IDENTIFICATION OF DISCRETE CHOICE MODELS WITH LATENT VARIABLES: THE ROLE OF CONSTRAINTS IN THE ESTIMATION

Sebastián Raveau, Pontificia Universidad Católica de Chile, sraveau@puc.cl

María Francisca Yáñez, Pontificia Universidad Católica de Chile, mfyanezc@puc.cl

Juan de Dios Ortúzar, Pontificia Universidad Católica de Chile, jos@ing.puc.cl

ABSTRACT

The formulation of hybrid discrete choice models, including both observable alternative attributes and latent variables associated with attitudes and perceptions has become a topic of discussion in recent years. Even though there have been developments related to model estimation the *parameter identifiability* issue has not been treated yet, but it is known that as the model estimates are not unique, it is necessary to impose some constraints in the estimation. In this paper we analyse the impact of different normalization criteria on the recovery of model parameters in a simulated environment. We identify advantages and disadvantages related to different normalizations, especially when arbitrary values are used on the constraints.

Keywords: Hybrid discrete choice models, Latent variables, Identifiability

INTRODUCTION

The inclusion of subjective elements in discrete choice models has re-emerged as an analysis and discussion topic, after losing some of the importance that made it an interesting subject in the 1980s (see for example, Ortúzar & Hutt, 1984). In the last decade hybrid choice models have been proposed considering not only tangible attributes of the alternatives (classic explanatory variables) but also more intangible elements associated with user perceptions and attitudes, expressed through latent variables (McFadden, 1986; Ashok *et al.*, 2002). To estimate these hybrid models, the most popular estimation method is the *sequential approach* where the latent variables are constructed before entering the discrete choice model as further regular variables (Ashok *et al.*, 2002; Vredin Johansson *et al.*, 2005).

Previous studies (Raveau *et al.*, 2010) have shown that when a hybrid discrete choice model is estimated sequentially, the estimators are unbiased.

One of the main characteristics of latent variable models is that the estimators are not unique. Therefore, a subset of parameter values must be constrained to allow estimation. While this identifiability issue has been empirically treated, general necessary and sufficient conditions have not been developed. In fact, the identification hybrid discrete choice model parameters must be analyzed case-by-case (Ben-Akiva *et al*., 2002), generally based separately on latent variable model identifiability rules (Bollen, 1989) and discrete choice model identifiability rules (Ben-Akiva & Lerman, 1985).

In this paper we analyse the impact of different normalization criteria on the capability of models to recover "true parameters" on a hybrid discrete choice modelling context. In particular, we analyze the performance of the sequential estimation method. In our analysis, we consider model efficiency and goodness-of-fit, as well as forecasting capacity. For this, synthetic datasets were generated. Our controlled experiment allowed us to test the convenience of choosing different normalizations, and to provide recommendations for latent variable modelling.

The rest of the paper is organised as follows. Section 2 briefly describes the theoretical framework for modelling with latent variables and their inclusion in discrete choice models. Section 3 reports and discusses the results obtained from the application of hybrid choice models in a controlled synthetic experiment, analyzing the effects of constraints in the estimation. Finally, section 4 presents our main conclusions.

HYBRID DISCRETE CHOICE MODELS

Latent variables are factors that, although influencing individual behaviour and perceptions, cannot be quantified in practice. This is because of their intangibility, as these variables do not have a measurement scale, or their intrinsic subjectivity, because different people may perceive them differently. Identification of latent variables requires supplementing a standard preference survey with questions that capture users' perceptions about some aspects of the alternatives (and choice context). The answers to these questions generate perception indicators that serve for identifying and measure the latent variables.

