Qc^*!æ^åÁ ã^åÁ[* ãÁð åÁæ^} óæãè|^Á [å^•
Á

ÁXã @æÁ ÈÖð @|^àð åææÁ æð æXð å^ài[^\ Áð åÁÁÄ~

Integrated mixed logit and latent variable models

Vishva M. Danthurebandara

Faculty of Business and Economics, Katholieke Universiteit Leuven, Naamsestraat 69, B-3000
Leuven, Belgium.

Email: Vishva.Danthurebandara@econ.kuleuven.be

Tel: +32 (0)16 32 69 63

Fax: +32 (0)16 32 66 24

Martina Vandebroek

Faculty of Business and Economics & Leuven Statistics Research Centre, Katholieke Universiteit
Leuven, Naamsestraat 69, B-3000 Leuven, Belgium.

Email: Martina.Vandebroek@econ.kuleuven.be

Tel: +32 (0)16 32 69 75

Fax: +32 (0)16 32 66 24

Jie Yu

Faculty of Business and Economics, Katholieke Universiteit Leuven, Naamsestraat 69, B-3000
Leuven, Belgium.

Email: Jie.Yu@econ.kuleuven.be

Tel: +32 (0)16 32 69 62

Fax: +32 (0)16 32 66 24

Integrated mixed logit and latent variable models *

Vishva M. Danthurebandara[†], Martina Vandebroek[‡] and Jie Yu[§]

Abstract

A traditional discrete choice model assumes that an individual's decision-making process is based on utility maximization and that the systematic part of the utility function depends on some observable attributes and covariates. These attributes and covariates however can only explain part of the utility and a large part remains unexplained. In recent years, researchers have recognized that psychological factors such as *attitudes*, *lifestyle* and *values* can affect an alternative's utility and hence the individual's choices. Therefore, extending the traditional discrete choice model by incorporating those latent or unobservable factors, can help to better understand the individual's decision making process and therefore to yield more reliable part-worth estimates and market share predictions.

Several integrated choice and latent variable (ICLV) models which merge the conditional logit model with a structural equation model exist in the literature. They assume homogeneity in the part-worths and use latent variables to model the heterogeneity among the respondents. The current research uses the mixed logit model that describes the heterogeneity in the part-worths and uses the latent variables to decrease the unexplained part of the heterogeneity. The empirical study that we conducted to assess student preferences on mobile phone features shows these ICLV models perform very well with respect to model fit and prediction.

Furthermore, we compare the different estimation procedures that exist in the literature. Results show that, as expected, the simultaneous procedure where we estimate the structural equation model (SEM) and the choice model simultaneously, gives better model fit and more accurate predictions compared to the sequential procedure that estimates the SEM first and then the choice model taking the estimated factor scores into account. But the gain of the simultaneous procedure is relatively small. Therefore one can use the easier sequential procedure without losing much efficiency if needed.

Keywords: Discrete choice models; Structural equation models; Hierarchical Bayesian estimation; Panel mixed logit model; Heterogeneity distribution

JEL Codes: C11, C25, C90

*This work was supported by the Katholieke Universiteit Leuven, Belgium.

[†]Katholieke Universiteit Leuven, E-mail: Vishva.Danthurebandara@econ.kuleuven.be

[‡]Katholieke Universiteit Leuven, E-mail: Martina.Vandebroek@econ.kuleuven.be

[§]Katholieke Universiteit Leuven, E-mail: Jie.Yu@econ.kuleuven.be

1 Introduction

In recent years, discrete choice models (DCM) have become increasingly popular for collecting and studying consumer preferences in many different disciplines such as marketing, transportation, and health economics. A traditional DCM assumes that an individual's decision-making process is based on utility maximization and that the utility can be modeled as a function of some observable attributes as well as the individuals' socio-demographics. For example, in their marketing application to explain the consumers choices of mobile phones, Head and Ziolkowski (2010) consider product attributes such as mobile phone functions (text messaging, web browsing, downloading ringtones) and features (calendar, alarm clock, personal notes) as well as the consumers' individual characteristics (income, age, gender). DCM link these observed inputs (attributes and covariates) to the observed outputs (choice data) by assuming that the individual's internal process during decision formation is implicitly captured well by the model (Ben-Akiva et al., 2002). However, in all disciplines, the attributes and covariates fail to capture the true inner process of a decision maker. Therefore, research has been devoted to using psychometric data that represent the true decision process of the respondent and, hence, to capture at least part of the previously unexplained utility.

Several researchers have illustrated how psychological factors such as *attitudes*, *lifestyle*, *perception* and *values* can affect an individual's decision making process (Ben-Akiva et al., 2002; Walker and Ben-Akiva, 2002; Timme et al., 2008). Olson and Zanna (1993) explain how these latent variables influence information processing and show the importance of taking these factors into account in order to improve the utility model. Ben-Akiva et al. (2002) discuss the importance of these factors in choice experiments and present a solid literature review. In their empirical application in transportation, Temme et al. (2008) extend the classical travel mode choice model to incorporate individuals' *attitudes* and *values* and show that the choice model that incorporates these latent factors clearly outperforms the traditional conditional logit model with respect to parameter estimation. La Paix et al. (2011) give an application in the travel domain, where they study the impact of the *propensity to travel* on the choice of different tour structures for three primary activities (home, work/study, shopping). Moreover, Ashok et al. (2002) present two empirical applications, one in marketing where they study the influence of *satisfaction* and *barriers* on the selection of cable TV providers and one in health care where they study the influence of the *satisfaction* with cost and *coverage* on the choice of a health care plan.

To model the impact of latent variables on choice, integrated choice and latent variable (ICLV) models have been developed which merge classical choice models with structural equation models (SEM). These ICLV models can be found in different model parametrizations, both with respect to the choice model and to the structural equation model. In all of these empirical applications, the ICLV model was based on the conditional logit model and the market heterogeneity was modelled by incorporating individual-specific latent factors. Nowadays, the mixed logit model is often used to model the market heterogeneity (for example, Bliemer et al., 2009; Yu et al., 2011; Danthurebandara et al., 2011). Though it models

individual specific part-worths, the mixed logit model can still be improved by including individual-specific latent factors that model the decision making process. In this paper we will compare ICLV models based on the conditional logit model and on the mixed logit model with models that do not include latent factors.

Furthermore, the current paper considers two different ways of incorporating the latent variables in the choice model. In the latent covariate model (Ben-Akiva et al., 2002; Temme et al., 2008), latent factors are included into the systematic part of the random utility model as extra individual covariates, while in the random scale latent variable model (Hess and Stathopoulos, 2011) the latent factors are assumed to influence the scale parameter. The first parametrization is a homoscedastic model since the scale factor is assumed constant. This model assumes that respondents have the same error variance or precision and uses individual latent factors to increase the explained part of the utility. The second parametrization yields a heteroscedastic model since the scale parameter now is a function of individual-specific latent factors. In this model, latent variables are used to explain differences in precision across respondents.

For the ICLV models, two prominent estimation methods can be found in the literature: sequential and simultaneous. The easier sequential method is a two stage approach where the SEM is estimated first and then the choice model incorporates the estimated factor scores (Ashok et al., 2002). However, Ben-Akiva et al. (2002) show that the sequential approach results in inconsistent and inefficient estimates. Alternatively, the integrated model can be estimated simultaneously resulting in efficient estimates but this approach is computationally much more demanding, especially for the ICLV models based on the mixed logit model and, as the number of latent variable increases, this approach gives convergence and identification problems. We will compare these different estimation methods and study the gain of using the full information simultaneous approach over the limited information sequential approach. We show how WinBUGS can be used to estimate the different ICLV models both sequentially and simultaneously.

