A hybrid choice model with a nonlinear utility function and bounded distribution for latent variables

Citation for published version (APA):

DOI:
10.1080/23249935.2016.1193567

Document status and date:
Published: 21/06/2016

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher’s website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.
A hybrid choice model with a nonlinear utility function and bounded distribution for latent variables: application to purchase intention decisions of electric cars

Jinhee Kim, Soora Rasouli & Harry Timmermans

To cite this article: Jinhee Kim, Soora Rasouli & Harry Timmermans (2016) A hybrid choice model with a nonlinear utility function and bounded distribution for latent variables: application to purchase intention decisions of electric cars, Transportmetrica A: Transport Science, 12:10, 909-932, DOI: 10.1080/23249935.2016.1193567

To link to this article: http://dx.doi.org/10.1080/23249935.2016.1193567
A hybrid choice model with a nonlinear utility function and bounded distribution for latent variables: application to purchase intention decisions of electric cars

Jinhee Kim, Soora Rasouli and Harry Timmermans

Department of the Built Environment, Eindhoven University of Technology, Eindhoven, The Netherlands

ABSTRACT

The hybrid choice model (HCM) provides a powerful framework to account for heterogeneity across decision-makers in terms of different underlying latent attitudes. Typically, effects of the latent attitudes have been represented assuming linear utility functions. In contributing to the further elaboration of HCMs, this study suggests an extended HCM framework allowing for nonlinear utility functions of choice alternatives including not only observed but also latent variables. Box–Cox transformations are used to represent the nonlinear utility function. Johnson's SB distribution is suggested to represent the random term of the latent variables, satisfying the constraint of the Box–Cox transformation. An empirical study using stated choice data about the intention to purchase electric cars is conducted. The empirical results show that the proposed framework can capture nonlinear effects of underlying variables including latent attitudes, thereby enhancing the explanatory power of the choice model.

1. Introduction

The hybrid choice model (HCM) can be viewed as an expanded discrete choice modeling framework, which integrates different types of models, such as latent class and latent variable models, into a structure that is simultaneously estimated (Ben-Akiva et al. 2002). It provides a powerful framework to account for heterogeneity across decision-makers, due to different latent attitudes, in modeling discrete choice behavior. Decision-makers’ latent attitudes may reflect their different values, social norms and lifestyles. The latent attitudes, thus, relate to subjective beliefs and tastes, which may influence the utility that people derive from their choice. Since these latent factors cannot be directly observed from revealed choices, they should be identified through a latent variable model based on a set of attitudinal indicators. HCMs allow simultaneously identifying the latent attitudes and their effects on discrete choice behavior.

Various studies have employed HCMs to explore the influence of latent attitudes on travel behavior. For instance, personal perceptions of comfort and convenience for different...
transportation modes have been taken into consideration as latent attitudes, influencing the utility of transportation mode (e.g. Walker and Ben-Akiva 2002; Temme, Paulssen, and Dannewald 2008; Paulssen et al. 2014). In addition, decision-makers’ environmentally friendly attitudes and environmental concerns have been treated as crucial latent variables to better understand and predict their choice of sustainable travel mode (e.g. Kim, Bae, and Chung 2012; Atasoy, Glerum, and Bierlaire 2013) and alternative fuel vehicles (e.g. Daziano and Bolduc 2013b; Hess, Shires, and Jopson 2013; Jensen, Cherchi, and Mabit 2013; Soto, Cantillo, and Arellana 2014). With respect to alternative fuel vehicles, latent attitudes toward vehicle features (e.g. design, spaciousness, technology, etc.) and leasing a vehicle have been taken into consideration (e.g. Glerum et al. 2014; Mabit et al. 2014). Personal distrust of government has been employed as a latent variable for investigating people’s stated preference for different rail security systems (e.g. Daly et al. 2012). Moreover, parents’ latent preference for walking has been considered for understanding walking preferences of their children (Kamargianni, Ben-Akiva, and Polydoropoulou 2014). Results of the aforementioned applications of HCMs provide evidence the additional insight in decision-making processes they can provide and their explanatory power. Kim, Rasouli, and Timmermans (2014b) provide a review of these models and recent progress.

People’s latent attitudes affecting their travel choice behavior can be revealed by their responses to a set of attitudinal indicators. Thus, HCMs simultaneously take into account the effects of latent attitudes on the preferences for alternatives (using a discrete choice model) and the responses to the indicators (using a latent variable model). Typically, these effects of latent variables have been investigated by assuming a linear-in-parameters model specification. Specification of a linear relationship is just a limited representation of all possible types of effects because a linear function is nothing but a special case of a nonlinear function. Assuming nonlinear relationships thus seems theoretically more appealing because they represent a more generalized approach to model specification. In terms of the relationship between latent attitudes and their indicators, recently Daly et al. (2012) suggested an ordered choice model by treating the indicators as ordered categorical variables. Several studies have employed this approach to represent attitudinal indicators in the HCM framework (e.g. Hess, Shires, and Jopson 2013; Bhat and Dubey 2014; Dekker et al. 2014; Soto, Cantillo, and Arellana 2014).

However, to date, nonlinear effects of latent attitudes on the utility of choice alternatives have not yet been examined. When the utility functions underlying the discrete choice model are assumed to be linear in nature, the model cannot represent any varying marginal utility of the levels of the explanatory variables. Different levels of personal latent attitudes can induce different marginal utilities of the alternative. For example, the utility of a private car may increase exponentially with an increasing positive personal latent attitude with respect to privacy. In general, without explicit consideration of either theoretical or empirical evidence of nonlinear relationships in terms of latent attitudes, researchers tend to assume linear-in-parameters relationships. Moreover, there is no general approach to deal with nonlinear relationships related to latent variables in discrete choice analyses. A model that does not take such nonlinear relationships into account might be expected to yield biased policy effects. Therefore, it is necessary to relax the linearity assumption and check for any nonlinearity in the utility functions of HCMs, reflecting the more general nature of travel choice behavior.
In terms of observable attributes affecting travel choice behavior (e.g. travel time and costs), a number of researchers have investigated nonlinear effects by using a power transformation such as the Box–Cox transformation (Box and Cox 1964) based on the classical discrete choice modeling framework (i.e. Logit model). The Box–Cox transformation is a power transformation, which can represent linear, logarithmic and power functions according to the transformation parameter. Thus, the application of this transformation is useful for investigating nonlinear functional forms in a linear-in-parameters formulation. Gaudry and Wills (1978) used the transformation to estimate diverse travel demand functions. Hensher and Johnson (1981), Koppelman (1982), McCarthy (1982), Gaudry, Jara-Diaz, and Ortuzar (1989), Mandel, Gaudry, and Rothengatter (1994, 1997) and Orro, Novales, and Benitez (2005) developed nonlinear utility functions of their travel mode choice models by applying the transformation rule. Picard and Gaudry (1998) and Rotaris et al. (2012) also applied the Box–Cox transformation for developing freight mode choice models. Stathopoulos and Hess (2012) investigated nonlinear effects of travel time and fare on commute type choice behavior by using the transformation. Their empirical evidence suggests that the marginal disutility for travel tends to increase with travel time and decrease with travel costs. Castillo and Benitez (2013) suggested using the Box–Cox transformation for improving the performance of a trip generation and attraction forecasting model.

The aim of the present study is to suggest a HCM framework allowing investigating nonlinear relationships between the utility and its explanatory variables, which includes not only observable variables but also latent variables, in an integrated modeling framework. In particular, this study focuses on how to represent the nonlinearity of the latent variables in the discrete choice model of HCMs based on the Box–Cox transformation rule. Furthermore, we conduct an empirical study based on stated choice data about the intention to purchase electric cars (ECs) (Rasouli and Timmermans 2013, 2016; Kim, Rasouli, and Timmermans 2014a). The data include responses to a set of attitudinal questions (i.e. indicators of latent attitudes), and allows investigating the effects of specific attributes of ECs as well as social influence of different subsets of social networks. To the best of our knowledge, this study is the first to suggest and apply an approach to deal with nonlinear relationships between the latent attitudes and travel choice behavior within the HCM framework.

This paper is organized as follows. The following sections present the methodologies to develop a classical HCM, whose utility functions are specified by linear-in-parameters, and the suggested nonlinear HCMs based on the Box–Cox transformation. There is a constraint in applying the transformation for the latent variables: the transformation does not allow taking zero or negative values as the attribute values of the variables. We suggest approaches to satisfy this constraint by considering the bounded nature of the attitudinal indicators, which are explained in the third section. In the fourth section, the data and the empirical models are presented. In addition, the estimation results are discussed by comparing the models with respect to their performance and behavioral aspects. The last section concludes the paper.

