

F.J. Bahamonde-Birke and J. de D. Ortúzar

**IS SEQUENTIAL ESTIMATION A SUITABLE SECOND BEST FOR THE  
ESTIMATION OF HYBRID CHOICE MODELS?**

**Francisco Bahamonde-Birke\***

Energy, Transportation, Environment Department, Deutsches Institut für Wirtschaftsforschung –  
Berlin  
and

Technische Universität Berlin  
Mohrenstr. 58, Berlin, 10117, Germany

Phone: +49-30-89789467

E-mail: [bahamondebirke@gmail.com](mailto:bahamondebirke@gmail.com)

**Juan de Dios Ortúzar**

Department of Transport Engineering and Logistics, Pontificia Universidad Católica de Chile  
Avda. Vicuña Mackena 4860, Macul, Santiago de Chile, 7820436, Chile

Phone: +56-2-23544822

E-mail: [jos@ing.puc.cl](mailto:jos@ing.puc.cl)

Paper prepared for the **presentation** at the 93<sup>th</sup> Annual Meeting of the Transportation Research Board, Washington, D.C., January 2013, **and** for the **publication** in the Transportation Research Record.

Word Count: Text = 5,717  
Tables (3 @ 250 words each) = 750  
Figures (3 @ 250 words each) = 750  
Total = 7,217

*\*Corresponding Author*

**ABSTRACT**

The simultaneous estimation method has overtaken the sequential approach as preferred estimation method for hybrid discrete choice models. Notwithstanding, the computational cost of the simultaneous estimation can still be prohibitive when models get more involved and in such cases sequential estimation can still be a potent option. In previous work we conducted a theoretical analysis that led them to identify a major bias affecting the sequential estimation method and proposed a correction term for the bias induced on the estimated parameters by the variability associated with the latent variables; however, they did not attempt to quantify this induced variability. In this paper, we attempt to determine the nature of the variability induced through the latent variables as well as the viability of relying on the sequential estimation method as an alternative (second-best) estimation tool, for cases when the complexity of the specification makes unfeasible to rely on simultaneous estimation. Our results show that the sequential method behaves in an acceptable way (the bias can be avoided through the correction), when the variability associated with the latent variables is low in comparison with the error term of the discrete choice model. On the contrary, when this variability is considerable the bias correction becomes an intricate matter and we cannot guarantee appropriate results.

*Keywords:* Hybrid discrete choice models, Latent variables, Variability, Sequential Estimation, Simultaneous Estimation.

## 1. INTRODUCTION

Discrete choice models are an essential element in contemporary travel demand modelling and forecasting. Their current state-of-practice considers objective characteristics of the alternatives and the individuals as explanatory variables, and yields as output the choice probabilities of the alternatives included in their choice sets (1), (2). However, it is well-known that attitudes and perceptions play also a role in the decision making process. The usual approach to take these into account considers the estimation of a Multiple Indicators Multiple Causes (MIMIC) model (3), (4). Here unobserved, latent, variables are explained by a set of characteristics from the users and the alternatives (through so called *structural equations*) while explaining, at the same time, a set of perception indicators obtained from the individuals (through so called *measurement equations*). The joint use of MIMIC models and discrete choice (DC) models leads to the state-of-the-art hybrid choice models (5), (6), (7), (8).

In the last years, the literature has provided abundant empirical and theoretical evidence about the advantages of this approach and the use of hybrid discrete choice (HDC) models has gained substantial popularity (9), (10), (11), among others. In their early days, the most usual form to estimate HDC models was the *sequential approach* (9), (12), (13), as this method requires significantly less computational resources and guarantees consistent estimators when integrating over the latent variables (6) or, at least, just a negligible bias when used without integration (according to the empirical results of Raveau *et al.* (14)). Nowadays, technical improvements and ever increasing computing power has allowed for extended use of the *simultaneous approach*, which guarantees consistent, unbiased and efficient estimators (6), (7). Nevertheless this method is considerably more demanding than the sequential approach and the computational cost can still be prohibitive, especially when working with more complex MIMIC models and a significant number of latent variables (as each latent variable adds a dimension over which the likelihood function must be integrated).

Bahamonde-Birke and Ortúzar (15) examined the increase in model variability associated with the direct inclusion of non-observed (estimated) variables, and their own error terms, into the utility function of DC models in a sequential estimation context. They discussed the problem theoretically concluding that although this variability induced bias on the estimated parameters, the bias could be determined and quantified as a function of the error associated with the utility function and the variability induced through the latent variables. However, they did not attempt to quantify this induced variability.

In this paper we attempt to analyse the nature of the variability induced through the latent variables, which was treated as an unknown variable by Bahamonde-Birke and Ortúzar (15). The aim is to analyse both the possibility of correcting the aforementioned bias as well as the viability of relying on the sequential estimation method as an alternative (second-best) estimation tool for cases when the complexity of the model detracts from applying the simultaneous estimation approach.

The rest of the paper is organized as follows. Section 2 summarises the models and estimation techniques considered. Section 3 extends the theory behind the aforementioned bias while section 4 sets up an experimental analysis (based on simulated data) to test the findings derived in section 3. Section 5 discusses the results of the experiment, and section 6 reports our conclusions.