To wrap up, this paper provides methodological and empirical contributions. We extend the classical ICLV models to the mixed logit model which models individual specific part-worths. We also introduce hierarchical Bayesian estimation and illustrate how this can be implemented in WinBUGS. Finally, in the empirical application we compare the different models and estimation procedures in a study on mobile phone features.

The following section introduces the generalized ICLV model and the different estimation approaches. Section 3 presents the results of an empirical study we conducted to compare the different models and estimation procedures, and Section 4 summarizes the key findings of the paper.

2 Methodology

The ICLV model consists of two major components: a structural equation model and a choice model. In this section we first introduce these components in detail and then present the different parametrizations of the combined ICLV model.

2.1 The structural equation model (SEM)

The standard SEM, also called LISREL model (Joreskog and Sorbom, 1996), consist of two sub-models. The first one is the measurement model which is a confirmatory factor model that relates the endogeneous and exogeneous latent variables to their corresponding indicators or manifest variables while taking the measurement error into account. The second is a structural model that represents the interrelationships among the latent endogeneous variables and also relates the latent endogeneous variables to the latent exogeneous variables and the observed exogeneous variables if present.

Consider a SEM with q endogenous latent variables $\boldsymbol{\eta}_{(q \times 1)}$, r exogenous latent variables $\boldsymbol{\xi}_{(r \times 1)}$ and c individual covariates $\mathbf{c}_{(c \times 1)}$. The endogenous latent variables are measured by the i manifest variables in $\mathbf{I}_{(i \times 1)}$ and the exogenous latent variables are measured by the j manifest variables in $\mathbf{J}_{(j \times 1)}$. Then the measurement model is defined by

$$\mathbf{I} = \boldsymbol{\Lambda}_I \boldsymbol{\eta} + \boldsymbol{\nu}, \quad (1)$$

$$\mathbf{J} = \boldsymbol{\Lambda}_J \boldsymbol{\xi} + \boldsymbol{\delta}, \quad (2)$$

where $\boldsymbol{\Lambda}_{I(i \times q)}$ and $\boldsymbol{\Lambda}_{J(j \times r)}$ are matrices of factor loadings and $\boldsymbol{\nu}_{(i \times 1)}$ and $\boldsymbol{\delta}_{(j \times 1)}$ are vectors of errors. The structural model can be formulated as

$$\boldsymbol{\eta} = \boldsymbol{\Pi} \boldsymbol{\eta} + \boldsymbol{\Gamma}_1 \boldsymbol{\xi} + \boldsymbol{\Gamma}_2 \mathbf{c} + \boldsymbol{\zeta}, \quad (3)$$

where $\boldsymbol{\Pi}_{(q \times q)}$, $\boldsymbol{\Gamma}_1_{(q \times r)}$ and $\boldsymbol{\Gamma}_2_{(q \times c)}$ contain regression coefficients and $\boldsymbol{\zeta}_{(q \times 1)}$ is a vector of disturbances. The vectors $\boldsymbol{\nu}$, $\boldsymbol{\delta}$ and $\boldsymbol{\zeta}$ are assumed to be multivariate normal distributed with zero mean and covariance matrices $\boldsymbol{\Sigma}_\nu$, $\boldsymbol{\Sigma}_\delta$ and $\boldsymbol{\Sigma}_\zeta$. The disturbances $\boldsymbol{\nu}$, $\boldsymbol{\delta}$ and $\boldsymbol{\zeta}$ are assumed uncorrelated. The vectors of endogenous and exogenous latent variables $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ are assumed to be normally distributed and therefore the latent variable indicators $\mathbf{I}$ and $\mathbf{J}$ also follow normal distribution. Then their densities are given by $\phi(\boldsymbol{\eta}|\boldsymbol{\mu}_\eta, \boldsymbol{\Sigma}_\zeta)$, $\phi(\boldsymbol{\xi}|\boldsymbol{\mu}_\xi, \boldsymbol{\Sigma}_\xi)$, $\phi(\mathbf{I}|\boldsymbol{\mu}_I, \boldsymbol{\Sigma}_\nu)$ and $\phi(\mathbf{J}|\boldsymbol{\mu}_J, \boldsymbol{\Sigma}_\delta)$ respectively.

For example, consider the SEM presented in Figure 1 with two endogenous latent variables (η_1 and η_2) and five corresponding indicators ($I_1, \dots, I_5$), one exogenous latent variable (ξ_1) with three indicators ($J_1, \dots, J_3$) and four exogenous observed variables or covariates ($c_1, \dots, c_4$). This system can be described by the expressions (1) - (3) with $q = 2$, $r = 1$, $c = 4$, $i = 5$ and $j = 3$.

Figure 1: A structural equation model

For standard SEM, the implied covariance matrix of the observed covariates and manifest variables, i.e. $\mathbf{I}$, $\mathbf{J}$ and $\mathbf{c}$, contains all the unknown parameters in the model (Bollen, 1989; Lee, 2008). The usual maximum likelihood (ML) and generalized least squares (GLS) methods estimate the parameters such that the distance between the observed and the implied covariance matrix is small. On the other hand, Bayesian methods have been developed and used to analyze SEM (for example, Lee and Song, 2004; Ansari and Jedidi, 2000; Ansari et al., 2000). The Bayesian procedures are convenient to use in many situations as they avoid complex optimization problems and they also have some desirable estimation properties such as consistency and efficiency that can be attained under relaxed conditions compared to the classical estimation methods (Train, 2003). Moreover, Bayesian methods allow to utilize prior information that researchers may have in practical situations. If a researcher does not have any informative prior knowledge, Bayesian methods can still be used with non-informative priors and produce results close to the maximum likelihood estimates. We will estimate the structural equation model as well as the ICLV models using such a Bayesian framework.

Bayesian analysis is based on the posterior distribution of the model parameters. For a given SEM, let $\boldsymbol{\theta}$ be the vector of unknown model parameters combining the elements in $\boldsymbol{\Pi}$, $\boldsymbol{\Gamma}_1$, $\boldsymbol{\Gamma}_2$, $\boldsymbol{\Lambda}_I$, $\boldsymbol{\Lambda}_J$, $\boldsymbol{\Sigma}_v$, $\boldsymbol{\Sigma}_\delta$, $\boldsymbol{\Sigma}_\zeta$ and let $\mathbf{Z}$ be the observed data set, that is $\mathbf{Z} = (\mathbf{I}, \mathbf{J})$. Then the posterior distribution $p(\boldsymbol{\theta}|\mathbf{Z}, \mathbf{c})$ is proportional to the product of the likelihood $p(\mathbf{Z}|\boldsymbol{\theta}, \mathbf{c})$ and the prior distribution $p(\boldsymbol{\theta}|\mathbf{c})$ (Lee, 2008):

$$p(\boldsymbol{\theta}|\mathbf{Z}, \mathbf{c}) \propto p(\mathbf{Z}|\boldsymbol{\theta}, \mathbf{c})p(\boldsymbol{\theta}|\mathbf{c}) \quad (4)$$

