Author's Pre-print version
Submitted to Transportation

Misspecified Hybrid Choice Models:
An empirical study of parameter bias and model selection.

Juan Manuel Lorenzo Varela
KTH Royal Institute of Technology / Universidade da Coruña, Observatorio Ocupacional
Address: Teknikringen 10a, 100 44, Stockholm, Sweden.
E-mail address: jmlv@kth.se
Tel.: +46 8 790 66 95.
URL: https://www.linkedin.com/in/jmlorenzovarela/
OrcID: 0000-0003-4512-9054

Abstract

Model misspecification is likely to occur when working with real datasets. However, previous studies showing the advantages of hybrid choice models to account for measurement errors have mostly used models where structural and measurement equations match the functions employed in the data generating process, especially when parameter estimate biases were discussed.

The aim of this study is to investigate the extent of parameter bias in misspecified hybrid choice models, assess if different modelling assumptions required to make the hybrid choice models operative impact the parameter estimates of the choice model, and evaluate the prediction accuracy of misspecified hybrid choice models in comparison with a simpler, also misspecified, multinomial logit. For these tasks, a mode choice model is estimated on 100 synthetic datasets. The synthetic datasets were designed to mimic the conditions present in real datasets; hence the postulated structural and measurement equations of the hybrid choice models are less flexible than the functions used for the data generating process.

Results show that hybrid choice models, even if misspecified, manage to recover better parameter estimates than a multinomial logit. However, hybrid choice models are not unbeatable, as results indicate that misspecified hybrid choice models will still yield biased parameter estimates. Moreover, results suggest that all models, the multinomial logit and the hybrid choice models, successfully isolate the source of model bias, preventing its propagation to other parameter estimates. Furthermore, results indicate that parameter estimates from hybrid choice models are robust to modelling assumptions. Finally, results show that a simple multinomial logit provides higher out-of-sample prediction accuracy than the hybrid choice models, highlighting that better parameter estimates, do not always translate into better model predictions.

Keywords: Hybrid Choice Models (HCM); Integrated Choice and Latent Variable models (ICLV); Mode choice; Latent variables; Model misspecification; Parameter bias; Synthetic dataset; Out-of-sample prediction.

1. INTRODUCTION

Hybrid choice models (HCMs) appeared in the last two decades, [Walker, 2001; Ben-Akiva et al., 2002], and since then their use has grown exponentially. These models are motivated by findings in the social sciences, where evidence supports that latent variables (attitudes, norms, perceptions, affects and/or beliefs) can often override the influence of observable variables on disaggregate behaviour. Making the HCM operative requires several modelling assumptions, which include the specification of structural and measurement equations. Through these assumptions, researchers postulate theories in the form of statistical models and test how well the theories model the observed data. Frequently, and especially when working with real datasets, the postulated theories may not be accurate and result in misspecified models. Hence, we should ask ourselves how the estimation results might be affected by model misspecification.

Nearly all models are to some degree structurally misspecified, [Browne & Cudeck, 1993]. As a result, most of the ideal properties of the maximum likelihood estimator need not hold in the real world. See [Cragg, 1968; Yuan and Hayashi, 2006; Bollen, et al., 2007]. Following this line of reasoning, [Kolenikov, 2011], dealt with the problem of quantifying the degree to which parameter estimates in a structural equation model (SEM) can be biased when structural relationships were not correctly specified, and proposed a framework to assess the degree to which the parameter estimates may be biased.

Although considerable research has been devoted to structural misspecification bias in SEM, rather less attention has been paid to its effects in HCMs. For instance, [Walker et al., 2010], studied how to estimate travel demand models when the underlying quality of level of service data is poor. In this study, the authors used synthetic data to test the capabilities of the HCM framework to correct measurement errors in explanatory variables; and found that the HCM framework was able to accurately estimate the true value of the parameters without knowing the true travel time. Nevertheless, the estimated HCM was correctly specified and matched the formulation used for the data generating process. A more recent study, [Vij and Walker, 2016], systematically evaluated the benefits of the HCM framework in comparison with a more traditional choice model without latent variables. In this study, the authors carried out a detail analysis of goodness-of-fit and bias of the parameter estimates through different Monte Carlo experiments with synthetic data. Among these experiments, the HCM in *experiment III* had a misspecified utility function, whilst the HCM in *experiment IV* had misspecified structural equations. Unfortunately, the focus of those experiments was the goodness-of-fit of the HCM compared to a reduced form mixed logit, and parameter estimates were not discussed, neither provided. Furthermore, in their discussion of measurement error bias, results from another Monte Carlo experiment were given, but the structural and measurement equations of the HCM tested were not only correctly specified, but also the structural equation included an extra parameter. In other words, the model tested was more flexible than the one used for the data generation. In their analysis, the hypothesis that the mean parameter estimates were equal to the true values could not be rejected.