## 2. THEORETICAL BACKGROUND

In Random Utility Theory (16), (17), it is assumed that individuals  $q$  (belonging to a given market segment  $Q$ ) are rational decision makers who choose an alternative  $i$  (in their set of available alternatives  $A(q)$ ) that maximises their perceived utility ( $U_{iq}$ ). In turn, this utility can be described as the sum of a representative component ( $V_{iq}$ ), considering all attributes that can be observed by a modeller, and an error term ( $\varepsilon_{iq}$ ) describing unknown elements that affect utility but cannot be measured by the observer. In a discrete choice modelling framework, this leads to the following expression (1):

$$U_{iq} = V_{iq} + \varepsilon_{iq} \quad [2.1]$$

Under the assumption that the error terms are independent and identically distributed following an Extreme Value Type 1 (EV1) distribution (with the same mean and variance  $\sigma^2$ ), the differences between the utilities associated with the alternatives follow a Logistic distribution with mean zero and scale  $\lambda$ , leading to the well-known Multinomial Logit (MNL) model (18); in this case, the probability of choosing alternative  $i$  is given by:

$$P_{iq} = \frac{e^{\lambda \cdot V_{iq}}}{\sum_{j \in A(q)} e^{\lambda \cdot V_{jq}}} \quad [2.2]$$

and  $\lambda$  is inversely related to the standard deviation of the error terms:

$$\lambda = \frac{\pi}{\sigma\sqrt{6}} \quad [2.3]$$

However, this scale parameter cannot be estimated, since any parameters in the representative utility function are multiplied by it (i.e. assuming a linear function as usual), so it is customary to *normalise* it to one (19).

As mentioned above, the representative utility  $V_{iq}$  is a function of attributes that can be measured by the modeller. Usually, DC models just consider level-of-service attributes ( $X_{kij}$ , where  $k$  represents the  $k$ th attribute) that can be directly observed by the analyst (i.e. travel times and fares) as well as socioeconomic characteristics of the individual. However, when dealing with a HDC model, latent variables ( $\eta_{lij}$ , where  $l$  represents the  $l$ th latent variable) are also included, but these are immaterial constructs that cannot be directly observed. Assuming a linear specification of the attributes in  $V_{iq}$ , so that the estimated parameters  $\theta_{ik}$  and  $\beta_{li}$  (related to the tangible attributes and latent variables, respectively) can be interpreted as marginal utilities, the representative utility function can be expressed as [2.4].

$$V_{iq} = \sum_k \theta_{ik} \cdot X_{kij} + \sum_l \beta_{li} \cdot \eta_{lij} \quad [2.4]$$

and the usual approach to identify the latent variables relies on a MIMIC model. This requires additional information about the attitudes and/or perceptions of the individuals (normally gathered in the form of *indicators*). The MIMIC model considers that a group of latent variables, representing attitudes or perceptions, are explained by a set of observable characteristics of the individuals and the alternatives ( $s_{iqr}$ ), while explaining a set of perceptual indicators. In this manner, the MIMIC model consists of a set of *structural equations* such as [2.5], explaining the latent variables ( $\eta_{liq}$ ), and a set of *measurement equations* such as [2.6], which consider the latent variables as inputs to explain the perception/attitudinal indicators ( $y_{ziq}$ ).

$$\eta_{liq} = \sum_r \alpha_{lri} \cdot s_{riq} + v_{liq} \quad [2.5]$$

$$y_{ziq} = \sum_l \gamma_{lzi} \cdot \eta_{liq} + \zeta_{ziq} \quad [2.6]$$

where the indices  $i, q, r, l$  and  $z$  refer to alternatives, individuals, exogenous variables, latent variables and indicators, respectively. The error terms  $v_{liq}$  and  $\zeta_{ziq}$  can follow any distribution, but they are typically considered to be Normal distributed with mean zero and a certain covariance matrix. Finally,  $\alpha_{lri}$  and  $\gamma_{lzi}$  are parameters to be jointly estimated.

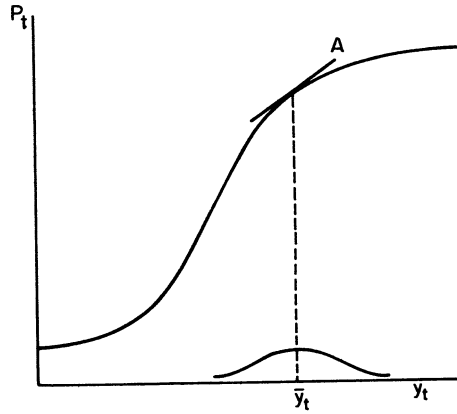
Two approaches have been reported in the literature for the estimation of HDC models. In the *simultaneous estimation* method (5), (20), both structures (the DC model and the MIMIC model) are considered jointly. As mentioned above, this methodology yields unbiased, consistent and efficient estimators (6), and for this reason this approach should be preferred. Unfortunately, the method is highly demanding in terms of computational resources and its cost can be prohibitive when dealing with a significant number of latent variables.