2. The classic HCM

HCMs incorporate a latent variable model into a discrete choice model in order to improve the explanatory power of the choice model by considering the effects of decision-makers’
latent attitudes. A basic HCM incorporates a latent variable model specified by linear-in-parameter relationships into a discrete choice model based on linear utility functions. The discrete choice model is composed of a structural relationship and a measurement relationship. The structural relationship is represented by utility functions based on linear-in-parameters functions, which consist of the latent variables and the observed exogenous variables. The measurement relationship is derived under the assumption of utility-maximizing behavior. In order to express the choice as a function of the utility, an individual’s observed choice is used as a single nominal indicator. The structural Equation (1) and measurement Equation (2) can be expressed as follows:

\[
U_{in} = \beta_Z X_{Zn}^T + \beta_M X_{Mn}^T + \beta_L X_{Ln}^T + \varepsilon_{in}, \quad \varepsilon_{in} \sim G(0, \sigma_{\varepsilon}),
\]

\[
y_{in} = \begin{cases} 
1, & \text{if } U_{in} = \max_j U_{jn} \\
0, & \text{otherwise}
\end{cases}
\]

where \( U_{in} \) is the random utility of alternative \( i \) of individual \( n \). \( X_{Zn}^T \) and \( X_{Mn}^T \) indicate the observed exogenous variables, which are a \((Z \times 1)\) vector of socio-demographic variables and a \((M \times 1)\) vector alternative-attribute variables corresponding alternative \( i \), respectively. \( X_{Ln}^T \) indicates a \((L \times 1)\) vector of the latent-attitude variables. The \((1 \times Z)\) vector \( \beta_Z \), the \((1 \times M)\) vector \( \beta_M \) and \((1 \times L)\) vector \( \beta_L \) are unknown parameters to be estimated for socio-demographics, alternative-attributes and latent-attitudes in terms of alternative \( i \), respectively. \( \varepsilon_{in} \) represents the random disturbance term with zero mean and standard deviation \( \sigma_{\varepsilon} \), which is assumed as independently and identically distributed (IID) extreme value type I (i.e. Gumbel distribution).

HCMs can be expanded to deal with heterogeneities due to latent market segmentation and individual taste variation by introducing a latent class model, respectively, a mixed-logit model (e.g. Walker and Ben-Akiva 2002; Hess, Shires, and Jopson 2013). Furthermore, recently, several researchers (e.g. Dekker et al. 2014; Hess and Stathopoulos 2014) suggested replacing the utility-maximizing choice model with the random regret minimization model (Chorus, Arentze, and Timmermans 2008) or with a hybrid utility/regret model (Chorus, Rose, and Hensher 2013).

In all these models, since the latent attitudes cannot be directly observed, they should be identified through a latent variable model based on a set of attitudinal indicators. The latent variable model is composed of a set of structural and measurement relationships. The indicators are responses to survey questions regarding different attitudes. This kind of structural equation model is called the multiple indicators, multiple causes (MIMIC) model. The structural Equation (3) and the measurement Equation (4) can be expressed as follows:

\[
X_{Ln} = \Gamma X_{Zn} + \zeta_n, \quad \zeta_n \sim N(0, \sigma_\zeta),
\]

\[
I_n = \Lambda^C + \Lambda X_{Ln} + \xi_n, \quad \xi_n \sim N(0, \sigma_\xi),
\]

where the \((L \times Z)\) matrix \( \Gamma \) contains unknown parameters, and the random part \( \zeta_n \) indicates a multivariate normal distribution with zero mean and an \((L \times L)\) covariance matrix \( \sigma_\zeta \). In order to identify latent variables, a \((D \times 1)\) vector of observable indicator variables
\( \mathbf{I}_n \) is utilized, which are responses of individual \( n \) to \( D \) survey questions regarding different latent attitudes. This measurement relationship can be expressed by a \((D \times 1)\) vector of unknown constants \( \mathbf{A}^C \), a \((D \times L)\) matrix of unknown parameters \( \mathbf{A} \) and random disturbance term \( \xi_n \), which is also a multivariate normal distribution with zero mean and an \((D \times D)\) covariance matrix \( \sigma_{\xi}^\mathbf{C} \). When the disturbance terms are assumed to be IID, the covariance matrices \( \sigma_{\xi}^z \) and \( \sigma_{\xi}^e \) are diagonal matrices whose off-diagonal elements are equal to zero.

The joint likelihood function of the HCM can be defined by assuming that the random disturbance terms are independent. The likelihood of an individual consists of the likelihood function of the discrete choice model and the measurement and structural components of the latent variable model (i.e. MIMIC model) as follows:

\[
\mathbb{L}_n(\mathbf{y}_n, \mathbf{I}_n) = \int_{\mathbf{X}} f_y(\mathbf{y}_n | \mathbf{X}_n^Z, \mathbf{X}_n^M, \mathbf{X}_n^L; \mathbf{\beta}_Z, \mathbf{\beta}_M, \mathbf{\sigma}_\varepsilon) f_I(\mathbf{I}_n | \mathbf{X}_n^L; \mathbf{\Lambda}^C, \mathbf{\Lambda}, \mathbf{\sigma}_\varepsilon) f_L(\mathbf{X}_n^L | \mathbf{X}_n^Z; \mathbf{\Gamma}, \mathbf{\sigma}_\xi) \, d\mathbf{X}_n^L, \tag{5}
\]

where \( f_y(\cdot) \) indicates the likelihood function of the discrete choice model, \( f_I(\cdot) \) and \( f_L(\cdot) \) indicate the distribution functions with regard to the latent variable model, which are corresponding to the measurement relationship and the structural relationship, respectively. Since the latent variables are unknown, the likelihood can be obtained by integrating over the latent constructs \( f_L(\cdot) \). Thus, the dimension of the integral becomes the same as the number of latent variables. By jointly constructing the distribution function of the measurement relationship, the indicators in the latent variable model not only allow identifying the latent variables, but also provide efficiency in estimating the full model (Walker 2001).

The functional forms of these sub-functions depend on the assumptions of the probability distributions of their random disturbance terms. According to the assumptions in the classical HCM (i.e. Equations (1), (3) and (4)), the sub-functions can be induced as the following equations:

\[
f_y(\mathbf{y}_n | \mathbf{X}_n^Z, \mathbf{X}_n^M, \mathbf{X}_n^L; \mathbf{\beta}_Z, \mathbf{\beta}_M, \mathbf{\sigma}_\varepsilon) = \prod_{i \in \mathcal{J}} P_n(\mathbf{y}_n = 1 | \mathbf{X}_n^Z, \mathbf{X}_n^M, \mathbf{X}_n^L; \mathbf{\beta}_Z, \mathbf{\beta}_M)^{y_{in}},
\]

\[
= \prod_{i \in \mathcal{J}} \left[ \frac{\exp(\mathbf{\beta}_{Z,i} \mathbf{X}_n^Z + \mathbf{\beta}_{M,i} \mathbf{X}_n^M + \mathbf{\beta}_{L,i} \mathbf{X}_n^L)}{\sum_{j \in \mathcal{J}} \exp(\mathbf{\beta}_{Z,j} \mathbf{X}_n^Z + \mathbf{\beta}_{M,j} \mathbf{X}_n^M + \mathbf{\beta}_{L,j} \mathbf{X}_n^L)} \right]^{y_{in}}, \tag{6}
\]

\[
f_I(\mathbf{I}_n | \mathbf{X}_n^L; \mathbf{\Lambda}^C, \mathbf{\Lambda}, \mathbf{\sigma}_\varepsilon) = \prod_{d \in \mathcal{D}} \frac{1}{\sigma_{\xi, d}} \Phi \left( \frac{\mathbf{X}_n^L - \mathbf{\Lambda}^C \mathbf{I} \mathbf{X}_n^Z}{\sigma_{\xi, d}} \right), \tag{7}
\]

\[
f_L(\mathbf{X}_n^L | \mathbf{X}_n^Z; \mathbf{\Gamma}, \mathbf{\sigma}_\xi) = \prod_{l \in \mathcal{L}} \frac{1}{\sigma_{\xi, l}} \Phi \left( \frac{\mathbf{X}_n^L - \mathbf{\Gamma} \mathbf{X}_n^Z}{\sigma_{\xi, l}} \right), \tag{8}
\]

where \( P_n(\cdot) \) indicates the probability to choose alternative \( i \) of individual \( n \), and \( \Phi(\cdot) \) indicates the standard normal density function. While the covariance matrix \( \sigma_{\xi} \) need not to be estimated according to the normalization with IID error for logit model, the covariance matrices \( \sigma_{\xi}^d \) and \( \sigma_{\xi} \) have to be estimated. \( \sigma_{\xi, d} \) and \( \sigma_{\xi, l} \) are the standard deviations of the disturbance terms indicating the \( d \)th and \( l \)th diagonal elements of \( \sigma_{\xi}^d \) and \( \sigma_{\xi} \), respectively. \( \mathbf{I}_n \)

and $\Lambda^{dC}$ indicate the $d$th elements of their corresponding vectors $I_n$ and $\Lambda^C$, and $\Lambda^d$ indicates the $d$th row vector of the $(D \times L)$ unknown parameter matrix $\Lambda$. Likewise, $X^l_n$ is the $l$th latent variable of individual $n$, and $\Gamma^l$ is the $l$th row vector of the unknown parameter matrix $\Gamma$.

