The Journal of Choice Modelling 17 (2015) 10–27

Contents lists available at ScienceDirect

The Journal of Choice Modelling

journal homepage: www.elsevier.com/locate/jocm

An integrated latent variable and choice model to explore the role of privacy concern on stated behavioural intentions in e-commerce

Dimitris Potoglou [a,*], Juan-Francisco Palacios [b], Claudio Feijóo [c]

[a] School of Planning and Geography, Cardiff University, Cardiff, United Kingdom
[b] Technical University of Madrid, Madrid, Spain
[c] Technical University of Madrid, Madrid, Spain - Tongji University, Shanghai, China

ARTICLE INFO

Article history:
Received 8 July 2015
Received in revised form
17 October 2015
Accepted 9 December 2015

Keywords:
Behavioural intentions
Discrete choice experiment
Privacy concern
E-commerce
Integrated choice and latent variable model

ABSTRACT

Consumers' privacy concerns remain the primary barrier for future growth of e-commerce. Research to date has so far considered privacy concerns either as an independent variable to explore consumers' actual (or stated) behavioural intentions (e.g. the decision to purchase goods online) or as dependent variable explained through a number of antecedents (e.g. privacy awareness). However, there has not been a formal link across antecedents, latent constructs and (stated) behavioural intentions. This study establishes this link through a stated choice experiment, and an integrated latent variable and choice model. The proposed approach simultaneously explains individuals' perceptions of privacy and general caution through observed individual characteristics and explores how these perceptions, in the form of latent constructs, may be associated with consumers' decisions to engage with an online transaction. The stated choice experiment is designed to collect consumers' choices across online retailers, a conventional store and an opt-out option in which online retailers are presented with varying levels of personal-information requirements. The data was collected come from over 500 respondents representing the online-user population in the UK. Model estimation results show that the higher an individual's privacy concern, general caution and technical protection, the less likely a consumer is to purchase a product online. In a joint model, the privacy concern variable is found to outweigh the effect of general caution and technical protection. Finally, consumers with higher levels of general caution are more sensitive towards an online retailer that shares their personal data with third parties.

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 enable e-businesses to uncover purchase patterns, which help them tailor products, identify consumer segments and better predict consumer behaviour. E-shoppers also

* Corresponding author.
E-mail address: potogloud@cardiff.ac.uk (D. Potoglou).
[1] Prediction based on data from in the UK, Germany, France, Sweden, The Netherlands, Italy, Poland and Spain.

http://dx.doi.org/10.1016/j.jocm.2015.12.002
1755-5345/

benefit from lower prices and personalised consumer experiences such as purchase recommendations. However, access to such services requires consumers to disclose personal data thus raising threats around what and how information is collected, how it is stored and managed. Potential risks of information disclosure may for example, be in the form of adverse price discrimination or invasive advertising (Goldfarb and Tucker, 2011). As a result, privacy concerns have remained high and are reported to be one of the main factors inhibiting e-commerce (TNS Opinion and Social, 2011).

There have so far been two streams of disjoint research efforts on consumers' privacy concerns. In the first stream, privacy concern is the dependent variable, which is determined by a number of antecedents. These antecedents may include socio-economic characteristics, for example, age (Goldfarb and Tucker, 2011) and gender (Bartel Sheehan, 1999); culture/climate (Smith et al., 1996); personality differences such as social awareness (Dinev and Hart, 2006); past distressing experiences related to disclosing personal information or privacy awareness (Smith et al., 1996). Privacy concerns are indirectly captured using psychometric scales; for example, such scales include the concern for information privacy (CFIP, Smith et al., 1996) and Internet users' information privacy concerns (IUIPC, Malhotra et al., 2004).

Pavlou (2011), in a review of Smith et al. (2011) and Bélanger and Crossler (2011), notes that there is a 'general consensus among researchers that information privacy concerns correspond to an individual's willingness to render personal information, transaction activity and government regulation'. Indeed, a second stream of research considers privacy concern as the determinant of several online-related behavioural reactions or 'stated intentions'–i.e., privacy concern is regarded as the independent variable and behavioural reactions are hypothesised to be guided by the so-called Privacy Calculus (Culnan and Bies, 2003). The concept of Privacy Calculus reflects the assumption that individuals' behavioural reactions (or stated intentions) occur following a calculus (i.e., a trade-off) between risks/costs and benefits as a consequence of that reaction (Dinev and Hart, 2006). For example, several studies in the subject area of e-commerce explore the role of privacy concerns on consumers' intention to engage in e-commerce transactions (Dinev and Hart, 2006), the selection of online retailers (e.g. Tsai et al., 2010) or their willingness to disclose information to retailers under certain privacy conditions or scenarios (e.g. Hui et al., 2007; Preibusch et al., 2013).

As shown in Fig. 1, the above two streams of research can be consolidated into the 'Antecedents→Privacy Concerns→Outcomes' (APCO) model in which 'measurement of privacy concerns is the central construct' (see, also Smith et al. (2011)). Privacy Concern as the dependent variable is captured on the left-hand side (i.e., Antecedents→Privacy Concerns) and Privacy Concern as independent variable and its potential association with behavioural intentions is represented on the right-hand side, respectively. Yet, with the exception of Dinev and Hart (2006) and Heirman et al. (2013) there have been no other studies that link consumers' willingness to disclose personal information in e-commerce with latent constructs (e.g. privacy concern, trust, perceived risk) and how these constructs are associated with their antecedents (e.g. beliefs, socio-economic characteristics of consumers).

Dinev and Hart (2006) elicit consumers' behavioural intentions using statements measuring 'willingness to transact/disclose information' on a 5-point Likert scale. The statements describe circumstances under which the respondent would be, on a scale from 'not at all' to 'very much': (a) willing "to purchase goods or services from websites that require consumers to submit accurate and identifiable information (i.e., credit card information)" and (b) "conduct sales transactions in e-commerce sites that require them to provide credit card information (e.g. using sites for purchasing goods or software)". Consumers' privacy concerns and its antecedents (e.g. Internet trust) are also captured through a set of 5-point Likert scales. The analysis to uncover potential associations across antecedents, privacy concerns and behavioural intentions (i.e., the decision to transact online) is then conducted using a structural equation model (SEM). Heirman et al. (2013) also estimate an SEM model to explore the association between adolescents' willingness to disclose information to commercial websites and trust.

Following Dinev and Hart's (2006) approach entails a number of shortcomings. The use of psychometric scales to capture the decision to transact leads into a simplified way of presenting a hypothetical scenario. This hypothetical scenario leaves little room for representation of the market in which there are multitude online retailers with varying requirements of personal information, ways of storing and sharing consumer information or adopting different strategies to attract consumers. Moreover, eliciting behavioural intentions through psychometric scales means that the Privacy Calculus model in Fig. 1 can only be empirically applied for a limited–usually fixed–set of risks and benefits involved in the transaction. It is then not possible to examine consumers' level of sensitivity for changes against retailers' privacy policies; that is examine how consumers trade-off across different online retailers offering varying levels of privacy-related terms and conditions in order to purchase a good or a service online.

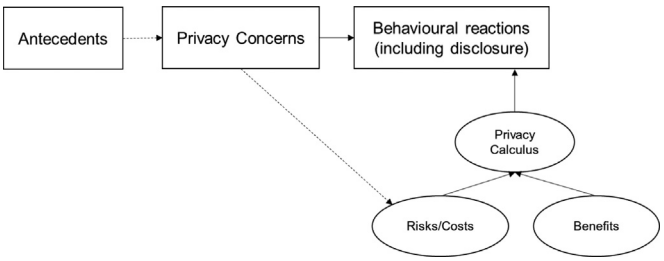

Fig. 1. A simplified version of the APCO Macro Model (adapted from Smith et al. (2011)).