For most SEM the posterior $p(\boldsymbol{\theta}|\mathbf{Z}, \mathbf{c})$ is complicated. The major difference between SEM and regression

and simultaneous equation models is the presence of latent variables. If the factor scores in a SEM were not random but observed data, then the SEM would simplify to a simultaneous equation model and the posterior distribution simplifies. The key idea in the Bayesian estimation procedure for SEM is data augmentation, initially proposed by Tanner and Wong (1987), which is an iterative procedure which starts with some initial values for the latent variables. Tanner and Wong (1987) give a complete description about calculating posterior distributions by data augmentation. By using this method in SEM, we treat the latent variables as hypothetical missing data and augment the observed data with imputed values for the latent variables. Then the posterior distribution based on the complete data set is relatively easy to deal with. Specifically, instead of working with $p(\boldsymbol{\theta}|\mathbf{Z}, \mathbf{c})$, we work with the posterior distribution $p(\boldsymbol{\theta}, \boldsymbol{\Omega}|\mathbf{Z}, \mathbf{c})$ where $\boldsymbol{\Omega}$ contains the latent variables, $\boldsymbol{\Omega} = (\boldsymbol{\eta}, \boldsymbol{\xi})$. Lee (2008) provides a detailed theoretical explanation of this Bayesian method for SEM. The posterior distribution is then given as:

$$p(\boldsymbol{\theta}, \boldsymbol{\Omega}|\mathbf{Z}, \mathbf{c}) \propto p(\mathbf{Z}|\boldsymbol{\theta}, \mathbf{c})p(\boldsymbol{\Omega}|\boldsymbol{\theta}, \mathbf{c})p(\boldsymbol{\theta}|\mathbf{c}). \quad (5)$$

In order to approximate the mean and other useful statistics of this posterior distribution, we use a Gibbs sampler (Geman and Geman, 1984) where we use full conditional Monte-Carlo draws. The Gibbs sampler at the $(j+1)^{th}$ iteration is:

- (i) generate $\boldsymbol{\Omega}^{j+1}$ from $p(\boldsymbol{\Omega} | \boldsymbol{\theta}^j, \mathbf{Z}, \mathbf{c})$
- (ii) generate $\boldsymbol{\theta}^{j+1}$ from $p(\boldsymbol{\theta} | \boldsymbol{\Omega}^{j+1}, \mathbf{Z}, \mathbf{c})$

More details on the conditional distributions $p(\boldsymbol{\Omega}|\boldsymbol{\theta}, \mathbf{Z}, \mathbf{c})$ and $p(\boldsymbol{\theta}|\boldsymbol{\Omega}, \mathbf{Z}, \mathbf{c})$, based on the model and the distributional properties of the model parameters and the observed and unobserved data, can be found in Lee (2008). The conditional distributions used in this Gibbs sampler are non-standard in most of the cases. Therefore we implement these models in WinBUGS (Lunn et al., 2009) such that we do not have to specify these conditional distributions explicitly. WinBUGS uses the Gibbs sampling method to sample from a given joint posterior distribution and it uses the Metropolis-Hastings sampler (Metropolis et al., 1953; Hastings, 1970) when the full conditional distribution is not available in closed form. For more details about Bayesian methods for SEM refer for example to Lee (2008).

2.2 The mixed logit model

The mixed logit model is nowadays often used to model the market heterogeneity in conjoint choice studies. This section introduces the mixed logit model without considering the latent factors, and in the next section we extend it to take the factors into account. In case all respondents are assumed to have the same part-worths, the mixed logit model simplifies to the conditional logit model. For more details about the conditional logit model, refer for example to Sándor and Wedel (2001) and Yu et al.

(2008).

Consider the random utility model:

$$u_{ksn} = V_{ksn}(\mathbf{x}) + \varepsilon_{ksn} = \mathbf{x}'_{ksn} \boldsymbol{\beta}_n + \varepsilon_{ksn} \quad (6)$$

where u_{ksn} is the utility that respondent n obtains by choosing alternative k from choice set s , $V_{ksn}(\mathbf{x})$ is called the systematic part of the utility which is assumed to be a linear function of the attribute values and corresponding utility coefficients or part-worths and $\mathbf{x}_{ksn}$ is the p -dimensional vector containing the attribute values of alternative k in choice set s for respondent n . The preferences of respondent n are represented by the individual-specific utility coefficient vector $\boldsymbol{\beta}_n$. The probability that respondent n chooses alternative k ($k = 1, \dots, K$) in choice set s ($s = 1, \dots, S$) is

$$p_{ksn}(\boldsymbol{\beta}_n) = \frac{\exp(\mu \mathbf{x}'_{ksn} \boldsymbol{\beta}_n)}{\sum_{i=1}^K \exp(\mu \mathbf{x}'_{isn} \boldsymbol{\beta}_n)}, \quad (7)$$

where μ is the scale factor. The scale factor is derived as $\pi/\sqrt{6}\sigma$, where σ is the standard deviation of ε . To avoid identification problems we usually set $\mu = 1$.

Denote by $\mathbf{y}_n$ the $K \times S$ -dimensional individual-specific vector containing the S choices from respondent n that correspond to the S choice sets of K alternatives. Conditional on $\boldsymbol{\beta}_n$, the likelihood function for a given $\mathbf{y}_n$ can be written as

$$L(\mathbf{y}_n | \mathbf{X}, \boldsymbol{\beta}_n) = \prod_{s=1}^S \prod_{k=1}^K (p_{ksn}(\boldsymbol{\beta}_n))^{y_{ksn}}, \quad (8)$$

where y_{ksn} , element of $\mathbf{y}_n$, is 1 if respondent n chooses alternative k in choice set s and 0 otherwise and $\mathbf{X}$ is a matrix containing the attribute values of each alternative in the S choice sets.

We assume that the coefficient vector $\boldsymbol{\beta}_n$ is randomly drawn from a p -variate normal distribution with mean $\boldsymbol{\mu}_\beta$ and covariance matrix $\boldsymbol{\Sigma}_\beta$, that is $N(\boldsymbol{\beta}_n | \boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta)$. Then the likelihood of a given sequence of choices $\mathbf{y}_n$, unconditional on $\boldsymbol{\beta}_n$, for respondent n is

$$\begin{aligned} L(\mathbf{y}_n | \mathbf{X}, \boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta) &= \int L(\mathbf{y}_n | \mathbf{X}, \boldsymbol{\beta}_n) \phi(\boldsymbol{\beta}_n | \boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta) d\boldsymbol{\beta}_n, \\ &= \int \left(\prod_{s=1}^S \prod_{k=1}^K (p_{ksn}(\boldsymbol{\beta}_n))^{y_{ksn}} \right) \phi(\boldsymbol{\beta}_n | \boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta) d\boldsymbol{\beta}_n. \end{aligned} \quad (9)$$

The probabilities of a single respondent in multiple choice situations will be correlated and the above formulation takes this dependency into account. That is, the coefficient vector $\boldsymbol{\beta}_n$ for a given respondent n appears in all choice sets and ensures that the model captures the correlation across repeated choices.