Whilst these exercises provided useful insights into the HCM framework capabilities to correct for measurement errors, I argue that misspecification effects expected from the complexity of real datasets, where the specified HCM is not flexible enough to model the data generation process, might not have been captured.

The aim of this study is to investigate the extent of parameter bias in misspecified HCMs, assess if different modelling assumptions required to make the HCMs operative impact the parameter estimates of the choice model, and evaluate the prediction accuracy of misspecified HCMs in comparison with a simpler, also misspecified, multinomial logit. Also, biases in parameters are discussed in relation to goodness-of-fit measurements frequently used when testing models with latent variables.

The paper is structured as follows: Section 2 describes the synthetic data generation process. Section 3 gives a description of the HCM framework as well as the models and goodness-of-fit measurements used in this study. Section 4 presents the results, and Section 5 concludes.

2. DATA

For this study, 100 synthetic datasets are generated for 1000 and 500 observations (200 datasets in total). Synthetic datasets are especially useful for bias quantification because the functions and true parameters underlying the data generating process are known. Furthermore, to facilitate out-of-sample tests of prediction accuracy, each dataset is split in a train and a test set, with sample sizes equivalent to 2/3 and 1/3 of the total dataset observations respectively.

Synthetic observed decision makers, for a mode choice model, are hypothesised to behave according to the decision-making process described below. I construct a hypothetical, multinomial Logit (MNL), mode choice model with 3 alternatives and explanatory variables of time and cost, along with alternative specific constants. The utilities are specified as defined by equations (1)-(3):

$$U_1 = \beta_{t1} t_1 + \beta_c c_1 + asc_1 + \varepsilon_1 \quad (1)$$
$$U_2 = \beta_{t2} t_2 + \beta_c c_2 + asc_2 + \varepsilon_2 \quad (2)$$
$$U_3 = \beta_{t3} t_3 + \beta_c c_3 + asc_3 + \varepsilon_3 \quad (3)$$

where, U_i is the utility for alternative i = {1,2,3}, β_{ti} the parameter for travel time, β_c the parameter for travel cost, t_i the travel time, c_i the travel cost, asc_i the alternative specific constant, and ε_i are independent and identical distributed (iid) Extreme Value errors. Time and cost variables are drawn independently from a mixture of gaussian and lognormal distributions described in Table 1, with mixing proportions 50-50%; parameters β_{ti} and β_c

are set to the values reported in Table 3 and ε_i are drawn independently from the standard Gumbel distribution.

Table 1. Distributions used to generate the synthetic variables.

		Time1	Time2	Time3	Cost1	Cost2	Cost3	Time1_disturbances
Normal distribution								
	μ	55	40	25	10	25	15	0
	σ	30	20	15	5	10	5	5
Lognormal distribution								
	$\log(\mu)$	3.5	3.15	2.95	2.2	2.5	3.1	0
	$\log(\sigma)$	0.3	0.4	0.6	0.46	0.55	0.39	1

In addition, to better mimic the conditions of real datasets, an imperfect measurement for t_1 , $t_{1\,measured}$, was generated. Measurement errors introduced in $t_{1\,measured}$ consist of two components; one multiplicative and other additive. These components represent typical sources of measurement error in our models. For instance, multiplicative disturbances (γ) are expected in level of service time variables from assignment models where the assumed average speed is incorrect; and additive measurement errors (α) represent the aggregation effect into zone centroids. Equation (4) shows the function generating this variable.

$$t_{1\,measured} = time_1 * \gamma + \alpha \tag{4}$$

where γ is lognormally distributed with parameters $\log(\mu)$ = 0 and $\log(\sigma)$ = 1; and, α is normally distributed with zero mean and $\sigma = 5$.

