relative to a conventional discrete choice model without latent attitudinal variables.

We have also included a set of psychometric scales in the survey in order to capture a latent variable reflecting an individuals' general trust. However, we were unable to develop a latent construct out of these scales as shown by the explanatory factor analysis. Therefore, we were unable to test hypotheses between willingness to disclose personal information online and trust, as it has been previously reported by other researchers (Dinev et al, 2006). Finally, subsequent analysis will seek to refine the findings presented in the paper in terms of selecting a 'best-fit' model and explore potential interactions between attributes describing online retailers and latent constructs.

## Acknowledgements

The authors are grateful to Dr Patil Sunil (Rand Europe) for his advice on the design of the experiments and Prof. Stephane Hess (ITS Leeds), Prof. Andrew Daly (ITS Leeds and RAND Europe) and Bhanu Patruni (Rand Europe) for their suggestions in model estimations. Any errors are our own.

## References

Acquisti, A. (2010). The economics of personal data and the economics of privacy. Background Paper for OECD Joint WPISP-WPIE Roundtable 1.

Acquisti, A., John, L. K., & Loewenstein, G. (2013). What Is Privacy Worth? *The Journal of Legal Studies (42:2)*, *42*(2), 249–274. doi:10.1086/671754

Arentze, T., Borgers, A., Timmermans, H., & DelMistro, R. (2003). Transport stated choice responses: effects of task complexity, presentation format and literacy. *Transportation Research Part E: Logistics and Transportation Review, 39 (3)*, 229–244.

Bélanger, F., & Crossler, R. E. (2011). Privacy in the Digital Age: A Review of Information Privacy Research in Information Systems. *MIS Quarterly*, *35*(4), 1017–1041.

Ben-Akiva, M. E., & Lerman, S. R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press.

Ben-Akiva, M., Walker, J., Bernardino, A. T., Gopinath, D. A., Morikawa, T., & Polydoropoulou, A. (1999). Integration of Choice and Latent Variable Models. *Massachusetts Institute of Technology, Cambridge*.

Beresford, A. R., Preibusch, S., & Kübler, D. (2010). Unwillingness to Pay for Privacy: A Field Experiment. *IZA Discussion Paper No 5017*, *117*(5017), 25–27. doi:10.1016/j.econlet.2012.04.077

Bierlaire, M. (2003). BIOGEME: A free package for the estimation of discrete choice models. In *Proceedings of the 3rd Swiss Transportation Research Conference*. Ascona, Switzerland.

Bierlaire, M., & Fetiarison, M. (2009). Estimation of discrete choice models : extending BIOGEME Estimation of discrete choice models : extending BIOGEME. *Swiss Transport Research Conference*, (September).

Bolduc, D., Ben-Akiva, M., Walker, J., & Michaud, A. (2005). Hybrid choice models with logit kernel: applicability to large scale models. In Elsevier (Ed.), *Lee-Gosselin, M., Doherty, S. (eds.) Integrated land-use and transportation models: behavioural foundations,* (pp. 275–302). Oxford.

Buchanan, T. of measures of online privacy concern and protection for use on the I., Paine, C., Joinson, A. N., & Reips, U. D. (2007). Development of measures of online privacy concern and protection for use on the Internet. *Journal of the American Society for Information Science and Technology*, *58*(2), 154–165.

Centre for Retail Research. (2015). Online Retailing: Britain, Europe, US and Canda 2015. Retrieved February 2, 2015, from http://www.retailresearch.org/onlineretailing.php

ChoiceMetrics. (2014). Ngene 1.1.2. User Manual and Reference Guide: The cutting edge in experimental design, ChoiceMetrics, Sydney.

Daly, A., Hess, S., Patruni, B., Potoglou, D., & Rohr, C. (2012). Using ordered attitudinal indicators in a latent variable choice model: A study of the impact of security on rail travel behaviour. *Transportation*, *39*(2), 267–297.

Dinev, T., & Hart, P. (2006). An Extended Privacy Calculus Model for E-Commerce Transactions. *Information Systems Research (17:1)*. doi:10.1287/isre.1060.0080

Doornik, J. A. (2001). Ox: An Object-Oriented Matrix Language. *Timberlake Consultants Press, London*.

EC, E. commission. (2011). *SPECIAL EUROBAROMETER 359 Attitudes on Data Protection and Electronic Identity in the European Union*.

Godel, M., Litchfield, A., & Mantovani, I. (2012). The Value of Personal Information Evidence from Empirical Economic studies. In *Communications & Strategies, No. 88, 4th Quarter 2012* (pp. 41–60).