The likelihood of the choices of all N respondents, unconditional on $\boldsymbol{\beta}_n$, is denoted by $L(\mathbf{y}_{full} | \mathbf{X}, \boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta)$

and defined as

$$L(\mathbf{y}_{full}|\mathbf{X}, \boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta) = \prod_{n=1}^N L(\mathbf{y}_n|\mathbf{X}, \boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta) \quad (10)$$

where $\mathbf{y}_{full}$ contains the responses of all N respondents.

The choice model can also be considered as a SEM. In the random utility model, the utility is a latent variable and the observed choices are indicators or manifest variables of this utility. The measurement model is here given by a non-linear relationship between the observed choices and the latent utility, i.e. for a given choice set s and a respondent n ,

$$y_{isn} = \begin{cases} 1, & \text{if } u_{isn} = \max_j(u_{jsn}), \quad j = 1, \dots, K \\ 0, & \text{otherwise} \end{cases}$$

This indicates that respondent n chooses alternative i from choice set s containing K alternatives if and only if the alternative i yields maximum utility. The random utility model in equation (6) can then be seen as a structural model that represents the relationship between observed attribute levels and the latent utility.

The mixed logit model is frequently estimated through a Hierarchical Bayes approach which also avoids the difficult optimization of the likelihood function (Toubia et al., 2004; Arora and Huber, 2001; Rossi et al., 1996). Here also we use the Gibbs sampler to draw from the posterior density. The Gibbs sampler at the $(j+1)^{th}$ iteration is:

- (i) generate $\boldsymbol{\mu}_\beta^{j+1}$ from $p(\boldsymbol{\mu}_\beta | \boldsymbol{\beta}_n^j, \boldsymbol{\Sigma}_\beta^j)$
- (ii) generate $\boldsymbol{\Sigma}_\beta^{j+1}$ from $p(\boldsymbol{\Sigma}_\beta | \boldsymbol{\beta}_n^j, \boldsymbol{\mu}_\beta^{j+1})$
- (iii) generate $\boldsymbol{\beta}_n^{j+1}$ from $p(\boldsymbol{\beta}_n | \boldsymbol{\mu}_\beta^{j+1}, \boldsymbol{\Sigma}_\beta^{j+1})$

Train (2003) contains the derivation of conditional distributions based on non-informative priors. We use these non-informative priors for choice model parameters in our WinBUGS programs. A detailed discussion on HB estimation for mixed logit models can be found in Train (2003).

2.3 Integrated choice and latent variable (ICLV) models

Now that we have introduced the two major components of the ICLV model, we can present the general form of the ICLV model and look at two parametrizations: the latent covariate model and the random scale latent variable model. To express that the factors are individual specific we introduce the index n in the measurement and structural equations of the SEM. Then the measurement and the structural equations become:

$$\boldsymbol{\eta}_n = \boldsymbol{\Pi}\boldsymbol{\eta}_n + \boldsymbol{\Gamma}_1\boldsymbol{\xi}_n + \boldsymbol{\Gamma}_2\mathbf{c}_n + \boldsymbol{\zeta}_n,$$

Figure 2: Path diagram of the latent covariate model

$$\mathbf{I}_n = \mathbf{\Lambda}_I \boldsymbol{\eta}_n + \boldsymbol{\nu}_n,$$

$$\mathbf{J}_n = \mathbf{\Lambda}_J \boldsymbol{\xi}_n + \boldsymbol{\delta}_n,$$

The key in defining an integrated model is the inclusion of latent factors into the choice model. In the latent covariate model, the latent variables are included into the systematic part of the utility as extra covariates. In the random scale latent variable model it is hypothesized that the scale parameter varies across respondents and the latent factors are used to model the scale parameter. Figure 2 and Figure 3 clearly show how these latent constructs influence the utility in these two parametrization.

In the next two sections we give the corresponding models.

2.3.1 The latent covariate model

In this parametrization, the factors are considered as latent covariates and are included into the systematic part of the utility. The random utility model is then defined as

$$u_{ksn} = \mathbf{x}'_{ksn} \boldsymbol{\beta}_n + \boldsymbol{\eta}_n' \boldsymbol{\alpha}_k + \varepsilon_{ksn},$$

where $\boldsymbol{\alpha}_k$ is the k^{th} column of a $q \times K$ matrix $\boldsymbol{\alpha}$ of alternative specific factor effects. Each column expresses the effect of the q latent factors on the corresponding alternative. In this parametrization only alternative-specific factor effects make sense because otherwise, those effects drop out when calculating the choice probabilities. These factor effects are assumed to be fixed and therefore not individual

Figure 3: Path diagram of the random scale latent variable model

specific in the sequel. The probability that respondent n chooses alternative k ($k = 1, \dots, K$) from choice set s ($s = 1, \dots, S$) is

$$p_{ksn}^{lc}(\beta_n, \alpha_k) = \frac{\exp(\mathbf{x}'_{ksn}\beta_n + \eta_n'\alpha_k)}{\sum_{i=1}^K \exp(\mathbf{x}'_{isn}\beta_n + \eta_n'\alpha_i)}, \quad (11)$$

The individual likelihood of the latent covariate model is the integral of the choice model over the distribution of the latent constructs. By assuming independent error components (ε, ζ) , the likelihood of the latent covariate model is given by

$$L_{lc}(\mathbf{y}_n | \mathbf{X}; \mu_\beta, \Sigma_\beta, \Pi, \Gamma, \Sigma_\zeta) = \int L(\mathbf{y}_n | \mathbf{X}, \eta_n; \mu_\beta, \Sigma_\beta) \phi(\eta_n | \mathbf{X}; \Pi, \Gamma, \Sigma_\zeta) d\eta_n \quad (12)$$

where

$$L(\mathbf{y}_n | \mathbf{X}, \eta_n; \mu_\beta, \Sigma_\beta) = \int \left(\prod_{s=1}^S \prod_{k=1}^K (p_{ksn}^{lc}(\beta_n, \alpha_k))^{y_{ksn}} \right) \phi(\beta_n | \mu_\beta, \Sigma_\beta) d\beta_n. \quad (13)$$

Considering the measurement equations that take the indicators into account, and by assuming independent error components $(\varepsilon, \zeta, \delta, \nu)$, the joint likelihood for individual n of the observed variables $\mathbf{y}_n$, $\mathbf{I}_n$ and $\mathbf{J}_n$ is

$$L_{lc}(\mathbf{y}_n, \mathbf{I}_n, \mathbf{J}_n | \mathbf{X}; \mu_\beta, \Sigma_\beta, \Pi, \Gamma_1, \Gamma_2, \Lambda_I, \Lambda_J, \Sigma_\zeta, \Sigma_\nu, \Sigma_\delta) = \int L(\mathbf{y}_n | \mathbf{X}, \eta_n; \mu_\beta, \Sigma_\beta) \phi(\mathbf{I}_n | \eta_n, \Lambda_I, \Sigma_\nu) \phi(\mathbf{J}_n | \eta_n, \Lambda_J, \Sigma_\delta) \phi(\eta_n | \mathbf{X}; \Pi, \Gamma, \Sigma_\zeta) d\eta_n, \quad (14)$$

where $L(\mathbf{y}_n | \mathbf{X}, \eta_n; \mu_\beta, \Sigma_\beta)$ is defined in (13). Obviously, obtaining maximum likelihood estimates in

this case is difficult. We use this likelihood together with the prior distributions of the model parameters to obtain Bayesian estimates. The estimation method and the software implementation of the latent covariate model will be presented in section 2.4.