Finally, observations where time and/or cost variables have negative values are excluded from the datasets. This was done to maintain theoretical consistency with the types of variables being modelled, as time and cost cannot yield negative values. Note that removing observations with negative time and/or cost variables will not bias the estimation results, as these observations are removed prior to the calculation of the utilities and setting of the chosen alternatives. In essence, this process is equivalent to use truncated distributions during the data generating process. I decided to take this approach to prevent negative values of the variable $\mathrm{t}_{1\,\mathrm{measured}}$, as it might occur that the additive component of the measurement error (α), can be larger than the component $time_1 * \gamma$, hence a negative travel time could appear. Note that because of this procedure, the number of observations for the train and test sets is not constant, but it oscillates slightly.

Using the remaining observations for each dataset, utilities are calculated for each of the synthesized observations, and the ‘chosen’ alternative is set to be that with the highest utility.

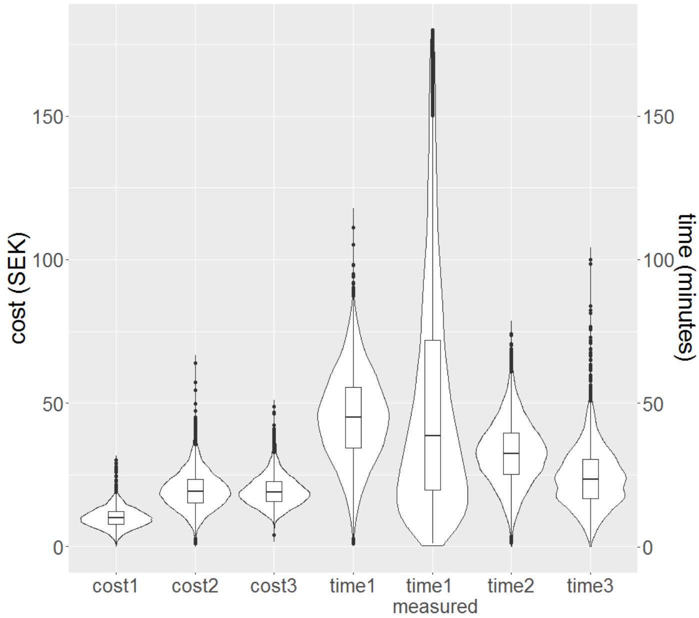


Figure 1 shows the distributions of the different variables for the first of the synthetic datasets. Here we can observe how measurement errors modify the shape of the variable *time1*, where the resulting variable with measurement errors (*time1 measured*) has thicker tales and a lower mean.

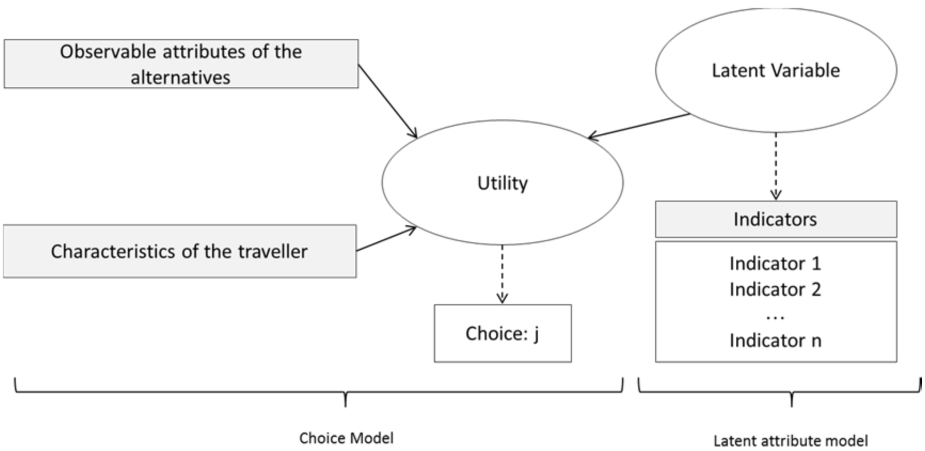
Figure 1. Distribution of synthetic variables

