



Endogeneity and Measurement Bias of the Indicator Variables in Hybrid Choice Models: A Monte Carlo Investigation

Wiktor Budziński¹ · Mikołaj Czajkowski¹

Accepted: 22 May 2022 / Published online: 8 August 2022
© The Author(s) 2022

Abstract

We investigate the problem of endogeneity and measurement bias arising from incorporating indicator variables (e.g., measures of attitudes) into discrete choice models. We demonstrate that although a hybrid choice framework can resolve both endogeneity and measurement problems, the former requires explicit accounting for in the model, which has not typically been done in applied studies to date. By conducting a Monte Carlo experiment, we demonstrate the extent of the bias resulting from measurement and endogeneity problems. We propose two novel solutions to address the endogeneity problem: explicitly accounting for correlation between structural and discrete choice component error terms (or with random parameters in a utility function), or introducing additional latent variables. Using simulated data, we demonstrate that these approaches work as expected, i.e. they successfully recover the true values of all model parameters.

Keywords Attitudinal variables · Endogeneity · Hybrid choice models · Indicator variables · Measurement error

JEL Classification C35 · C51 · Q51 · R41

1 Introduction

A hybrid choice (HC) model is a flexible tool that incorporates perceptions and cognitive processes into a random utility framework commonly used to model individuals' choices. Indicator variables used to measure psychological or sociological constructs enter the model through latent variables, rather than being directly interacted with choice attributes. The HC model can therefore be viewed as a combination of a classical discrete choice model, such as the mixed logit model (MXL, Revelt and Train 1998), with a Multiple

✉ Wiktor Budziński
wbudzinski@wne.uw.edu.pl

Mikołaj Czajkowski
mc@uw.edu.pl

¹ Faculty of Economic Sciences, University of Warsaw, Długa 44/50, 00-241 Warsaw, Poland

Indicators, Multiple Causes (MIMIC) model (Jöreskog and Goldberger 1975). The former links some assumed decision process (e.g., utility maximization) and observed explanatory variables (attributes of alternatives, socio-demographics) with observed choices, whereas the latter identifies latent factors linked with observed indicator variables, for example, answers to attitudinal survey questions.

The HC framework has been extensively used over the last decade to better understand the attitudes and psychological factors that drive individuals' preferences toward non-market goods and policies. In the environmental context, applications include stated preference studies on individuals' choices regarding coastal water quality improvements, land-use policies, conservation policies, and recycling rules (Hess and Beharry-Borg 2012; Hoyos et al. 2015; Lundhede et al. 2015; Mariel et al. 2015; Bartczak et al. 2016; Czajkowski et al. 2017b; Boyce et al. 2019; Zawojka et al. 2019). The latent factors can represent a wide range of psychological measures, such as attitudes toward chargeable policy, awareness of consequences, outcome uncertainty, risk preferences, social norms, morals, personality, and perceived survey consequentiality.

The popularity of the HC model stems from the fact that including indicator variables directly in a choice model is considered methodologically flawed because of the lack of causality, dependence on survey question framing (Ben-Akiva et al. 2002), as well as potential measurement error and endogeneity (Guevara 2015). With regard to endogeneity, the usually acknowledged cause of this issue is that responses to attitudinal questions are likely correlated with other unobserved factors, which, if not accounted for, will end up in the error term of the random utility model (Daly et al. 2012; Hess and Beharry-Borg 2012; Kløjgaard and Hess 2014; Hoyos et al. 2015). It is widely believed that the HC model resolves both issues – measurement error and endogeneity (Daly et al. 2012; Hess et al. 2013; Hess and Stathopoulos 2013; Kløjgaard and Hess 2014; Bello and Abdulai 2015). We show that while HC models do indeed account for measurement error, endogeneity can arise from other sources (e.g., omitted variables, simultaneous determination) and requires explicit accounting for in the specification of the model – something that has not typically been addressed in the applied choice modeling literature.

Endogeneity arises when explanatory variables in the model are not independent of the stochastic term, which leads to biased parameter estimates. Incorporating indicator variables directly into the choice model, may lead to the correlation of explanatory variables and the stochastic term, because indicator variables are usually not direct measures of latent constructs but rather their functions (Guevara 2015). Walker et al. (2010) and Vij and Walker (2016) show that measurement error, leading to endogeneity, can be accounted for in a straightforward way by using the HC framework and incorporating measurement errors directly into the measurement equation (see also the equation (4) in Sect. 2.2). Campbell and Sandorf (2020) employ simulations to further investigate the performance of HC models while controlling for different strengths of relationship between latent factors, indicator variables, and choices. They find that if these relationships are weak then the advantage of using HC models is greatly diminished. In some cases, putting indicator variables directly into the choice model works just as well, measurement errors notwithstanding.

However, endogeneity of indicator variables can also arise from other sources than measurement error. For example, Ben-Akiva et al. (2002) and Guevara (2015) show that endogeneity and inconsistent estimates in HC models may be caused by the omission of important attitudes or explanatory variables. Guevara and Polanco (2016) discuss the endogeneity of indicator variables caused by missing covariates and simultaneous determination. Chorus and Kroesen (2014) list reasons why attitudes or perceptions can be rendered endogenous (other than indicator variables) and discuss the implications for the practical

use of HC models. In what follows we investigate whether the HC framework can address such issues by accounting for correlations between error terms in different parts of the model. For clarity, we use the term endogeneity – in line with the general tenor of the literature – to refer to the impact of factors other than measurement error (such as omitted variables).

Previous studies have utilized HC models to address the issue of omitted variables and the endogeneity that arises because of it. The usual example given is the case of the endogenous price of a good when the quality attribute is missing (Palma et al. 2016). The usefulness of the HC framework in such cases was also demonstrated by simulations reported by Vij and Walker (2016). However, the setting studied in this paper is different. In the example above, the price is endogenous, and the HC framework is used to solve the endogeneity by imputing the missing attribute (quality) as a latent factor. The latent factor is nevertheless assumed to be exogenous. In the current paper, we analyze the situation where the latent factor, or the indicators that are used to measure it, are endogenous. We, therefore, address the concerns raised by Chorus and Kroesen (2014). Our focus in this study also differs from Campbell and Sandorf (2020). Rather than evaluating the performance of the HC model with data of varying quality (for example, indicators that are weakly correlated with a latent factor), we assume that the data are well fitted for the application of the HC framework – that is the latent variable strongly affects individuals' choices, the latent variable is well measured by the indicator variables, and there is sufficient data for reliable estimation. We investigate how HC models are affected by the endogeneity bias under such conditions.

