_ _

Hybrid Choice Models : Estimation Using Canned SEM Software

_ _

Dannewald, Till; Paulssen, Marcel; Temme, Dirk; Walker, Joan

This publication URL: https://archive-ouverte.unige.ch/unige:77057

Hybrid Choice Models
Estimation Using Canned SEM Software

Till Dannewald[1], Marcel Paulssen[1], Dirk Temme[1]* and Joan Walker[2]

[1]*Institute of Marketing, Humboldt-University Berlin, 10099 Berlin, Germany*
[2]*Department of Geography and Environment, Boston University, Boston, MA 02215, USA*

Abstract: Hybrid choice models represent a new class of models which merge classical choice models with the structural equation approach (SEM) for latent variables. Even though hybrid choice models allow for a more realistic explanation of choice behavior by incorporating latent constructs such as attitudes and values, applications in marketing are scarce. The present study on travel mode choice clearly underlines the value of hybrid choice models to enhance our understanding of choice processes. In addition to the usually studied directly observable variables such as travel time and cost, abstract motivations such as power as well as latent choice criteria such as flexibility strongly impact on travel mode choice. Moreover, we can show that it is possible to estimate hybrid choice models with the widely available structural equation modelling package Mplus.

---

* Corresponding author. E-Mail address: temme@wiwi.hu-berlin.de

1. Introduction

Recent work in modelling discrete choice has emphasized the need to incorporate unobservable psychological factors in addition to directly observed variables such as time and cost that are traditionally in the focus of interest (e.g., Ben-Akiva et al., 1994; Morikawa et al., 2002). Extending choice models (e.g., logit models) with latent variables like values or attitudes leads to a more realistic representation of the choice process taking place in the consumer's "black box" and thus should provide greater explanatory power (Ben-Akiva et al., 2002a; Walker & Ben-Akiva, 2002). These so called hybrid choice models represent a promising new class of models which merge classical choice models with the structural equation approach (SEM) for latent variables.

Although conceptually appealing, there are hardly any applications of hybrid choice models in the marketing field. The major reason for their lack of popularity is most likely the fact that estimation of these models is rather involved. None of the standard routines for traditional logit or probit models implemented in software packes like SAS, LIMDEP or STATA are suitable if latent variables are involved. Thus, researchers need to develop their own programs in order to apply hybrid choice models or, alternatively, have to follow a sequential procedure. The latter consists of estimating scores for the latent factors in the first step using SEM software like LISREL, EQS or AMOS. In the second step, factor scores enter traditional choice models as observed variables. Unfortunately, the sequential approach suffers from not providing fully efficient estimators (Morikawa et al., 2002) and may lead to inconsistent results (Ben-Akiva et al., 2002a). In this paper, we present another option for dealing with hybrid choice models which has not been considered so far. The proposed approach makes use of the program Mplus (Muthén & Muthén, 2006), one of the most comprehensive software packages for SEM. Of interest in our context is the fact that Mplus is able to handle ordinal observed data as dependent variables. This makes it fairly easy to specify choice models which simultaneously include observed as well as latent exogenous variables. In addition, estimation gains from the efficient programming of the routines implemented in Mplus. Compared to customized software written in programming languages like R, GAUSS or MATLAB this leads to enormous advantages in terms of the running time needed for the estimation of hybrid choice models. So far, however, the approach is restricted to binary choice variables.

The remaining part of the paper is structured as follows. First, the general structure of hybrid choice models is presented. Second, an empirical study on travel mode choice illustrates the applicability of Mplus to estimate binary hybrid choice models. Finally, we conclude by summing up the main findings of our study and by providing some avenues for further research.

2. General Specification of Hybrid Choice Models

We follow Walker and Ben-Akiva (2002) in describing the hybrid choice model. As Fig. 1 depicts, the integrated model consists of two different parts: a discrete choice and a latent variable model, each defined by one structural equation and one measurement equation.
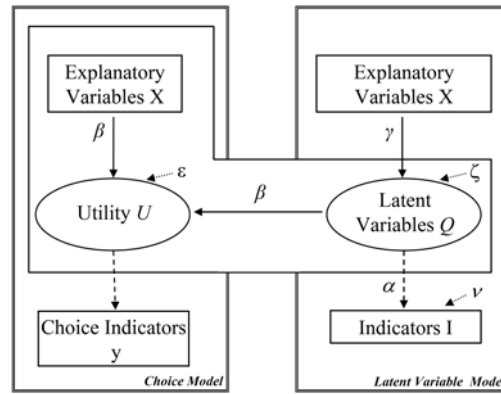