2.3.2 The random scale latent variable model

Hess and Stathopoulos (2011) use latent factors to explain the variation across respondents in the scale parameter of the choice model. For example, in their empirical research, they consider the scale factor as a function of the respondents' *survey engagement*. The model hypothesizes that differences across respondents in survey engagement yield differences in precisions and therefore scale heterogeneity.

In this case, the probability that respondent n chooses alternative k ($k = 1, \dots, K$) in choice set s ($s = 1, \dots, S$) can be formulated as

$$p_{ksn}^{rslv}(\beta_n, \alpha) = \frac{\exp(\mu_n \mathbf{x}'_{ksn} \beta_n)}{\sum_{i=1}^K \exp(\mu_n \mathbf{x}'_{isn} \beta_n)}, \quad (15)$$

where $\mu_n = e^{\eta_n' \alpha}$ and α is a $q \times 1$ vector of fixed factor effects.

Similar to the latent covariate model, the individual joint likelihood of the observed variables $\mathbf{y}_n$, $\mathbf{I}_n$ and $\mathbf{J}_n$ is

$$L_{rslv}(\mathbf{y}_n, \mathbf{I}_n, \mathbf{J}_n | \mathbf{X}; \mu_\beta, \Sigma_\beta, \Pi, \Gamma_1, \Gamma_2, \Lambda_I, \Lambda_J, \Sigma_\zeta, \Sigma_\nu, \Sigma_\delta) = \int L(\mathbf{y}_n | \mathbf{X}, \eta_n; \mu_\beta, \Sigma_\beta) \phi(\mathbf{I}_n | \eta_n, \Lambda_I, \Sigma_\nu) \phi(\mathbf{J}_n | \xi, \Lambda_J, \Sigma_\delta) \phi(\eta_n | \mathbf{X}; \Pi, \Gamma, \Sigma_\zeta) d\eta_n, \quad (16)$$

where

$$L(\mathbf{y}_n | \mathbf{X}, \eta_n; \mu_\beta, \Sigma_\beta) = \int \left(\prod_{s=1}^S \prod_{k=1}^K (p_{ksn}^{rslv}(\beta_n, \alpha))^{y_{ksn}} \right) \phi(\beta_n | \mu_\beta, \Sigma_\beta) d\beta_n. \quad (17)$$

The Bayesian estimation method where we use this individual joint likelihood together with prior distributions of model parameters and the software implementation of the random scale latent variable model will be presented in the next section.

2.4 Estimation

The sequential approach is the simplest way to estimate the ICLV models. In this case we estimate the structural equation model first and then estimate the choice model incorporating the estimated factor scores. This method is known to be easy and straightforward because researchers can use readily

available software for SEM (for example LISREL, AMOS, EQS, SAS) and for the choice model (for example SAS, NLOGIT, STATA). The sequential method considers the estimated factor scores as non-stochastic variables and includes them into the utility model as exogenous variables. This brings measurement error into the model and therefore results in inconsistent and inefficient estimates (Ben-Akiva et al., 2002). The alternative simultaneous approach provides consistent and efficient estimates but, as the number of latent factors increases, this approach becomes numerically cumbersome. To perform simultaneous estimation, one can program the joint likelihood in a flexible software package such as GAUSS (Ashok et al., 2002), Mplus (Temme et al., 2008), Ox (Hess and Stathopoulos, 2011), BIOGOME (La Paix et al., 2011) and also in SAS, R or Matlab.

In this research, the Hierarchical Bayes method is used to estimate the latent covariate and the random scale latent variable models. For each model parametrization, the posterior distribution is proportional to the product of the corresponding likelihood given by the expression (14) or (16) and the prior distributions of the corresponding model parameters. Obviously, for the integrated model, this posterior is much more complex than the posteriors for SEM and for the choice model. A large number of non-standard conditional distributions have to be calculated during the estimation process. Therefore, we implement these models in WinBUGS which is freely available software for Bayesian analysis (Spiegelhalter et al., 2003; Lunn et al., 2009). The key advantage of using WinBUGS here is that we do not have to specify the conditional distributions explicitly. WinBUGS uses Markov chain Monte Carlo (MCMC) methods such as the Gibbs sampler (Geman and Geman, 1984) and the Metropolis Hastings algorithm (Metropolis et al., 1953; Hastings, 1970), which help to avoid the complex integration problems in ML methods. For each model, we perform 100,000 iterations in total, 70,000 which are used for burn-in and the remaining 30,000 draws, which are thinned by selecting every 10th draw to remove the auto-correlation, are used to determine the parameter estimates.

3 Student preference on mobile phone features

In this section we present the choice experiment we conducted to compare the different integrated models introduced previously. In the next section, the choice experiment is explained in detail: the study domain, the data collection and the models estimated, and section 3.2 contains the results.

3.1 Data and models

Mobile phone use has grown rapidly throughout the world. Srivastava (2005) gives a rich overview of how the mobile phone has moved beyond being a simple technical device to becoming a key social object. In Belgium, the number of mobile phone subscribers has increased by 68% from 2000 to 2005. This rapid quantitative growth of consumers leads to a significant qualitative improvement of mobile

phones. A large number of mobile phone features and applications are launched everyday to capture the users' attention. Our conjoint experiment is designed to assess the relative preferences for these mobile phone features. Obviously, teenagers and young adults constitute the largest consumer group in the mobile phone market (McClatchy, 2006) and hence is the major target group of mobile phone manufactures. Therefore, the experiment is conducted with undergraduate and graduate students at the Faculty of Business and Economics of the Katholieke Universiteit Leuven, Belgium. A near orthogonal design with five attributes and nine choice sets of size three is used. Five mobile phone features are used as attributes and the descriptions and levels of the attributes are displayed in Table 1.

Table 1: Attribute descriptions and levels

Attribute	Description	Level		
GPS (x_1)	whether there is an inbuilt GPS (Global Positioning System)	Yes	No	
WiFi (x_2)	whether there is a wireless internet access	Yes	No	
Camera (x_3)	resolution of the inbuilt camera	8 Mp	10 Mp	12 Mp
M & Mp3 (x_4)	whether there is an inbuilt media and Mp3 player	Yes	No	
Memory (x_5)	inbuilt memory size	4 GB	8 GB	16 GB

We consider two endogenous latent factors to assess how students use these mobile phone features to express themselves (*expressiveness*) and to assess how useful these features are in their day to day life (*usefulness*), and one exogenous latent factor to assess whether they enjoy these features (*enjoyment*). Use of these latent factors to assess the respondents' attitude within the context of mobile devices is motivated by several other researches, for example, Nysveen et al. (2005) and Head and Siolkowski (2010). The latent variables are observed indirectly by two or more manifest variables (indicators) as shown in Table 2. All manifest variables are measured on a five point Likert scale from *strongly disagree* to *strongly agree*.

In total, 309 respondents participated in this *paper and pencil* survey and each respondent had to evaluate nine choice sets of size three. Four covariates are recorded for each respondent: gender, whether the respondent has a student job, whether the respondent still lives at home or lives in a student room or an apartment in the university area (residence type) and the type of personal computer.

These survey data are used to estimate different ICLV models and to assess their performance in terms of model fit and prediction accuracy. Only eight choice sets are used to estimate the models and the last choice set is used to assess the prediction accuracy of each model.