3. The nonlinear HCM

3.1. Nonlinear utility functions based on the Box–Cox transformation

There are several ways to allow nonlinear specifications of the utility function (i.e. Equation (1)) to investigate possible nonlinear effects of the explanatory variables on choice behavior. First, nonlinear relationships can be approximated by linear-in-parameter specifications. Kim, Rasouli, and Timmermans (2014b) reviewed two possible approaches in this context. One approach is using piecewise linear functions. The attribute levels of each explanatory variable can be divided into certain intervals. By estimating separate linear functions for each interval (i.e. piecewise linear functions), we can obtain different parameters for different attribute ranges. The other approach is to calculate new values for variables by using predetermined nonlinear functions such as logarithmic, exponential or power functions. In other words, the values of variables are exogenously manipulated through a selected functional form, and the manipulated values are used as the explanatory variables based on the linear-in-parameter specifications. However, these approaches have some drawbacks. In the first approach, the researcher has to make several operational decisions such as functional form, the number of piecewise linear functions and range of each segment. The optimal combination of these components may be found heuristically, but it involves extremely time-consuming task. In case of the piecewise linear approach, the utility function is still linear-in-parameters within a selected range. In particular, it is difficult to directly and properly handle the value of a latent variable because the latent variables are treated as random variables and their values are estimated endogenously in the HCM framework. It means it is varied iteratively according to parameters estimated at each iteration of the estimation procedure.

A more general way to consider the nonlinear effects involves applying a power transformation to each explanatory variable and estimating regression parameters. In the case of applying the Box–Cox transformation, the nonlinear utility functions can be represented as follows:

$$U_{in} = \beta_{Z_i} X^{Z,(\lambda_{Z_i})}_n + \beta_{M_i} X^{M,(\lambda_{M_i})}_n + \beta_{L_i} X^{L,(\lambda_{L_i})}_n + \epsilon_{in}, \quad \epsilon_{in} \sim G(0, \sigma_{\epsilon_i}), \quad (9)$$

where

$$X^{k,(\lambda_{k,j})}_n = \begin{cases} \frac{X^{k+1}_{n,j} - 1}{\lambda_{k,j}}, & \lambda_{k,j} \neq 0 \\
\ln X^{k}_{n,j}, & \lambda_{k,j} = 0 \end{cases} \quad X^{k}_{n,j} > 0, \quad \forall k \in \{Z, M, L\}. \quad (10)$$

Equation (10) indicates a transformed variable corresponding the $k$th explanatory variable $X^{k}_{n,j}$. $\lambda_{k,j}$ is a transformation parameter with respect to $X^{k}_{n,j}$. According to the transformation parameter, the function can approximate different types of nonlinear functions. It becomes linear when $\lambda_{k,j} = 1$, and is transformed into a logarithmic function when $\lambda_{k,j} = 0$. 
Otherwise, it takes the form of a power function. This approach provides several advantages in modeling nonlinear relationships. First, the transformed equations can eliminate the collinearity between the regression parameter ($\beta_{k,i}$) and the transformation parameter ($\lambda_{k,i}$), which exists when directly estimating the original power function (Box and Cox 1964; Koppelman 1981). Secondly, the functional form for each variable can be decided endogenously based on the estimation result of the transformation parameter. As explained above, depending on the estimated $\lambda_{k,i}$, the function can approximate linear, logarithmic and power functions. Finally, this transformation allows carrying out a statistical diagnostic test for functional form. Based on the estimated $\lambda_{k,i}$ and its standard error, we can statistically infer whether $\lambda_{k,i}$ is significantly different from 0, 1 or both (e.g. Mandel, Gaudry, and Rothengatter 1994; Stathopoulos and Hess 2012).

### 3.2. Limitation of application of the Box–Cox transformation for latent variables

There is one constraint in applying the Box–Cox transformation. It is only valid for positive $X_k^n$. In other words, every variable corresponding to every individual must have only positive values because of the logarithmic and power components of the utility specification. In case of the observed exogenous variables, this constraint can be easily satisfied by exogenously manipulating zero or negative values while the scale differences among the original values are maintained. Several examples of such manipulation will be shown in the next section. Another approach is including a location parameter $\mu_k$ in the Box–Cox transformation, which is known as the Box–Tukey transformation. According to this approach, $X_k^n$ is replaced by $X_k^n + \mu_k$. If the location parameter $\mu_k$ is estimated or fixed as a value to assure that $X_k^n + \mu_k$ is always greater than zero for all observations, the transformation can be effected for zero and negative $X_k^n$. For instance, Hensher and Johnson (1981) and McCarthy (1982) applied the Box–Tukey transformation because some observations take on zero values for travel cost (e.g. parking cost) variables. They fixed exogenously their location parameters as a very small value to avoid zero values for all their observations.

However, it is difficult to exogenously handle the value of latent variables and find or estimate proper location parameters with respect to latent variables because a latent variable in HCM is dealt with as an endogenous random variable which includes the random disturbance term and has to be estimated simultaneously and endogenously through the latent variable model. Based on the classical HCM, the $l$th latent variable transformed by the Box–Cox transformation can be represented as follows:

$$X_{n,l}^{l(\lambda_{l,l})} = \begin{cases} \frac{(\Gamma^l X_{n}^l + \zeta_{n}^l)^{\lambda_{l,l}} - 1}{\lambda_{l,l}}, & \lambda_{l,l} \neq 0, \\ \ln(\Gamma^l X_{n}^l + \zeta_{n}^l), & \lambda_{l,l} = 0 \end{cases} \quad (\Gamma^l X_{n}^l + \zeta_{n}^l) > 0, \quad \zeta_{n}^l \sim N(0, \sigma_{\zeta}^l). \quad (11)$$

The scale of a latent variable depends on both the scale of the systematic part (i.e. $\Gamma^l X_{n}^l$) and the random part (i.e. $\zeta_{n}^l$). Because of the random part, it cannot guarantee that the scale of the latent variable is always positive for the entire space of the latent variable when the normal distribution is chosen for the random part, which has infinite ranges in both positive and negative directions. This limitation makes it impossible to apply the Box–Cox
transformation to latent variables because the entire space of their density functions must be considered in order to evaluate the joint likelihood function (i.e. Equation (5)). In addition, it means that the Box–Tukey transformation is not applicable either because the location parameter can assure a positive $X_l + \mu_l$ for all observations and the entire space of the distribution only if $\mu_l$ is a positive infinite number. Therefore, it is necessary to develop a method to deal with this limitation by including an explicit mechanism in the latent variable model of the HCM framework.

3.3. Normalization and bounded distribution for latent variables

In the latent variable model, the latent variables are measured through their indicator variables as shown in Equation (4). At this point, to identify latent variables, it is necessary to normalize the structural model or the measurement model by fixing either the scale of $\sigma_\zeta$ or several parameters of $\Lambda^C_1$ and $\Lambda$, respectively. Theoretically, the estimation results should be identical (Daly et al. 2012; Raveau, Yáñez, and Ortúzar 2012). Normalizing the measurement model is a more popular approach than normalizing the variances of the structural equations. For example, when there are three indicator variables for one latent variable, the following specification can be assumed for the normalization: $\Lambda^{d=1,C} = 0$ and $\Lambda^{d=1,l=1} = 1$. Accordingly, the other measurement parameters (i.e. $\Lambda^{d=2,C}, \Lambda^{d=3,C}, \Lambda^{d=2,l=1}$ and $\Lambda^{d=3,l=1}$) can be estimated as the relative effects of the first relationship (i.e. $p^{d=1} = X_l^{d=1} + \epsilon^{d=1}$). In this context, the first indicator can be referred as the normalized indicator. At the same time, this specification also implies that the expected scale of the latent variable is almost the same as the scale of the normalized indicator because $E[\epsilon^{d=1}]$ is zero. Thus, the scales of latent variables depend on what types of indicators are normalized for each latent variable.