In the recent literature, the most popular approach to include the effect of latent variables relies on a MIMIC (Multiple Indicator Multiple Cause) model (Bollen, 1989), where the latent variables (η_{ilq}) are explained by characteristics s_{iqr} from the users and from the alternatives through *structural equations* as (2.1); at the same time, the latent variables explain the perception indicators (y_{ipq}), which are observed by the modeller from the survey, through *measurement equations* as (2.2):

$$\eta_{ilq} = \sum_r \alpha_{ilr} \cdot s_{iqr} + \nu_{ilq} \tag{2.1}$$

$$y_{ipq} = \sum_l \gamma_{ilp} \cdot \eta_{ilq} + \zeta_{ipq} \tag{2.2}$$

where the index i refers to an alternative, q to an individual, l to a latent variable, r to an explanatory variable and p to an indicator; α_{ilr} and γ_{ilp} are parameters to be estimated, while ν_{ilq} and ζ_{ipq} are error terms with mean zero and a certain covariance matrix. As the η_{ilq} terms are unknown, both equations must be considered jointly in the parameter estimation process.

Traditionally, in discrete choice modelling it is assumed that people are rational decision makers maximising their perceived utility U_{iq} over the alternatives; the modeller, who is an observer, defines a representative utility V_{iq} and (as he does not have perfect information) an error term ε_{iq} associated with each alternative (Ortúzar and Willumsen, 2001), such that:

$$U_{iq} = V_{iq} + \varepsilon_{iq} \tag{2.3}$$

The representative utility V_{iq} is a function of the objective attributes X_{ikq} (i.e. travel times and fare, as well as socioeconomic characteristics of the individual); if latent variables are also included, a utility function as (2.4) is postulated, where θ_{ik} and β_{il} are parameters to be estimated associated with the tangible attributes and the latent variables, respectively.

$$V_{iq} = \sum_k \theta_{ik} \cdot X_{ikq} + \sum_l \beta_{il} \cdot \eta_{ilq} \tag{2.4}$$

Finally, to characterise the individuals' decisions over their set of available alternatives (defined as A_q), binary variables d_{iq}, that take values according to equation (2.5), are defined:

$$d_{iq} = \begin{cases} 1 & \text{if} \quad U_{iq} \geq U_{jq}, \quad \forall j \in A_q \\ 0 & \text{in other case} \end{cases} \tag{2.5}$$

Using the *sequential estimation* method, the problem is treated in two stages, separating the MIMIC model and discrete choice model interactions. First, the MIMIC model is solved to obtain parameter estimators for the equations relating the latent variables with the explanatory variables and perception indicators. Using these parameters in the structural equation (2.1), expected values for the latent variables of each individual and alternative are obtained. Later, the latent variables are added to the set of variables of the discrete choice model and their parameters can be estimated together with those of the traditional variables in a second stage.

To allow the estimation of the MIMIC model, a subset of parameter values must be constrained. Generally, as many parameters as latent variables must be fixed, but more constraints could be necessary depending on the structure of the relationships. Stapleton

(1978) recommends fixing as many measurement equations parameters (y_{ipq}) as needed, but also shows that the model can be estimated if the variances of the structural equations (v_{ilq}) are fixed. Constraining the parameters of the measurement equations is a more popular approach (Ben-Akiva *et al*., 2002; Vredin Johansson *et al.*, 2005) than constraining the variances of the structural equations (Raveau *et al.*, 2010), even though there is no evidence of advantages in either case.

IDENTIFIABILITY ANALYSIS

To study the effect of constraints on the estimation of hybrid discrete choice models in a context free of undesired unknown effects, synthetic datasets were generated to test the empirical performance of the estimation methods. Following Williams and Ortúzar (1982), a collection of 10 datasets of 3,000 pseudo-observed individuals, who behaved according to a predetermined choice process, were simulated. The samples were generated with three alternatives i = {*1,2,3*} described by two generic attributes: *time* and *cost*. The systematic utility function was built assuming an incremental specification, linear on both the attributes and the latent variables, considering also specific constants associated with alternatives 2 and 3. Latent variables, identified as *safety* and *reliability,* were considered; the former was present in all alternatives, while the later was only introduced in alternatives 1 and 2. Thus, for each individual q we had:

$$\begin{aligned}
V_{1q} &= \theta_{Cost} \cdot Cost_{1q} + \theta_{Time} \cdot Time_{1q} + \beta_{Safety} \cdot Safety_{1q} + \beta_{Reliability} \cdot Reliability_{1q} \\
V_{2q} &= \theta_{Cost} \cdot Cost_{2q} + \theta_{Time} \cdot Time_{2q} + \beta_{Safety} \cdot Safety_{2q} + \beta_{Reliability} \cdot Reliability_{2q} + ASC_2 \\
V_{2q} &= \theta_{Cost} \cdot Cost_{3q} + \theta_{Time} \cdot Time_{3q} + \beta_{Safety} \cdot Safety_{3q} + ASC_3
\end{aligned} \tag{2.6}$$