3. METHODOLOGY

3.1. – Hybrid Choice Model framework

The HCM framework has two main components: a discrete choice model and a latent variable model. According to the components included, this structure has also been referred to as the integrated choice and latent variable (ICLV) model. A HCM allows us to introduce observable variables in the utility functions, as well as unobservable ones. The key concept is that while the value of a latent variable is unknown, an approximation is available in the value measured with error, the indicator. The approximated and true values are then connected via a measurement relationship. The model schematics are shown by Figure 2,

Figure 2. Hybrid Choice Model framework



where observed variables, indicators and choices are represented by rectangular boxes, whilst unobserved variables such as utilities and latent variables are represented by ellipses. In addition, structural equations are represented by continuous lines and measurement equations by dashed lines.

Equations for the HCM framework were presented in Walker et al. 2010 as follows; First, we choose a variable to be treated as a latent variable (X), known only to a distribution f_X and a set of estimated parameters θ:

$$f_X(X;\theta) \tag{5}$$

In terms of the mode choice model, if the explanatory variable were known, we would have a typical mode choice model that depicts the probability of an individual choosing a mode i conditional on a set of estimated parameters β and explanatory variables (which includes the true value of X):

$$P(i|\beta, X) \tag{6}$$

However, since X is unknown, it is necessary to integrate the conditional choice probability over the distribution of X:

$$P(i|\beta,\theta) = \int P(i|\beta, X)\, f_X(X;\theta) dX \tag{7}$$

The measurement equation now comes into play to incorporate the measured value as an indicator of X. For this, the distribution of the indicator (I) conditional on X and a set of estimated parameters λ is necessary:

$$f_M(I|X;\lambda) \tag{8}$$

Assembling all elements together, the likelihood function of the entire framework shown in Figure 1 is:

$$L(i, I|\beta,\theta,\lambda) = \int P(i|\beta, X)\, f_X(X;\theta) f_M(I|X;\lambda) dX \tag{9}$$

Finally, the unknown parameters (β,θ,λ) can be estimated using maximum likelihood estimation from observed modal choices.

3.2. – Models and goodness-of-fit measurements

Making the HCM framework operative requires assumptions on the functional form of the choice model, equation (6), the prior distribution of latent variables, equation (5), and the definition of the measurement models, equation (8); hence, models with different assumptions will typically result in different specifications, which some may be more appropriate than others under certain contexts. In this paper we define two competing HCM specifications as described below.

HCM1: latent variables with normal distribution and additive error model

This is how most of HCMs are specified in practice. First, the latent variable (LV) is assumed to be normally distributed, $LV \sim N(\mu_{normal}, \sigma^2_{normal})$. Second, the measurement equation is modelled as additive, and residuals are assumed to follow a normal distribution with zero mean. This formulation implies that indicators are centred on the true value of the latent variable; that errors are introduced by the measuring device; and that the magnitude of the error is independent from the value being measured. Mathematically, this model is expressed by equations (6)-(11).

$$U_1 = \beta_{t1} LVt_1 + \beta_c c_1 + asc_1 + \varepsilon_1 \quad (6)$$
$$LVt_1 = \mu_{normal} + \sigma_{normal}\phi \quad with\ \phi \sim N(0, 1^2) \quad (7)$$
$$t_{1\ \mathrm{measured}} = LVt_1 + \eta \quad (8)$$
$$\eta = \sigma_{error\_n}\phi' \quad with\ \phi' \sim N(0,1^2) \quad (9)$$
$$U_2 = \beta_{t2} t_2 + \beta_c c_2 + asc_2 + \varepsilon_2 \quad (10)$$
$$U_3 = \beta_{t3} t_3 + \beta_c c_3 + asc_3 + \varepsilon_3 \quad (11)$$

where U_i is the utility for alternative i={1,2,3}; β_{ti} the parameter for travel time, β_c the parameter for travel cost; LVt_1 the Latent Variable modelling t_1 with mean μ_{normal} , standard deviation σ_{normal}, and indicator $t_{1\ \mathrm{measured}}$; t_i the travel time; c_i the travel cost; asc_i the alternative specific constant; ε_i are iid errors from a standard Gumbel distribution; η the measurement error with parameter σ_{error_n} ; and, ϕ and ϕ' are draws from a standard normal distribution.