The objective of this paper is therefore to demonstrate how an analytical framework comprising a stated choice experiment and analysis of the data using an Integrated Choice and Latent Variable (ICLV) model can be applied to appropriately capture and explain consumers' behavioural intentions in relation to latent constructs and their antecedents as described within the APCO model. A stated choice experiment embodies well the Privacy Calculus model as it allows exploration of how consumers would respond to varying levels of risks/costs and benefits. The ICLV model has been suggested by researchers as an appropriate way of capturing the effects of latent constructs onto consumers' choices by avoiding issues of inconsistent or inefficient model estimates (Ashok et al., 2002; Daly et al., 2012). For example, the ICLV allows for an unbiased exploration of hypotheses that examine how latent constructs (e.g. privacy concerns) may be related to dimensions of personal information (e.g. sharing of personal information) in an online transaction. It then follows that it is possible to conduct segmentation analysis and quantify in monetary terms whether, for example, different segments in sample place a different value on different dimensions of their personal information.

Currently, there still exists a gap in terms of an analytical framework that establishes a simultaneous link and estimates of the effects of factors within the APCO model, namely offers a link across antecedents, privacy concerns and stated intentions as well as addresses the abovementioned shortcomings. Therefore, this paper attempts to address this gap in the literature both in substantive and methodological terms. The empirical application focuses on consumers' decision to engage in e-commerce and thus the paper contributes to a growing body of literature that has used different experimental survey and laboratory approaches to study e-shopper behaviour and their willingness to engage and disclose their personal information in online transactions (see for example, Beresford et al., 2012; Hann et al., 2007; Hui et al., 2007; Jentzsch et al., 2012; Tsai et al., 2010).

The remainder of this paper is organised as follows. Section 2 provides details on the sample, the experimental design and the psychometric scales used in the survey. Section 3 describes the analytical approach based on ICLV. Section 4 presents the results and finally, the paper concludes with the discussion of results and practical implications of the findings and the proposed framework.

2. Data and methods

2.1. Data

The data comes 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 (mean = 43). The sample includes 248 male (49.6%) and 269 female (52.0%) respondents. More than half of the participants

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			

in the sample worked full or part time, 7% of individuals were students, and the rest of participants were unemployed or retired. The majority of participants reported an individual income below £50,000. Finally, most respondents were frequent internet users; 50% of respondents in the sample spent more than 14 hours per week on the Internet and the majority of them were regular e-shoppers. Sixty-six percent (66%) of participants searched for products and 59% reported that they buy products online at least once a month, respectively.

2.2. Choice of online retailers involving varying levels of information privacy

Respondents' choice for online retailers and subsequently, their decision to engage in online transactions is captured through a stated preference discrete choice experiment, a core component of the online survey. The advantage of the SPDCE approach in this case is that it can capture the Privacy Calculus concept (Dinev and Hart, 2006) as it can incorporate both potential risks to privacy and benefits (see Fig. 1).

Consumers, for example, may receive benefits in the form of discounts or additional services for the same product price. Monetary incentives have previously been found to positively affect individuals' willingness to disclose information (Hui et al., 2007 cited in, Bélanger and Crossler, 2011). Also, the additional-services shown in Table 2 reflect potential benefits offered by online retailers in exchange of the personal information provided by respondents. These attributes are combined such that participants weigh up between restricting the conditions under which retailers can collect and handle their information, and access to benefits such as discount or additional services (e.g. priority shipping). 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 receive discounts off the market price of the product.

With regard to risks, Smith et al. (1996) identify four dimensions of individuals' concerns about organisational information privacy practices: collection of personal information, handling errors, unauthorized secondary use and improper access. Therefore, we include an attribute reflecting the amount of information required in order to complete the transaction. As shown in Table 2, there are four possible versions based on the information collected ranging from a single identifier (email) to a complete range of purchase activity including purchase history, browsing, navigation history, email and additional personal details.

Online retailers also vary in terms of the amount of restrictions they place on the information collected from individuals, namely: (1) sharing of customers' data with other entities, and (2) the duration of storage of such information. In the most restricted privacy-protective case, retailers are legally not allowed to transfer the information collected to any other organisation. Also, they may be obliged to remove customers' information following a specific time period (i.e., 1, 2 or 5 years). These two restrictions describe the perceived degree of risk involved in the online purchase. Retaining personal information for longer of time or sharing it with third parties increases the risk of adverse price discrimination (Acquisti and Varian;

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

Attribute	Levels
Premium/discount against security costs	(1) Discount £4.00 (2) Discount £2.00 (3) No charge (4) Premium £2.00 (5) Premium £4.00
Additional information saved and linked to your account	(1) Only email [Reference Level] (2) Purchase history and email (3) Purchase history, browsing and navigation history and email (4) Purchase history, browsing, navigation history, email and additional personal details
Permission of sharing personal information with third parties	(1) Yes (2) No [Reference Level]
Time your personal information is stored for	(1) 1 year [Reference Level] (2) 2 years (3) 5 years (4) Without an explicit temporal limit
Availability of product or service at a conventional store/outlet (Only available in the Conventional store/outlet alternative)	(1) This item can also be easily purchased in your neighbourhood at a conventional retailer (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.) (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] (2) Faster checkout (one-click order) (3) Detailed reviews of products/seller (4) Priority shipping of product

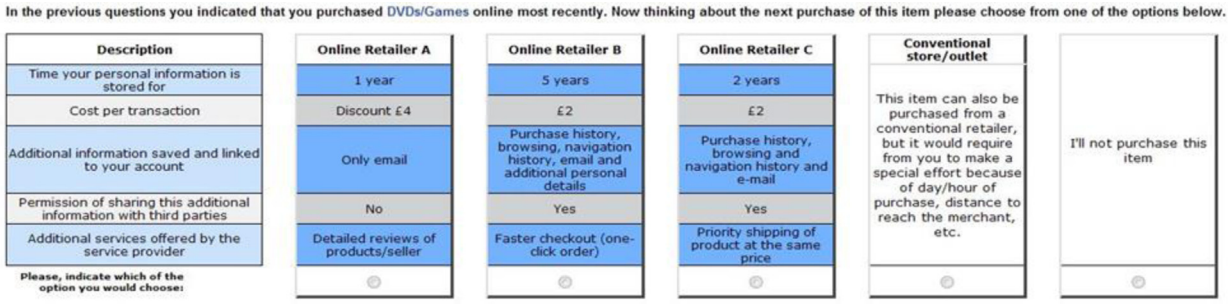

Fig. 2. An example choice exercise.

2005; Taylor, 2004), or invasive advertising (Tucker, 2012).

In the survey, respondents were firstly asked to recall an online purchase they had recently completed[2]. This strategy of linking the experiment to a real-life experience helps respondents to frame the choice scenario (Hess and Rose, 2009). As shown in Fig. 2, respondents were then offered the choice to buy the product from three different online retailers, a conventional store and an opt-out option.