Overall, our study contributes to the current literature by analyzing the problem of the endogeneity of indicator variables in HC models. We present a Monte Carlo simulation to demonstrate how different types and sources of endogeneity and measurement errors are, or are not, accounted for in a typical HC model and how this affects the results (bias). We show that unless a correlation of the error terms is explicitly accounted for, the HC framework by itself does not solve the problem of the endogeneity of indicator variables (or latent factors) caused by omitted variables, and the resulting estimates are biased. We then propose two methods of accounting for endogeneity in HC models and demonstrate that they successfully recover the true values of the coefficients. We also contribute to the non-market valuation literature that utilizes the HC framework to investigate how certain attitudes or perceptions affect individuals' willingness to pay (WTP) for a given policy approach. We show that accounting for measurement error and unobserved heterogeneity is important for the estimation of mean WTP. We also confirm the results of previous research indicating that HC models can account for measurement error.¹ In addition, we demonstrate that identifying the true relationship between latent variables and WTP requires the endogeneity problem to be addressed. Overall, our results are important for applied studies which use HC models to discover how psychological factors such as social norms, consequentiality, and environmental attitudes affect welfare measures and policy support.

¹ In contrast to previous research (Walker et al. 2010, Vij and Walker 2016), we employ a specification that is more common in environmental economics, in which latent variables are interacted with choice attributes. Previous studies have considered latent factors which could be treated as attributes (e.g., travel costs). Furthermore, we show that measurement error cannot be accounted for by incorporating unobserved preference heterogeneity (i.e. random parameters).

The remainder of the paper is structured as follows. Section 2 presents the general econometric framework of the hybrid choice model. Section 3 describes the design of our Monte Carlo experiment, the data generating process (DGP), the models we compare, and the methodology of comparisons. Next, the results are presented and interpreted in detail. The last section provides a summary and discussion, acknowledges the limitations of our study, and concludes with recommendations for future research. In addition, the software codes and supplementary materials are available online to make the use of hybrid choice models for empirical studies easier and to facilitate future research.

2 The Hybrid Choice Model Framework

HC models (Ben-Akiva et al. 2002) can consist of up to three parts: a discrete choice model, measurement component, and structural component. We describe each part in detail to set the scene for the empirical illustration that follows.

2.1 Discrete Choice Model

The discrete choice component of the HC describes individuals' decision processes when making a choice. It is usually based on random utility theory, although other decision processes are proposed in the literature, such as random regret (Kim et al. 2017) or a mixture of random regret and random utility (Hess and Stathopoulos 2013). In what follows we employ the random utility model, as it is the most common approach. The utility (V) gained by individual i from choosing alternative j in choice situation t depends on the vector of observed characteristics (\mathbf{X}) and unobserved idiosyncrasies, represented by the stochastic component e :

$$V_{ijt} = X_{ijt}\beta_i + e_{ijt}, \quad (1)$$

where β_i denotes a vector of individual-specific parameters, thus allowing for heterogeneous preferences amongst respondents and leading to a mixed logit model.² The stochastic component of the utility function (e_{ijt}) is assumed to follow an i.i.d. type I extreme value distribution with constant variance $\text{var}(e_{ijt}) = \pi^2/6$. This normalization of the variance term is required for identification.

The HC model allows random parameters to be a function of individual-specific latent variables, denoted by \mathbf{LV}_i (e.g., pro-environmental attitude), socio-demographic (e.g., income) or other directly observable variables (such as information treatments in the survey) collected in the vector \mathbf{SD}_i .³ For a normally distributed β_i , this dependence can be specified in the following way:

² Is it typically assumed that individual parameters follow a particular distribution (possibly multivariate distribution allowing for non-zero correlation of model parameters), rather than being separately estimated for each parameter. The distributions can be continuous, leading to a known random parameter model, or discrete, resulting in a latent class model. Assuming instead that parameters are the same for all respondents implies homogenous preferences and leads to a multinomial logit model (MNL) as a special case.

³ There are other possible specifications, in which latent variables enter the choice model differently. For example, they can explain class probabilities in the latent class model or a variance of the error term (scale). We use this specification as we find it to be the most straightforward way of connecting attitudes with individuals' tastes.

$$\beta_i = \beta_i^* + \mathbf{LV}_i\boldsymbol{\gamma} + \mathbf{SD}_i\boldsymbol{\varphi}, \quad (2)$$

where $\boldsymbol{\gamma}$ and $\boldsymbol{\varphi}$ are matrices of estimable coefficients and β_i^* has a multivariate normal distribution with a vector of means and a covariance matrix to be estimated.⁴ As a result, the conditional probability of individual i making choices y_i , for all T_i choice tasks, is given by:

$$P(y_i|\mathbf{X}_i, \beta_i^*, \mathbf{LV}_i, \boldsymbol{\gamma}, \boldsymbol{\varphi}) = \prod_{i=1}^{T_i} \frac{\exp(\mathbf{X}_{ij}\beta_i)}{\sum_{k=1}^C \exp(\mathbf{X}_{ik}\beta_i)}. \quad (3)$$

2.2 Measurement Component

The main purpose of including latent variables in an HC model is the belief that they are describing some behavioral or other factors, which cannot be measured directly (unlike, e.g., age or gender). Instead, various indicators are used, which are assumed to be determined by the latent variables. The choice of the model for the indicator variables depends on the particular application. The measurement equations could be linear, ordered, binary, multinomial, or count regressions, whatever best fits the interpretation of each indicator. Throughout the simulation that follows, we will use continuous indicator variables and therefore we assume a linear specification of the form:

$$\mathbf{I}_i = \mathbf{LV}_i\boldsymbol{\Gamma} + \mathbf{X}_i^{Mea}\boldsymbol{\Phi} + \boldsymbol{\eta}_i, \quad (4)$$

where \mathbf{I}_i is a vector of indicator variables, \mathbf{X}_i^{Mea} is a vector of additional variables that influence indicator variables, but not through the latent variable itself (Ben-Akiva et al. 2002),⁵ $\boldsymbol{\Gamma}$ and $\boldsymbol{\Phi}$ are matrices of coefficients and $\boldsymbol{\eta}_i$ denotes a vector of error terms assumed to come from a multivariate normal distribution with 0 means. Essentially, we assume that indicators \mathbf{I}_i are driven by (and hence used to measure) unobserved latent variables \mathbf{LV}_i and potentially also by some other observed individual-specific characteristics \mathbf{X}_i^{Mea} while allowing for measurement errors, represented by the error component $\boldsymbol{\eta}_i$.