All relevant attribute differences ($Cost_{2q}$ – $Cost_{1q}$, $Cost_{3q}$ – $Cost_{1q}$, $Time_{2q}$ – $Time_{1q}$ and $Time_{3q}$ – $Time_{1q}$) were built taking random draws from independent truncated Normal distribution functions with arbitrary lower and upper bounds (Table I). The error terms distributed identically and independently Gumbel (0,1) leading to a multinomial logit (MNL) kernel for the discrete choice model (conditional on the unobserved latent variables). By setting the Gumbel scale factor to 1, we were able to directly analyse the estimated parameters without scale concerns.

Table I – Discrete choice model variables

Difference	Mean	Standard Error	Range
$Cost_2 - Cost_1$	- 2.70	0.70	[- 4.0 ; - 1.0]
$Cost_3 - Cost_1$	- 2.50	0.80	[- 4.0 ; - 1.0]
$Time_2 - Time_1$	2.85	2.80	[1.0 ; 6.0]
$Time_3 - Time_1$	2.70	2.70	[1.0 ; 4.0]

Regarding the latent variables model, we had a set of structural equations where *safety* and *reliability* were explained by four socio-economic variables: *income* (continuous variable), *age* (dummy variable that indicates if the individual is older than 30 years), *children* (dummy variable that indicates if the individual has any child) and *education level* (dummy variable that indicates if the individual has higher education studies) according to (2.7), where the random terms $v_{Saf,iq}$ and $v_{Rel,iq}$ were distributed Normal.

$$\begin{aligned} Safety_{jq} &= \alpha_{Inc\text{-}Saf,i} \cdot Income_q + \alpha_{Age\text{-}Saf} \cdot Age + \alpha_{Chi\text{-}Saf} \cdot Children + v_{Saf,iq} \\ Reliability_{jq} &= \alpha_{Inc\text{-}Rel,i} \cdot Income_q + \alpha_{Edu\text{-}Rel,i} \cdot Education_q + v_{Rel,iq} \end{aligned} \quad (2.7)$$

The socio-economic explanatory variables were generated according to Table II.

Table II – Structural equations variables

Attribute	Distribution	Mean	Standard Error	Range
Income	Truncated Normal	1.00	0.50	[0.1 ; 2.0]
Age	Bernoulli	0.50	0.50	[0.0 ; 1.0]
Children	Bernoulli	0.50	0.50	[0.0 ; 1.0]
Education	Bernoulli	0.50	0.50	[0.0 ; 1.0]

As the latent variables are unobservable we needed to include measurement equations. Thus, we assumed perception indicators for every latent variable, defined by (2.8), where the error terms $\zeta_{Com,iq}$ and $\zeta_{Rel,iq}$ were distributed Normal.

$$\begin{aligned} y_{Accidents,iq} &= \gamma_{Acc\text{-}Saf} \cdot Safety_{iq} + \zeta_{Acc,iq} \\ y_{Wait,iq} &= \gamma_{Wai\text{-}Saf} \cdot Safety_{iq} + \gamma_{Wai\text{-}Rel} \cdot Reliability_{iq} + \zeta_{Wai,iq} \\ y_{Travel,iq} &= \gamma_{Tra\text{-}Rel} \cdot Reliability_{iq} + \zeta_{Tra,iq} \end{aligned} \quad (2.8)$$

To obtain simulated data reflecting the virtual choice process, we considered fixed taste parameters as shown in (2.6). Using these taste parameters, the attribute values and the individual characteristics, we were able to construct the individual latent variables as well as the utility function for each individual. Then, according to a discrete maximisation exercise, we obtained the chosen alternative for each individual.