As a second best alternative, several researchers (9), (12), (13), have appealed to the *sequential estimation* method, which divides the problem into two stages, considering first the MIMIC component of the model as an isolated problem to evaluate the expected values of the latent variables. After that, these variables are incorporated directly into the DC model for estimation.

The sequential estimation can be performed in two ways. First, acknowledging that the estimated latent variables are in fact random variables, so that it is necessary to integrate the likelihood of the DC model over the domain of the latent variables for estimation. These approach guarantees consistent but inefficient results (6). Even so there are no major reasons to favour this approach over the *simultaneous estimation* method, as both require integrating the likelihood function yielding computational costs of the same order of magnitude (though the sequential approach is slightly less demanding).

The second way (which is the most popular one to conduct the sequential estimation of HDC models) assumes that the estimated latent variables are in fact deterministic variables. Under this assumption the estimation of the DC model is straightforward (both the specification of the likelihood function and of the required processing power), but it leads to biased estimates for the parameters. As the probability function associated with the logit model is non-linear, it cannot be assumed that the slope of the probability curve is constant over the space over which the density function of the latent variables distributes; hence the probability associated with the expected value of the latent variables is not representative of the probability for the domain, given that similar changes in the value of the latent variables (but in opposite directions), will

have a different effect over the choice probabilities (“naïve approach”), as shown in the Figure 1 (21).



**FIGURE 1 Bias from estimating HDC models using expected values for the latent variables without integration (21).**

When working with this approach, it is usually assumed that increasing the size of the sample can be sufficient to reduce the magnitude of the error, providing acceptable estimators, as long as the variance of the latent variable’s random error is small (6). In the same line, the empirical evidence suggests, that there are no major discrepancies regarding the ratios of the estimated parameters as well as concerning the marginal rate of substitution between the attributes (10), (14).

Nevertheless Bahamonde-Birke and Ortúzar (15) were able to identify a major bias, which is not related to sample size but exclusively to the magnitude of the latent variable model’s variability. They found that when estimating a HDC model without considering the variability of the latent variables, an external error source is added directly into the DC model, deflating all estimated parameters according to the following proportion:

$$\tau = \left( 1 + \frac{6\sigma_{LV}^2 \cdot \lambda_{DC}^2}{\pi^2} \right)^{-1/2} \quad [2.7]$$

where  $\sigma_{LV}^2$  represents the error added through the inclusion of estimated non-observed parameters as non-stochastic variables and  $\lambda_{DC}$  stands for model variability related to own error terms of the DC model. They did not propose a way to assess the magnitude of this extra variability (except when dealing with *a priori* known parameters in a controlled environment) and, therefore, expression [2.7] cannot be used to correct the estimates.

### 3. QUANTIFYING THE EXTRA VARIABILITY

When considering only the expected values of the latent variables the analyst is also adding an external source of error directly into the DC model to be estimated, so that the total discrepancy between the representative and perceived utility corresponds to the sum of the error term underlying the DC model and an extra error coming from the MIMIC model. In fact, replacing [2.4] and [2.5] in [2.1] we get:

$$U_{iq} = \sum_k \theta_{ki} \cdot X_{kqi} + \sum_l \beta_{li} \cdot \left( \sum_r \alpha_{lri} \cdot s_{riq} + v_{liq} \right) + \varepsilon_{iq} \quad [3.1]$$

When estimating a HDC model using the *sequential estimation* method, the analyst assumes the existence of a single error term but this is, in fact, greater than the usual error term associated with the DC model. This new error term can be represented in the following manner:

$$\varepsilon_{(HDC)iq} = \sum_l \beta_{li} \cdot v_{liq} + \varepsilon_{(DC)iq} \quad [3.2]$$

where  $\varepsilon_{(HDC)iq}$  is the total error considered in the sequential estimation, while  $\varepsilon_{(DC)iq}$  stands for the error term associated with the underlying discrete choice component of the model, as required when following the simultaneous approach. As a consequence, the model variability can be expressed as follows:

$$\sigma_{HDC}^2 = \sigma_{DC}^2 + \sum_l \beta_{li}^2 \cdot \sigma_l^2 \quad [3.3]$$

where  $\sigma_{HDC}^2$  represents the model variability considered in the estimation,  $\sigma_l^2$  is the variability associated with the error terms of the MIMIC model's structural equations and  $\sigma_{DC}^2$  stands for the variability of the underlying discrete choice model. To simplify the notation, from [3.3] onwards we assume, without loss of generality, that a given latent variable only affects the utility of a single alternative. If we consider that the parameters  $\beta_l^2$  are deterministic and known *a priori*,  $\sum_l \beta_{li}^2 \cdot \sigma_l^2$  stands for the induced variability  $\sigma_{LV}^2$  considered by Bahamonde-Birke and Ortúzar (15) in equation [2.7].