2.3 Structural Component

Latent variables can also directly depend on exogenous factors, such as socio-demographic variables, which are stacked in the vector \mathbf{X}_i^{str} . This relationship is described by the following structural equation:

$$\mathbf{LV}_i = \mathbf{X}_i^{str}\boldsymbol{\Psi} + \boldsymbol{\xi}_i, \quad (5)$$

with a matrix of coefficients $\boldsymbol{\Psi}$ and error terms $\boldsymbol{\xi}_i$, which are typically assumed to come from a multivariate normal distribution. The vector \mathbf{X}_i^{str} should overlap with vectors \mathbf{SD}_i

⁴ In the simulation that follows we assume that the coefficient for cost follows log-normal distribution, namely $\beta_i = -\exp(\beta_i^* + \mathbf{LV}_i\boldsymbol{\gamma} + \mathbf{SD}_i\boldsymbol{\varphi})$. This is a standard assumption in studies which calculate willingness to pay.

⁵ For example, some individuals may have a tendency to overstate (or understate) their real attitudes.

and \mathbf{X}_i^{Mea} to account for direct and indirect (i.e. through latent factors) effects of socio-demographic variables on individuals' choices and indicators.

2.4 Identification

In order to make an identification of hybrid choice models possible, the scale of latent variables needs to be normalized (Daly et al. 2012). This can be done by normalizing variances of the error terms in structural equations or by normalizing some coefficients in the $\mathbf{\Gamma}$ matrix for each latent variable (Raveau et al. 2012). In this study, we adopt the former approach. In contrast to most studies conducted to date, we do not normalize the variance of ξ_i to one. Instead, we use normalization to ensure that the (unconditional) variance of every latent variable in \mathbf{LV}_i is equal to one. Although such an approach introduces additional nonlinearities into the model, it is quite useful. As all latent variables now have the same scale, assessing their relative importance in the choice model and measurement equations is straightforward. Furthermore, as the scale of the latent factor is fixed at 1 even with socio-demographic variables in structural equations, the effect of the latent variable on preferences should remain stable when covariates are added to, or removed from, the structural equation. We find this convenient for testing the robustness of the HC model specification in practical applications. We have not observed any additional issues with convergence due to this normalization.

We formally define $\mathbf{LV}_i^* = \mathbf{X}_i^{str} \mathbf{\Psi}^* + \xi_i^*$, with $\mathbf{\Psi}^*$ being a matrix of parameters to be estimated and ξ_i^* being a vector of independent normally distributed variables with mean zero and unit standard deviation. For $\mathbf{LV}_{\bullet k}^*$, representing a vector of values of the k -th non-normalized latent variable for all individuals and $\delta_k = std(\mathbf{LV}_{\bullet k}^*)$ representing its standard deviations, we have $\mathbf{LV}_{\bullet k} = \mathbf{LV}_{\bullet k}^* / \delta_k$, $\mathbf{\Psi}_k = \mathbf{\Psi}_k^* / \delta_k$ and $\xi_{\bullet k} = \xi_{\bullet k}^* / \delta_k$.⁶

Unfortunately, the exact conditions for the identification of the HC model are not yet known; they depend on the number of latent variables and measurement equations (Bahamonde-Birke et al. 2015) and need to be analyzed on a case-by-case basis (Ben-Akiva et al. 2002). We follow Bollen and Davis (2009) to ensure that the necessary condition for the identification of structural equation models holds; our specifications satisfy the "2+ emitted paths rule" (we assume that each latent variable has two unique indicators in the measurement component and is interacted with three attributes in the discrete choice component).⁷

2.5 Estimation

Finally, we combine the discrete choice model specified in (3), the measurement equations defined in (4), and structural equations described in (5) to obtain the full-information likelihood function for the HC model (for ease of exposition, we stack the parameters $\gamma, \varphi, \Gamma, \Phi, \Psi$, as well as parameters of the assumed distribution of (β_i^*, ξ_i^*) denoted by θ , into Ω):

⁶ ψ_k denotes k -th row of Ψ matrix, and $\xi_{\bullet k}$ denotes stacked values of the random term in the k -th structural equation for all individuals.

⁷ One exception is Model 9 (see below), which uses two latent variables and two indicators. In this case we have also used simulation to confirm that the model is identified and produces stable results.

$$L_i = \int P(y_i | \mathbf{X}_i, \mathbf{X}_i^{str}, \boldsymbol{\beta}_i^*, \boldsymbol{\xi}_i^*, \boldsymbol{\Omega}) P(I_i | \mathbf{X}_i^{str}, \boldsymbol{\xi}_i^*, \mathbf{X}_i^{Mea}, \boldsymbol{\Omega}) f(\boldsymbol{\beta}_i^*, \boldsymbol{\xi}_i^* | \boldsymbol{\theta}) d(\boldsymbol{\beta}_i^*, \boldsymbol{\xi}_i^*). \tag{6}$$

As random disturbances of $\boldsymbol{\beta}_i^*$ and (non-normalized) error terms in structural equations $\boldsymbol{\xi}_i^*$ are not directly observed, they must be integrated out of the conditional likelihood. This multidimensional integral can be approximated using a simulated maximum likelihood approach. As can be seen, we use one-step estimation. This approach has two main advantages over a two-step (or multi-step) method. First, the two-step method can lead to inefficient or even inconsistent estimates. In order to obtain consistent estimates, researchers would need to account for measurement error and integrate the choice probability over its distribution (Ben-Akiva et al. 2002). Second, one-step estimation allows for the identification of more flexible specifications because it has more degrees of freedom.

3 The Setup of the Monte Carlo Investigation into the Effects of Endogeneity

In this section, we first describe the Monte Carlo simulation employed in the study. We then illustrate how endogeneity and measurement bias can arise when utilizing indicators or latent factors in the choice model using our data generating process as an example. This ties directly into the model’s specifications which we will compare using simulated data to investigate the extent of measurement and endogeneity bias. Lastly, we describe the methods used to compare the results of different model specifications.