In this context, to satisfy the constraint of the Box–Cox transformation, we assume that the scale of a latent variable is bounded by minimum and maximum values of its normalized indicator when the latent attitude is measured directly through linear functions of the indicators (i.e. Equation (4)), and one of the functions is selected as the normalized indicator. The indicators are responses to attitudinal questions and psychological dispositions included in surveys. A respondent judges the degree of agreement of each statement. Likert scales are the most widely used to measure attitudes, which involve a finite number of scales. For example, a five-point Likert scale can be used to rate the indicators using ‘strongly disagree’, ‘disagree’, ‘neutral’, ‘agree’ and ‘strongly agree’. These scales are re-scaled based on a finite number of integers with equal interval (e.g. 1–5) in the modeling stage. It implies that different scales of each respondent in terms of a specific latent attitude are adjusted to a common scale. Thus, we assume that the range of the latent scale can be represented by a bounded and continuous domain if the domain can properly approximate the infinite domain of the normal distribution in terms of the shape of distribution and the performance in estimating parameters. In the present study, it is assumed that the boundary of the domain is strictly limited by the range of the normalized indicator, which means every level of a latent attitude can be represented by a value between the minimum and maximum values of its normalized indicator. The assumption of the bounded scale for latent variables makes it possible to apply the Box–Cox transformation to latent variables. If the minimum value is greater than zero, it can assure that $X_l$ is always greater than zero for all observations.
The bounded scale of latent variables can be dealt with by considering a bounded probability distribution as its random distribution. From a mathematical perspective, the conception of this approach is in line with the approaches for representing bounded random parameters in mixed-logit models. For instance, Bhat (1998, 2000), Reveit and Train (2000) and Train (2001) employed the lognormal distribution, which is a bounded probability distribution whose logarithm has a normal distribution, in order to assure that the range of a random parameter is located only on the positive domain. By assuming the lognormal distribution, the range is anchored at zero and has an infinite tail in the positive direction. On the one hand, Hensher and Greene (2003) suggested using a constrained triangular distribution to avoid negative values and limit the boundary of random parameters. According to their specification, the shape of distribution is triangular, and the spread or standard deviation is a function of the mean of each random parameter. By doing so, the boundary of a random parameter is doubly bounded and increases with the absolute value of the mean of the parameter. Train and Sonnier (2005), Hess, Bierlaire, and Polak (2005) and Hess, Axhausen, and Polak (2006) used the Johnson’s SB distribution to limit both the sign and the boundary of random parameters. The Johnson’s SB distribution (Johnson and Kortz 1970) can be obtained as a logistic transformation of the normal distribution and includes four parameters, which are mean, variance and minimum and maximum values of its domain. It can represent a symmetric or asymmetric probability distribution within a finite range of the domain.

Among the bounded probability distributions, the Johnson’s SB distribution is considered to reflect the assumption of the bounded scale of latent variables in the present study. There are three reasons to employ the Johnson’s SB distribution. First, the distribution is a doubly bounded, which matches the assumption that the boundary of a latent variable is strictly limited by the range of the normalized indicator. Second, the minimum and maximum values of the domain can be controlled exogenously because they should be consistent with the minimum and maximum values of the normalized indicator. Third, it can properly approximate the normal distribution except for the infinite tales of the normal distribution. Thus the Johnson’s SB distribution is not only able to represent the bounded scale of latent variables, but it can also perform like the normal distribution. The empirical evidence will be discussed in Section 4.3. A latent variable with the Johnson’s SB distribution \( X_{ln}^{fb} \) can be represented as transformations of the normal distribution as follows:

\[
X_{ln}^{fb} = B_L + (B_U - B_L) \frac{\exp(X_{ln}^f)}{1 + \exp(X_{ln}^f)},
\]

where \( X_{ln}^f = \Gamma X_n^Z + \zeta_n^f \) and \( \zeta_n^f \) is normally distributed with zero mean and the standard deviation \( \sigma_{\zeta_n^f} \), which is based on the specification of the classical HCM (i.e. Equation (3)). \( B_L \) and \( B_U \) are the lower and upper boundaries of the scale of the latent variables, respectively, which indicate the minimum and the maximum values of the normalized indicator variable. The functional form of the Johnson’s SB distribution is like a logistic distribution, which can be shaped depending on the location and scale parameters, and the boundary values. The boundary values can be estimated according to the model specification, but, in this study, they are dealt with as fixed parameters in order to be consistent with the range of the indicator variable. When the indicator variables include only positive values, this assumption and specification make it possible to apply the Box–Cox transformation to the latent variables.
by replacing the normally distributed latent variable \( X_n^l \) with the bounded latent variable \( X_n^{l*} \) in Equations (9) and (11).

### 3.4. The maximum simulated likelihood approach for the nonlinear HCM

There are several estimation approaches to simultaneously estimate the HCM; the maximum simulated likelihood (MSL) approach (e.g. Walker and Ben-Akiva 2002; Daly et al. 2012; Prato, Bekhor, and Pronello 2012; Jensen, Cherchi, and Mabit 2013; Kamargianni, Ben-Akiva, and Polydoropoulou 2014; Soto, Cantillo, and Arellana 2014), the Bayesian approach (e.g. Daziano and Bolduc 2013a, 2013b; Dekker et al. 2014) and the maximum approximate composite marginal likelihood approach (e.g. Bhat and Dubey 2014). In the present study, the MSL approach was employed for estimating parameters of the suggested nonlinear HCMs, which results in consistent and fully efficient estimates and can be used without any concern about the number of latent variables. The MSL method is the same as the maximum likelihood method, except that simulated probabilities are used instead of the exact probabilities (Train 2009). The simulated likelihood function of the nonlinear HCM can be developed by adapting the function of the classical HCM. The function of the classical HCM is formulated by using a reformulated structural equation of the latent variable model (Equation (3)) as follows:

\[
\tilde{X}_n^L = \Gamma X_n^Z + \sigma_\xi \tilde{\xi}_n, \quad \tilde{\xi}_n \sim N(0, I), \tag{14}
\]

where \( \sigma_\xi \) indicates an \((L \times L)\) matrix whose off-diagonal elements are zero and whose diagonal elements are \( \{\sigma_{\xi 1}, \ldots, \sigma_{\xi j, \ldots}, \sigma_{\xi l}\} \). By introducing Equation (14) into Equation (5), the distribution function for the structural relationship (i.e. Equation (8)) is reduced as

\[
\prod_{l \in L} \Phi(\tilde{\xi}_l^r) \text{ for each latent variable and each observation in sample and takes the average value of the likelihood values corresponding to all random draws as an unbiased simulator, which means the simulated likelihood function, as follows:}
\]

\[
\hat{L}_{CL}^n(y_n, I_n) = \frac{1}{R} \sum_{r=1}^{R} \prod_{i \in J} \left\{ \frac{\exp(\beta_{Z, n} X_n^Z + \beta_{M, n} X_n^M + \beta_{L, n} \tilde{X}_n^{l, r})}{\sum_{j \in J} \exp(\beta_{Z, n} X_n^Z + \beta_{M, n} X_n^M + \beta_{L, n} \tilde{X}_n^{l, r})} \right\}^{y_{in}} 
\times \prod_{d \in D} \frac{1}{\sigma_{\xi d}} \Phi \left( \frac{d_n - \Lambda d C - \Lambda d \tilde{X}_n^{l, r}}{\sigma_{\xi d}} \right), \tag{15}
\]

where

\[
\tilde{X}_n^{l, r} = \Gamma x_n^Z + \sigma_\xi \tilde{\xi}_n^r, \tag{16}
\]

where \( \tilde{\xi}_n^r \) indicates the \( r \)th random draw from the standard normal distribution in terms of the \( l \)th latent variable.