Fig. 1. Integrated Choice and Latent Variables Model

While the left hand side in Fig. 1 describes a traditional choice model (e.g., logit or probit), the right hand side represents a structural equation model for latent variables. Although in principle the latent variable part could be estimated first by SEM software and factor scores could than be used as exogenous variables in the estimation of the choice model, such an approach has been shown to be inefficient and may therefore lead to insignificant parameter estimates (Morikawa et al., 2002). Thus, both parts of the hybrid choice model should be estimated simultaneously. This approach will now be described in more detail.

Structural Equations: For the latent variable part of the model the structural equation becomes:[1]

$$Q = h(X; \gamma) + \zeta, \tag{1}$$

where Q are latent constructs, $h(\cdot)$ is a function of explanatory variables X as well as the unknown path coefficients γ. ζ represents random error. Explanatory variables can be both observed or latent. The structural equation of the choice model is given by the random utility function:

$$U = V(X, Q; \beta) + \varepsilon, \tag{2}$$

which can be decomposed into a deterministic part $V(\cdot)$ and a random part ε. As usual, the latter represents all influences not observed by the researcher. In contrast to traditional choice models, the unobserved variables Q in equation (1) enter the deterministic part of the utility function.

Measurement Equations: In the latent variable model each observed indicator I is explained by

$$I = g(Q; \alpha) + v, \tag{3}$$

where α are factor loadings and v are measurement errors. For the choice model, the agent's decision is expressed as a function of the utility a certain alternative j provides. Assuming utility maximizing behavior, the measurement equation for the observed choices becomes (Ben-Akiva et al., 2002b):[2]

$$y = \begin{cases} 1 & \text{if} \quad U = \max_{j}(U_j) \\ 0 & \text{else} \end{cases}. \tag{4}$$

Note that so far we have not defined concrete specifications for the functions $h(\cdot)$, $g(\cdot)$, and $V(\cdot)$. In the following, we assume that these functions are linear in parameters.

---

[1] As already mentioned, our notation follows Walker & Ben-Akiva (2002) and therefore partly deviates from the notation typically used in the SEM literature (e.g., Bollen 1989).

[2] The observed choice y can be interpeted as an error-free indicator of the "latent" utility.

Likelihood Function: Estimation of the integrated model requires the formulation of a common likelihood function containing (a) the likelihood of the choice model $P_a(\cdot)$, (b) the distribution $f_b(\cdot)$ of the latent variables given the explanatory variables, and (c) the distribution $f_c(\cdot)$ of the indicators conditional on the values of the latent variables. For a binary logit model, for example, this leads to the well-known likelihood function (e.g., Franses & Paap, 2001)

$$P\left(y \mid X; \beta\right) = \frac{1}{1 + \exp\left\{-X\beta\right\}} . \tag{5}$$

If we add latent variables and their indicators, the likelihood function becomes more complex. Assuming for simplicity that all error components are independent from each other, the joint probability for the integrated model in its general form is given by

$$P\left(y, I \mid Q, X; \alpha, \beta, \gamma, \Sigma\right) = \int_Q P_a(y \mid X, Q; \beta, \Sigma_\varepsilon) f_c(I \mid X, Q; \alpha, \Sigma_\nu) f_b(Q \mid X; \gamma, \Sigma_\zeta) dQ . \tag{6}$$

For the special case of a binary choice, this equation can be expressed as follows:

$$P\left(y, I \mid Q, X; \alpha, \beta, \gamma, \Sigma\right) = \int_Q \left(\frac{1}{1 + \exp\left\{-(X\beta_1 + Q\beta_2)\right\}}\right) \cdot \prod_{n=1}^{N} \phi \left|\frac{I_n - Q\alpha_n}{\sigma_{\nu_n}}\right| \cdot \prod_{r=1}^{R} \phi \left|\frac{Q_r - X\gamma_r}{\sigma_{\zeta_r}}\right| dQ . \tag{7}$$

where ϕ denotes the probability density function of the standard normal. The total number of latent variables is given by R and the number of indicators is given by N. Maximizing this common likelihood over the latent variables Q yields estimates of the unknown parameters α, β, γ and the covariance matrices of the error components Σ.