Table 3 presents the different models we compare in this study. We estimate the latent covariate model and the random scale latent variable model based on the mixed logit model as well as the conditional logit model with both sequential and simultaneous estimation methods. In order to assess to what

Table 2: Latent variables and their indicators

Latent Variable	Manifest variable
Expressiveness (η_1) (<i>Head and Ziolkowski, 2010</i>)	I often talk with others about my mobile phone features ($\mathbf{I}_1$) Other people are often impressed by the way I use my mobile phone features ($\mathbf{I}_2$)
Usefulness (η_2) (<i>Lund, 2001</i>)	My mobile phone features help me to be more productive ($\mathbf{I}_3$) My mobile phone features give me more control over the activities in my life ($\mathbf{I}_4$) My mobile phone features meet my needs ($\mathbf{I}_5$)
Enjoyment (ξ_1) (<i>Lin et al., 2008</i>)	While using my mobile phone features, I am concentrated ($\mathbf{J}_1$) I'm happy to use my mobile phone features ($\mathbf{J}_2$) I'm satisfied with my mobile phone features ($\mathbf{J}_3$)

extent the latent factors improve the model fit and the prediction accuracy, we estimate a classical mixed logit model (*MixL*) as a benchmark. The last two models in Table 3 can also be considered benchmark models as they assess whether including the raw indicator variables directly into the model does a better job than including their corresponding latent factors. The choice probability for the LC_{Mix}^{ind} model is given by

$$p_{ksn}^{lc_{Mix}^{ind}}(\beta_n, \alpha_k) = \frac{\exp(\mathbf{x}'_{ksn}\beta_n + \mathbf{I}'_n\alpha_{1k} + \mathbf{J}'_n\alpha_{2k})}{\sum_{i=1}^K \exp(\mathbf{x}'_{isn}\beta_n + \mathbf{I}'_n\alpha_{1i} + \mathbf{J}'_n\alpha_{2i})},$$

where α_{1k} and α_{2k} are 5×1 and 3×1 matrices of alternative specific indicator effects.

And the choice probability for the $RSLV_{Mix}^{ind}$ model is given by

$$p_{ksn}^{rslv_{Mix}^{ind}}(\beta_n, \alpha) = \frac{\exp(\mu_n \mathbf{x}'_{ksn}\beta_n)}{\sum_{i=1}^K \exp(\mu_n \mathbf{x}'_{isn}\beta_n)},$$

where $\mu_n = e^{\mathbf{I}'_n\alpha_1 + \mathbf{J}'_n\alpha_2}$ and α_1 and α_2 are 5×1 and 3×1 vectors of indicator effects.

3.2 Results

3.2.1 Structural equation model

We first estimated the structural equation model shown in Figure 1. The individual covariates and the structural relationship between *expressiveness* and *usefulness* turned out to be insignificant. Therefore, the final structural equation model does not contain these parameters and Table 4 shows the parameter estimates and standard errors of the factor loadings and regression coefficients of the SEM estimated

Table 3: Different models and estimation procedures included in the comparison study

	Integrated model	Choice model	Estimation procedure
LC_{Mix}^1	Latent covariate model	Mixed logit model	Simultaneous (1 stage)
LC_{Mix}^2	Latent covariate model	Mixed logit model	Sequential (2 stages)
LC_{Cl}^1	Latent covariate model	Conditional logit model	Simultaneous (1 stage)
LC_{Cl}^2	Latent covariate model	Conditional logit model	Sequential (2 stages)
$MixL$	No factors	Mixed logit model	-
$RSLV_{Mix}^1$	Random scale latent variable model	Mixed logit model	Simultaneous (1 stage)
$RSLV_{Mix}^2$	Random scale latent variable model	Mixed logit model	Sequential (2 stages)
$RSLV_{Cl}^1$	Random scale latent variable model	Conditional logit model	Simultaneous (1 stage)
$RSLV_{Cl}^2$	Random scale latent variable model	Conditional logit model	Sequential (2 stages)
LC_{Mix}^{ind}	Indicators as covariates	Mixed logit model	-
$RSLV_{Mix}^{ind}$	Scale parameter is a function of indicators	Mixed logit model	-

with both sequential and simultaneous estimation method. The second and third columns give the parameter estimates and standard errors obtained by estimating only the SEM as the first step of a sequential method. The fourth and fifth columns show the parameter estimates for the SEM obtained by estimating the latent covariate model with the simultaneous method and the last two columns give parameter estimates obtained by estimating the random scale latent variable model with the simultaneous method. The parameters (λ_{I1} , λ_{I3} and λ_{J1}) are set equal to 1 in order to avoid identification problems. Results show that the three approaches presented in Table 4 give similar parameter esti-

Table 4: Estimates of factor loadings and regression coefficients

Parameter	Sequential (SEM)		Simultaneous (LC_{Mix}^1)		Simultaneous ($RSLV_{Mix}^1$)	
	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
γ_{11}	0.18	0.0199	0.29	0.0280	0.13	0.0114
γ_{12}	1.20	0.0314	1.16	0.0310	1.13	0.0298
λ_{I1}	1.00	-	1.00	-	1.00	-
λ_{I2}	2.12	0.3048	1.44	0.0238	2.30	0.4008
λ_{I3}	1.00	-	1.00	-	1.00	-
λ_{I4}	1.01	0.0239	1.26	0.0299	1.26	0.0350
λ_{I5}	1.30	0.0292	1.50	0.0374	1.55	0.0398
λ_{J1}	1.00	-	1.00	-	1.00	-
λ_{J2}	1.26	0.0308	1.21	0.0234	1.46	0.3000
λ_{J3}	1.50	0.0376	1.28	0.0300	1.59	0.3775

mates for the SEM. For all three approaches, the exogenous latent variable *enjoyment* has a positive impact on both *expressiveness* and *usefulness* (γ_{11} and γ_{12}), and the effect on *usefulness* is larger than on *expressiveness*. All indicator variables have a positive effect on their corresponding latent factor. The indicator variable I_2 shows a high impact on *expressiveness* compared to the indicator I_1 , which means that students use their mobile phone features more to impress their friends and colleagues than to talk about these features with others. All three indicator variables corresponding to *usefulness* show similar effects. Among its three indicators, J_3 shows the highest effect on *enjoyment*, which means that, on average, students are satisfied with their mobile phone features.