In the context of e-commerce, geographic proximity to conventional stores is a determining aspect in order to complete e-commerce transactions (Dinev and 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 were combined to form the Online Retailer and Conventional Store alternatives using a D-efficient experimental design based on the Multinomial Logit (MNL) model with zero priors (Rose and Bliemer, 2009). The final design contained 60 rows (scenarios) which were divided into 12 blocks so that each respondent was presented with a set of five choice tasks. Each orthogonal block was randomly assigned to respondents to ensure that any heterogeneity retrieved in the parameter estimates was not affected by the attribute levels presented to respondents (Arentze et al., 2003). In this study, we used responses from 502 individuals, which resulted into a total of 2510 observations. The experimental design matrix was generated using the software Ngene (ChoiceMetrics, 2010).

2.3. *Privacy and information protection psychometric scales*

As part of the survey, respondents were also asked to indicate their strength of preference to a series of statements relating to attitudes towards privacy and information protection. As shown in Table 3, these statements correspond to three sets of attitudinal scales previously validated by other studies (Buchanan et al., 2007; Kumaraguru and Cranor, 2005).

Table 4 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 significant indicator loads onto the corresponding expected factor. A significant indicator was identified using as a threshold value of a loading greater than 0.4 (see, bold figures in Table 4). Thus, the results from the explanatory factor analysis provide evidence on the existence of three latent variables capturing respondents': (a) General Caution (Cronbach's alpha=0.877), (b) Technical Protection (Cronbach's alpha=0.804) and (c) Privacy Concern (Cronbach's alpha=0.611). The explanatory factor analysis used is a principal-axis factor analysis with promax rotation (Tabachnick and Fidell, 2007).

The first two factors reflect two latent variables, which measure individuals' attitudes towards privacy protection. The corresponding scales were derived and validated by Buchanan et al. (2007). The latent variable General Caution reflects respondents' actions taken to protect their personal information. The latent variable Technical Protection comprises 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 asks respondents to indicate their strength of preference regarding privacy of personal information, security and liberty, which form an adapted version of the Privacy Concern Index (Kumaraguru and Cranor, 2005; Robinson et al., 2010).

3. Analytical approach: the integrated latent variable choice model

The analytical framework is based on the ICLV model (Bolduc et al., 2005; Walker and Ben-Akiva, 2002). This model structure involves the simultaneous estimation of a choice model and attitudinal-structural equation (SEM) model. The ICLV

[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."*

Table 3
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: Kamaguru and Cranor (2005)	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%

Table 4
Rotated factor loadings against psychometric scales.

Psychometric indicator	Factor 1	Factor 2	Factor 3
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? [Indicator 1]	**0.625**		0.118
Do you read a website's privacy policy before you register your information? [Indicator 2]	**0.883**		
Do you look for a privacy certification on a website before you register your information? [Indicator 3]	**0.784**		
Do you read license agreements fully before you agree to them? [Indicator 4]	**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? [Indicator 6]	0.103	**0.707**	−0.137
Do you use a pop up window blocker? [Indicator 7]	−0.213	**0.656**	0.107
Do you check your computer for spy ware? [Indicator 8]		**0.646**	0.107
Do you clear your browser history regularly? [Indicator 9]	0.103	**0.751**	−0.185
Do you block messages/emails from someone you do not want to hear from? [Indicator 9]		**0.612**	
Protecting the privacy of my personal information is… [Indicator 10]		−0.123	**0.680**
Taking action against important security risks (e.g. international terrorism, organised crime) is… [Indicator 11]			**0.589**
Defending current liberties and human rights is… [Indicator 12]			**0.602**

model assumes that both choice and responses to attitudinal indicators are influenced by latent constructs (e.g. privacy concerns); modelling together both outcomes offers 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. Then, the ICLV model can be 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 while they engage in an online transaction.

In this study, the choice model is employed to explain how respondents make choices between three online retailers, the conventional-store and the opt-out option given the different attributes describing each option and respondents' latent (unobserved) attitudinal constructs. In the specification of the utilities of the choice model, therefore, this research incorporates 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, section below for discussion on the Latent Variable Model).

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

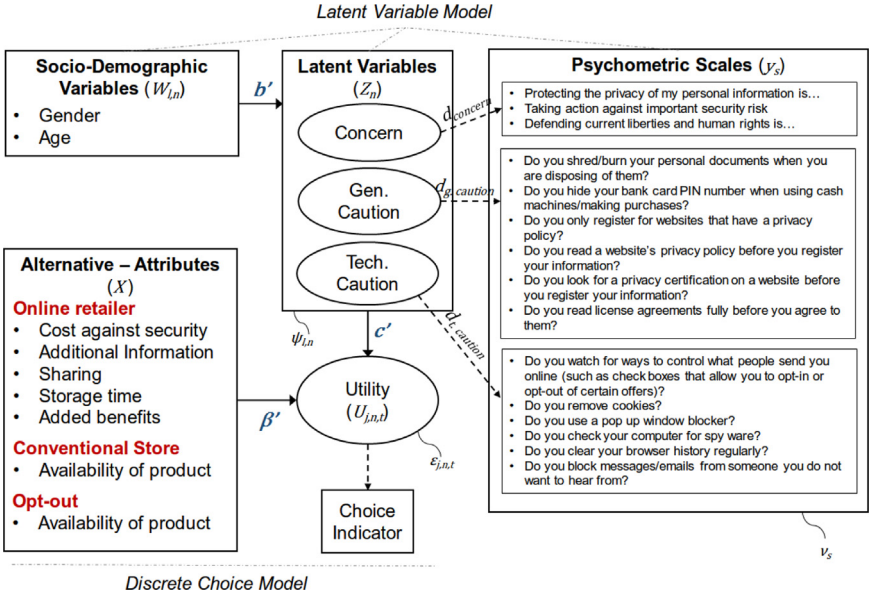

Fig. 3. The Integrated Choice and Latent Variable Modelling approach (adapted from: Walker and Ben-Akiva (2002); Bolduc et al. (2005).

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 Fig. 3).

On the right-hand side of the graph in Fig. 3, there is a set of psychometric indicators capturing respondents' attitudes. Participants' responses to these indicators are explained using the latent variables General Caution, Technical Protection and Concern, which are in turn a function of socio-demographic characteristics of individuals (see, top-left hand side of Fig. 3). 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.

3.1. 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:

$$U_{j,n,t} = \beta_1 \text{Cost}_{j,n,t} + \beta_2' \text{Add. Information}_{j,n,t} + \beta_3 \text{Sharing}_{j,n,t} + \beta_4' \text{StorageTime}_{j,n,t} + \beta_5' \text{Add. Services}_{j,n,t} + \epsilon_{j,n,t} \quad \forall\ 1 \le j \le 3 \tag{1}$$

where Cost, Add. Information, Sharing, StoreageTime, Add. Services correspond to the attributes describing the online retailer j and βs are unknown coefficients to be estimated. Specifically, β_1 is the coefficient representing the effect of the cost attribute (premium or discount) in choosing an online retailer.

All privacy-related attributes (i.e., except the cost attribute) are dummy coded. For each attribute, we set a "reference level" (see, Table 2), which represents the 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 reference 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.

β_2'–β_5' are vectors of coefficients capturing the effect of the associated attribute level (against its reference 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 β_3 captures the effect of unrestricted access to respondents' personal information on the odds of choosing a retailer, relative to the reference level, in which the retailer guarantees that the information will not be shared with third parties.

The random error $\epsilon_{j,n,t}$ 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 (Ben-Akiva and Lerman, 1985).