If we assume that the error terms of the MIMIC model's structural equations ( $v_{liq}$ ) follow a distribution which is equal to the difference between two IID EV1 distributions with different variance, it can be shown that both the underlying DC model and the HDC model to be sequentially estimated can be represented as Logit models. If this is not so, the crux of the argument does not change but the mathematics and interpretation of results would get much more involved. Also, theoretically the whole estimation of HDC models using the sequential estimation method would neglect the hypotheses of the Logit model; even so, empirical experience (10), (14), provides evidence sustaining that this neglect does not have major implications over the estimates. Then, under the above assumption, equation [3.3] can be simply written as:

$$\frac{\pi^2}{6 \cdot \lambda_{HDC}^2} = \frac{\pi^2}{6 \cdot \lambda_{DC}^2} + \sum_l \beta_{li}^2 \cdot \sigma_l^2 \quad [3.4]$$

where  $\lambda_{DC}$  and  $\lambda_{HDC}$  are the scale parameters of the Logit models associated with underlying DC model and the HDC model to be estimated, respectively. The  $\beta_l$  parameters, in turn, stand not for the parameters associated with the underlying DC model, but for the estimated parameters of the

HDC model estimated sequentially; therefore, they are also deflated by its scale parameter, so that:

$$\beta_l^* = \frac{\lambda_{HDC}}{\lambda_{DC}} \cdot \beta_l \quad [3.5]$$

where  $\beta_l^*$  are the parameters associated with the underlying DC model. Hence, equation [3.4] can be rewritten in terms of the parameters associated with the underlying DC model (again under the assumption that the  $\beta_l$  parameters are deterministic and *a priori* known variables), which are those that would be recovered if the problem was approached properly (without the induction of bias):

$$\frac{\pi^2}{6 \cdot \lambda_{HDC}^2} = \frac{\pi^2}{6 \cdot \lambda_{DC}^2} + \frac{\lambda_{HDC}^2}{\lambda_{DC}^2} \cdot \sum_l \beta_l^{*2} \cdot \sigma_l^2 \quad [3.6]$$

Working on equation [3.6], the scale parameter  $\lambda_{HDC}$  can be isolated as a function of  $\lambda_{DC}$  and  $\beta_l^*$ :

$$\lambda_{HDC} = \sqrt{\frac{-\frac{\pi^2}{6 \cdot \lambda_{DC}^2} + \sqrt{\frac{\pi^4}{6^2 \cdot \lambda_{DC}^4} + \frac{4 \cdot \pi^2}{6 \cdot \lambda_{DC}^2} \cdot \sum_l \beta_l^{*2} \cdot \sigma_l^2}}{\frac{2}{\lambda_{DC}^2} \cdot \sum_l \beta_l^{*2} \cdot \sigma_l^2}} \quad [3.7]$$

Figure 2 presents a graphical representation of the relation between  $\lambda_{HDC}$  and  $\lambda_{DC}$  (which actually represents the deflation of the parameters) following [3.7] as a function of the artificially induced variability  $\sum_l \beta_l^{*2} \cdot \sigma_l^2$  for different values of  $\lambda_{DC}$ .

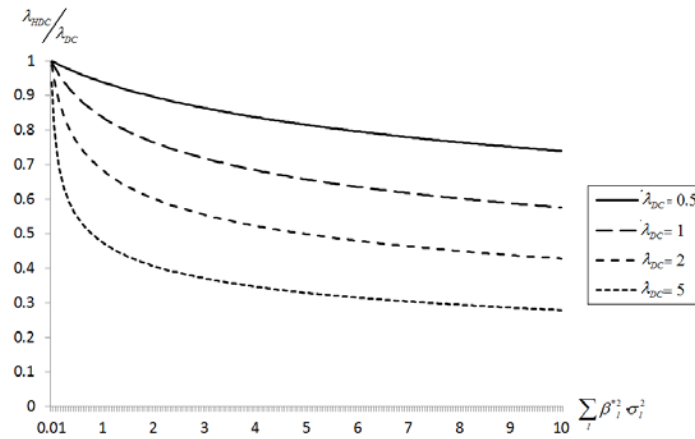


FIGURE 2 Relation between  $\lambda_{HDC}$  and  $\lambda_{DC}$ .



As can be seen,  $\lambda_{HDC}$  is equal to  $\lambda_{DC}$  when no variability is added (as expected), and gets smaller in comparison with  $\lambda_{DC}$  when the induced error increases, both in relative (smaller  $\lambda_{DC}$ ) and in absolute terms. The relation tends asymptotically to zero for all values of  $\lambda_{DC}$ .

As stated in the previous section, all estimated parameters of a DC model are deflated by the scale parameter  $\lambda$ . As this parameter is inversely related with the standard deviation of the error terms [2.3], it is clear that higher variability should imply smaller estimates, which is consistent with our findings and with the fact that a model affected by greater error terms is less informative (the smaller the parameter estimates, the more the estimated model tends to the equiprobable model). Hence, the sequential estimation of HDC models increases model variability and deflates the estimates, affecting the choice probabilities and decreasing artificially the model's goodness-of-fit.