Since the special case of a binary logit model is equivalent to the logistic function for ordered-categorical dependent variables (Train, 2003, p. 166), the corresponding maximum-likelihood estimator implemented in Mplus can be applied to derive values for the free parameters in equation (7).

3. Travel Mode Choice – An Application of the Hybrid Choice Approach

For our empirical application of hybrid choice models we chose travel mode choice because traditional discrete choice models have been extensively applied in this area (e. g., Ben-Akiva & Lerman, 1985). In that paradigm individual travel mode choice is modelled both as a function of individual characteristics of the decider such as income, employment status etc. and of attributes of the different travel mode choice alternatives such as travel time, travel cost, availability, etc. Behaviour is conceptualized as a function of these solely directly measurable variables. Researchers have criticized this approach and its underlying assumption of rational behaviour under complete information (e.g., Bamberg & Schmidt, 1994). Specifically, Bamberg and Schmidt (1994) have shown that many car drivers neither know the cost per driven kilometer nor do they possess adequate knowledge about prices, connections or schedules of public transportation alternatives. They conclude that variables that are not directly observable such as „Fahrspaß" or „Desire for Flexibility" might have an important impact on travel mode choice, but are neglected in traditional choice models (Bamberg & Schmidt 1994). Due to these shortcomings, researchers have turned to behavioral models such as the theory of planned behavior that incorporate latent variables such as attitudes, social norms and perceived behavioral control to explain travel mode choice (Bamberg, 1995; Bamberg & Ludemann, 1998). Results of their studies confirm the utility of explaining travel mode choice with the mentioned latent variables attitudes, social norms and perceived behavioral control. Thus the marriage of discrete choice models with latent variable approaches as realized in hybrid choice models offers great potential to enhance our understanding of travel mode choice.

Several studies have indicated that values might have an impact on travel mode choice (Bamberg, 1996; Bamberg & Kühnel, 1998). This proposition has been confirmed in a recent study by Collins and Chambers (2005). They could show that values impact on travel mode choice and that their impact is mediated by beliefs about environmental threats. Furthermore, they were able to show that the usually employed, directly observable, situational criteria such as time, cost and accessibility possess an additional impact on travel mode choice. Unfortunately, they estimated their model with the deficient two-step approach. We will build on their findings and incorporate values and choice criteria, as suggested by Bamberg and Schmidt (1994), next to directly observable situational factors in a hybrid choice model. Due

to space limitations we cannot provide more detail on the selection of values from Schwartz's (1999) framework and hypotheses concerning their impact on choice criteria.

3.1 Method

We analyze travel mode choice in a representative sample of German consumers between 14 and 75 years of age. 907 respondents that were recruited from a consumer panel of a major international market research provider participated in a telephone survey. Panellists were recruited based on a demographic quota sampling approach. The sample distribution of demographic variables such as social status, size of household etc. equals the population distribution. In addition to variables employed in traditional discrete choice models such as accessibilty, travel time etc. we surveyed choice criteria and the motivational orientations of respondents. Choice criteria were developed in an initial qualitative study through focus groups that were supplemented with repertory grid interviews and include constructs such as privacy, time, flexibility etc. Respondents had to indicate the importance of these choice criteria on five-point rating scales ranging from completely unimportant to completely important. In order to measure respondents' motivational orientation we employ the PQ-Questionnaire from Schwartz et al. (1999). Respondents had to indicate their similarity to 40 persons descriptions (portraits) on six-point rating scales ranging from very unalike to very much alike.

3.2 Results

Detailed information on respondents' trip mode choice (e.g., car, bus, underground, bicycle) for daily trips to work, education or shopping as well as short travels were available. We analyzed choice of the car versus alternative transportation means for daily trips. We started with a traditional binary logit model including several observed variables deemed to be relevant for explaining the respondents' choice: (1) age, (2) number of children below the age of ten, (3) number of cars in the household, (4) ownership of a railcard (in Germany, BahnCard owners can travel half-price throughout the country for one year), (5) distance to next bus stop, and finally (6) size of the respondents' place of residence. Although additional information on the distance to various means of public transport, the travel time needed to get to work/shopping by public transportation or car, as well as monthly income was available, we could not include these variables in the model because of a substantial number of missing values. The large number of missing values on these variables support Bamberg and

Schmidt's (1994) contention that many car drivers do not possess sufficient information about public transport alternatives.