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

$$U_{\text{CS,n,t}} = \text{ASC}_{\text{CS}} + \beta_6 \text{Spec. Effort}_{\text{CS,n,t}} + c' Z_{\text{n}} + \epsilon_{\text{CS,n,t}} \tag{2}$$

$$U_{\text{OO,n,t}} = \text{ASC}_{\text{OO}} + c' Z_{\text{n}} + \epsilon_{\text{OO,n,t}} \tag{3}$$

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, 2009, 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 found at a nearby store and takes the value of one when special effort is required by respondents to purchase it. The alternative becomes unavailable when the Spec.Effort attribute is presented with its third level, 'This item is not available to purchase from a conventional retailer in your neighbourhood' (see, Table 2).

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 $k = \{k_1,\ldots,k_T\}$ made by respondent n, conditional on Z_n is given by:

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

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

3.2. Latent Variable Model

As mentioned in the previous section, the analysis focuses on three latent variables: (a) General Caution, $z_{\text{Gen\_caution,n}}$, (b) Technical Protection, $z_{\text{Tech\_protection,n}}$ and (c) Privacy Concern, $z_{\text{Concern,n}}$. 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:

$$y_s = \delta_s + d_s Z + v_s \tag{5}$$

where y_s is the observed response to the sth 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:

$$Z_{l,n} = b_l W_{l,n} + \sigma_l \psi_{l,n}$$

$$For\ l = \text{General Concern, General Caution, Technical Protection} \tag{6}$$

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 σ_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:

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

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

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

where $y_{s,n}$ is the observed response to the sth 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 sth 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. (1997) 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 Fig. 3, 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 estimate an error-component multinomial logit (ECMNL) or mixed multinomial logit model without latent variables. The ECMNL takes into account potential violations of the Independence of Irrelevant Alternatives (IIA) property and the correlation of the five choices obtained from the same respondent in each experiment known as agent effect (Walker, 2001). The utility for each of the alternatives in the ECMNL model is specified as in Eqs. ((1)–(3)) but with an added error component, $\sigma_j \zeta_{jn}$, with mean zero, standard deviation σ_j and ζ_{jn} is standard normal random variable (Cantillo et al., 2007). The random variables ζ_{jn} and standard deviation σ_j of the online retailers are constrained to be equal as these alternatives are unlabelled and there should be no unobserved differences across them (Walker, 2001).

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 5–8 present a summary of model statistics and the estimated parameters of an ECMNL and two ICLV models, respectively. All models were estimated in Python-Biogeme (Bierlaire, 2003) and results were confirmed by re-estimation of the models in Ox (Doornik, 2001).

The ICLV estimates are split into their corresponding components, namely: the choice (Table 6), structural equation (Table 7) and the measurement-equation (Table 8) models. The Joint ICLV model is specified with the three latent variables Concern, General Caution and Technical Protection in the utilities of the Conventional Store and Opt-out alternatives. The Joint ICLV and Interaction model also examines how latent constructs (e.g. General Caution) may reflect respondents' sensitivity to potential risks from information disclosure through the specification of interactions between the attributes describing the online retailers and latent constructs.

Table 5
Models statistics.

	ECMNL	Joint ICLV	Joint ICLV and Interaction
Number of individuals	502	502	502
Number of observations	2510	2510	2510
Number of Halton draws	1000	1000	1000
Log-likelihood (DCM)	−3099.33	−3177.76	−2826.38
Number of parameters (DCM)	16	16	15
Number of parameters (ICLV)	–	82	54
Log-likelihood (ICLV)	–	−10764.7	−6748.968

4.1. Choice model component

The ECMNL model is only used as a starting point of our analysis and as benchmark for testing the stability of the estimated parameters as we add the latent variables in the ICLV model. With regard to estimation of the ICLV, initial model specifications separately tested the hypotheses that the higher an individual's Concern, General Caution and Technical Protection the more likely these consumers were to opt-out from an online transaction, respectively. All else being equal, the estimated coefficients of Concern, General Caution and Technical Protection each in the corresponding models, satisfied the above hypotheses (for details, see Palacios et al. (2015)). 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 strongly significant (t-ratio=4.43) whereas the effect of the General Caution

Table 6
Parameter estimates in the choice model.

	ECMNL		Joint ICLV		Joint ICLV & Interaction	
	Est.	t-ratio	Est.	t-ratio	Est.	t-ratio
Online retailer alternatives						
Cost per transaction (in £)	−0.130	−13.33	−0.130	−14.58	−0.129	−14.42
Additional information saved and linked to your account						
Only mail			Reference level			
Purchase history and email	−0.187	−2.73	−0.181	−2.59	−0.186	−2.65
Purchase history, browsing, navigation history and email	−0.482	−6.21	−0.478	−6.47	−0.475	−6.40
Purchase history, browsing, navigation history, additional personal details and email	−0.411	−5.05	−0.411	−5.54	−0.416	−5.59
Information is shared with third parties	−0.724	−10.38	−0.723	−13.79	−0.565	−7.75
Time your personal information is stored for						
1–2 years			Reference level			
5 years	−0.349	−4.98	−0.353	−5.36	−0.366	−5.53
Without an explicit temporal limit	−0.363	−5.13	−0.363	−5.53	−0.363	−5.52
Additional services free of charge						
None			Reference level			
Faster checkout (one-click order)	0.381	4.81	0.385	5.06	0.372	4.86
Detailed reviews of products/seller	0.442	5.70	0.435	5.70	0.427	5.56
Priority shipping of product	0.469	6.28	0.469	6.10	0.464	6.02
Conventional store alternative						
Alternative Spec. Constant-Conventional Store	−1.700	−6.21	−3.11	−5.92	−3.05	−5.81
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.)	−0.941	−4.11	−0.834	−5.09	−0.830	−5.05
Opt-out alternative						
Alternative Spec. Constant-Opt-Out	−2.90	−9.87	−3.70	−7.08	−3.64	−6.97
Influence of latent variables						
Concern × (Conventional Store; Opt-out Alternatives)	–		2.970	4.43	3.26	4.26
General Caution × (Conventional Store; Opt-out Alternatives)			0.095	0.99	–	
Technical Protection × (Conventional Store; Opt-out Alternatives)			−0.119	−1.31		
Information is shared with third parties × General Caution					−0.128	−4.00
Error components						
Standard deviation $\sigma_{1,2,3}$ (alternatives corresponding to online retailers 1, 2, 3)	−2.560	−10.20	–			
Standard deviation $\sigma_{Conv.Store.}$(alternatives corresponding to conventional retailer)	1.280	2.52				
Standard deviation $\sigma_{Opt\text{-}out}$ (alternatives corresponding to the "Opt-out' alternative)	2.490	7.30				

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

	ECMNL	Joint ICLV		Joint ICLV & Interaction	
		Est.	t-ratio		
Increased concern	Age	0.00852	2.27	0.00749	2.17
	Gender (female)	0.339	2.78	0.314	2.75
	Standard deviation	0.926	4.82	0.843	4.46
Increased general caution	Age	0.0268	4.43	0.0287	4.86
	Gender (female)	0.362	2.03	0.418	2.26
	Standard deviation	1.87	11.97	1.84	12.06
Increased technical protection	Age	0.00972	1.43	–	
	Gender (female)	−0.670	−3.04		
	Standard deviation	2.13	11.05		

and Technical Protection diminishes. This is an indication that the latent variable Concern outweighs the other two latent variables, General Caution and Technical Protection.