HCM2: latent variables with lognormal distribution and multiplicative error model.

Due to the nature of the latent variable being modelled (travel time), we define a second specification that exploits the fact that time variables must be positive. For this, we select the lognormal distribution, which has support on the required interval (0,+∞), and the additive measurement equation is replaced by a multiplicative one. This formulation implies that the error measurements are proportional to the values being measured, and can be interpreted as scaling factors; hence, it is critical to guarantee that the error term will always be positive. Again, this is achieved by modelling the error term with a lognormal distribution.

Finally, the formulation assumes that measurement errors will have an expected value of 1. This condition is similar to the additive errors having mean zero. In our particular case, the expected value of a lognormally distributed variable equals one as long as its parameters fulfil the following condition: $\sigma_{error\_ln} = \sqrt{-2\mu_{error\_ln}}$. This model is expressed by equations (12)-(17).

$$U_1 = \beta_{t1} LVt_1 + \beta_c c_1 + asc_1 + \varepsilon_1 \quad (12)$$
$$LVt_1 = \exp(\mu_{lognormal} + \sigma_{lognormal}\phi\ ) \quad with\ \phi \sim N(0, 1^2) \quad (13)$$
$$t_{1\ \mathrm{measured}} = LVt_1 * \eta \quad (14)$$

$$\eta \;= \exp\left(\mu_{error\_ln} + \sqrt{-2\mu_{error\_ln}}\phi'\right) \;\; with\; \mu_{error\_ln} < 0 \; and \; \phi' \sim N(0,1^2) \quad (15)$$

$$U_2 \;= \beta_{t2}t_2 + \beta_c c_2 + asc_2 + \varepsilon_2 \quad (16)$$

$$U_3 \;= \beta_{t3}t_3 + \beta_c c_3 + asc_3 + \varepsilon_3 \quad (17)$$

where again U_i is the utility for alternative i={1,2,3}; β_{ti} the parameter for travel time, β_c the parameter for travel cost; LVt_1 the Latent Variable modelling t_1 with mean μ_{normal} , standard deviation σ_{normal}, and indicator $t_{1\ measured}$; t_i the travel time; c_i the travel cost; asc_i the alternative specific constant; ε_i are iid errors from a standard Gumbel distribution; η the measurement error with parameter μ_{error_ln}; and, ϕ and ϕ' are draws from a standard normal distribution.

Goodness-of-fit and model section

Despite the increasing popularity of the HCM framework, discussion of the accuracy of measurement of latent variables in the HMC framework is largely absent in transportation research, [Motoaki and Daziano, 2015]. Unfortunately, traditional goodness-of-fit measurements of transport demand models, (e.g. likelihood ratio test and ρ^2) cannot be used to assess model fit, reliability, and validity of hybrid choice models; hence, researchers face difficulties to evaluate the accuracy of measurement of the latent variables, as there still is no consensus about how to test the HCM goodness-of-fit.

Common current practices to evaluate goodness-of-fit of HCMs come from studies in structural equation modelling, where there are established goodness-of-fit measurements such as the Bayesian Information Criterion (BIC) [Schwarz et al., 1978], and the Akaike Information Criterion (AIC) [Akaike, 1974] fit indexes. Hauber et al. (2016), presents this two indexes as

$$BIC = -2LL + ln[sample\ size]K, \quad (18)$$

$$AIC = -2LL + 2K, \quad (19)$$

where LL is the log-likelihood of the full model, and K is the number of parameter estimates corresponding to the number of explanatory variables in the model. These are comparative measures of the relative quality of statistical models for a given dataset, and they evaluate the plausibility of the models focusing on minimizing information loss. An advantage of using these indexes is that comparing multiple models becomes trivial, as the model with the lowest value is preferred. On the other hand, these goodness-of-fit measurements are based on the model's final likelihood; hence, when applied to HCM with different number of latent variables and or different measurement equation formulations, the goodness-of-fit measurements might provide counterintuitive results.

Some authors have expressed their concerns about these goodness-of-fit measurements. For instance [Barrett, 2007] and [Ropovik, 2015] state that fit indexes add anything to the analysis, and [Hayduk et al., 2007] argue that fit index thresholds can be misleading and subject to misuse. Moreover [Motoaki and Daziano, 2015] showed through a Monte Carlo experiment that the behaviour of SEM fit assessment tools did not work as expected for the HCM.