Table 1
Traditional binary logit model of transportation mode choice

Explanatory variable	Coefficient	*t*-statistic
Age	0.027	3.725
# of children < 10 years	0.132	0.649
# of cars in household	1.787	8.878
Railcard holder (dummy)	–2.227	–3.351
Distance to next bus stop	0.001	0.710
Size of place of residence	–0.199	–1.995
Number of observations	413	

Mode: 1 = other transportation mean, 2 = car

The baseline logit model has been estimated using the maximum likelihood estimator implemented in Mplus 4.2 (Muthén & Muthén, 1998-2006). The estimated parameters possess face-validity (see Table 1). Respondents' age has a significant positive effect on choosing the car for trips to work or shopping. Since age can be considered to be a proxy for various factors not included in the model (e.g., income) such a result seems plausible. By far the strongest influence emerges for the number of cars the household owns. More cars imply greater availability and thereby greater probability of usage for daily trips. As can be expected, railcard ownership reduces the probability to use a car. Distance to the next bus stop shows no significant effect. A reason for that might be that the variable is only a weak proxy for how easy it is to get to work by public transport in general (e.g., by underground, city train). In contrast, size of the respondents' residence decreases the propensity to use a car. Larger towns or cities typically offer a well developed public transport system thus increasing the propensity to use it.

The analyzed observed variables are to some extent able to predict transport mode choice. However, important latent explanatory variables might be missing. A hybrid choice model was therefore specified to include individual choice criteria (flexibility, possession, passivity, and environment protection) as well as respondents' values (power, hedonism, and security) from Schwartz's (1999) framework. Choice criteria and values (except environment protection) have been measured by multiple indicators. Since basic values are known to influence observed behaviour only indirectly (e.g., McCarthy & Shrum, 1994), we specified a

hierarchical model: Respondents' values determine their choice criteria which in turn influence their trip mode choice.

Table 2
Hybrid binary logit model of transportation mode choice

Dependent variable	Explanatory variable	Coefficient	*t*-statistic
Trip mode choice	Flexibility	0.865	2.472
	Possession	1.312	2.876
	Passivity	–0.260	–0.652
	Environment protection	–0.428	–2.338
	Age	0.030	3.101
	# of children < 10 years	0.256	1.058
	# of cars in household	1.973	7.273
	Railcard holder (dummy)	–2.105	–2.839
	Distance to next bus stop	0.001	0.818
	Size of place of residence	–0.240	–1.960
Flexibility	Power	0.125	2.846
	Hedonism	0.177	2.674
	Security	0.121	1.633
Possession	Power	0.074	1.079
	Hedonism	0.267	2.788
	Security	0.354	3.218
Passivity	Security	0.595	5.841
Environment protection	Power	–0.070	–1.384
	Hedonism	0.076	1.054
Number of observations	405		

Parameter estimates for the hybrid choice model (see Table 2) show that three out of four choice criteria have a strong influence on trip mode choice. The more important it is for a respondent that he possesses the transport mode and has it at its own disposal (flexibility), the higher the probability to use the car. In contrast, environmental concerns induce the respondents to choose public transport or any of the other alternative trip modes. Passivity has a negative but insignificant effect. For the observed variables only minor changes in the parameters estimates occur compared to the traditional logit model. Parameter estimates further reveal that respondents' choice criteria are determined by the underlying value orientation. Flexibility is driven by striving for pleasure (hedonism) as well as power. Respondent's for whom hedonism and security are particularly salient values put a higher

relevance on possessing the transportation mean. The extent to which passivity is considered important is likewise determined by the security value. Environmental concerns do not seem to be influenced by the value orientation considered in our model. Since comparing the traditional logit model with our hybrid choice model is rather difficult because of the relationships specified among the latent variables, we performed the following chi-square difference test: The parameters for the influence of the latent choice criteria on trip mode choice have been fixed to zero in the constrained hybrid choice model. Estimation of this model leads to a significant increase in chi-square by 41.56 (df = 4, $p < .001$). In addition to the diagnostic insight generated by such models through uncovering the motivational sources of behavior, the substantial increase in explanatory power clearly underscore their value.