The Joint ICLV and Interaction model shows how latent constructs may interact with privacy-related attributes in the SPDCE experiment rather than decisions to purchase a product from a conventional store or opt-out. After extensive hypothesis testing, this study shows that only the valuation of sharing of personal data is associated with attitudes towards General Caution when using the Internet. In particular, if $Sharing_{j,n,t}$ is equal to 1, it means that the online retailer *shares information with third parties*; in the ECMNL and Joint ICLV models, the contribution of this attribute to the utility would then be given by $\beta_3{}^{*}Sharing_{j,n,t}$ as shown in Eq. (1) while in the specification of the Joint ICLV and Interaction model it will be given by $(\beta_3+\beta^{*}\ General\_Caution_n)$ which incorporates the increased General Caution latent variable for respondent n. We observe that consumers with increased general caution place a higher value on controlling the access to their personal data, and in turn, are significantly less likely to buy products from online retailers that share customer data with other parties.

The parameter estimates across the ECMNL, Joint ICLV and Joint ICLV & Interaction models in Table 6 reveal the effect of each attribute describing the online retailers and the Conventional Store alternatives. 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 the *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 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, this research considers 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 of 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 'without an explicit temporal limit' are valued (statistically) equally negative–i.e., there is no 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

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

	Joint ICLV		**Joint ICLV & Interaction**	
	Est.	t-ratio	Est.	t-ratio
Concern indicators (d parameters)				
Privacy (Indicator 10)	1	–	1	–
Security (Indicator 11)	0.607	3.39	0.620	3.17
Liberty (Indicator 12)	0.544	3.56	0.563	3.35
General caution indicators (d parameters)				
Register only if privacy policy (Indicator 1)	1	–	1	–
Read privacy policy (Indicator 2)	2.650	6.63	2.88	6.10
Privacy certification (Indicator 3)	1.450	9.64	1.41	9.75
License agreements (Indicator 4)	1.500	8.95	1.49	9.11
Technical protection indicators (d parameters)				
Cookies (Indicator 5)	1	–		
Pop-up blocker (Indicator 6)	0.718	7.54		
Spy ware (Indicator 7)	1.190	7.22		
Clear browsing history (Indicator 8)	1.240	8.58		
Block messages (Indicator 9)	0.747	7.88		
Constants (δs)				
Privacy	6.170	6.07	6.120	6.04
Security	4.960	8.45	4.960	8.45
Liberty	4.970	8.48	4.960	8.47
Register only if privacy policy	3.100	8.23	3.010	7.89
Read privacy policy	2.800	3.44	2.690	3.00
Privacy certification	1.840	4.08	1.680	3.77
License agreements	1.830	3.95	1.690	3.60
Cookies	4.050	9.04	–	
Pop-up blocker	4.220	11.03		
Spy ware	5.340	9.25		
Clear browsing history	5.030	8.58		
Block messages	3.770	10.61		
Thresholds ($\mu_{y,s,n}$)				
Concern				
Privacy Thresholds				
Threshold 1	0	–	0	–
Threshold 2	2.350	2.46	2.350	2.46
Threshold 3	2.926	2.99	2.916	2.98
Threshold 4	4.856	4.81	4.806	4.77
Security Thresholds				
Threshold 1	0	–	0	–
Threshold 2	1.720	3.27	1.720	3.27
Threshold 3	2.242	4.07	2.239	4.06
Threshold 4	3.902	6.71	3.889	6.69
Liberty Thresholds				
Threshold 1	0	–	0	–
Threshold 2	2.490	4.48	2.490	4.48
Threshold 3	2.838	5.02	2.836	5.02
Threshold 4	4.578	7.83	4.566	7.82
General Caution				
Register only if privacy policy				
Threshold 1	0	–	0	–
Threshold 2	1.650	6.99	1.660	6.97
Threshold 3	4.160	13.70	4.150	13.66
Threshold 4	6.030	17.81	6.020	17.79
Read privacy policy				
Threshold 1	0	–	0	–
Threshold 2	3.810	7.68	4.090	7.04
Threshold 3	8.270	11.22	8.850	10.19
Threshold 4	11.360	13.40	12.190	12.21
Privacy certification				
Threshold 1	0	–	0	–
Threshold 2	2.140	9.82	2.100	9.91
Threshold 3	4.420	15.04	4.320	15.21
Threshold 4	6.490	18.68	6.350	18.61
License agreements				
Threshold 1	0	–	0	–
Threshold 2	2.640	11.05	2.620	11.05
Threshold 3	5.160	16.32	5.130	16.37
Threshold 4	7.400	19.27	7.37	19.31
Technical Protection				

Table 8 (*continued*)

	Joint ICLV		Joint ICLV & Interaction	
	Est.	t-ratio	Est.	t-ratio
Cookies				
Threshold 1	0	–	–	
Threshold 2	1.780	8.56		
Threshold 3	4.340	15.19		
Threshold 4	6.460	18.76		
Pop-up blocker				
Threshold 1	0	–	–	
Threshold 2	1.090	5.05		
Threshold 3	2.860	10.22		
Threshold 4	4.440	14.41		
Spy ware				
Threshold 1	0	–	–	
Threshold 2	1.170	4.74		
Threshold 3	3.340	9.76		
Threshold 4	5.540	14.01		
Clear browsing history				
Threshold 1	0	–	–	
Threshold 2	2.000	7.46		
Threshold 3	4.340	12.61		
Threshold 4	6.210	15.87		
Block messages				
Threshold 1	0	–	–	
Threshold 2	1.350	7.07		
Threshold 3	3.070	12.61		
Threshold 4	4.310	16.01		

the likelihood of an online retailer being selected by respondents. These services are equally valued as there are no significant differences in the coefficients of 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 fairly consistent across the DCM parts of the ICLV models reported in Table 6.

4.2. Latent Variable Model Component

As mentioned in the beginning of this section, we explore the impact of the latent constructs Concern, General Caution and 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 a joint model and a joint model with interaction. 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 these models.

The parameter estimates of the latent variable are better understood in conjunction with the structural and measurement equations shown in Tables 7 and 8. Table 7 presents 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 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 and 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

Table 9
Willingness to pay estimates.

	ECMNL (£)	Joint ICLV & Interaction (£)		
Additional information saved and linked to your account				
Purchase history and email	−1.44	−1.44		
Purchase history, browsing, navigation history and email	−3.71	−3.68		
Purchase history, browsing, navigation history, additional personal details and email	−3.16	−3.22		
Information is shared with third parties	−5.57	N/A		
Time your personal information is stored for				
5 years	−2.68	−2.84		
Without an explicit temporal limit	−2.79	−2.81		
Additional services free of charge				
Faster checkout (one-click order)	2.93	2.88		
Detailed reviews of products/seller	3.40	3.31		
Priority shipping of product	3.61	3.60		
		10th Perc.	Median	90th perc.
Information is shared with third parties × general caution	N/A	−4.28	−6.43	−8.58

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 3.

4.3. Willingness to pay to protect personal information

Table 9 shows the results of respondents' marginal willingness to pay (WtP) computed from the estimates of the choice model (Table 6) as the ratio of each of the estimates to the cost per transaction. These values provide an estimation of (maximum) monetary valuations of the different aspects of personal information (amount, sharing and time of retention) and of potential offers when purchasing goods online relative to the reference level of each attribute (Table 6), respectively. Negative WtP values imply that attribute levels in the *additional information, sharing* and *time information is stored* are seen by consumers as being less favourable than their corresponding reference level. For example, in a choice task, a consumer would prefer to only disclose their email (reference level) and would be willing to pay £1.44 in order to prevent disclosure of their purchase history. On the other hand, positive WtP values imply that attribute levels in the *additional services free of charge* attribute are seen more favourably than 'no additional services', the reference level of the attribute.