3.1 Data Generating Process

The DGP we selected is relatively simple and mimics the usual settings of stated preference-based discrete choice data. The discrete choice consists of three choice alternatives and six choice tasks per respondent. It includes three attributes: a binary variable SQ_{ijt} representing an alternative specific constant for the first (status quo) alternative, and two continuous attributes $Quality_{ijt}$ and $Cost_{ijt}$, assumed to always equal 0 for the status quo alternative, and distributed (independently) uniformly between 0 and 2. Each artificial sample consists of 1,000 individuals. The individual-specific explanatory variables X_i^{SD} and X_i^{Miss} were assumed to have a standard normal distribution. Table 1 describes the details of the DGP.

We assume that the preference heterogeneity is driven by the latent factor and the individual-specific X_i^{Miss} variable. We did not include additional unobserved heterogeneity in the form of random parameters for two reasons. First, it helps to demonstrate that incorporating unobserved heterogeneity helps neither with measurement bias nor with endogeneity. Second, it facilitates the simulation, as the benchmark model can be estimated in less time.

In our simulation, endogeneity is caused by omitting the X_i^{Miss} variable, which causes the error terms to become correlated. Specifically, because X_i^{Miss} enters not only the choice model but also the structural component of the latent variable, excluding it causes the error terms of structural and discrete choice components to become correlated ($cor(\xi_i^*, e_{ijt}^*) \neq 0$)

Table 1 Description of the data-generating process used for Monte Carlo simulations

Choice model	
Utility function	$V_{ijt} = \beta_{1i}SQ_{ijt} + \beta_{2i}Quality_{ijt} - \beta_{3i}Cost_{ijt} + e_{ijt}$ $SQ_{ijt} \in \{0, 1\}$ $Quality_{ijt} \sim U(0, 2)$ $Cost_{ijt} \sim U(0, 2)$ $e_{ijt} \sim EV_I(0, 1)$
	$\beta_{1i} = \alpha_{11} + \alpha_{12}LV_i + \alpha_{13}X_i^{Miss}$ $\beta_{2i} = \alpha_{21} + \alpha_{22}LV_i + \alpha_{23}X_i^{Miss}$ $\beta_{3i} = \exp(\alpha_{31} + \alpha_{32}LV_i + \alpha_{33}X_i^{Miss})$ $X_i^{Miss} \sim N(0, 1)$
Latent variable (structural component)	Indicator variables (measurement component)
$LV_i^* = \alpha_{61}X_i^{SD} + \alpha_{62}X_i^{Miss} + \xi_i$ $\xi_i \sim N(0, 1)$ $X_i^{SD} \sim N(0, 1)$	$I_{i1} = \alpha_{41} + \alpha_{42}LV_i + \alpha_{43}\eta_{i1}$ $I_{i2} = \alpha_{51} + \alpha_{52}LV_i + \alpha_{53}\eta_{i2}$ $\eta_{i1} \sim N(0, 1)$ $\eta_{i2} \sim N(0, 1)$

⁸ Note that even though for convenience we use missing variables to cause endogeneity and analyze its effects, our results are not limited to this scenario. The results would be qualitatively the same in cases where there are other reasons for the correlation of these error terms.

To investigate the effects of endogeneity under various model specifications we designed and conducted a Monte Carlo simulation. Essentially, for the DGP presented in Table 1 we simulated artificial data and investigated how well different model specifications perform in terms of recovering the original parameters. The process was repeated multiple times to make sure the results were not coincidental. This exercise allowed us to clearly illustrate the theory and demonstrate that some specifications suffer from the endogeneity problem (and hence result in biased estimates). We were also able to demonstrate how the problem can be controlled for.

The literature is not always clear on what exactly is meant by the endogeneity of indicator variables. Studies that consider the endogeneity of indicator variables rarely discuss the underlying latent factors. A notable exception is Chorus and Kroesen (2014), who consider underlying attitudes or perceptions, rather than the indicator variables themselves, to be endogeneous. This is also a framework that we follow in the current study. To highlight the fact that in our simulations the latent factor is the source of endogeneity, we label it as “LV-endogeneity,” which means that LV is correlated with an error term in the choice model. Furthermore, as noted in the Introduction, in what follows we consider endogeneity caused by a measurement error as a separate case. Clearly, this arises only when using indicator variables directly in the choice model and is distinct from the LV-endogeneity induced by

⁸ In what follows, unless stated otherwise, we denote error terms in models with a missing variable by adding a star.

some omitted variables.⁹ The measurement errors should be accounted for by use of the HC framework, however LV-endogeneity could still cause estimates to be biased.

There is also another way in which omitted variables could cause the HC model to be endogenous. Specifically, if an omitted variable entered the measurement equations instead of the structural equation (as in Table 1), it would cause the error terms in the measurement equations to be correlated with the error terms in the discrete choice model. We denote this as M-endogeneity. It occurs when the same unobserved factor influences measurement errors and individual choices. For example, in stated preference studies “yea-saying” could make individuals overstate their real attitudes (whether the indicator questions are framed positively or negatively) as well as make individuals more likely to choose costly improvement alternatives. To streamline the argument we limit the presentation of the results in the main text to LV-endogeneity. We suspect that LV-endogeneity may be more prevalent in practice, and therefore we make it the focus of the current study. Nonetheless, the question of how often M-endogeneity occurs in applied research is of course empirical and beyond the scope of this study. For this reason, we report the results and conclusions for M-endogeneity in Appendix C.

Finally, we note that our study relates to endogeneity caused by indicator variables or latent factors. If we were to estimate a regular discrete choice model with no indicators and latent factors (such as mixed logit), it would result in unbiased estimates of preferences and welfare measures (e.g., mean WTP). Such a model would constitute a reduced form specification of the DGP presented in Table 1. It would, however, have the limitation of not allowing us to determine the effect of the latent variables under consideration on individuals’ preferences and choices.

3.2 Endogeneity and Measurement Bias in Indicator and Latent Variables

We use different model specifications to examine whether the presence of endogeneity affects results and to what extent this can be controlled for. Table 2 below provides a summary of the model specifications that we consider. In what follows we describe each model in more detail and provide some intuition behind the expected biases. We provide a full mathematical formulation of each model specification in Appendix A.