4. Conclusion

From a substantial point of view, hybrid choice models can be considered one of the most interesting advances in discrete choice modeling during the last decade. However, estimation of such models is rather difficult since standard software is not suitable if latent variables are involved. This issue seems to prevent the diffusion of hybrid choice models in the marketing field. In this paper we have shown that at least for the special case of a binary choice situation the SEM software Mplus can be used to estimate hybrid choice models. This offers the researcher an extreme amount of flexibility in the specification of his/her models. The empirical application on travel mode choice data has shown that choice criteria included as latent variables are important factors in respondents' decisions. In our application including latent variables provided valuable insight into the causes of choice behavior and additional explanatory power**.**

References

Bamberg, S. & Schmidt, P. (1994). Auto oder Fahrrad? Empirischer Test einer Handlungstheorie zur Erklärung der Verkehrsmittelwahl. *Kölner Zeitschrift für Soziologie und Sozialforschung*, (46)1, 80-102.

Bamberg, S. & Ludemann, Ch. (1996). Eine Überprüfung der Theorie des geplanten Verhaltens in zwei Wahlsituationen: Rad versus PKW und Container versus Hausmüll. *Zeitschrift für Sozialpsychologie*, 27(1), 32-46.

Bamberg, S. (1995). Wie bekommt man den/die Autonutzer/in in den Bus? Probleme und Ergebnisse einer Anwendung der Theorie des geplanten Verhaltens im Kontext praktischer Verkehrsplanungsfragestellungen. *Zeitschrift für Sozialpsychologie*, 26(4), 243-262.

Bamberg, S. (1996a). Allgemeine oder spezifische Einstellungen bei der Erklärung umweltschonenden Verhaltens?. *Zeitschrift für Sozialpsychologie*, 27(1), 47-60.

Bamberg, S. & Kühnel, St. (1998). Überzeugungssysteme in einem zweistufigen Modell rationaler Handlungen. *Zeitschrift für Soziologie*, 27(4), 256-270.

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

Ben-Akiva, M., Bradley, M., Morikawa, T., Benjamin, J., Novak, T., Oppewal, H., & Rao, V. (1994). Combining Revealed and Stated Preferences Data. *Marketing Letters*, *5*(4), 335-349.

Ben-Akiva, M., McFadden, D., Train, K., Walker, J., Bhat, C., Bierlaire, M., Bolduc, D., Boersch-Supan, A., Brownstone, D., Bunch, D. S., Daly, A., De Palma, A., Gopinath, D., Karlstrom, A., & Munizaga, M. A. (2002). Hybrid Choice Models: Progress and Challenges. *Marketing Letters*, *13*(3), 163-175. (a)

Ben-Akiva, M., Walker, J., Bernardino, A., Gopinath, D., Morikawa, T., Polydoropoulou, T. (2002). Integration of Choice and Latent Variable Models. In H. Mahmassani (Ed.), *In Perpetual Motion: Travel Behaviour Research Opportunities and Application Challenges* (pp. 431-470). Elsevier. (b)

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York: Wiley.

Collins, C. M. & Chambers, S. M. (2005). Psychological and Situational Influences on Commuter-Transport-Mode-Choice. Environment and Behavior, 37(5), 640-661.

Franses, P. H. & Paap, R. (2001). *Quantitative Models in Marketing Research*. Cambridge: Cambridge University Press.

Morikawa, T., Ben-Akiva, M., & McFadden, D. L. (2002). Discrete Choice Models Incorporating Revealed Preferences and Psychometric Data. In P. Franses & A. Montgomery (Eds.), *Econometric Models in Marketing* (Vol. 16, pp. 29-55). Elsevier.

Muthén, L. K. & Muthén, B.O. (1998-2006). Mplus User's Guide. 4th ed., Los Angeles, CA: Muthén & Muthén.

Schwartz, Sh. H., Lehmann, A. & Roccas, S. (1999). Multimethod Probes of Basic Human Values. In J. Adamopolos & Y. Kashima (Eds.), *Social psychology and cultural context: Essays in Honour of Harry C. Triandis*, Newbury Park, CA: Sage, pp. 107-123.

Train, K. E. (2003). *Discrete Choice Methods with Simulation*. Cambridge: Cambridge University Press.

Walker, J. & Ben-Akiva, M. (2002). Generalized Random Utility Model. *Mathematical Social Sciences*, *43*(2), 303-343.