For both the ECMNL and Joint ICLV & Interaction models, we compute point estimates for the WtP measures by dividing the corresponding coefficient of each attribute level with the cost per transaction coefficient. For the interaction between the latent variable General Caution and *sharing* we report the median, the 10th and 90th percentile points, which are computed by generating 1000 individual-specific draws for the latent variable General Caution for each respondent as discussed in Hess and Beharry-Borg (2012).

The ICLV model shows that there is high variation in the value respondents place on data sharing depending on how vigilant they are when using the Internet. That is consumers who exercise less caution when using the Internet, place a lower value on the potential sharing of the data (£4.28) whereas consumers who are highly cautious (90th percentile of the latent variable General Caution). Being able to uncover such finding is a direct result of incorporating the latent variable into the choice models through the ICLV.

Finally, it is worth noting that WtP point valuations are fairly stable across all attributes and between the ECMNL and the ICLV model. The highest valuation in the ECMNL model concerns sharing of personal information with third parties followed by the level of collected information. Free services are valued from £2.93 for faster check-out, £3.40 for detailed reviews of products/seller and the highest valuation is £3.61 for priority shipping.

5. Discussion and conclusions

This paper develops and tests a framework for establishing a link across antecedents, privacy concerns and related latent constructs and how these influence stated behavioural intentions in e-commerce. This is the first time (to the knowledge of the authors) that the above effects are simultaneously incorporated into an ICLV model capturing the choice of online retailers through a stated choice experiment.

The added value of employing a stated choice experiment is that behavioural intentions to engage in e-commerce are captured through trade-offs between varying levels of *risk* such as the conditions under which retailers collect and manage customer information and share it with third parties and different types of *benefits* including discounts and/or free services (e.g. priority shipping). This approach fits well with the Privacy Calculus research strand in the Information Management literature. Note that in previous studies, there has been a tendency to incorporate behavioural intentions to disclose

information against a single (fixed) scenario and capture responses on a Likert-scale. For example, Dinev and Hart (2006) ask respondents to *what extent they would be willing to use the Internet to purchase books from websites that require credit card information*. By employing a stated choice experiment of the type described in this paper, we comprehensively capture stated intentions involving choices across multiple dimensions - risks and benefits - and levels within them thus offering more granular insights on how choices for retailers are associated with varying aspects of the usage of personal information (storage, sharing) and on how these choices are related with privacy concerns and latent constructs.

Within the stated choice experiment, 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.). Overall, respondents are more likely to buy products from online retailers that require minimum amount of information–i.e., their email, store personal information for one year and do not share consumer information with third parties. This finding is in line with Hui et al. (2007) who report that consumers try to minimise the amount of information they disclose to online retailers. We also find that consumers are against paying premiums in order to introduce control of their personal data, which confirms previous work reporting that consumers are open to 'sell' their information but against the option to 'shield their information' (Grossklags and Acquisti, 2007).

The ICLV model is in line with the Antecedents, Privacy Concern Outcomes theoretical model summarised by Smith et al. (2011), as it simultaneously captures the effects: (a) between the latent variables of Privacy Concern and General Caution with antecedents (e.g. age, gender) where the psychometric scales are used as manifestations of the latent variables and (b) between the utility of each alternative and observable (e.g. alternative attributes) as well as latent variables (Dekker et al., 2014). Previous studies have only incorporated consumers' Privacy Concern and General Caution to explain willingness to disclose information in online transactions (Dinev and Hart, 2006) or willingness to accept for providing personal data (Motiwalla et al., 2014) but with no attempt to link latent constructs with their antecedents (Pavlou, 2011). The ICLV modelling framework makes use of responses to psychometric scales while recognising that actual attitudes are unobserved and the responses to psychometric scales are dependent variables associated with the latent variables (Hess and Beharry-Borg, 2012). Adopting a simpler modelling framework without latent constructs, we would only be able to uncover relationships between stated behavioural intentions and the psychometric scales as explanatory variables. However, the shortcomings of directly introducing psychometric scales in the utility of each alternative, namely inconsistency and inefficiency of parameter estimates, have been extensively discussed elsewhere (Daly et al., 2012; Dekker et al., 2014; Vij and Walker, 2015; Walker and Ben-Akiva, 2002).

Privacy Concern and General Caution are significantly associated with consumers' decision to engage in e-commerce. Increased Privacy Concern is positively associated with the consumers' decision to purchase a product from a conventional store or opt-out than buying the same product online. The results agree with Dinev and Hart (2006) who report that Privacy Concern is negatively associated with consumers' decision to disclose personal information online. The influence of privacy concerns remains statistically significant even after we control for the combination of factors directly related to personal information requirements and how information may be handled (storage, sharing) by online retailers.

Findings also demonstrate that consumers exhibit different levels of sensitivity to the potential sharing of their personal information with third parties. Specifically, all else being equal, cautious consumers place a higher monetary value in order to prevent a retailer from sharing their personal information with third parties. Given the differences in valuation between a selection of free-services and the valuation of restricting sharing of consumer data, the value of a free service is currently not enough to compensate the disutility generated by secondary use of customer information. This result together with next ones, as explained further below, has important consequences for action both in the case of online retailers and policy makers.

Finally, the analysis shows how Privacy Concerns, General Caution and Technical Protection are all correlated with their antecedents. This paper primarily focuses on demographic differences, namely age and gender, and shows that women and older consumers are more privacy concerned and vigilant when it comes to their personal information compared to men and younger consumers. These findings also agree with previous studies. For example, Sheehan (2002) reports that women are more concerned about the impact of personal information disclosure may have upon their privacy. Also, Culnan (1995) finds that young consumers are less likely to be concerned about their privacy. A previously uncovered effect from our study is that women are less likely to seek ways to enhance their technical protection when using the Internet.

5.1. *Strengths and limitations*

One of the strengths of this paper is that the empirical study was undertaken on consumer preferences from a nationally representative sample of the Online User population in the UK. Therefore, this study has directly addressed previous calls for research on personal information to focus on representative and particularly, non-student samples (Bélanger and Crossler, 2011, p. 1026). Another critique of previous studies is that they are primarily US-based (Bélanger and Crossler, 2011, p. 1019). Our study draws findings from a nationally representative sample of online users in the UK, ensuring the representativeness of internet users and thus results are 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 and purchasing goods or services online is of the most popular activities (ONS, 2014). Thus, this paper widens the existing empirical evidence-base departing from narrow socio-geographic-based samples generally found in the literature (Bélanger and Crossler, 2011) by gathering nationally representative data.

Given that technologies surrounding online transactions in e-commerce and consumers' attitudes are subject to change over time, one may argue that the empirical data collected in 2012 may be outdated. However, the dimensions related to personal information examined in the stated choice experiment do remain timely. During 2012 the e-commerce market was already mature in the UK with most Internet users having experienced online purchasing. Therefore, independently of technology progress from that date, the level of personal information requirements, potential sharing of personal information with third parties and the length of time that information may be retained by the retailer constituted core elements in the context of privacy in the online world as they are relevant in 2015. We also appreciate that consumer attitudes might also be subject to change. Our study inevitably inherits the limitations of using a cross-sectional dataset in which only a single attitude-related observation per respondent is available and that observation is measured at a single point in time when respondents complete the survey. However, this limitation is not prohibitive of comparisons across consumers based on their differences in latent constructs and antecedents e.g. comparing the sensitivity against data sharing between consumers who exercise general caution and those who do not when purchasing goods online (Borsboom et al., 2003 cited in; Chorus and Kroesen, 2014). Having said that, there is certainly scope in developing a longitudinal study to explore whether attitudes related to online intentions to engage in commerce are stable or dynamic variables.