3.2.2 ICLV models

We compare the different integrated models based on model fit and predictions. The model fit is assessed by the deviance information criterion (DIC) (Spiegelhalter et al., 2002). To assess the prediction accuracy, we computed how often the choice of the individual respondents for the last choice set is predicted correctly by each model. Table 5 shows the DIC values and hit rates of each model in increasing order of DIC (note that the smaller DIC the better). These results show that the integrated

Table 5: Model fit and prediction accuracy of different ICLV models

Integrated model	Choice model	Estimation procedure	Model	DIC	HIT
LC model	Mixed logit model	Simultaneous	LC_{Mix}^1	2654.55	84%
RSLV model	Mixed logit model	Simultaneous	$RSLV_{Mix}^1$	2662.49	83%
LC model	Mixed logit model	Sequential	LC_{Mix}^2	2668.01	83%
RSLV model	Mixed logit model	Sequential	$RSLV_{Mix}^2$	2678.72	81%
Indicators as covariates	Mixed logit model	-	LC_{Mix}^{ind}	2737.25	78%
Scale parameter is a function of indicators	Mixed logit model	-	$RSLV_{Mix}^{ind}$	2748.28	76%
No factors	Mixed logit model	-	$MixL$	2764.68	75%
RSLV model	Conditional logit model	Simultaneous	$RSLV_{Cl}^1$	3890.18	66%
LC model	Conditional logit model	Simultaneous	LC_{Cl}^1	3892.87	60%
LC model	Conditional logit model	Sequential	LC_{Cl}^2	3905.14	68%
RSLV model	Conditional logit model	Sequential	$RSLV_{Cl}^2$	3968.37	60%

models based on the mixed logit model perform much better than the models based on conditional logit models. Among them, as expected, models estimated with the simultaneous procedure (LC_{Mix}^1

and $RSLV_{Mix}^1$) show better fit and better predictions than the models estimated with the sequential method. However, the gain of using the simultaneous method over the sequential method is rather small. Therefore, we can conclude that, one can use the easier sequential approach to estimate the integrated model without losing much efficiency. Although these two models (LC_{Mix}^1 and $RSLV_{Mix}^1$) have rather different interpretations they have very similar fit and predictions. We can also conclude that including all observed indicators into the model directly is worse than including the latent factors both with respect to the model fit and the prediction accuracy. Furthermore, the mixed logit model without latent variables can model the heterogeneity better than the integrated models based on the conditional logit model.

Next we present the parameter estimates of the two best models LC_{Mix}^1 and $RSLV_{Mix}^1$ in Table 5. The latent covariate model contains alternative specific factor effects. In each choice set of the design three alternatives are arranged according to the camera resolution: the first alternative with low resolution, the second with medium resolution and the third with high resolution. The utilities of the three mobile alternatives are given by:

$$u_{1sn} = \beta_n^{gps} x_{1sn}^{gps} + \beta_n^{wifi} x_{1sn}^{wifi} + \beta_n^{cam} x_{1sn}^{cam} + \beta_n^{mp3} x_{1sn}^{mp3} + \beta_n^{mem} x_{1sn}^{mem} + \alpha_1^{expres} \eta_n^{expres} + \alpha_2^{useful} \eta_n^{useful} + \varepsilon_{1sn},$$

$$u_{2sn} = \beta_n^{gps} x_{2sn}^{gps} + \beta_n^{wifi} x_{2sn}^{wifi} + \beta_n^{cam} x_{2sn}^{cam} + \beta_n^{mp3} x_{2sn}^{mp3} + \beta_n^{mem} x_{2sn}^{mem} + \alpha_3^{expres} \eta_n^{expres} + \alpha_4^{useful} \eta_n^{useful} + \varepsilon_{2sn},$$

$$u_{3sn} = \beta_n^{gps} x_{3sn}^{gps} + \beta_n^{wifi} x_{3sn}^{wifi} + \beta_n^{cam} x_{3sn}^{cam} + \beta_n^{mp3} x_{3sn}^{mp3} + \beta_n^{mem} x_{3sn}^{mem} + \varepsilon_{3sn}.$$

The factor effects are interpreted as differential effects: for example, α_1^{expres} is the differential effect of expressiveness on the utility of the first alternative (low camera resolution) compared to the alternative with high resolution.

The random utility for the random scale latent variable model is given by:

$$u_{ksn} = e^{(\alpha^{expres} \eta_n^{expres} + \alpha^{useful} \eta_n^{useful})} (\beta_n^{gps} x_{ksn}^{gps} + \beta_n^{wifi} x_{ksn}^{wifi} + \beta_n^{cam} x_{ksn}^{cam} + \beta_n^{mp3} x_{ksn}^{mp3} + \beta_n^{mem} x_{ksn}^{mem}) + \varepsilon_{ksn}.$$

For this model parametrization, the factor effects are interpreted as effects of latent variables on the scale parameter, hence on the precision.

Table 6 presents the estimated population means of the attributes and the estimated factor effects for the LC_{Mix}^1 and $RSLV_{Mix}^1$ models.

Both models indicate that *WiFi*, *M&Mp3* and *GPS* are the most preferred mobile phone features on average among students. The camera resolution is the least important attribute. Notice that, in the latent covariate model, the effects of *expressiveness* on the utilities have positive signs. Since the alternatives are arranged from low to high resolution, these parameters were expected to be negative. This would represent that a mobile with higher camera resolution is preferred over a mobile with lower camera resolution. Though the model gives unexpected results about the effects of the latent factors on the model, it has improved the prediction accuracy by 10% compared to the mixed logit

Table 6: Population mean and factor effects estimates

	LC_{Mix}^1		$RSLV_{Mix}^1$	
Parameter	Estimate	Std. Error	Estimate	Std. Error
μ_{GPS}	1.9283	0.1820	1.2053	0.3478
μ_{WiFi}	4.1280	0.2907	5.3640	0.9674
μ_{Camera}	0.4142	0.1249	0.2995	0.0849
μ_{Mp3}	2.1654	0.2557	2.3237	0.7709
μ_{Memory}	0.6635	0.0876	1.3940	0.1962
$\alpha_1^{express}$	0.2273	0.0197	-	-
α_2^{useful}	-0.3792	0.0219	-	-
$\alpha_3^{express}$	0.1618	0.0152	-	-
α_4^{useful}	-0.1295	0.0138	-	-
$\alpha^{express}$	-	-	1.6404	0.1010
α^{useful}	-	-	0.4333	0.0910

model estimated without latent factors. For the $RSLV_{Mix}^1$ model both *expressiveness* and *usefulness* has positive effect on the scale and the *expressiveness* has the highest effect. This indicates that an increase in the latent variable leads to an increase in scale, hence an increase in precision. This is expected, since if a respondent sees these mobile phone features as highly useful to express himself and to be more productive in his day to day life then the respondent will make more consistent choices and this will be reflected by a small error variance.

Figure 4 shows the heterogeneity distributions of the attribute coefficients estimated by LC_{Mix}^1 and $RSLV_{Mix}^1$ models. It can be concluded that the WiFi attribute which is on average most preferred, exhibits the highest heterogeneity. The camera resolution on the other hand is least preferred on average and exhibits only a small heterogeneity. We can therefore conclude that respondents have neglected the camera resolution attribute in their decision making process which may explain the unexpected sign of the factor effects.

Figure 4: Heterogeneity distributions of attribute coefficients estimated by (a) LC_{Mix}^1 and (b) $RSLV_{Mix}^1$

4 Discussion and Conclusions

The current research proposes an integrated mixed logit and latent variable model to model market heterogeneity based on choice experiments. Researchers often use a conditional logit model that assumes homogeneity in the part-worths and consider individual specific latent factors to model the market heterogeneity. We extend the classical mixed logit model which assumes market heterogeneity in the part-worths by incorporating latent factors and study whether this improves model fit and predictions. We also considered two parametrization of the ICLV model: the latent covariate model where we consider these latent quantities as covariates and include them into the systematic part of the utility and the random scale latent variable model where the latent factors affect the scale factor.

This paper compares different ICLV models based on a empirical marketing experiment. Results for this particular experiment show that the proposed integrated mixed logit and latent variable models outperform all other models we considered. In both model parametrization, the integrated mixed logit models estimated with the simultaneous procedure yield best model fit and the highest prediction accuracy. It was also shown that the classical mixed logit model can model the market heterogeneity better than the ICLV models based on the conditional logit model which uses the factors to capture the heterogeneity.