One could suggest correcting the estimated parameters using equation [3.7] and fixing  $\lambda_{DC}$  to one (to emulate the results of the simultaneous estimation). However this strategy suffers from theoretical problems and equation [3.7] is not useful in practice. First, it must be acknowledged that as the  $\beta^*_l$  parameters are unknown (they should be estimated), they are not available to perform a correction (in contrast, the  $\sigma_l$  values are known deterministic variables, as the modeller has to fit the variances associated with the MIMIC model's structural equations to guarantee identification, (22)). Further, the  $\beta^*_l$  parameters, which we had considered as fixed known deterministic variables, are in fact stochastic. This further increases the variability induced into the model and the increment depends on the nature of the model and the dataset. Hence, the result presented in [3.7] can only be understood as an upper limit and the real deflation associated with the use of the sequential estimation should probably be larger.

#### 4. AN EXPERIMENTAL ANALYSIS

To analyse our findings and to test how much the real deflation differs from the result derived above, we devised an experimental analysis based on simulated data. This allows to examine the research subject in a context free of undesired effects, while at the same time enabling to determine the magnitude of the theoretically expected deflation (the upper limit of  $\lambda_{HDC}$  or the lower limit of the deflation), as the real parameters are an input of equation [3.7]. Following the tradition of Williams and Ortúzar (23), we generated 15 different samples, each of 25,000 simulated individuals, behaving in a compensatory manner in accordance with different utility functions.

We considered a MIMIC model specification based on three explanatory variables, two latent variables and three perception indicators, though certain parameters were fixed at zero in some specifications, excluding latent variables or perception/attitudinal indicators from the modelling. Regarding the specification of the utility function, we considered two alternatives, each represented as the sum of one observed variable, a latent variable and an error term. The structure used in the generation of the dataset was the following:

$$\begin{aligned} U_{iq} &= \theta_i \cdot X_{iq} + \beta_i \cdot \eta_{iq} + \varepsilon_{iq} \\ \eta_{iq} &= \alpha_{1i} \cdot s_{1q} + \alpha_{2i} \cdot s_{2q} + \alpha_{3i} \cdot s_{3q} + v_{iq} \\ y_{zq} &= \gamma_{zi} \cdot \eta_{iq} + \gamma_{zi} \cdot \eta_{iq} + \zeta_{zq} \end{aligned} \tag{4.1}$$

which is a simplified form of the structure presented in the equations [2.4], [2.5] and [2.6], where the sub-index  $i$  stands for an alternative,  $z$  for an indicator and  $q$  for an individual. This notation is consistent with the sub-indices in Table 1.

All three explanatory variables ( $s_{ri}$ ) were generated taking random draws from independent continuous uniform distributions, between zero and one, for each individual. The error terms of the measurement equations ( $\zeta_{zq}$ ) are distributed Normal with zero mean and unit variance. To vary the magnitude of the error induced into the utility function, the error terms associated with the structural equations of the MIMIC model ( $v_{iq}$ ) have different variances, distributed Normal with zero mean and standard deviation  $\sigma_i$ . The observed variable ( $X_{iq}$ ) taking part on the utility function was generated taking draws from a Normal distribution with mean 3.0 and standard deviation 1.4 for alternative one, and mean 4.0 and standard deviation 1.2 for alternative two. We fixed to one all parameters associated with the utility function and the scale parameters of the error terms associated with them, to simplify the evaluation of the bias. The values of the MIMIC model's parameters for each sample, as well as the standard deviations  $\sigma_i$  of the error terms  $v_{iq}$  are also presented in Table 1. To dismiss potential misspecifications in the data generation process we estimated the model for all samples following the simultaneous approach, observing that the data was indeed properly recovered.

**Table 1 Parameters used in the generation of the MIMIC model**

Sample	$\alpha_{11}$	$\alpha_{21}$	$\alpha_{31}$	$\alpha_{12}$	$\alpha_{22}$	$\alpha_{32}$	$\gamma_{11}$	$\gamma_{12}$	$\gamma_{21}$	$\gamma_{22}$	$\gamma_{31}$	$\gamma_{32}$	$\sigma_1$	$\sigma_2$
1	3	2	-1	0	0	0	0.7	0	0.5	0	0	0	1	0
2	3	2	-1	0	0	0	0.7	0	0.5	0	0	0	5	0
3	3	2	-1	0	0	0	0.7	0	0.5	0	0	0	2	0
4	3	2	-1	0	0	0	0.7	0	0.5	0	0	0	0.5	0
5	3	2	-1	0	0	0	0.7	0	0.5	0	0	0	0.2	0
6	0.5	0.2	0.3	0	0	0	3	0	1.5	0	0	0	1	0
7	0.5	0.2	0.3	0	0	0	3	0	1.5	0	0	0	5	0
8	0.5	0.2	0.3	0	0	0	3	0	1.5	0	0	0	2	0
9	0.5	0.2	0.3	0	0	0	3	0	1.5	0	0	0	0.5	0
10	0.5	0.2	0.3	0	0	0	3	0	1.5	0	0	0	0.2	0
11	3	2	0	0	2	3	0.7	0	0.5	0.5	0	0.7	1	1
12	3	2	0	0	2	3	0.7	0	0.5	0.5	0	0.7	2	2
13	3	2	0	0	2	3	0.7	0	0.5	0.5	0	0.7	0.5	2
14	3	2	0	0	2	3	0.7	0	0.5	0.5	0	0.7	2	0.5
15	3	2	0	0	2	3	0.7	0	0.5	0.5	0	0.7	0.5	0.5