Goldfarb, A., & Tucker, C. (2011). Online Display Advertising: Targeting and Obtrusiveness. *Marketing Science*, *30*(3), 389–404. doi:10.1287/mksc.1100.0583

Hann, I.-H., Hui, K.-L., Lee, S.-Y. T., & Png, I. P. (2007). Overcoming Online Information Privacy Concerns: An Information-Processing Theory Approach. *Journal of Management Information Systems*, *24*, 13–42. doi:10.2753/MIS0742-1222240202

Hess, S., & Rose, J. M. (2009). Should reference alternatives in pivot design SC surveys be treated differently? *Environmental and Resource Economics*, *42*(3), 297–317.

Hui, K.-L., Teo, H. H., & Lee, S.-Y. T. (2007). The Value of Privacy Assurance: An Exploratory Field Experiment. *MIS Quarterly*, *31*(1), 19–33. doi:Article

Jentzsch, N., Preibusch, S., & Harasser, A. (2012). Study on monetising privacy An economic model for pricing personal information. *Agenda*, 76.

Kjaer, T. (2005). A review of the discrete choice experiment- with emphasis on its application in health care. *Denmark: Syddansk Universitet*.

Kumaraguru, P., & Cranor, L. F. (2005). Privacy Indexes: A Survey of Westin's Studies.

Louviere, J. J., Hensher, D. A., & Swait, J. D. (2000). *Stated choice methods: analysis and applications*. *Analysis and Applications* (Vol. 12). Cambridge University Press.

Monteleone, S. (2013). The Privacy Paradox between Users' Attitudes, Stringent Legal Framework and (the Lack of) Adequate Implementation Tools. *Human Aspects of Information Security, Privacy, and Trust, L. Marinos and I. Askoxylakis (eds.). Springer Berlin Heidelberg*, 284–294.

Motiwalla, L. F., Li, X. (Bob), & Liu, X. (2014). Privacy paradox: does stated privacy concerns translate into the valuation of personal information? In *PACIS 2014 Proceedings*.

OECD. (2012). *Internet Economy Outlook*.

Office for National Statistics. (2014). *Internet Access-Household and Individuals*.

ONS. (2011). Internet Access - Households and Individuals. Retrieved February 19, 2015, from http://www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%3A77-226727

Pavlou, P. A. (2011). State of the Information Privacy Literature : Where Are We Know and Where Should We Go? *MIS Quarterly*, *35*, 977–988.

Preibusch, S. (2013). Guide to measuring privacy concern: Review of survey and observational instruments. *International Journal of Human Computer Studies*, *71*(12), 1133–1143.

Robinson, N., Potoglou, D., Kim, C. W., Burge, P., & Warnes, R. (2010). Security at What Cost? Quantifying People's Trade-Offs across Liberty, Privacy, and Security. Retrieved October 28, 2014, from http://www.rand.org/pubs/technical_reports/TR664.html

Rose, J. M., & Bliemer, M. C. J. (2009). Constructing Efficient Stated Choice Experimental Designs. *Transport Reviews*.

Smith, H. J., Dinev, T., & Xu, H. (2011). Information privacy research: an interdisciplinary review. *MIS Quarterly*, *35*(4), 989–1015. doi:10.1126/science.1103618

Smith, H. J., Milberg, S. J., & Burke, S. J. (1996). Information Privacy: Measuring Individuals' Concerns about Organizational Practices. *MIS Quarterly*, *20*(2), 167–196. doi:10.2307/249477

Spiekermann, S., Grossklags, J., & Berendt, B. (2001). E-privacy in 2nd generation E-commerce: privacy preferences versus actual behavior. In M. P. Wellman & Y. Shoham (Eds.), *Proceedings of the 3rd ACM conference on Electronic Commerce* (pp. 38–47). ACM. doi:10.1145/501158.501163

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics*. (Pearson International, Ed.)*Using multivariate statistics* (5th edit.). Boston. doi:10.1037/022267

Train, K. E. (2003). Discrete choice methods with simulation, Kenneth E. Train, Cambridge University Press, 2003, ISBN: 0-521-81696-3, pp. 334. *Journal of Applied Econometrics*, *18*(3), 379–383. doi:10.1002/jae.719

Tsai, J. Y., Egelman, S., Cranor, L., & Acquisti, A. (2010). The Effect of Online Privacy Information on Purchasing Behavior: An Experimental Study. *Information Systems Research*, *22*(June), 254–268. doi:10.1287/isre.1090.0260

Walker, J., & Ben-Akiva, M. (2002). Generalized random utility model. *Mathematical Social Sciences*, *43*(3), 303–343.