Nevertheless, is common to find fit indexes reported in current literature despite of these criticisms.

Other practice that is becoming increasingly popular is the use of out-of-sample prediction tests. This practice might be considered the most critical metric to evaluate how well the model does, when developing models for prediction. Therefore, I present in the next section the out-of-sample prediction accuracy percentages for the different models, calculated as the ratio of the sum of correctly predicted choices to the total number of choices in the test set.

4. RESULTS

For each of the 100 datasets, I estimate three different models using pythonBiogeme, [Bierlaire, M., 2016]. These models include one multinomial logit (MNL) model, and the two different HCMs defined in Section 3. All three models assume that the time variable for alternative 1 suffers from measurement errors. During these estimations, 13 HCMs with an additive error formulation and 11 HCMs with a multiplicative error formulation failed to converge when estimated on the 500 observations datasets. This is likely to be a parameter initialisation and simulation issue. Figures (3) and (4) show the empirical cumulative distribution function (ECDF) of the final log-likelihood for the three different models.

Figure 3. ECDF of the final log-likelihood. 100 datasets of ~2/3*1000 observations

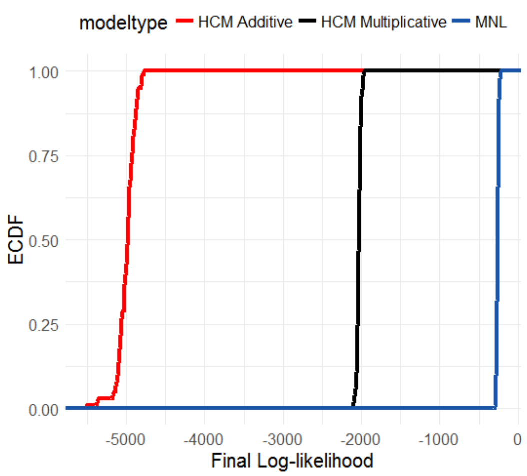
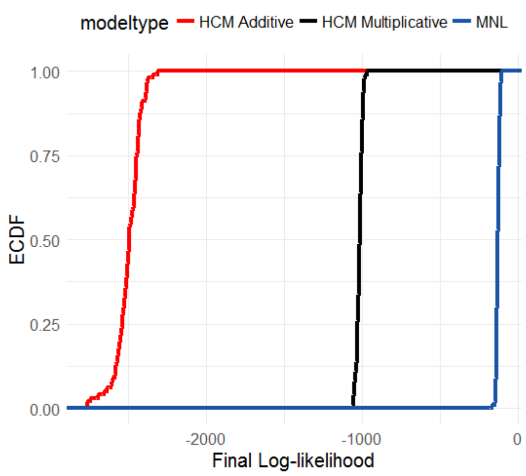


Figure 4. ECDF of the final log-likelihood. 100 datasets of ~2/3*500 observations



From Figures (3) and (4), we observe large differences in the final model log-likelihood depending on the type of model, the assumptions regarding the modelling assumptions of the HCMs, as well as the number of observations in the dataset. These differences are explained by the fact that HCMs are trying to explain not only the observed choice but also the probability of the latent variable model.

Parameter bias

Parameter bias is discussed around the ability of the models to recover the true values of time (*VTT*) for the different alternatives. I focus on the ability to recover ratios of parameters because differences among the estimated model parameter due to the model scales will cancel out. *Table 2* shows the sample mean and standard error of parameter estimates for the different alternatives and models that converged.

Table 2. Sample mean and standard error of parameter estimates

			Number of observations			
			~1000 * 2/3		~500 * 2/3	
Parameter	**True value**	**Model Type**	**Estimate**	**Std err**	**Estimate**	**Std err**
VTT_1	0.25	*MNL*	0.009	0.003	0.009	0.005
		HCM1	0.190	0.100	0.203	0.137
		HCM2	0.171	0.059	0.186	0.127
VTT_2	0.25	*MNL*	0.252	0.031	0.251	0.049
		HCM1	0.248	0.031	0.248	0.046
		HCM2	0.247	0.031	0.249	0.047
VTT_3	0.3	*MNL*	0.306	0.038	0.307	0.062
		HCM1	0.303	0.036	0.301	0.057
		HCM2	0.304	0.035	0.305	0.058

As can be seen in *Table 2*, the MNL model struggles to recover the true value of time for alternative 1 (*VTT_1)*. The mean of the estimated parameter for the 100 datasets is 0.009, which is heavily diluted when compared with its true value (0.25). Note that the variable *time1* was assumed to suffer from measurement errors.