Model Constraining

For the MIMIC model four different normalizations were tested fixing five parameters (one for each latent variable) each time. In the first case, the variances of the structural equations were fixed using their real (simulated) values. In the second case, the structural equation variances were fixed using arbitrary values. In the third case, five measurement equation parameters were fixed using their real (simulated) value, and in the fourth case these parameters were fixed using arbitrary values. The results reported below correspond to the mean estimates of the 10 repetitions of the sub-sampling process.

The “true” parameter values for the structural equations of the MIMIC model used in the simulation, are presented in Table III; the estimated parameters (presented as a percentage of the target values) and their respective t-tests, are also shown in each case. As expected, better results are obtained when using the real values on the constraints (i.e. parameters closer to 100% of their targets). It would appear that when using real values for constraining, fixing the measurement equation parameters is better than fixing the structural equation variances (see for example $\alpha_{Age\text{-}Saf,3}$ and $\alpha_{Chi\text{-}Saf,3}$). On the other hand, when using arbitrary values for constraining, fixing the structural equation variances results on more accurate estimates than fixing the measurement equation parameters. As the data and the degrees-of-freedom are the same, the statistical significance of all parameters does not change when accurate or arbitrary values are used.

Table III – Structural equations estimates

Parameter	Target	Right Variances		Wrong Variances		Right Parameters		Wrong Parameters	
$\alpha_{Edu\text{-}Rel,1}$	1.00	101.0%	(13.9)	81.6%	(13.9)	99.9%	(20.1)	99.9%	(20.1)
$\alpha_{Inc\text{-}Rel,1}$	2.00	101.5%	(16.9)	81.9%	(16.9)	100.5%	(34.6)	100.5%	(34.6)
$\alpha_{Edu\text{-}Rel,2}$	0.70	99.4%	(10.9)	68.2%	(10.9)	99.6%	(10.9)	79.7%	(10.9)
$\alpha_{Inc\text{-}Rel,2}$	2.50	100.2%	(27.1)	68.8%	(27.1)	100.6%	(31.1)	80.5%	(31.1)
$\alpha_{Inc\text{-}Saf,1}$	0.90	108.8%	(17.4)	112.5%	(17.4)	104.8%	(17.5)	41.9%	(17.5)
$\alpha_{Age\text{-}Saf,1}$	1.50	102.9%	(25.1)	106.4%	(25.1)	99.2%	(25.9)	39.7%	(25.9)
$\alpha_{Chi\text{-}Saf,1}$	2.00	103.9%	(28.0)	107.4%	(28.0)	100.2%	(29.1)	40.1%	(29.1)
$\alpha_{Inc\text{-}Saf,2}$	0.70	100.9%	(9.1)	106.9%	(9.1)	96.3%	(9.2)	89.9%	(9.2)
$\alpha_{Age\text{-}Saf,2}$	1.80	100.8%	(24.6)	106.7%	(24.6)	96.1%	(33.7)	89.7%	(33.7)
$\alpha_{Chi\text{-}Saf,2}$	1.50	99.5%	(22.3)	105.4%	(22.3)	94.8%	(28.5)	88.5%	(28.5)
$\alpha_{Inc\text{-}Saf,3}$	0.60	81.1%	(16.9)	173.4%	(16.9)	103.5%	(19.4)	103.5%	(19.4)
$\alpha_{Age\text{-}Saf,3}$	1.20	77.5%	(26.9)	165.7%	(26.9)	98.9%	(42.6)	98.9%	(42.6)
$\alpha_{Chi\text{-}Saf,3}$	1.10	76.1%	(25.8)	162.6%	(25.8)	97.2%	(38.5)	97.2%	(38.5)
$v_{Rel,1}$	1.53	100.0%	**	65.2%	**	98.5%	(9.5)	98.5%	(9.5)
$v_{Rel,2}$	2.12	100.0%	**	47.1%	**	101.0%	(18.9)	64.7%	(18.9)
$v_{Saf,1}$	0.94	100.0%	**	106.8%	**	92.9%	(16.2)	14.9%	(16.2)
$v_{Saf,2}$	0.89	100.0%	**	112.1%	**	91.1%	(14.9)	79.4%	(14.9)
$v_{Saf,3}$	0.22	100.0%	**	456.6%	**	163.3%	(16.3)	163.3%	(16.3)