The first model (Model 1) reflects the data-generating process presented in Table 1 with no missing variables. It is used to test if we are able to correctly recover the parameters when there is no endogeneity present and measurement error is accounted for by the HC framework.

Model 2 is a simple multinomial logit (MNL) intended to capture the effect of measurement error on parameter estimates. Measurement error arises when the measurement of a certain independent variable is not exact. Consider the data generating process presented in Table 1 as an example. There, the true variable of interest is a latent variable, LV_i , which affects the marginal utilities of the choice attributes. Indicator variables, I_{i1} and I_{i2} , can be considered approximations of this latent factor, although they contain measurement errors,

⁹ Chorus and Kroesen (2014) list missing variables that influence both latent variables and an individual’s choices as one of the most probable causes of endogeneity in latent variables. Other causes include learning effects, which may cause a bi-directional dependence between experience with attributes and their perceptions, and the aligning of individuals’ attitudes with their actual choices in order to appear consistent (Ariely, Loewenstein and Prelec, 2003). Furthermore, simultaneous determination and model misspecification may also render a variable endogenous.

η_{i1} and η_{i2} . Due to these errors, imputing I_{i1} and/or I_{i2} into the model instead of LV_i would result in biased parameter estimates. For Model 2 we consider the following simple choice model with an indicator variable put directly in the utility function:

$$\begin{cases} V_{ijt} = \beta_{1i}SQ_{ijt} + \beta_{2i}Quality_{ijt} - \beta_{3i}Cost_{ijt} + e_{ijt}^*(\eta_{i1}) \\ \beta_{1i} = \alpha_{11} + \alpha_{12}I_{i1} + \alpha_{13}X_i^{Miss} \\ \beta_{2i} = \alpha_{21} + \alpha_{22}I_{i1} + \alpha_{23}X_i^{Miss} \\ \beta_{3i} = \exp(\alpha_{31} + \alpha_{32}I_{i1} + \alpha_{33}X_i^{Miss}) \end{cases} \quad (7)$$

In (7), the error term $e_{ijt}^*(\eta_{i1})$ becomes a function of the measurement error to accommodate it in the model.¹⁰ As both I_{i1} and e_{ijt}^* contain η_{i1} , they become correlated, which leads to endogeneity. In such a setting the measurement error will lead to incorrect coefficient estimates in Model 2. We primarily expect coefficients α_{12} , α_{22} , and α_{32} to be biased, as they control for the effect of the latent factor on marginal utilities. Nonetheless, as choice models are highly nonlinear, and the indicator variable enters the model as an interaction with choice attributes, all the coefficients could be biased due to endogeneity. Our simulation allows us to investigate which coefficients of the model are actually affected.

Model 3 is an extension of Model 2, in which we add random parameters for all attributes, with a full correlation matrix.¹¹ A mixed logit (MXL) model like this has become state-of-the-practice in the choice modeling literature, it is therefore of interest to investigate whether it could be used to control for measurement bias that arises upon incorporating indicator variables directly into the model.

In Models 4 – 9 we omit X_i^{Miss} as if it was unobserved, and hence we induce LV-endogeneity. Model 4 is then MNL analogous to Model 2 considered in (7) but with X_i^{Miss} missing. It, therefore, suffers from both measurement and endogeneity bias. Comparison with Model 2 will allow us to observe how endogeneity affects the results when measurement error is present. Model 5 adds fully correlated random parameters for all attributes which can capture some of the measurement error as in Model 3, but the model can also capture unobserved preference heterogeneity caused by the omitted X_i^{Miss} variable.

The next four models that we consider account for measurement errors by directly incorporating them into measurement equations in the HC framework (see equation (4)). Because of that, we expect that they will not suffer from the endogeneity bias caused by measurement errors (Walker et al. 2010).

Model 6 is an HC model which suffers from endogeneity due to a missing variable. The specification is the same as in Model 1, but without the X_i^{Miss} variable:

$$\begin{cases} V_{ijt} = \beta_{1i}SQ_{ijt} + \beta_{2i}Quality_{ijt} - \beta_{3i}Cost_{ijt} + e_{ijt}^*(X_i^{Miss}) \\ \beta_{1i} = \alpha_{11} + \alpha_{12}LV_i \\ \beta_{2i} = \alpha_{21} + \alpha_{22}LV_i \\ \beta_{3i} = \exp(\alpha_{31} + \alpha_{32}LV_i) \\ LV_i = \alpha_{61}X_i^{SD} + \xi_i^*(X_i^{Miss}) \end{cases} \quad (8)$$

¹⁰ Because the coefficient for Cost is nonlinear function of the latent factor (cf. Table 1), $e_{ijt}^*(\eta_{i1})$ will also be a nonlinear function of the measurement error.

¹¹ In all models with random parameters, the random parameters for *SQ* and *Quality* follow normal distribution, whereas the parameter for *Cost* follows log-normal distribution.

Table 2 Overview of model specifications used to investigate endogeneity bias in controlled and uncontrolled cases

	Model type	Measurement error	LV-endogeneity
Model 1	Hybrid MNL	Controlled	No
Model 2	MNL	Yes	No
Model 3	MXL	Yes	No
Model 4	MNL	Yes	Yes
Model 5	MXL	Yes	Yes
Model 6	Hybrid MNL	Controlled	Yes
Model 7	Hybrid MXL	Controlled	Yes
Model 8	Hybrid MXL	Controlled	Controlled
Model 9	Hybrid MNL	Controlled	Controlled

Model 1 corresponds to the DGP presented in Table 1

Because of the missing covariate, error terms $e_{ijt}^{**}(X_i^{Miss})$ and $\xi_i^*(X_i^{Miss})$ will become functions of X_i^{Miss} , which will cause them to become correlated. As $\xi_i^*(X_i^{Miss})$ enters the structural equations for the latent variable, the latent factor will become correlated with e_{ijt}^{**} , which will make it endogenous in the model. Again, we expect the coefficients which account for the effect of latent factor (α_{12} , α_{22} , and α_{32}) to be the most affected, but because of the nonlinearity of the choice model, it is likely that we will also be unable to recover the true values for some other coefficients.

Model 7 is an extension of Model 6, with fully correlated random parameters added for each attribute. As in Model 5, random parameters can capture unobserved preference heterogeneity caused by the missing variable. We note that Model 7 is probably the most common specification of the HC models that are used in applied studies, as it combines latent factors with random parameters.