Looking at the HCM parameter estimates, we can observe that the variance of the estimated parameters is larger for the datasets with fewer observations.

Out-of-sample prediction accuracy

Results from accuracy prediction on the test sets show that the two competing HCM specifications have a similar performance despite of their large difference in the final log-likelihood values observed in Figures (3) and (4). The mean and standard deviation of out-of-sample predictions for the different models and test sets is shown in *Table 3*.

Table 3. Out-of-sample prediction accuracy (sample mean and standard error).

	Number of observations			
	~1000 * 1/3		~500 * 1/3	
Model	**mean**	**Std err**	**mean**	**Std err**
MNL	0.82	0.02	0.82	0.03
HCM1	0.77	0.03	0.77	0.03
HCM2	0.78	0.02	0.77	0.04

Surprisingly, the out-of-sample prediction accuracy for the simpler MNL model surpasses both HCM specifications. MNL average accuracy is 82%, whilst HCM specifications have an accuracy of 77% and 78%. A possible explanation is included in [Vij and Walker, 2016], where the authors argue that the latent variables help introduce correlation between the choice and measurement indicators, but the correlation thus imposed may lead to a loss in the ability of the model to predict outcomes to the choice indicators.

5. CONCLUSIONS

Model misspecification is likely to occur when working with real datasets. Despite of this fact, previous empirical studies showing the advantages of the HCMs have mostly used correctly specified models, especially when parameter estimate biases are discussed. However, the paper at hand explores the extent of parameter bias in misspecified HCMs, and evaluates the prediction accuracy of misspecified HCMs in comparison with a simpler, also misspecified, multinomial logit through Monte Carlo experiments using synthetic data. Efforts are made on mimicking the conditions present when real datasets are used, and the assumed structural and measurement equations are less flexible than the functions used for the data generating process.

The study has shown that parameter estimates of the MNL models are heavily diluted in the presence of measurement errors. This finding is contrary to the current stream of knowledge regarding the robustness of MNL to misspecification, [Cramer, 2007]. A potential explanation might be that part of the measurement errors in this study comes from a lognormal distribution that has thicker tales, and expected value different from zero, as opposed to the majority of studies where measurement errors are symmetric with zero mean. Regarding the ability of HCMs to recover the true parameter estimates, we see how HCMs manage to recover parameter estimates that are closer to the true values than the ones provided by the MNL, given that the number of observations is sufficient. Hence, further analysis should be carried out to understand what requirements the dataset should satisfy, in order to benefit from a HCM formulation.

In terms of robustness of parameter estimates to modelling assumptions of the HCMs, results show that parameter estimates of the choice model seem robust to modelling assumptions required to estimate the HCM. Furthermore, we see that all models, -the multinomial logit and the hybrid choice models-successfully isolate the source of model bias preventing its propagation to other parameter estimates.

Regarding overall goodness-of-fit measurements, results show that both HCM formulations recover similar parameter estimates despite of having large differences in the final model log-likelihood. Also, parameters of the HCMs present smaller biases than parameters from the MNL despite of having larger final log-likelihood values. This is explained by the fact that the HCMs are not only trying to explain the observed choices, but also the probability of the latent variable model. The final log-likelihood of HCMs is impacted by the number of latent variables, the functional forms adopted for the latent variable and measurement errors, and the number of measurement equations per latent variable [Varela et al., 2018]; hence goodness-of-fit measurements such as the BIC and AIC are not always informative.

Out-of-sample prediction accuracy results show that the simpler MNL has a higher accuracy (82%) than any of the more advanced HCM formulations (77% and 78%). This finding supports [Vij and Walker, 2016] statement, where the authors argue that the correlation between the choice and measurement indicators introduced by the latent variable may lead to a loss in the ability of the model to predict outcomes to the choice indicators.