Constraints have significant implications on the estimation of the structural equations, either overestimating parameters (e.g. $\alpha_{Inc\text{-}Saf,3}$ when fixing variances) or underestimating them (e.g. $\alpha_{Age\text{-}Saf,1}$ when fixing parameters). As the latent variables are constructed from the estimates using (2.1), potential bias product of a given normalization could have repercussions on the discrete choice model as introducing a latent variable with measurement error could result in inconsistent estimators (Ben-Akiva *et al.*, 2002).

For each normalization case, Table IV presents the “true” parameter values for the measurement equations of the MIMIC model, the estimated parameters (again presented as a percentage of the target values) and their respective t-tests. As the structural and measurement equations of the MIMIC model are estimated jointly, the results presented in

Table IV have the same general patterns as those presented in Table III. When using the real values on the constraints, more accurate estimates are obtained. Again, it would appear that when using real values for constraining, fixing the measurement equation parameters is better, but when using arbitrary values fixing the structural equation variances is preferable.

Table IV – Measurement equations estimates

Parameter	Target	Right Variances		Wrong Variances		Right Parameters		Wrong Parameters	
$\gamma_{Tra\text{-}Rel,1}$	1.00	99.2%	(19.0)	122.8%	(19.0)	100.0%	**	100.0%	**
$\gamma_{Wai\text{-}Rel,1}$	0.30	102.1%	(19.0)	126.5%	(19.0)	103.2%	(13.0)	103.2%	(13.0)
$\gamma_{Tra\text{-}Rel,2}$	0.80	100.5%	(37.8)	146.4%	(37.8)	100.0%	**	125.0%	**
$\gamma_{Wai\text{-}Rel,2}$	1.20	100.1%	(37.8)	145.8%	(37.8)	99.7%	(26.7)	124.6%	(26.7)
$\gamma_{Wai\text{-}Saf,1}$	0.60	96.4%	(32.4)	93.2%	(32.4)	100.0%	**	250.0%	**
$\gamma_{Acc\text{-}Saf,1}$	2.00	95.6%	(32.4)	92.5%	(32.4)	99.3%	(33.0)	248.1%	(33.0)
$\gamma_{Wai\text{-}Saf,2}$	1.40	95.4%	(29.7)	90.1%	(29.7)	100.0%	**	107.1%	**
$\gamma_{Acc\text{-}Saf,2}$	0.90	97.7%	(29.7)	92.2%	(29.7)	102.3%	(32.9)	109.6%	(32.9)
$\gamma_{Wai\text{-}Saf,3}$	1.50	127.7%	(32.7)	59.8%	(32.7)	100.0%	**	100.0%	**
$\gamma_{Acc\text{-}Saf,3}$	0.75	128.8%	(32.7)	60.3%	(32.7)	100.9%	(36.6)	100.9%	(36.6)
$\zeta_{Tra,1}$	0.28	126.4%	(2.4)	126.4%	(2.4)	126.4%	(2.4)	126.4%	(2.4)
$\zeta_{Tra,2}$	0.98	100.0%	(14.4)	100.0%	(14.4)	100.0%	(14.4)	100.0%	(14.4)
$\zeta_{Wai,1}$	1.48	97.7%	(33.5)	97.7%	(33.5)	97.7%	(33.5)	97.7%	(33.5)
$\zeta_{Wai,2}$	0.72	200.9%	(8.8)	200.9%	(8.8)	200.9%	(8.8)	200.9%	(8.8)
$\zeta_{Wai,3}$	0.56	97.3%	(11.7)	97.3%	(11.7)	97.3%	(11.7)	97.3%	(11.7)
$\zeta_{Acc,1}$	1.12	106.1%	(6.1)	106.1%	(6.1)	106.1%	(6.1)	106.1%	(6.1)
$\zeta_{Acc,2}$	1.83	98.0%	(28.5)	98.0%	(28.5)	98.0%	(28.5)	98.0%	(28.5)
$\zeta_{Acc,3}$	0.76	101.3%	(33.6)	101.3%	(33.6)	101.3%	(33.6)	101.3%	(33.6)