The last two specifications represent different ways to control for endogeneity and show that if the correlation of the error terms is accounted for, the HC model can recover the true values of the parameters even where there is unobserved X_i^{Miss} .¹² Model 8 is the hybrid MXL that allows for estimable correlation between the error term in the structural component (ξ_i^{**}) and the random parameters for the attributes (β_{1i} , β_{2i} and β_{3i}). Consider again the specification in Table 1, with the X_i^{Miss} variable missing. With Model 8 we would estimate the following set of equations:

¹² We note that in those models X_i^{SD} affects the latent variable only through the structural equation, and does not have a direct effect on the random utility. This exclusion restriction works in much the same way here as it does in the instrumental variables and sample selection models. From our auxiliary simulations we found that without X_i^{SD} , Model 8 provides estimates that are closer to the true values, but still biased, whereas Model 9 leads to estimates which are not significantly different from the true values, but standard errors increase by about 100%. We therefore recommend having such a variable in the structural equation for better identification.

$$\begin{cases} V_{ijt} = \beta_{1i}SQ_{ijt} + \beta_{2i}Quality_{ijt} - \beta_{3i}Cost_{ijt} + e_{ijt} \\ \beta_{1i} = \alpha_{11} + \alpha_{12}LV_i + \beta_{1i}^* \\ \beta_{2i} = \alpha_{21} + \alpha_{22}LV_i + \beta_{2i}^* \\ \beta_{3i} = \exp(\alpha_{31} + \alpha_{32}LV_i + \beta_{3i}^*) \\ LV_i = \alpha_{61}X_i^{SD} + \xi_i^{**} \end{cases},$$

where β_{1i}^* , β_{2i}^* , and β_{3i}^* are random parameters with zero mean, such that they are fully correlated with each other, but also with ξ_i^{**} . These random parameters account for the unobserved heterogeneity caused by X_i^{Miss} (as in Model 7). In this way X_i^{Miss} does not enter the error term of the choice model (e_{ijt}) as it did in the example in (8). By additionally modeling the correlation with the error term in the structural equation of the latent variable, it recognizes that X_i^{Miss} is also present in the ξ_i^{**} . We expect this model to recover the true values of all the coefficients.

Model 9 is a hybrid MNL that uses a different approach to control for endogeneity. It does not employ random parameters to account for unobserved heterogeneity caused by the missing covariate, but instead, it assumes there exists an additional latent variable that enters both measurement equations. By using this additional latent variable, we impute the unobserved X_i^{Miss} variable into the model. To see how this model works in the case of LV-endogeneity, consider that the two specifications below are equivalent.

$$\begin{array}{l} \text{(A)} \\ \left\{ \begin{array}{l} V_{ijt} = \beta_{1i}SQ_{ijt} + \beta_{2i}Quality_{ijt} - \beta_{3i}Cost_{ijt} + e_{ijt} \\ \beta_{1i} = \alpha_{11} + \alpha_{12}LV_i + \alpha_{13}X_i^{Miss} \\ \beta_{2i} = \alpha_{21} + \alpha_{22}LV_i + \alpha_{23}X_i^{Miss} \\ \beta_{3i} = \exp(\alpha_{31} + \alpha_{32}LV_i + \alpha_{33}X_i^{Miss}) \\ LV_i = \alpha_{61}X_i^{SD} + \alpha_{62}X_i^{Miss} + \xi_i \\ I_{i1} = \alpha_{41} + \alpha_{42}LV_i + \alpha_{43}\eta_{i1} \\ I_{i2} = \alpha_{51} + \alpha_{52}LV_i + \alpha_{53}\eta_{i2} \end{array} \right. \end{array} \quad \begin{array}{l} \text{(B)} \\ \left\{ \begin{array}{l} V_{ijt} = \beta_{1i}SQ_{ijt} + \beta_{2i}Quality_{ijt} - \beta_{3i}Cost_{ijt} + e_{ijt} \\ \beta_{1i} = \alpha_{11} + \alpha_{12}LV_i + \alpha_{13}^*X_i^{Miss} \\ \beta_{2i} = \alpha_{21} + \alpha_{22}LV_i + \alpha_{23}^*X_i^{Miss} \\ \beta_{3i} = \exp(\alpha_{31} + \alpha_{32}LV_i + \alpha_{33}^*X_i^{Miss}) \\ LV_i = \alpha_{61}X_i^{SD} + \xi_i \\ I_{i1} = \alpha_{41} + \alpha_{42}LV_i + \alpha_{44}^*X_i^{Miss} + \alpha_{43}\eta_{i1} \\ I_{i2} = \alpha_{51} + \alpha_{52}LV_i + \alpha_{54}^*X_i^{Miss} + \alpha_{53}\eta_{i2} \end{array} \right. \end{array}$$

Specification (A) on the left is the same as in Table 1. In specification (B), X_i^{Miss} was taken out of the structural equation, and instead put in both the utility function and the measurement equation, such that $\alpha_{j3}^* = \alpha_{j3} + \alpha_{j2}\alpha_{62}/\sqrt{1 + \alpha_{61}^2 + \alpha_{62}^2}$, for $j \in \{1, 2, 3\}$, and $\alpha_{j4}^* = \alpha_{j2}\alpha_{62}/\sqrt{1 + \alpha_{61}^2 + \alpha_{62}^2}$, for $j \in \{4, 5\}$.¹³ Model 9 is based on specification (B). Specifically, if we treat X_i^{Miss} as a missing variable causing endogeneity, then Model 9 accounts for it by estimating

¹³ The denominators in the formulas for α_{23}^* and α_{43}^* are based on normalization of the error term in the structural equation. See Sect. 2.4 and Appendix B for more detailed discussion on normalization.

$$\left\{ \begin{array}{l} V_{ijt} = \beta_{1i}SQ_{ijt} + \beta_{2i}Quality_{ijt} - \beta_{3i}Cost_{ijt} + e_{ijt} \\ \beta_{1i} = \alpha_{11} + \alpha_{12}LV_i + \alpha_{13}^*LV_{2,i} \\ \beta_{2i} = \alpha_{21} + \alpha_{22}LV_i + \alpha_{23}^*LV_{2,i} \\ \beta_{3i} = \exp(\alpha_{31} + \alpha_{32}LV_i + \alpha_{33}^*LV_{2,i}) \\ LV_i = \alpha_{61}X_i^{SD} + \xi_i \\ I_{i1} = \alpha_{41} + \alpha_{42}LV_i + \alpha_{44}^*LV_{2,i} + \alpha_{43}\eta_{i1} \\ I_{i2} = \alpha_{51} + \alpha_{52}LV_i + \alpha_{54}^*LV_{2,i} + \alpha_{53}\eta_{i2} \end{array} \right. ,$$

where $LV_{2,i}$ is an additional latent factor, which accounts for both the unobserved preference heterogeneity in the choice model and the effect of the unobserved X_i^{Miss} on the indicator variables.