These results advocate for a careful use of advance modelling techniques such as the HCM framework, and highlights that it is necessary to further understand the implications of these types of models on policy analysis. For instance, after accounting for measurement errors in travel times, users appear to be more sensitive to travel time changes, and have higher values of time. Similar findings have been reported by [Walker et al.,2010], and [Varotto et al., 2017]. Hence, it is important that practitioners and decision makers keep a critical attitude towards parameter estimates and predictions from Hybrid Choice Models.

Finally, this paper draws attention into current practices used on testing hybrid choice models, where there is lack of a rigorous evaluation of its results, and it is not unusual to find selective reporting of fit indices: hence, the development of a common framework to assess the performance of HCM should be prioritised.

6. COMPLIANCE WITH ETHICAL STANDARDS

Conflict of interest: On behalf of all authors, the corresponding author states that there is no conflict of interest.

7. REFERENCES

Ben-Akiva, M., McFadden, D., Train, K., Walker, J., Bhat, C., Bierlaire, M., & Daly, A., 2002. Hybrid choice models: progress and challenges. Marketing Letters, 13(3), 163-175.

Bierlaire, M., 2016. pythonbiogeme: a short introduction. report transp-or 160706, series on biogeme. transport and mobility laboratory, school of architecture, civil and environmental engineering, ecole polytechnique fédérale de lausanne, switzerland.

Bollen, K. A., Kirby, J. B., Curran, P. J., Paxton, P. M., and Chen, f., 2007. Latent variable models under misspecification two-stage least squares (2sls) and maximum likelihood (ml) estimators. sociological methods and research, 36(1), 48-86. doi: 10.1177/0049124107301947.

Browne, M., and Cudeck, R., 1992. Alternative ways of assessing model fit. Sociological methods and research, vol. 21, no. 2, pp. 230–258.

Barrett, P., 2007. Structural equation modeling: adjudging model fit. pers. individ. dif. 42, 815–824.doi:10.1016/j.paid.2006.09.018.

Cramer, J. S. 2007. Robustness of Logit Analysis: Unobserved Heterogeneity and Mis-specified Disturbances. Oxford Bulletin of Economics and Statistics, 69(4), 545-555.

Hauber AB, González JM, Groothuis-Oudshoorn CG, Prior T, Marshall DA, Cunningham C, et al. 2016. Statistical methods for the analysis of discrete choice experiments: a report of the ISPOR Conjoint Analysis Good Research Practices Task Force. Value Health.

Hayduk, L.A., Cummings, G., Boadu, K., Pazderka-Robinson, H., and Boulianne, S., 2007. Testing! testing! one, two, three – testing the theory in structural equation models! pers. individ. dif. 42, 841–850.doi:10.1016/j.paid. 2006.10.001.

Kolenikov, S., 2011. Biases of parameter estimates in misspecified structural equation models. Sociological methodology, 41(1), 119–157. doi:10.1111/j.1467-9531.2011.01236.x

Motoaki, Y. and Daziano, R.A., 2015. Assessing goodness of fit of hybrid choice models. transp. res. rec. j. Transp. res. board. 2495, 131–141.

Roeckelein, J.E., 2000. The Concept of Time in Psychology: A Resource Book and Annotated Bibliography. Greenwood Press, Westport.

Ropovik, I., 2015. A cautionary note on testing latent variable models. Front. psychol.6:1715. doi:10.3389/fpsyg.2015.01715.

Varela, J. M. L., Börjesson, M., & Daly, A., 2018. Quantifying errors in travel time and cost by latent variables. Transportation Research Part B: Methodological, 117, 520-541.

Varotto, S. F., Glerum, A., Stathopoulos, A., Bierlaire, M., and Longo, G., 2017. Mitigating the impact of errors in travel time reporting on mode choice modelling, in journal of transport geography, volume 62, 2017, pages 236-246, issn 0966-6923, https://doi.org/10.1016/j.jtrangeo.2017.05.016.

Vij, A. and Walker, J., 2016. How, when and why integrated choice and latent variable models are latently useful. TR part B. 90, 192–217.

Walker, J., Li, J., Srinivasan, S., and Bolduc, D., 2010. Travel demand models in the developing world: correcting for measurement errors. transportation letters 2, 231–243.