Based on this specification, the simulated likelihood function of the nonlinear HCM can be represented by including the nonlinear utility function based on the Box–Cox transformation (Equation (9)) and the transformation function for the Johnson’s SB distribution.
(Equation (12)), as follows:

\[
\hat{g}^{NL}_n(y_n, I_n) = \frac{1}{R} \sum_{r=1}^{R} \left[ \prod_{i \in J} \left\{ \exp(\beta_{Z_i} \tilde{X}^{L, r}_{n} + \beta_{M_i} \tilde{X}^{M, r}_{n}) + \beta_{L_i} \tilde{X}^{L, r, (\lambda_{L_i})}_{n} \right\} \right]_{y_n} \\
\times \prod_{d \in D} \frac{1}{\sigma_{\xi_d}} \Phi \left( \frac{j_d - \Lambda_d C - \Lambda_d \tilde{X}^{L, r}_{n}}{\sigma_{\xi_d}} \right),
\]

(17)

where

\[
\tilde{X}^{L, r}_{n} = B_L + (B_U - B_L) \frac{\exp(\Gamma^L \tilde{X}^{Z}_{n} + \sigma_{\xi_d} \tilde{X}^{L}_{n})}{1 + \exp(\Gamma^L \tilde{X}^{Z}_{n} + \sigma_{\xi_d} \tilde{X}^{L}_{n})},
\]

(18)

\[
\tilde{X}^{L, r, (\lambda_{L_i})}_{n} = \begin{cases} 
\frac{(\tilde{X}^{L, r}_{n})^{\lambda_{L_i}} - 1}{\lambda_{L_i}} & , \lambda_{L_i} \neq 0, \\
\ln \tilde{X}^{L, r}_{n} & , \lambda_{L_i} = 0
\end{cases}
\]

(19)

where \(\tilde{X}^{L, r}_{n}\) indicates the \(i\)th latent variable with the Johnson’s SB distribution in terms of the \(r\)th random draw, which is evaluated based on the normally distributed latent variable \(\tilde{X}^{L}_{n}\) (i.e. Equation (16)). In other words, the \(\tilde{X}^{L}_{n}\) is also drawn from the standard normal distribution and used for evaluating the latent variable like the classical HCM, but the latent variable is transformed to represent its bounded nature. \(\tilde{X}^{L, r, (\lambda_{L_i})}_{n}\) represents the nonlinear specification of the bounded latent variable \(\tilde{X}^{L, r}_{n}\) based on the Box–Cox transformation. Accordingly, most of the estimating procedures including generating the random draws and calculating the latent variables are similar in both the classical and nonlinear HCMs, except that the nonlinear HCM needs the additional procedures for the transformations of the variables: applying the bounded distribution function to the latent variables, and applying the Box–Cox transformation to the observable variables and the transformed (i.e. bounded) latent variables in the utility function.

4. Application

4.1. Data used

The proposed model was applied to investigate the effects of latent attitudes and social influence on the intention to purchase ECs. The data used in this study are based on a stated choice experiment administered in June 2012 in the Netherlands among 726 respondents. Each respondent indicated his or her purchase intentions for ECs under 16 different hypothetical situations. The respondents were shown binary choice problems, and were asked whether to choose ‘the purchase of EC’ or not. In addition, the data set includes socio-demographics and responses to a set of attitudinal indicator questions. Rasouli and Timmermans (2013) provide a more detailed description of the design and administration of the stated choice experiment. The indicators and the underlying variables which are the specific attributes of ECs and the social influence variables are described in Kim, Rasouli, and Timmermans (2014a). The variables and their attribute levels used in this study are listed in Table 1. Thus, there are two subsets of attributes varied in the experiment: alternative-attributes and social influence variables. The attributes include the price of the EC relative...
Table 1. Observable variables and their attribute levels.

**Exogenous explanatory variables in the discrete choice model: attribute levels**

| Alternative-attributes variable (X_{inn}^{att}) | Purchase price of the EC relative to an equivalent standard car (x_{inn}^{pri}): +35%, +25%, +15%, +5%, −5%, −15%, −25%, −35%  
| Costs of electricity relative to gas (x_{inn}^{cos}): +35%, +25%, +15%, +5%, −5%, −15%, −25%, −35%  
| Cruising range of the ECs (x_{inn}^{cru}): 100, 250, 400, 550 km  
| Time to charge battery (x_{inn}^{tim}): 5 min (=0.083 h), 1, 4, 7 h  
| Maximum speed of the car (x_{inn}^{spe}): 80, 120, 160, 200 km/h  
| Distance to charging station (x_{inn}^{dis}): At home, 1, 5, 10 km  

**Social influence variable (X_{inn}^{soc})**

- Reviews (x_{inn}^{rev}): Only positive, mainly positive but some criticism, mainly negative, but some positive, only negative  
- Share of the EC among friends and acquaintances (x_{inn}^{fam}): 0%, 5%, 50%, 75%  
- Share of the EC among members of larger family (x_{inn}^{rel}): 0%, 5%, 50%, 75%  
- Share of the EC among colleagues (x_{inn}^{col}): 0%, 5%, 50%, 75%  
- Share of the EC among peers (x_{inn}^{peer}): 0%, 5%, 50%, 75%

**Exogenous explanatory variables in the latent variable model: descriptive statistics**

| Socio-demographic variables (X_{inn}^{demo}) | Gender (x_{inn}^{gen}): 1 = Male, −1 = Female: Mean = 0.0220, Std. = 0.9998  
| Age (x_{inn}^{age}): 1 = 18–25 year, 2 = 26–35 year, 3 = 36–50 year, 4 = 51–65 year, 5 = over 65 year: Mean = 3.0987, Std. = 1.2004  
| Marital status (x_{inn}^{mar}): 1 = Couple, −1 = Single: Mean = 0.4298, Std. = 0.9029  
| Education level (x_{inn}^{edu}): 1 = Higher vocational education and university, −1 = lower: Mean = −0.2645, Std. = 0.9644  
| Income (x_{inn}^{inc}): 1 = under 625 Euros/month, 2 = 625–1250 Euros/month, 3 = 1251–1875 Euros/month, 4 = 1876–2500 Euros/month, 5 = over 2500 Euros/month: Mean = 3.1711, Std. = 1.2570

| Latent variables (X_{inn}^{lat}) | Indicator variables for each latent variable: Descriptive statistics  
| Environmental aspects (X_{inn}^{env}) | ECs are a perfect solution for people who want to drive a car and at the same time wish to reduce CO2 emission (L^{env1}): Mean = 4.8623, Std. = 1.4385  
| ECs do not make any contribution to sustainable transport because they still depend on non-green energy (L^{env2}): Mean = 4.0579, Std. = 1.2940  
| I am highly interested in an EC to contribute to a better environment (L^{env3}): Mean = 4.0634, Std. = 1.2493

| Battery aspects (X_{inn}^{bat}) | Charging the batteries takes too long (L^{bat1}): Mean = 4.2975, Std. = 1.3256  
| I find the risk of driving with a low battery too high (L^{bat2}): Mean = 4.6529, Std. = 1.3093  
| I travel too much to consider buying an EC: I would need to charge the battery too often (L^{bat3}): Mean = 3.9573, Std. = 1.4062

| Innovation value (X_{inn}^{inn}) | I am always among the first to buy new high tech products (L^{inn1}): Mean = 3.5620, Std. = 1.2699  
| I am very interested in the latest developments in cars (L^{inn2}): Mean = 4.0358, Std. = 1.3540  
| I prefer to buy cars with the latest technological gadgets (L^{inn3}): Mean = 3.7755, Std. = 1.2412

The choice of social networks’ type (i.e. friends, relatives, colleagues and peers) systematically varied across to the price of an equivalent standard car, the costs of electricity relative to gas, the cruising range of the car (with fully-charged battery), the time required to (re-)charge the battery, the maximum speed of the car and the distance to a charging station. In terms of the distance to a charging station, the respondents were instructed that distance should be interpreted in terms of closest distance from their main activity location (home, work, etc.). The social influence variables were included to investigate the effects of general public opinions about ECs and social networks’ decisions about purchasing ECs. The level of market shares of each social network type (i.e. friends, relatives, colleagues and peers) systematically varied across
different choice situations in a stated choice experiment. An orthogonal fractional factorial design of the $8^2 \times 4^9$ full factorial design in 128 runs was created. The 8-level attributes appear 16 times in the design, whereas the 4-levels attributes are listed 32 times.