Neither of the proposed specifications (8 and 9) has been used in applied research. Model 8 requires the correlation of the respective error terms to be explicitly accounted for, while Model 9 only requires an additional latent variable that enters the same measurement equations as the latent variable suspected of endogeneity. To the best of our knowledge, previous studies have always used at least one measurement equation with a single latent variable, which simplifies both identification and interpretation.

In summary, we expect that Model 1 will recover the DGP parameters correctly, while the estimates of Models 2–7 will be biased due to measurement error and/or endogeneity. Models 8 and 9 should be able to control for both of these issues. A summary of the specifications we used is presented in Table 2.

3.3 Methodology of the Comparisons

To compare the estimates derived from the different models to the true values, we require a method that not only looks at the expected values but also penalizes for variance. Consider the usual case when one tests if x_i is statistically equal to its true value x_{true} (e.g., using the standard t-test). The larger the variance associated with x_i , the more difficult it is to reject the equality hypothesis. As a result, models that result in a high variation of the estimates (high standard errors) make it easier to falsely conclude that an estimate is not statistically significantly different from its true value.

To address this problem, we base our comparisons on equivalence tests (Hauck and Anderson 1984; Kristofersson and Navrud 2005). Equivalence tests reverse the null hypothesis and the alternative hypothesis; instead of testing if x_i is equal to x_{true} , we test if the absolute difference between them is higher than an a priori defined “acceptable” level. Czajkowski and Ščasný (2010) and Czajkowski et al. (2017a) argue that equivalence tests can be operationalized by a Minimum Tolerance Level (*MTL*), that is, the minimum “acceptable” difference that allows us to conclude that two values are equivalent at the required level of statistical significance.

For a random variable ω , *MTL* is formally defined as the minimum $\theta \geq 0$ that satisfies:

$$P(|\omega - \omega_{true}| > \theta \cdot |\omega_{true}|) = \alpha, \quad (9)$$

where α is the required significance level (e.g., 0.05). In our case,¹⁴ the probability can be evaluated using Two One-Sided T-Tests, while *MTL* can be found as:

$$MTL_{\alpha} = \underset{\theta \in (0, +\infty)}{\operatorname{argmin}} \{ \theta \} \text{ s.t. } P(|\omega - \omega_{true}| > \theta \cdot |\omega_{true}|) \leq \alpha. \quad (10)$$

MTL has an intuitive interpretation. For example, $MTL_{0.05} = 0.01$ means that, with 95% probability, the deviation of the estimated coefficients from the true values will not be larger than 1%.

4 Results

We generated 1,000 datasets following the DGP described in Table 1. For each dataset, we estimated nine models introduced in Sect. 3.2.¹⁵ For each model, we use the *MTL* approach to test if the estimates obtained are different from the true (DGP) parameters (or their rescaled equivalent – see Appendix B for details). In Appendix C we present parallel results for M-endogeneity.

Table 3 presents average parameter estimates for the nine models. To save the space we limited the presentation to the parameters of the utility function, as they are usually of most interest to researchers.¹⁶ For reference, the third column reports the true values of the coefficients, as assumed in the DGP. The results of Model 1 indicate that our modeling framework works well. If there are no missing variables and an HC model is used, the true parameter values are recovered with satisfactory precision.¹⁷

If indicator variables are used directly as interactions with choice attributes (Model 2 and Model 3), we observe that parameter estimates are substantially different from the true values due to measurement error, even though no variables are missing.¹⁸ It seems that these effects are underestimated by roughly 30%. We do not observe significant changes

¹⁴ A collection of MATLAB functions that are useful for calculating *MTL* is available at <https://github.com/czaj>.

¹⁵ The models were estimated using maximum simulated likelihood techniques, using 1,000 scrambled Sobol draws (Czajkowski and Budziński 2019). The software used here (estimation package for DCE data) was developed in Matlab and is available at <https://github.com/czaj/DCE> under CC BY 4.0 license. The data, software codes and supplementary materials are available from <http://czaj.org/research/supplementary-materials>.

¹⁶ Full results are available in the supplementary material to this paper available online.

¹⁷ Cf. Campbell and Sandorf (2020), who observe that even under correct specification the parameters of HC models can be severely biased. The possible explanation for this difference is that we use more indicator variables which may help to better identify a latent variable and its associated coefficients. Furthermore, we included a socio-demographic variable in the structural equation, which accounts for the observed variation in the latent factor, which may further facilitate identification.

¹⁸ In Models 2–5, the indicator variable I_i enters the choice model directly, with a mean normalized to 0. In our DGP both indicators have the same level of correlation with the latent factor so it does not matter which one we put into the model. We chose the first one as it has a positive correlation with the latent factor. In a real-life application one would have to either choose one indicator variable, calculate a function of all the indicators (e.g., mean) or incorporate all the indicator variables into the model. We found that including both indicator variables in the model actually makes the results worse.

Table 3 Monte Carlo simulation of the effects of LV-endogeneity for parameter estimates under different specifications – mean estimate values in 1,000 simulations (*, **, *** indicate $MTL_{0.05}$ at the 0.10, 0.05, 0.01 level, respectively.)