## 5. RESULTS AND DISCUSSION

Using the sequential approach (without integration over the domain of the latent variables) we estimated HDC models for all the samples, following the exact specification used in the generation of the dataset. As expected, the parameters associated with the MIMIC part of the model were properly recovered, and there is no statistical evidence to reject the hypothesis of equality between the estimates and the target values for any parameter of all 15 samples (at a confidence level of 5%). The results obtained from the estimation of the DC models for the 15

samples are presented on Table 2. The notation is that used in equation (4.1) and the standard deviation of the estimates is shown in brackets. We have also included the expected induced variability  $\sum_i \beta_i^{*2} \sigma_i^2$  as well as the value for  $\lambda_{HDC}$  calculated in accordance to equation [3.7],

considering that the  $\lambda_{DC}$  parameter was fixed to one. The calibration of the DC models was performed using BIOGEME (24).

From Table 2 it is clear that the estimates obtained from the sequential approach are affected by the variability associated with the estimates of the latent variables, and it is not possible to recover the target values without performing a correction. Acknowledging this issue is important as, in some cases, the estimates are deflated by as much as three times their real values, affecting substantially the choice probabilities and the predictive capability of the models.

**Table 2 Estimation results of the DC model**

Sample	$\beta_1$	$\beta_2$	$\eta_1$	$\eta_2$	$\sum_i \beta_i^{*2} \sigma_i^2$	$\lambda_{HDC}$
1	0.870 (0.0135)	0.876 (0.0125)	0.866 (0.0173)	-	1	0.837
2	0.306 (0.00884)	0.317 (0.00796)	0.321 (0.0114)	-	25	0.475
3	0.636 (0.0112)	0.628 (0.0103)	0.629 (0.0140)	-	4	0.684
4	0.968 (0.0145)	0.954 (0.0135)	0.956 (0.0174)	-	0.25	0.939
5	0.988 (0.0139)	1.00 (0.0139)	1.05 (0.0182)	-	0.04	0.988
6	0.871 (0.0176)	0.846 (0.0125)	0.840 (0.0894)	-	1	0.837
7	0.295 (0.0135)	0.291 (0.00869)	0.183 (0.0993)	-	25	0.475
8	0.645 (0.0155)	0.623 (0.0107)	0.642 (0.0864)	-	4	0.684
9	0.922 (0.0183)	0.963 (0.0135)	1.01 (0.0863)	-	0.25	0.939
10	0.979 (0.0186)	0.983 (0.0138)	0.918 (0.0791)	-	0.04	0.988
11	0.747 (0.0127)	0.763 (0.0125)	0.783 (0.0166)	0.772 (0.0166)	2	0.765
12	0.504 (0.0103)	0.514 (0.0102)	0.502 (0.0141)	0.482 (0.0141)	8	0.602
13	0.618 (0.0114)	0.634 (0.0112)	0.637 (0.0153)	0.621 (0.0153)	4.25	0.677
14	0.639 (0.0115)	0.638 (0.0113)	0.630 (0.0154)	0.612 (0.0149)	4.25	0.677
15	0.901 (0.0320)	0.884 (0.0312)	0.832 (0.0388)	0.871 (0.0387)	0.5	0.896

To ease following the situation discussed, in Figure 3 we provide a graphic representation of the estimated parameters together with the proposed correction curve, given the different induced variabilities. The horizontal axis of the graph uses a logarithmic scale.

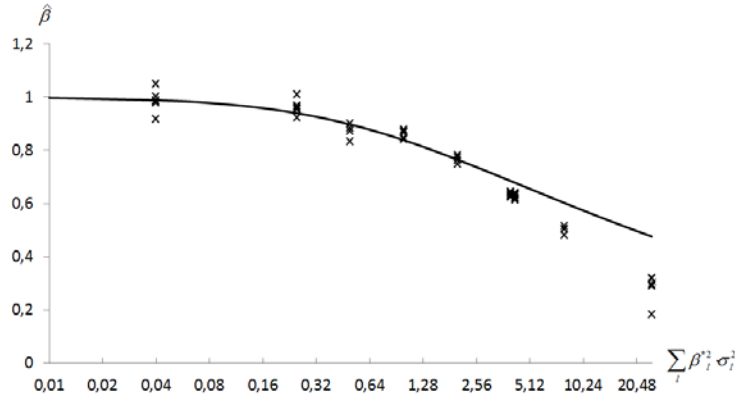


FIGURE 3 Estimated parameters and expected deflation

In relation to the proposed expression for the  $\lambda_{HDC}$  parameter, the empirical results indicate that there are not major discrepancies between the suggested upper limit for  $\lambda_{HDC}$  and the real deflation affecting the estimates. In fact, it is not possible to detect significant differences between this upper limit and the real deflation, as long as the induced variability does not exceed a magnitude of two. Over this value the differences tend to increase in conjunction with the added variability and our proposed value cannot be considered a proper predictor of the deflation.