4.2. Model specifications and scenarios

The specification of the model included nonlinear relationships with regard to alternative-attribute and latent-attitude variables. The other variables such as social influence and socio-demographic variables were assumed to be linear-in-parameters in the utility function. In the alternative-attributes variables, several variables have zero or negative values. These values had to be manipulated in order to apply the Box–Cox transformation. Since the purchase price of the EC and the costs of electricity indicate relative values, we were able to manipulate those levels from 0.65 to 1.35, implying the same scale as the original levels. In the case of distance to charging station, the ‘at home’ level was substituted by a relatively small value (0.001). In addition, the scales of the range and the maximum speed of ECs are too large to be applied in the power function. In other words, those scales may cause infinite numbers at certain parameters levels, making it impossible to find the optimal solution in the estimating procedure. Therefore, every level of these variables was scaled down by dividing them by one hundred.

With regard to latent-attitude variables, each variable has three indicator variables. In order to identify the magnitudes of the latent variables, the parameters in the measurement model for the first indicator of each variable were normalized to 0 for constants and 1 for coefficients. Moreover, for convenience of interpretation and reduction of complexity in the estimation stage, the scale of second indicator about the environmental aspects was converted to the opposite direction: higher values indicate a more environmentally positive perception of ECs. Finally, since the choice indicator of discrete choice model includes only two options, indicating ‘purchase ($i$)’ or ‘not purchase ($j$)’, the choice probability can be expressed as follows:

$$P_{nq}(i) = \frac{1}{1 + \exp\left\{-V\left(X_n^Z, X_{nq}^S, X_{nq}^M, X_{nq}^L, \alpha_{in}ight)\right\}}, \quad (20)$$

$$P_{nq}(j) = 1 - P_{nq}(i), \quad (21)$$

where

$$V(\cdot) = \beta_{Z,n}X_n^Z + \beta_{S,n}X_{nq}^S + \beta_{M,n}X_{nq}^M + \beta_{L,n}X_{nq}^L + \alpha_{in}. \quad (22)$$

$q$ indicates a choice situation of 16 different situations in our stated choice experiment. In order to consider the random panel effect caused by the correlations over the multiple choice situations made by each individual $n$, the random parameter $\alpha_{in}$ is included, which varies over individuals but is constant over choice situations for each individual. We assumed that the probability distribution of $\alpha_{in}$ is the normal distribution with $\beta_{C,i}$ for mean and $\eta_i$ for standard deviation. Thus, $\beta_{C,i}$ represents the alternative specific constant in terms of ‘purchase of EC’.

Two different models were estimated: (1) the classical HCM and (2) the nonlinear HCM with the Johnson’s $S_B$ distribution. As mentioned in the previous section, the MSL approach was employed for estimating the models. In the present study, the scrambled Halton draws
approach was used for generating the random draws from the standard normal density function. This method properly provides coverage even in high dimensions and induces negative correlation over observations (Bhat 2003). Finding optimal parameters of the simulated log-likelihood functions was carried out by using an optimization routine for an unconstrained multivariable function in MATLAB. In order to validate the suggested model framework, about 10% of the observations (i.e. 72 respondents, 1152 choice outcomes) were set aside as a validation sample for comparing the performance of the different models. Accordingly, the choice data of 654 respondents (i.e. 10,464 choice outcomes) were used for estimation.

4.3. Estimation results

The estimation results of the models are summarized in this section. In terms of model 2, a preliminary model was estimated, and specification tests based on the estimated transformation parameters were carried out in order to find proper nonlinear specifications. Accordingly, the estimates indicated that the transformation parameters corresponding to the purchase price ($\lambda_{pri,i}$) and the innovation value ($\lambda_{inn,i}$) are significantly different from both 0 and 1, which means that the power function is appropriate to represent these variables. In addition, the specification tests revealed that the log function was the best function to represent the effect of several alternative-attribute variables such as $X_{\text{cos} \text{inq}}$, $X_{\text{rag} \text{inq}}$, $X_{\text{tim} \text{inq}}$, and $X_{\text{spe} \text{inq}}$. On the other hand, for the other alternative-attribute and latent-attitude variables, the linear function was suitable to represent their effects on the stated purchase intention.

The estimation results of the discrete choice model part are shown in Table 2. Figure 1 graphs the estimated utility functions with respect to the nonlinear variables. Five attributes variables were found to have nonlinear effects. The utility of ECs decreases with increasing price, relative to a standard car with a gasoline engine. In addition, the marginal disutility of the relative price increases with increasing difference between the price of an EC and the price of a standard car. It suggests that the evaluation of ECs relative to the standard car increasingly deteriorates by increasing their relative price. The estimation results also show that increasing the relative cost of electricity to gas reduces the utility, while its marginal disutility decreases. Since the expenditure for charging electricity takes a large part of travel cost using ECs, this result may be consistent with the general expectation of decreasing marginal perceptions of travel cost. In the case of the range of ECs, utility increases with increasing range, but the marginal increment of the utility is diminishing with increasing range. In contrast, the estimated function for time to charge battery indicates that utility decreases with increasing time. In particular, the utility decreases quickly from 5 min to 1 h charging time and decreases slowly after 1 h. This result shows that people are not willing to spend time charging the battery. The estimation result for the maximum speed of ECs shows that speed also affects utility in a nonlinear fashion. The marginal utility of maximum speed decreases with increasing speed. In contrast, the linear-in-parameter specification best represents the relationship between the utility of purchasing ECs and the distance to their charging station. The estimation results indicate that distance has a negative effect on the utility of purchasing ECs.

Before exploring the effects of the latent attitudes on the intention to purchase ECs, the form of each latent-attitude variable should be identified. Tables 3 and 4 show the
Table 2. Estimation results of the discrete choice sub-model.