This way, if previous knowledge about the magnitude of the parameters was available, it would be recommended to normalize the measurement equation parameters. Contrariwise, if the parameters are completely unknown, it would be recommended to normalize the structural equation variances, as the bias risk is lower. On the other hand, as the data and degrees-of-freedom are the same, the statistical significance of all parameters in each normalization case does not change when less or more precise values are used for the constraints; also, the measurement equation variances are insensitive to the normalization used. The estimates of the measurement equations are directly related to the estimates of the structural equations. Note, for example, that when constraining variances with arbitrary values all the structural equation parameters related to reliability are underestimated (Table III), and so all the measurement equation parameters related to reliability are overestimated (Table IV) to compensate.

Latent variables were constructed from the parameters shown in Table III, using (2.1), for the different constraining cases. The “true” parameter values simulated for the discrete choice model, the estimated parameters (presented as a percentage of the target values), their respective t-tests and the discrete choice model log-likelihood at convergence, are presented in Table V. The statistical significance of the parameters for *time* and *cost* (tangible attributes) are always high, and their point estimates are always close to the targets. The same could be said about the specific constant for alternative 2; however, the specific constant for alternative 3 presents systematic overestimation in all cases. This way it would appear that the traditional attributes (tangible variables and specific constants) of the discrete choice model are not affected by the normalization used on the MIMIC model.

The latent variable parameters are close to the simulated ones when real values are used for the constraints; once again the most accurate results are obtained when the measurement equation parameters are normalized. When arbitrary values are used for the constraints, constraining the structural equation variances results in better results; wrongly normalizing the measurement equation parameters results on a wrong sign for the safety parameter (even though it is not statistically significant).

Table V – Discrete choice model estimates

Parameter	**Target**	**Right Variances**		**Wrong Variances**		**Right Parameters**		**Wrong Parameters**	
θ_{Cost}	- 2.16	98.2%	(30.1)	98.1%	(30.1)	98.2%	(30.1)	98.1%	(30.1)
θ_{Time}	- 1.23	99.2%	(27.0)	99.1%	(27.1)	99.2%	(27.1)	99.3%	(27.1)
$\beta_{Reliability}$	0.20	95.7%	(3.8)	156.0%	(4.4)	100.6%	(4.2)	122.4%	(4.1)
β_{Safety}	0.15	71.6%	(1.5)	106.4%	(1.1)	108.2%	(1.6)	- 3.4%	(0.1)
ASC_2	- 1.70	99.1%	(8.2)	92.4%	(7.3)	98.7%	(7.8)	96.1%	(7.2)
ASC_3	- 0.80	128.8%	(4.6)	133.5%	(4.9)	129.1%	(4.6)	136.6%	(4.7)

Table VI presents the goodness-of-fit of the MIMIC and discrete choice models when the different normalizations are used. As the data and degrees-of-freedom are the same in all cases, the MIMIC model's goodness-of-fit is not affected. On the other hand, even though slightly affected, the goodness-of-fit of the discrete choice model has no significant variation in all the constraining cases. This way, it seems clear that goodness-of-fit is not a criterion for choosing one normalization approach over another.

Table VI – Models' log-likelihoods

Model	**Right Variances**	**Wrong Variances**	**Right Parameters**	**Wrong Parameters**
MIMIC	- 51,753.6	- 51,753.6	- 51,753.6	- 51,753.6
Discrete Choice	- 2,134.5	- 2,134.3	- 2,134.0	- 2,134.3

Given some of the differences obtained between the "true" values of parameters and the estimated values when arbitrary constraining the MIMIC model, serious problems may arise if valuation functions are sought (problems not only in magnitude, but also in sign). This may produce problems when the models are intended to be the base for forecasting and/or transport policy assessment. Therefore, it is clear that constraining must be taken into account when using hybrid choice models in practice.