As a consequence, a correction based on this upper limit can be attempted and it should provide acceptable results when working with a small induced variability. Although this correction is not 100% reliable – for instance the results obtained for sample 1 are slightly biased (in terms of the magnitude) even after correcting the estimates and standard deviations – it provides clearly better estimates than working directly with the estimation results, making it possible to recover most target values. Even in those cases affected by a high variability (more than two in our tests) a correction of the deflation using this upper limit offers clearly better results for forecasting (although still biased) than dealing with the original values.

As stated before, the result presented in [3.7] cannot be used in practice since the  $\beta_i^*$  parameters are unknown, but an alternative formulation based on the model estimates  $\beta_i$  can be proposed working on expression [3.4]. In that case, the  $\lambda_{HDC}$  parameter may be expressed as a function of  $\lambda_{DC}$  and the  $\beta_i$ .

$$\lambda_{HDC} = \frac{\lambda_{DC}}{\sqrt{1 + \frac{6 \cdot \lambda_{DC}^2 \cdot \sum \beta_i^2 \sigma_i^2}{\pi^2}}} \quad [5.1]$$

However, it is important to proceed very carefully when dealing with this formulation, as the  $\beta_i$  parameters are also deflated by the  $\lambda_{HDC}$  parameter, so that the inclusion of over-deflated estimates (as can be expected, since we are working with an upper limit for  $\lambda_{HDC}$ ) could lead to an underestimation of the general deflation. Moreover, it is important to state that in this

specification, the input variables are also stochastic estimates, implying that the  $\lambda_{HDC}$  parameter is of the same nature.

To illustrate this situation, we have computed  $\lambda_{HDC}$  following the alternative formulation for our 15 samples. The results are shown in Table 3. As expected, the proposed form provides a good proxy for  $\lambda_{HDC}$ , when the deflation is small, but underestimates the latter (even more), when the induced variability gets larger.

**Table 3 Comparison of deflation estimators**

Sample	$\lambda_{HDC}$ [3.7]	$\lambda_{HDC}$ [5.1]
1	0.837	0.829
2	0.475	0.624
3	0.684	0.714
4	0.939	0.937
5	0.988	0.987
6	0.837	0.837
7	0.475	0.814
8	0.684	0.707
9	0.939	0.930
10	0.988	0.990
11	0.765	0.759
12	0.602	0.793
13	0.677	0.808
14	0.677	0.806
15	0.896	0.905

Finally, it is important to acknowledge, that in line with the previous empirical evidence, the relation between the estimated parameters as well as the marginal rates of substitution between the attributes do not appear to be affected by any bias and the target values are properly recovered, despite the use of the sequential estimation method (the only exception could be the value associated with the latent variable in sample 7, but the large standard deviation associated with this estimate prevents rejecting any hypothesis of equality).

## 6. CONCLUSIONS

The estimation of HDC models following the sequential estimation approach is still widely used as a second-best estimation tool, because of the high computational costs of the simultaneous estimation method, especially when working with several latent variables. Notwithstanding, the estimators associated with this methodology are biased (but consistent), as the estimates for the latent variables are introduced into the DC model as deterministic (observed) variables, inducing an extra error into the model. As the estimators are biased, this problem cannot be reduced by increasing the size of the sample.

We expanded the theoretical analysis of Bahamonde-Birke and Ortúzar (15) to quantify the magnitude of this bias and to propose a correction term for the estimates. However, we were

only able to identify an upper limit for the scale parameter associated with the deflation caused by the direct inclusion of the latent variables into the DC model.

In the same line, we conducted an empirical experiment (based on simulated data) to analyse how much the real deflation differed from the quantifiable upper limit (bottom limit of the deflation), observing that the discrepancies were negligible for small induced variability. Hence, we argue that this upper limit is a good predictor for the real deflation when the induced variability is low and behaves appropriately as a correction term. When the induced error gets larger the upper limit underestimates the real deflation and therefore, the correction term is not able to guarantee unbiased estimators.

We argue that performing this correction is highly recommended when approaching the estimation problem sequentially, as it corrects or diminishes (depending on the magnitude of the induced error) a significant bias affecting the predictive capability of the estimated models. Notwithstanding, and in accordance with other studies, our findings show that the marginal rates of substitution between the attributes are not actually affected by the estimation technique. Therefore our study supports the thesis that sequential estimation of HDC models is a suitable second best alternative when the focus is centred on finding marginal rates of substitution or willingness-to-pay measures.

On the contrary, when the analyst expects to use the model for forecasting and for evaluation of choice probabilities, the suitability of the estimation technique must be properly evaluated. So, if the induced error associated with the variability of the latent variables is relatively small in comparison with the error terms intrinsic to the DC model, the estimation methodology should work in an acceptable manner if the correction term suggested in this paper is applied. If this is not the case, other alternatives should be favoured.

Finally, it is important to remember that the most important reason to opt for the sequential over the simultaneous estimation method are the high computational costs associated with the inclusion of several latent variables. Unfortunately, it can be expected that the inclusion of several latent variables will imply the introduction of higher induced variability into the DC model, causing the appearance of larger bias, detracting from the advantages of our simplified technique, making it a less suitable alternative.