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th></th>
<th></th>
<th></th>
<th>Model 2</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Alternative specific constant</td>
<td>( p_{C_i} )</td>
<td>-6.0517</td>
<td>0.8873</td>
<td>.0000</td>
<td>-3.1508</td>
<td>1.0189</td>
<td>.0021</td>
<td></td>
</tr>
<tr>
<td>Std. of the panel effect</td>
<td>( n_i )</td>
<td>2.2182</td>
<td>0.1039</td>
<td>.0000</td>
<td>2.2121</td>
<td>0.1090</td>
<td>.0000</td>
<td></td>
</tr>
<tr>
<td>Socio-demographic variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>( \beta_{\text{gen},i} )</td>
<td>-0.0131</td>
<td>0.1173</td>
<td>.9108</td>
<td>-0.0198</td>
<td>0.1060</td>
<td>.8516</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>( \beta_{\text{age},i} )</td>
<td>-0.2147</td>
<td>0.0948</td>
<td>.2400</td>
<td>-0.0976</td>
<td>0.1047</td>
<td>.3516</td>
<td></td>
</tr>
<tr>
<td>Marital status</td>
<td>( \beta_{\text{mar},i} )</td>
<td>-0.0483</td>
<td>0.1149</td>
<td>.6743</td>
<td>-0.0554</td>
<td>0.1243</td>
<td>.6561</td>
<td></td>
</tr>
<tr>
<td>Education level</td>
<td>( \beta_{\text{edu},i} )</td>
<td>0.3475</td>
<td>0.1119</td>
<td>.0020</td>
<td>0.3815</td>
<td>0.1227</td>
<td>.0020</td>
<td></td>
</tr>
<tr>
<td>Income level</td>
<td>( \beta_{\text{inc},i} )</td>
<td>0.1435</td>
<td>0.0939</td>
<td>.1270</td>
<td>0.1238</td>
<td>0.1094</td>
<td>.2582</td>
<td></td>
</tr>
<tr>
<td>Alternative-attribute variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Purchase price(^a)</td>
<td>( \beta_{\text{pri},i} )</td>
<td>-1.9842</td>
<td>0.1480</td>
<td>.0000</td>
<td>-1.9870</td>
<td>0.1649</td>
<td>.0000</td>
<td></td>
</tr>
<tr>
<td>Costs of electricity(^b)</td>
<td>( \lambda_{\text{pri},i} )</td>
<td>2.9562</td>
<td>0.7381</td>
<td>.0083</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cruising range of the ECs(^b)</td>
<td>( \beta_{\text{cos},i} )</td>
<td>-3.0559</td>
<td>0.1526</td>
<td>.0000</td>
<td>-2.8911</td>
<td>0.1434</td>
<td>.0000</td>
<td></td>
</tr>
<tr>
<td>Time to charge(^b)</td>
<td>( \beta_{\text{tim},i} )</td>
<td>0.1766</td>
<td>0.0200</td>
<td>.0000</td>
<td>0.4964</td>
<td>0.0546</td>
<td>.0000</td>
<td></td>
</tr>
<tr>
<td>Maximum speed(^b)</td>
<td>( \beta_{\text{spe},i} )</td>
<td>0.6346</td>
<td>0.0734</td>
<td>.0000</td>
<td>0.9361</td>
<td>0.0981</td>
<td>.0000</td>
<td></td>
</tr>
<tr>
<td>Distance to charging</td>
<td>( \beta_{\text{dis},i} )</td>
<td>0.0339</td>
<td>0.0083</td>
<td>.0000</td>
<td>-0.0349</td>
<td>0.0084</td>
<td>.0000</td>
<td></td>
</tr>
<tr>
<td>Latent-attitude variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Environmental aspects</td>
<td>( \beta_{\text{env},i} )</td>
<td>0.9047</td>
<td>0.1350</td>
<td>.0000</td>
<td>0.6397</td>
<td>0.1746</td>
<td>.0003</td>
<td></td>
</tr>
<tr>
<td>Battery aspects</td>
<td>( \beta_{\text{bat},i} )</td>
<td>-0.9070</td>
<td>0.1110</td>
<td>.0000</td>
<td>-0.9035</td>
<td>0.1290</td>
<td>.0000</td>
<td></td>
</tr>
<tr>
<td>Innovation value(^a)</td>
<td>( \beta_{\text{inn},i} )</td>
<td>0.5861</td>
<td>0.1179</td>
<td>.0000</td>
<td>0.0033</td>
<td>0.0051</td>
<td>.5243</td>
<td></td>
</tr>
<tr>
<td>Social influence variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Review</td>
<td>( \beta_{\text{rev1},i} )</td>
<td>0.1928</td>
<td>0.0554</td>
<td>.0005</td>
<td>0.1974</td>
<td>0.0556</td>
<td>.0004</td>
<td></td>
</tr>
<tr>
<td>Only positive</td>
<td>( \beta_{\text{rev2},i} )</td>
<td>0.2386</td>
<td>0.0552</td>
<td>.0000</td>
<td>0.2321</td>
<td>0.0555</td>
<td>.0000</td>
<td></td>
</tr>
<tr>
<td>Mainly positive</td>
<td>( \beta_{\text{rev3},i} )</td>
<td>-0.0723</td>
<td>0.0566</td>
<td>.2016</td>
<td>-0.0814</td>
<td>0.0570</td>
<td>.1533</td>
<td></td>
</tr>
<tr>
<td>Only negative</td>
<td>( \beta_{\text{rev4},i} )</td>
<td>-0.3591</td>
<td>-</td>
<td>-</td>
<td>-0.3480</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Share: friends</td>
<td>( \beta_{\text{fri},i} )</td>
<td>0.1130</td>
<td>0.1166</td>
<td>.3326</td>
<td>0.0889</td>
<td>0.1172</td>
<td>.4488</td>
<td></td>
</tr>
<tr>
<td>Share: family</td>
<td>( \beta_{\text{fam},i} )</td>
<td>0.2599</td>
<td>0.1170</td>
<td>.2627</td>
<td>0.2659</td>
<td>0.1174</td>
<td>.0238</td>
<td></td>
</tr>
<tr>
<td>Share: colleagues</td>
<td>( \beta_{\text{col},i} )</td>
<td>-0.1253</td>
<td>0.1162</td>
<td>.2811</td>
<td>-0.1327</td>
<td>0.1169</td>
<td>.2570</td>
<td></td>
</tr>
<tr>
<td>Share: peers</td>
<td>( \beta_{\text{pee},i} )</td>
<td>0.0746</td>
<td>0.1158</td>
<td>.5195</td>
<td>0.0935</td>
<td>0.1166</td>
<td>.4230</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)The variable is transformed to the power function in Model 2: the first parameter indicates its regression parameter and the second parameter indicates its transformation parameter.

\(^b\)The variable is transformed to the logarithm function in Model 2.

The estimation results of the latent variables. Although the random distributions of latent-attitude variables in models 1 and 2 are different, the measurement parameters, which determine the form of each latent-attitude variable, are almost the same for both models. The environmental aspects indicate people’s latent perceptions of ECs as environmentally friendly vehicles. The battery aspects are related to people’s negative perception that ECs are related to charging and battery use. The last latent attitude relates to personal innovation value, which represents people’s attitudes about the latest technological items. The models differ in terms of the constant and the standard deviation. Their estimates are much smaller in model 2 than in model 1 due to the logistic function and the minimum and maximum representing the bounded distribution (i.e. Equation (18)). On the other hand, the signs of the significant estimates of the effects of the socio-demographic variables are the same in both models. Aged and/or higher income earners tend to perceive ECs as environmentally friendly. Highly educated people tend to have more negative perceptions about using the battery. Males and/or higher income earners tend to be more interested in the latest technological items. These results suggest that the latent variable model based on Johnson’s S\(_B\) distribution can produce similar results as the model based on the normal distribution.
In order to compare the performance of the two models, the distribution of each latent-attitude variable across individuals was estimated using Equations (16) and (18). Figure 2 shows the distributions of the estimated latent-attitude variables for complete sample. Since model 1 assumes the normal probability distribution, the shape of the distribution of the estimated latent variable is symmetric and covers the negative domain as shown in Figure 2(b) and 2(c). In contrast, the domain of the estimated distribution of model 2 is restricted to 1–7 for all latent variables, and the shape of the distribution is asymmetric for the environmental aspects (Figure 2(a)) and the innovation value (Figure 2(c)) but almost symmetric for the battery aspects (Figure 2(b)). By definition, the estimated distribution of certain latent variables should properly represent the observed distribution of its indicators. From this perspective, model 2 can be regarded as the better model because
it can not only represent the symmetric but also the skewed distribution of observed indicators.

Table 2 shows the effects of the latent attitudes on the intention to purchase ECs. The signs of all estimates are in line with the identified nature of each latent-attitude variable. The result reveals that the environmental aspects have a positive effect on the intention to purchase ECs. It means that the personal belief that using an EC helps improving environmental quality increases the intention to purchase ECs. In contrast, the battery aspects have a significant negative effect, suggesting that concerns about the battery significantly reduce the intention to purchase ECs. The utility of purchasing ECs is linearly related to these two latent attitudes, indicating that the marginal utility of each latent attitude does not vary with changing attitude levels. In contrast, the estimation results show that latent attitudes with respect to innovation value are nonlinearly related with the utility of the intention to purchase. The utility increases exponentially with increasing innovation value (Figure 1(f)). In other words, people who have a strong preference for the latest technological items have a higher intention to buy ECs than people who have a relatively weak preference for the latest technology.

The last part of Table 2 presents the estimation results of the social influence variables. This result reflects the extent of people’s conformity behavior with respect to the...
Table 4. Estimation results of the latent variable model part: Structural relationship.

<table>
<thead>
<tr>
<th>Table 4. Estimation results of the latent variable model part: Structural relationship.</th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est.</td>
<td>Std.</td>
</tr>
<tr>
<td>Environmental aspects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>( \Gamma_{env}^{C} )</td>
<td>4.8711</td>
</tr>
<tr>
<td>Socio-demographic variables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>( \Gamma_{env}^{gen} )</td>
<td>0.0243</td>
</tr>
<tr>
<td>Age</td>
<td>( \Gamma_{env}^{age} )</td>
<td>-0.1271</td>
</tr>
<tr>
<td>Marital status</td>
<td>( \Gamma_{env}^{mar} )</td>
<td>0.0296</td>
</tr>
<tr>
<td>Education level</td>
<td>( \Gamma_{env}^{edu} )</td>
<td>-0.0656</td>
</tr>
<tr>
<td>Income level</td>
<td>( \Gamma_{env}^{inc} )</td>
<td>0.1121</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>( \sigma_{\zeta}^{env} )</td>
<td>0.8795</td>
</tr>
<tr>
<td>Battery aspects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>( \Gamma_{bat}^{C} )</td>
<td>4.4977</td>
</tr>
<tr>
<td>Socio-demographic variables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>( \Gamma_{bat}^{gen} )</td>
<td>0.0372</td>
</tr>
<tr>
<td>Age</td>
<td>( \Gamma_{bat}^{age} )</td>
<td>-0.0433</td>
</tr>
<tr>
<td>Marital status</td>
<td>( \Gamma_{bat}^{mar} )</td>
<td>0.0043</td>
</tr>
<tr>
<td>Education level</td>
<td>( \Gamma_{bat}^{edu} )</td>
<td>0.1443</td>
</tr>
<tr>
<td>Income level</td>
<td>( \Gamma_{bat}^{inc} )</td>
<td>-0.0112</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>( \sigma_{\zeta}^{bat} )</td>
<td>1.0272</td>
</tr>
<tr>
<td>Innovation value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>( \Gamma_{inn}^{C} )</td>
<td>3.2751</td>
</tr>
<tr>
<td>Socio-demographic variables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>( \Gamma_{inn}^{gen} )</td>
<td>0.2216</td>
</tr>
<tr>
<td>Age</td>
<td>( \Gamma_{inn}^{age} )</td>
<td>-0.0338</td>
</tr>
<tr>
<td>Marital status</td>
<td>( \Gamma_{inn}^{mar} )</td>
<td>0.0043</td>
</tr>
<tr>
<td>Education level</td>
<td>( \Gamma_{inn}^{edu} )</td>
<td>-0.0601</td>
</tr>
<tr>
<td>Income level</td>
<td>( \Gamma_{inn}^{inc} )</td>
<td>0.1205</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>( \sigma_{\zeta}^{inn} )</td>
<td>0.8631</td>
</tr>
</tbody>
</table>