Finally, we have included a set of psychometric scales in the survey in order to capture a latent variable reflecting individuals' general trust and not trust specifically related to the online retailers. This is a limitation of our study as we are unable to develop a latent construct out of the corresponding scales in order to capture 'trust to online retailers'. As a result, we are unable to test hypotheses between willingness to engage in e-commerce and trust to the retailer. Having said that, trust has only been previously found as having a mediating relationship between privacy concerns and willingness to transact (Dinev and Hart, 2006; Van Slyke et al., 2006).

5.2. Empirical and practical implications of the proposed framework

From an empirical perceptive, the ICLV model provides insights not only about the effect of latent constructs, but also about the way socio-demographic factors affect the decision to engage in online transactions in the context of e-commerce or whether certain aspects of personal information are valued higher by some segments in the sample. For example, the results of the ICLV model provide more insights with regard to the profile of users who are privacy concerned and cautious when using the Internet. While the structural equations of the ICLV model show that Privacy Concerned and General Cautious consumers are more likely to be women and older aged individuals, the responses to psychometric scales and the effects estimated in the measurement equations of the ICLV model show that Privacy Concerned consumers are more likely to respond that, it is important to: *protect their privacy*, *take action against security risks and defend current liberties and human rights*. On the other hand, cautious consumers are more likely to *register for websites that have a privacy policy*, *read the website's privacy policy* and *look for privacy certification on a website prior to registration* and *read license agreements fully before they agree to them*.

From a practical perspective, the application of the ICLV model helps in identifying segments of consumers with respect to their privacy concerns and their level of caution when using the Internet. Sample segmentation in this study looks beyond socio-economic characteristics and can provide useful information to e-retailers to design and create content that better reaches their corresponding market segment. For example, e-retailers having men and younger individuals as their market base may opt-in for shorter versions of their privacy-related content as this segment is less likely to be cautious and thus not seek for the website's privacy certification, terms and conditions, etc. On the other hand, online retailers with their market base being women and older consumers may present more privacy related information. Overall, the findings of the ICLV model point towards personalised interfaces and customer-centred delivery of content. Most importantly, it is evident that a 'one-size-fits-all' with regard to promoting online privacy content is impractical and e-retailers may consider multiple strategies in order to accommodate different market segments and levels of general caution.

It is also worth highlighting that e-retailers not only need to be transparent with regard to their privacy content as these alone will not encourage consumers to engage in online transactions. Especially for privacy concerned consumers the issue is not related with how information is presented but with the risk of how this information may be used. Privacy cautious consumers, on the other hand, would be willing to engage in an online purchase of goods only if their data is not shared with third parties. Therefore, e-retailers should be explicit about their personal information practices and allow customers to have more control over their personal information (Jai and King, 2016).

The ICLV model finally helps to uncover changes to WtP measures as a function of latent constructs in ways that would not have been possible to identify through a simpler choice model. Taking the broader definition of privacy and personal information as a commodity, this finding is quite useful for retailers or data controllers seeking for the 'optimal price' (value) of personal information in order to obtain their customers' consent to allow secondary use (sharing) of their personal information with third parties. Such finding sets the basis for companies to transparently request secondary use of customer data by offering them market-based returns (Spiekermann et al., 2015). These insights would have not been possible to obtain without estimating an ICLV model, which hold despite recent criticisms regarding the potential limitations in terms of statistical gains of the model (e.g. goodness of fit) (Vij and Walker, 2015).

Acknowledgements

The work reported in the paper was part of the research project "Analysis of the economic value of personal information" funded by the European Commission's Joint Research Centre, Institute for Prospective Technological Studies (Ref: 2011-J04-44-AMI). Earlier versions of this paper were presented at the 2015 International Choice Modelling Conference, Austin, Texas, US and the 2015 European Regional Conference of the International Telecommunications Society, San Lorenzo de El Escorial, Spain. The authors are grateful to Patil Sunil for his advice on the design of the experiment; Jonathan Cave, Andrea Collins and Neil Robinson for their useful comments and suggestions, and Stephane Hess, Andrew Daly and Bhanu Patruni for their suggestions in model estimations. Last but not least, we would like to thank the two anonymous reviewers for their constructive comments and help to improve the paper. Any errors are our own.