## ACKNOWLEDGMENTS

We gratefully acknowledge the support of Becas Chile given by the Chilean Council for Scientific and Technological Research (CONYCIT), the Millennium Institute in Complex Engineering Systems (ICM: P05-004F; FONDECYT: FB016), the Across Latitudes and Cultures BRT Centre of Excellence funded by the Volvo Research and Educational Foundations, the Alexander von Humboldt Foundation and the Centre for Sustainable Urban Development, CEDEUS (Conicyt/Fondap/15110020). The useful comments of several unknown reviewers are gratefully acknowledged.

## REFERENCES

- 1      Ortúzar, J. de D. and Willumsen, L.G (2011). *Modelling Transport*. Fourth Edition, John Wiley and Sons, Chichester.
- 2      Train, K. E. (2009). *Discrete Choice Methods with Simulation*. Second Edition, Cambridge University Press, Cambridge.
- 3      Zellner A. (1970). Estimation of regression relationships containing unobservable variables. *International Economic Review* **11**, 441-454
- 4      Bollen, K.A. (1989). *Structural Equations with Latent Variables*. John Wiley and Sons, Chichester.
- 5      Ashok, K., Dillon, W. and Yuan, S. (2002). Extending discrete choice models to incorporate attitudinal and other latent variables. *Journal of Marketing Research* **39**, 31-46.
- 6      Ben-Akiva, M.E., Walker, J.L., Bernardino, A.T., Gopinath, D.A., Morikawa, T. and Polydoropoulou, A. (2002). Integration of choice and latent variable models. In H.S. Mahmassani (ed.), *In Perpetual Motion: Travel Behaviour Research Opportunities and Challenges*. Pergamon, Amsterdam.
- 7      Bolduc, D. and Alvarez-Daziano, R. (2010). On the estimation of hybrid choice models. In S. Hess and A. Daly (eds.), *Choice Modelling: The State-of-the-Art and the State-of-Practice*, 259–288. Emerald, Bingley.
- 8      Raveau, S., Yáñez, M.F. and Ortúzar, J. de D. (2012). Practical and empirical identifiability of hybrid discrete choice models. *Transportation Research* **46B**, 1374-1383.
- 9      Yáñez, M.F., Raveau, S. and Ortúzar, J. de D. (2010). Inclusion of latent variables in mixed logit models: modelling and forecasting. *Transportation Research* **44A**, 744-753.
- 10     Bahamonde-Birke, F.J., Raveau, S., Yáñez, M.F. and Ortúzar, J. de D. (2010). The role of tangible attributes in hybrid discrete choice models. *European Transport Conference 2010*, Glasgow, 11-13, October, 2010.
- 11     Alvarez-Daziano, R. and Bolduc, D. (2013). Incorporating pro-environmental preferences towards green automobile technologies through a Bayesian hybrid choice model. *Transportmetrica* **9A**, 74-106.
- 12     Vredin-Johansson, M., Heldt, T. and Johansson, P. (2006). The effects of attitudes and personality traits on mode choice. *Transportation Research* **40A**, 507-525.
- 13     van Acker, V., Mokhtarian, P.L. and Witlox, F. (2011). Going soft: on how subjective variables explain modal choices for leisure travel. *European Journal of Transport and Infrastructure Research* **11**, 115-147.
- 14     Raveau, S., Alvarez-Daziano, R., Yáñez, M.F., Bolduc, D. and Ortúzar, J. de D. (2010). Sequential and simultaneous estimation of hybrid discrete choice models: some new findings. *Transportation Research Record* **2156**, 131-139.

- 15 Bahamonde-Birke, F.J. and Ortúzar, J. de D. (2012). On the variability of hybrid discrete choice models. *Transportmetrica* (DOI:10.1080/18128602.2012.700338).
- 16 Thurstone, L.L. (1927). A law of comparative judgment. *Psychological Review* **34**, 273-286.
- 17 McFadden, D. (1974). Conditional logit analysis of qualitative choice behaviour. In Zarembka, P. (ed.), *Frontiers in Econometrics*, 105-142. Academic Press New York.
- 18 Domencich, T. and McFadden, D. (1975). *Urban Travel Demand – A Behavioural Analysis*. North Holland, Amsterdam.
- 19 Walker, J.L. (2002). The mixed logit (or logit kernel) model: dispelling misconceptions of identification. *Transportation Research Record* **1805**, 86-98.
- 20 Morikawa, T., Ben-Akiva, M. and McFadden, D.L. (2002). Discrete choice models incorporating revealed preferences and psychometric data. *Econometric Models in Marketing* **16**, 29-55.
- 21 Train, K.E., McFadden, D.L. and Goett, A.A. (1987). Consumer attitudes and voluntary rate schedules for public utilities. *Review of Economics and Statistics* **64**, 383-91.
- 22 Stapleton, D.C. (1978). Analysing political participation data with a MIMIC model. *Sociological Methodology* **9**, 52-74.
- 23 Williams, H.C.W.L. and Ortúzar, J. de D. (1982). Behavioural theories of dispersion and the mis-specification of travel demand models. *Transportation Research* **16B**, 167-219.
- 24 Bierlaire, M. (2003). BIOGEME: A free package for the estimation of discrete choice models, *Proceedings of the 3rd Swiss Transportation Research Conference*, Ascona, Switzerland.