Public opinion at large and their social network members. It shows that reviewers’ positive opinions have a significant and positive effect on the intention to purchase ECs. However, the effect decreases and becomes negative when the reviews become more negative. Thus, this result confirms that the general public opinion may play a role in the intention to purchase an EC. Four social influence variables were used to investigate the effects of the choice behavior of social network members and differentiate between friends, family, colleagues and peers. The results suggest that the choices of social network members do not significantly affect the utility of purchasing ECs assuming a linear relationship.

Table 5 shows comparative goodness-of-fit of the two models for both the estimation and validation sample. In the case of the estimation sample, the maximum log-likelihood value of full model 2 is marginally smaller than the corresponding statistic for model 1. This difference may be caused by the assumption of the bounded distribution for the latent variables. As shown in Figure 2, although Johnson’s SB distribution can approximate the normal distribution, the infinite tales of the normal distribution cannot be covered by this distribution. Therefore, there can be a loss in the likelihood value of the latent variable model part, which in turn can influence the likelihood value of the full model. On the other hand, in terms of discrete choice model, various goodness-of-fit indices indicate that model 2 has a higher explanatory power than model 1. In order to statistically compare model’s performance, the adjusted likelihood ratio index \( \bar{\rho}^2 \) (Ben-Akiva and Swait 1986) for non-nested discrete choice models was applied. According to this approach, the following holds.
Figure 2. Comparison of distribution of estimated latent variables and their indicators. (a) Environmental aspects, (b) battery aspects and (c) innovation value.
### Table 5. Comparison of the models: Goodness-of-fit indexes and validation results.

<table>
<thead>
<tr>
<th></th>
<th>Full model (HCM)</th>
<th>Discrete choice sub-model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model 1</td>
<td>Model 2</td>
</tr>
<tr>
<td>Number of parameters ($N_p$)</td>
<td>65</td>
<td>67</td>
</tr>
<tr>
<td>Estimation sample ($N = 654$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log-likelihood value ($\ln \hat{L}$)</td>
<td>$-13045.44$</td>
<td>$-13049.93$</td>
</tr>
<tr>
<td>Goodness-of-fit index</td>
<td>CAIC$^a$</td>
<td>40.0931</td>
</tr>
<tr>
<td>$\rho^2$</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>$\bar{\rho}^2$</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Validation sample ($N = 72$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log-likelihood value ($\ln \hat{L}$)</td>
<td>$-1468.56$</td>
<td>$-1464.79$</td>
</tr>
<tr>
<td>Goodness-of-fit index</td>
<td>CAIC$^a$</td>
<td>42.5990</td>
</tr>
<tr>
<td>$\rho^2$</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>$\bar{\rho}^2$</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

$^a$A lower value of the consistent Akaike information criterion (CAIC) statistic indicates a better goodness-of-fit; CAIC $= -2\ln \hat{\chi}_m + N_p(N + 1)$.

Asymptotically under the null hypothesis that model 1 is the true specification:

$$\Pr(\bar{\rho}^2_2 - \bar{\rho}^2_1 > z) \leq \Phi(-\{2z \ln \hat{\chi}(0) + (N_{p,2} - N_{p,1})\}/2), \quad z > 0,$$

where $\bar{\rho}^2_m$ indicates the adjusted likelihood ratio of model $m$, $\hat{\chi}(0)$ is the log-likelihood value when all the parameters are zero, $N_{p,m}$ is the number of parameters in model $m$, and $\Phi(\cdot)$ indicates the standard normal cumulative distribution function. Accordingly, $\Pr(\cdot)$ indicates the probability that $\bar{\rho}^2_2$ is greater by some $z > 0$ than $\bar{\rho}^2_1$, which is asymptotically bounded above by the $\Phi(\cdot)$. In terms of the estimation sample, the difference between $\bar{\rho}^2_2$ and $\bar{\rho}^2_1$ is 0.0024, $\hat{\chi}(0) = -7253.09$, and the difference in the number of parameters is 2. As the result, the probability is almost zero, which means that, at the 0.05 significant level, introducing a nonlinear specification significantly enhance the explanatory power of the choice model.

In case of the validation sample, the maximum log-likelihood values of both the full model and the discrete choice sub-model part are greater for model 2 than for model 1. In addition, the explanatory power of the choice model is significantly enhanced: $\bar{\rho}^2_2 - \bar{\rho}^2_1 = 0.0047$; $\hat{\chi}(0) = -798.506$; $\Pr(\cdot) = 0.001$. Consequently, not only the result for the estimation sample analysis but also for the validation sample indicate that the explanatory power of model 2 to forecast people’s intention to purchase ECs is significantly greater than the explanatory power of model 1.

### 5. Conclusions and discussion

The present study suggested a generalized HCM framework allowing nonlinear relationships between the utility of choice alternatives and their explanatory variables. The suggested framework provides an integrated and simultaneous procedure that estimates latent variables through a latent variable model and incorporates them into a nonlinear utility function based on the Box–Cox transformation. In order to satisfy the constraint of the Box–Cox transformation, we suggested using the Johnson’s $S_B$ distribution to represent the bounded distribution of the latent variables. According to these specifications, the utility function can approximate various types of nonlinear functions, and the latent variable model becomes more flexible and realistic in representing the true nature of observed...
data. To simultaneously estimate the parameters of the suggested nonlinear HCM, the MSL approach was employed.

The empirical results show that the proposed framework significantly enhances the explanatory power of the nonlinear HCM compared to the model allowing only a linear specification. Furthermore, the results reveal that several underlying variables nonlinearly affect people’s intention to purchase ECs. By increasing the price of ECs relative to the price of a standard car, the utility of purchasing ECs decreases and its marginal disutility increases. The cost of electricity and the time to charge the battery also have negative effects on utility, but their marginal disutility decrease with increasing attribute levels. On the other hand, the range and the maximum speed of ECs have positive effects in that the utility increases with increasing attribute levels, but their marginal utilities diminish with increasing levels. Finally, both the utility of purchasing ECs and the marginal utility of personal attitudes about innovation value increase with increasing attitude levels. These results provide evidence that the classical HCM may overestimate or underestimate the sensitivity of the concerned variables for certain attribute levels. Therefore, researchers should test whether any nonlinearity exists in each variable and figure out the form of these nonlinear effects. From this perspective, our proposed framework helps finding empirical evidence about nonlinear effects of not only observable attributes but also unobservable latent attitudes. It is advisable to always estimate this nonlinear HCM because the classical HCM is a limiting case of this suggested specification.

The present study addressed only the nonlinear relationship with respect to the structural relationship of discrete choice model. In other words, the latent variable model part in the HCM is still specified in terms of linear-in-parameters relationships. Furthermore, some scholars may argue that Likert scales are ordinal in nature, implying that a linear model (e.g. MIMIC model) might not be the best specification. By employing an ordered choice model as the latent variable model, the HCM can allow for the ordinal nature of the indicator (e.g. Daly et al. 2012; Hess, Shires, and Jopson 2013; Bhat and Dubey 2014; Dekker et al. 2014; Soto, Cantillo, and Arellana 2014). However, in order to investigate the nonlinear effect of the latent variable based on an ordered model, a more sophisticated approach has to be developed. This problem will be addressed in a future study.

Disclosure statement

No potential conflict of interest was reported by the authors.

References