CONCLUSIONS

Hybrid choice models permit the inclusion of latent variables associated with perceptions and attitudes that are usually not considered by standard random utility choice models. Using simulated datasets, we analysed the effects of various normalization criteria on model estimation; for this, different normalization were considered in the latent variables model.

Results show that the MIMIC model estimators are sensitive to the normalization criteria. When arbitrary values are used to fix estimators (either the structural equation variances or the measurement equation parameters), problems arise in the form of biased parameters although the models' goodness-of-fit is not affected. These problems occur in both the structural and measurement equations. This is a serious drawback, as the structural equations are used to generate the latent variables that are eventually introduced in the discrete choice model.

The classical attributes of the MNL model (i.e. time, cost and specific constants) appear to be robust to the normalization used, but the latent variable parameters are highly affected. Even sign problems arise when arbitrary values are used to normalize some parameters. This could have severe consequences when trying to obtain attribute valuations, producing forecasts or evaluating transport policies. Nevertheless, given the robustness of the estimates of the traditional attributes, the uncertainty over which is the preferred normalization should not be a reason for not working with latent variables. In fact, the goodness-of-fit of the model shows no significant variation over the constraining cases.

Our results show that when "real" values are used for the constrains fixing measurement equation parameters results on better estimators than fixing the structural equation variances; nevertheless, in a real study situation this is not possible as all parameters are unknown. On the other hand, when arbitrary values are used fixing the measurement equation parameters could even result in wrong parameter signs in the discrete choice model. This way fixing the structural equation variances appears to be better. As mentioned above, as the model goodness-of-fit is not affected by the normalization used it is not possible to determine which values are "a good guess" and which are not. Constraining the measurement equation parameters could be better, but constraining the structural equation variances is safer. Therefore, our general suggestion would be to normalize the measurement equation parameters only if there is information about their relative magnitude (e.g. previous studies), and to normalize the structural equation variances if no information is available.

REFERENCES

Ashok, K., W.R. Dillon and S. Yuan (2002). Extending discrete choice models to incorporate attitudinal and other latent variables. Journal of Marketing Research, 39, 31-46.

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

Ben-Akiva, M.E., J.L. Walker, A.T. Bernardino, D.A. Gopinath, T. Morikawa and A. Polydoropoulou (2002). Integration of choice and latent variable models. In H.S. Mahmassani (ed.), In Perpetual Motion: Travel Behaviour Research Opportunities and Challenges, 431-470. Pergamon, Amsterdam.

Bollen, K.A. (1989). Structural Equations with Latent Variables. John Wiley and Sons, Chichester.

McFadden, D. (1986). The choice theory approach to market research. Marketing Science, 5, 275-297.

Ortúzar, J. de D. and G.A. Hutt (1984). La influencia de elementos subjetivos en funciones desagregadas de elección discreta. Ingeniería de Sistemas, IV, 37–54 (in Spanish).

Ortúzar, J. de D. and L.G. Willumsen (2001). Modelling Transport. John Wiley and Sons, Chichester.

Raveau, S., R. Álvarez-Daziano, M.F. Yáñez, D. Bolduc and J. de D. Ortúzar (2010). Sequential and simultaneous estimation of hybrid discrete choice models: some new findings. Transportation Research Record (forthcoming).

Stapleton, D.C. (1978). Analyzing political participation data with a MIMIC model. Sociological Methodology, 9, 52-74.

Vredin Johansson, M., T. Heldt and P. Johansson (2005). Latent variables in a travel mode choice model: attitudinal and behavioural indicator variables. Working Paper, Department of Economics, Uppsala University, Sweden.

Williams, H.C.W.L. and J. de D. Ortúzar (1982). Behavioural theories of dispersion and the mis-specification of travel demand models. Transportation Research Part B, 16, 167-219.