References

Acquisti, A., Varian, H.R., 2005. Conditioning prices on purchase history. Mark. Sci. 24 (3), 367–381.
Arentze, T., Borgers, A., Timmermans, H., DelMistro, R., 2003. Transport stated choice responses: effects of task complexity, presentation format and literacy. Transp. Res. E: Logist. Transp. Rev. 39 (3), 229–244.
Ashok, K., Dillon, W.R., Yuan, S., 2002. Extending discrete choice models to incorporate attitudinal and other latent variables. J. Mark. Res. 39 (1), 31–46.
Bartel Sheehan, K., 1999. An investigation of gender differences in on-line privacy concerns and resultant behaviors. J. Interact. Mark. 13 (4), 24–38.
Bélanger, F., Crossler, R.E., 2011. Privacy in the digital age: a review of information privacy research in information systems. MIS Q.: Manag. Inf. Syst. 35 (4), 1017–1041.
Ben-Akiva, M., Lerman, S.R., 1985. Discrete Choice Analysis: Theory and Application to Travel Demand. MIT Press, Cambridge.
Ben-Akiva, M., Walker, J., Bernardino, A.T., Gopinath, D.A., Morikawa, T., Polydoropoulou, A., 1997. Integration of Choice and Latent Variable Models. Massachusetts Institute of Technology, Cambridge, MA.
Beresford, A.R., Kübler, D., Preibusch, S., 2012. Unwillingness to pay for privacy: A field experiment. Econ. Lett. 117 (1), 25–27.
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.
Bolduc, D., Ben-Akiva, M., Walker, J., Michaud, A., 2005. Hybrid choice models with logit kernel: applicability to large scale models. In: Lee-Gosselin, M., Doherty, S. (Eds.), Integrated Land-Use and Transportation Models: Behavioural Foundations, Elsevier, Oxford, pp. 275–302.
Borsboom, D., Mellenbergh, G.J., Van Heerden, J., 2003. The theoretical status of latent variables. Psychol. Rev. 110 (2), 203–219.
Buchanan, T., Paine, C., Joinson, A.N., Reips, U.-D., 2007. Development of measures of online privacy concern and protection for use on the internet. J. Am. Soc. Inf. Sci. Technol. 58 (2), 157–165.
Cantillo, V., Ortuzar, J.D., Williams, H.C.W.L., 2007. Modeling discrete choices in the presence of inertia and serial correlation. Transp. Sci. 41 (2), 195–205.
Centre for Retail Research, 2015. Online retailing. Britain, Europe, US and Canada 2015, ⟨http://www.retailresearch.org/onlineretailing.php⟩ (accessed: 02.02.2015).
ChoiceMetrics, 2010. Ngene 1.0.2 user manual and reference guide: the cutting edge in experimental design, ⟨www.choice-metrics.com⟩ (accessed: 15.09.2015).
Chorus, C.G., Kroesen, M., 2014. On the (im-)possibility of deriving transport policy implications from hybrid choice models. Transp. Policy 36, 217–222.
Culnan, M.J., 1995. Consumer awareness of name removal procedures: implications for direct marketing. J. Direct Mark. 9 (2), 10–19.
Culnan, M.J., Bies, R.J., 2003. Consumer privacy: balancing economic and justice considerations. J. Soc. Issues 59 (2), 323–342.
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.
Dekker, T., Hess, S., Arentze, T., Chorus, C., 2014. Incorporating needs-satisfaction in a discrete choice model of leisure activities. J. Transp. Geogr. 38, 66–74.
Dinev, T., Hart, P., 2006. An extended privacy calculus model for e-commerce transactions. Inf. Syst. Res. 17 (1), 61–80.
Doornik, J.A., 2001. Ox: An Object-Oriented Matrix Language. Timberlake Consultants Press, London.
Goldfarb, A., Tucker, C., 2011. Online display advertising: targeting and obtrusiveness. Mark. Sci. 30 (3), 389–404.
Grossklags, J., Acquisti, A., 2007. When 25 Cents is too much: an experiment on willingness-to-sell and willingness-to-protect oersonal information. In: Online Proceedings of the Workshop on the Economics of Information Security Series (WEIS 2007), ⟨http://weis2007.econinfosec.org/papers/66.pdf⟩ (accessed: 28.10.2014).
Hann, I.H., Hui, K.L., Lee, S.Y.T., Png, I.P.L., 2007. Overcoming online information privacy concerns: an information-processing theory approach. J. Manag. Inf. Syst. 24 (2), 13–42.
Heirman, W., Walrave, M., Ponnet, K., Van Gool, E., 2013. Predicting adolescents' willingness to disclose personal information to a commercial website: testing the applicability of a trust-based model. Cyberpsychology: J. Psychosoc. Res. Cyberspace 7 (3). Article 3.
Hess, S., Beharry-Borg, N., 2012. Accounting for latent attitudes in willingness-to-pay studies: the case of coastal water quality improvements in tobago. Environ. Resour. Econ. 52, 109–131.
Hess, S., Rose, J., 2009. Should reference alternatives in pivot design SC surveys be treated differently? Environ. Resour. Econ. 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 Q. 31 (1), 19–33.
Jentzsch, N., Preibusch, S., Harasser, A., 2012. Study on monetising privacy: an economic model for pricing personal information, ⟨http://www.enisa.europa.eu/activities/identity-and-trust/library/deliverables/monetising-privacy⟩ (accessed: 28.10.2014).
Kjaer, T., 2005. A review of the discrete choice experiment-with emphasis on its application in health care, (accessed: 01.03.2015).
Kumaraguru, P., Cranor, L.F., 2005. Privacy Indexes: A Survey of Westin's Studies, Pittsburgh: Institute for Software Research International. Carnegie Mellon University ⟨http://repository.cmu.edu/cgi/viewcontent.cgi?article=1857&context=isr⟩.
Malhotra, N.K., Kim, S.S., Agarwal, J., 2004. Internet users' information privacy concerns (IUIPC): the construct, the scale, and a causal model. Inf. Syst. Res. 15 (4), 336–355.
Motiwalla, L.F., Li, X., Liu, X., 2014. Privacy paradox: does stated privacy concerns translate into the Valuation of Personal Information?. In: Proceedings of the 18th Pacific Asia Conference on Information Systems (PACIS 2014), Chengdu, China.
OECD, 2012. OECD Internet economy outlook 2012, ⟨http://www.oecd.org/sti/ieconomy/internet-economy-outlook-2012-highlights.pdf⟩ [accessed: 20/09/2015].
ONS, 2011. Internet access-households and individuals, ⟨http://www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%3A77-226727⟩ [accessed: 19/02/2015].
ONS, 2014. Internet access-household and individuals, ⟨http://www.ons.gov.uk/ons/rel/rdit2/internet-access—households-and-individuals/2014/stb-ia-2014.html⟩ [accessed: 20/02/2015].
Palacios, J., Potoglou, D. and Feijóo, C., 2015. Does privacy concern influence e-shoppers choices for online retailers? An application of the Integrated Latent Variable Choice Model. In: International Choice Modelling Conference, 10–13 May, 2015, Austin, Texas, USA.
Pavlou, P.A., 2011. State of the information privacy literature: where are we now and where should we go? Mis Q. 35 (4), 977–988.

Preibusch, S., Kübler, D., Beresford, A., 2013. Price versus privacy: an experiment into the competitive advantage of collecting less personal information. Electron. Commer. Res. 13 (4), 423–455.
Robinson, N., Potoglou, D., Kim, C.W., Burge, P. and Warnes, R., 2010. Security at what cost? Quantifying people's trade-offs across liberty, privacy, and security, ⟨http://www.rand.org/pubs/technical_reports/TR664.html⟩ [accessed: 28/10/2014].
Rose, J., Bliemer, M., 2009. Constructing efficient stated choice experimental designs. Transp. Rev. 29 (5), 587–617.
Sheehan, K.B., 2002. Toward a typology of internet users and online privacy concerns. Inf. Soc. 18 (1), 21–32.
Smith, H.J., Dinev, T., Xu, H., 2011. Information privacy research: an interdisciplinary review. Mis Q. 35 (4), 989–1016.
Smith, H.J., Milberg, S.J., Burke, S.J., 1996. Information privacy: measuring individuals' Concerns about organizational practices. Mis Q. 20 (2), 167–196.
Spiekermann, S., Acquisti, A., Böhme, R., Hui, K.-L., 2015. The challenges of personal data markets and privacy. Electron. Mark. 25 (2), 161–167.
Tabachnick, B.G., Fidell, L.S., 2007. Using Multivariate Statistics, 5th ed. Pearson International, Boston.
Taylor, C.R., 2004. Consumer privacy and the market for customer information. RAND J. Econ. 35 (4), 631–650.
TNS Opinion & Social, 2011. Special eurobarometer 359: attitudes on data protection and electronic identity in the European Union, ⟨http://ec.europa.eu/public_opinion/archives/ebs/ebs_359_en.pdf⟩ [accessed: 20/09/2015].
Train, K., 2009. Discrete Choice with Simulation, 2nd Edition Cambridge University Press, Cambridge.
Tsai, J.Y., Egelman, S., Cranor, L., Acquisti, A., 2010. The effect of online privacy information on purchasing behavior: an experimental study. Inf. Syst. Res. 22 (2), 254–268.
Tucker, C.E., 2012. The economics of advertising and privacy. Int. J. Ind. Organ. 30 (3), 326–329.
Van Slyke, C., Shim, J.T., Jiang, J.J., 2006. Concern for information privacy, risk perception and online consumer purchasing. J. Assoc. Inf. Syst. 7 (6), 415–444.
Vij, A., Walker, J.L., 2015. Statistical properties of integrated choice and latent variable models, ⟨http://www.joanwalker.com/uploads/3/6/9/5/3695513/vij&walker_2015_iclv_stat_props.pdf⟩ [accessed: 04/09/2015].
Walker, J., 2001. Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures and Latent Variables (Ph.D thesis). MIT ⟨https://its.mit.edu/sites/default/files/documents/WalkerPhD.pdf⟩.
Walker, J., Ben-Akiva, M., 2002. Generalized random utility model. Math. Soc. Sci. 43 (3), 303–343.