Regarding the parameter estimates, the best two models (LC_{Mix}^1 and $RSLV_{Mix}^1$) give similar estimates for the population means of the attribute coefficients and also give similar estimated heterogeneity distributions. Both models indicate that *WiFi*, *M&Mp3* and *GPS* are on average the most preferred mobile phone features for these respondents.

References

- Ansari, A. and K. Jedidi, 2000. Bayesian Factor Analysis for Multilevel Binary Observations. *Psychometrika*. 65 475-497.
- Ansari, A. and K. Jedidi, and S. Jagpal, 2000. A Hierarchical Bayesian Methodology for Treating heterogeneity in Structural Equation Models. *Marketing Science*. 19 328-347.
- Arora, N., J. Huber, 2001. Improving Parameter Estimates and Model Prediction by Aggregate Customization in Choice Experiments. *Journal of Consumer Research*. 28 273-283.
- Ashok, K., W.R. Dillon, and S. Yuan, 2002. Extending Discrete Choice Models to Incorporate Attitudinal and Other Latent Variables. *Journal of Marketing Research*. 39(1) 31-46.
- Ben-Akiva, M., J. Walker, A. Bernardino, D. Gopinath, T. Morikawa, and A. Polydoropoulou, 2002. Integration of Choice and Latent Variable Models, in (H. Mahmassani, Ed.) In Perpetual Motion: Travel Behaviour Research Opportunities and Application Challenges. *Elsevier Science*. 431-470.
- Bierlaire, M. and M. Ben-Akiva, 2010. Lecture notes in Discrete Choice Analysis. EPFL, Lausanne, Switzerland.
- Bliemer, M.C.J., J.M. Rose, 2010. Construction of Experimental Designs for Mixed Logit Models Allowing for Correlation Across Choice Observations. *Transportation Research Part B: Methodological*. 44 720-734.
- Bollen, K. A., 1989. Structural Equation Models with latent Variables. John Wiley & Sons, Inc.
- Danthurebandara, V. M., J. Yu, and M. Vandebroek, 2011. Sequential Choice Designs to Estimate the Heterogeneity Distribution of Willingness-to-pay. *Quantitative Marketing and Economics*. 9(4) 429-448.
- Geman, S. and D. Geman, 1984. Stochastic Relaxation, Gibbs Distribution and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 6 721-741.
- Hastings, W. K., 1970. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*. 57 97-109.
- Head, M., N. Ziolkowski, 2010. Understanding student attitudes of mobile phone applications and tools: A study using conjoint, cluster and SEM analyses. *Proceedings of the 18th European Conference on Information Systems (ECIS 2010), Pretoria, South Africa*.
- Hess, S., and A. Stathopoulos, 2010. Linking Response Quality to Survey Engagement: A Combined Random Scale and Latent Variable Approach. *Proceedings of the European Transport Conference (ETC 2010), Glasgow, Scotland, UK*.

- Joreskog, K. G., and D. Sorbom, 1996. LISREL 8: Structural Equation Modelling with the SIMPLIS Command Language. Hove and London: Scientific Software International.
- La Paix, L., M. Bierlaire, E. Cherchi, and A. Monzón, 2011. How Urban Environment Affects Travel Behaviour? Integrated Choice and Latent Variable Model for Travel Schedules. *Proceedings of the Second International Choice Modelling Conference (ICMC 2011), Leeds, North England, UK*.
- Lee, S. Y. and Song, X. Y., 2004. Evaluation of Bayesian and Maximum Likelihood Approaches in Analysing Structural Equation Models with Small Sample Sizes. *Multivariate Behavioural Research*. 39 653-686.
- Lee, S. Y., 2008. Structural Equation Modelling: A Bayesian Approach. John Wiley & Sons, Inc.
- Lin, A., S. Gregor and M. Ewing, 2008. Developing a scale to measure the enjoyment of web experiences. *Journal of Interactive Marketing*. 22.
- Lund, A., 2001. Measuring usability with the USE questionnaire. *STC Usability SIG Newsletter*. 8(2).
- Lunn, D., D. J. Spiegelhalter, A. Thomas and N. Best, 2009. The BUGS project: Evolution, critique and future directions. *Statistics in Medicine*. 28 3049 - 3067.
- McClatchey, S., 2006. The consumption of mobile services by Australian university students. *Australian College Students*. 1(1) 2-9.
- Mels, G., 2006. LISREL for Windows: Getting Started Guide. Lincolnwood, IL: Scientific Software International, Inc.
- Metropolis, N., 1953. Equations of State Calculation by Fast Computing Machine. *Journal of Chemical Physics*. 21 1087-1091.
- Nysveen, H., Per E. Pedersen and H. Thorbjørnsen, 2005. Intentions to use mobile services: antecedents and cross-service comparisons. *Journal of the Academy of Marketing Science*. 33(3) 330-356.
- Olson, J. M., and M. P. Zanna, 1993. Attitudes and attitude change. *Annual Review of Psychology*. 44 117-154.
- Rossi, P.E., R.E. McCulloch, and G.M. Allenby, 1996. The Value of Purchase History Data in Target Marketing. *Marketing Science*. 15 321-340.
- Rungie, C. M., L. V. Coote, and J. Louviere, 2011. Structural Choice Modeling: Theory and Applications to Combining Choice experiments. *Proceedings of the Second International Choice Modelling Conference (ICMC 2011), Leeds, North England, UK*.

- Sándor, Z., M. Wedel, 2001. Designing conjoint choice experiments using managers' prior beliefs. *Journal of Marketing Research*. 38 430-444.
- Spiegelhalter, D. J., N. G. Best, B. P. Carlin and A. van der Linde, 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B*. 64 583-639.
- Spiegelhalter, D. J., A. Thomas, N. G. Best, and D. Lunn, 2003. WinBUGS User Manual. Version 1.4. Cambridge, England. MRC Biostatistics Unit.
- Srivastava, L., 2005. Mobile phones and the evolution of social behaviour. *Behaviour and Information technology*. 24(2) 111-129.
- Tanner, M. A. and W. H. Wong, 1987. The calculation of posterior distributions by data augmentation (with discussion). *Journal of the American Statistical Association*. 82 528-550.
- Temme, D., M. Paulssen, and T. Dannewald, 2008. Incorporating Latent Variables into Discrete Choice Models - A Simultaneous Estimation Approach Using SEM Software. *Business Research*. 1(2) 220-237.
- Toubia, O., J. R. Hauser, and D. I. Simester, 2004. Polyhedral Methods for Adaptive Choice-Based Conjoint Analysis. *Journal of Marketing Research*. 41 116-131.
- Train, K., 2003. Discrete choice methods with simulation. Cambridge University Press.
- Walker, J. and M. Ben-Akiva, 2002. Generalized Random Utility Model. *Mathematical Social Sciences*. 43(3) 303-343.
- Yu, J., Goos, P., Vandebroek, M., 2008. Model Robust Design of Conjoint Choice Experiments. *Journal of Communications in Statistics - Simulations and Computations*. 37 1603-1621.
- Yu, J., Goos, P., Vandebroek, M., 2011. Individually adapted sequential Bayesian conjoint-choice designs in the presence of consumer heterogeneity. Submitted for publication.