Parameter description	Notation in Table 1	True value of the coef.	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Model			Hybrid MNL	MNL	MXL	MNL	MNL	Hybrid MNL	Hybrid MXL	Hybrid MXL	Hybrid MNL, 2LV
LV-Endogeneity			No	No	No	Yes	Yes	Yes	Yes	Yes, controlled	Yes, controlled
Measurement error			No	Yes	Yes	Yes	No	No	No	No	No
SQ (main effect)	α_{11}	-4	-4.054**	-3.915**	-4.135**	-0.839	-4.113**	-3.303	-4.116**	-4.070**	-4.079**
Quality (main effect)	α_{21}	5	5.014***	4.945**	5.050**	3.653	5.041**	4.105	5.001***	5.003***	5.002***
Cost (main effect)	α_{31}	-0.5	-0.512**	-0.476*	-0.518**	-0.544*	-0.535*	-0.461*	-0.513**	-0.513**	-0.512**
LV interaction (with SQ)	α_{12}	-2	-2.010**					-3.502	-3.172	-2.013**	-2.008**
LV interaction (with Quality)	α_{22}	-1	-1.001***				0.6	0.6	0.103	-0.999**	-1.005**
LV interaction (with Cost)	α_{32}	0.5	0.504**				0.32	0.32	0.23	0.507**	0.509**
I_{11} interaction (with SQ)		-2		-1.444	-1.47	-1.198					
I_{11} interaction (with Quality)		-1		-0.732	-0.727	0.374	0.131				
I_{11} interaction (with Cost)		0.5		0.357	0.367	0.235	0.170				

Table 3 (continued)

Parameter description	Notation in Table 1	True value of the coef.	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
X_i^{Miss} interaction (with SQ)	α_{13}	-2	-2.028**	-2.234	-2.387						
X_i^{Miss} interaction (with Quality)	α_{23}	2	2.002***	1.843*	1.841*						
X_i^{Miss} interaction (with Cost)	α_{33}	-0.5	-0.507**	-0.419	-0.436						
Standard dev. (for SQ)		2			0.833		2.204		1.857*	2.050**	
Standard dev. (for Quality)		2			0.488		1.644		1.649	1.999***	
Standard dev. (for Cost)		0.5			0.275		0.518*		0.405	0.508**	
Second LV interaction (with SQ)		-3.155									-3.218**
Second LV interaction (with Quality)		1.423									1.415**
Second LV interaction (with Cost)		-0.211									-0.212**
Mean log-likelihood			-4038.25	-1941.70	-1938.41	-2571.37	-2222.93	-4837.14	-4605.12	-4561.33	-4560.87

between Model 2 and Model 3, which shows that allowing for correlated random parameters does not help to address the bias caused by the measurement error.

In Models 4 and 5, which suffer from both measurement and endogeneity bias, coefficients diverge from the true values even more. In some cases, the coefficients have wrong signs (for example “ I_{i1} interaction (with *Quality*)”). Overall, we consider this to be convincing evidence against using attitudinal variables as direct interactions of the model parameters.¹⁹ Hybrid choice models have an obvious advantage in this regard by directly accounting for the measurement error.

The next two models (Model 6 and Model 7) correspond to cases where the measurement errors are controlled for by using the hybrid choice framework. However, it is apparently not enough to account for the endogeneity alone, as many estimates of the coefficients remain biased. Specifically, interactions of the latent variable with the attributes are quite far away from the true values. Similar to the two previous models, the coefficient of the interaction with *Quality* has the wrong sign. Note that accounting for unobserved heterogeneity in Model 7 helps somewhat, and brings the other coefficients closer to the true values (for example, the main attribute effects are correctly estimated).

Finally, the last two models represent attempts to control for endogeneity, either by explicitly allowing for correlation between error terms in the structural and discrete choice components (Model 8) or assuming the existence of an additional LV to compensate for the missing variable (Model 9). We find that both models perform well, recovering the expected coefficient values, although some of them (in Model 9) require rescaling, as described in Appendix B. On average, both specifications have similar log-likelihood, although we note that these models are not nested, and Model 9 utilizes more coefficients, so some difference in terms of the log-likelihood is expected.

The main purpose of non-market valuation studies is the estimation of WTP for a given change in the quality of an environmental good. To assess how measurement error and LV-endogeneity affect the mean WTP we compare it across the nine specifications considered in the Monte Carlo simulation. The results are reported in Table 4. We find that Models 1, 8, and 9 recover the true value of the mean WTP, although the precision of these estimates is lower than that of the parameter estimates reported in Table 3. This is likely because the formula for the WTP includes a coefficient ratio. Models 2–6 result in biased estimates of mean WTP, with Model 4 generating the highest errors. This model suffers from both measurement and endogeneity bias, and it does not account for any unobserved heterogeneity. Once we add random parameters (Model 5) the estimates are very similar to Models 2 and 3 which suffer only from measurement error. Similarly, Model 6, which suffers from endogeneity and does not account for unobserved heterogeneity, results in highly biased mean WTP estimates. However, once the unobserved heterogeneity is accounted for (Model 7) the WTP estimates are very close to the true values. These results indicate that for mean WTP estimates it is crucial to control for measurement error and unobserved heterogeneity. The increase in precision of WTP estimates as a result of directly controlling for endogeneity (Models 8 and 9) is rather limited.

Even though obtaining mean WTP estimates is usually the main goal of stated preference studies, HC models are usually employed to investigate how WTP is affected by attitudes, perceptions, or other unobserved factors of interest. We, therefore, investigated how this relationship is affected by model misspecification caused by not accounting for preference

¹⁹ It is likely that the problem is even more pronounced for ordinal rather than linear indicators, as then the relationship between the LV and indicator variables is not one-to-one, as assumed in the DGP used here.

Table 4 Comparison of mean WTP under different specifications— mean estimate values in 1,000 simulations

Model	True mean WTP		Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7		Model 8		Model 9		
	No	Yes	Hybrid	MNL	MNL	No	Yes	MNL	Yes	MNL	Yes	Yes	No	Yes	Yes	Yes	No	Yes	No	Yes	No
Endogeneity			No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	No	Yes	No
Measurement error			No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No	No	No
SQ	- 7.330		- 7.668*	- 9.184	- 10.678	- 0.851	- 0.851	- 0.851	- 8.34	- 8.34	- 8.34	- 3.616	- 3.616	- 3.616	- 7.874*	- 7.874*	- 7.445**	- 7.445**	- 7.443*	- 7.443*	
Quality	10.324		10.697**	14.171	16.374	6.358	6.358	6.358	11.506	11.506	11.506	6.601	6.601	6.601	10.790*	10.790*	9.979**	9.979**	10.732**	10.732**	

*, **, *** indicate $MTL_{0.05}$ at the 0.10, 0.05, 0.01 level, respectively

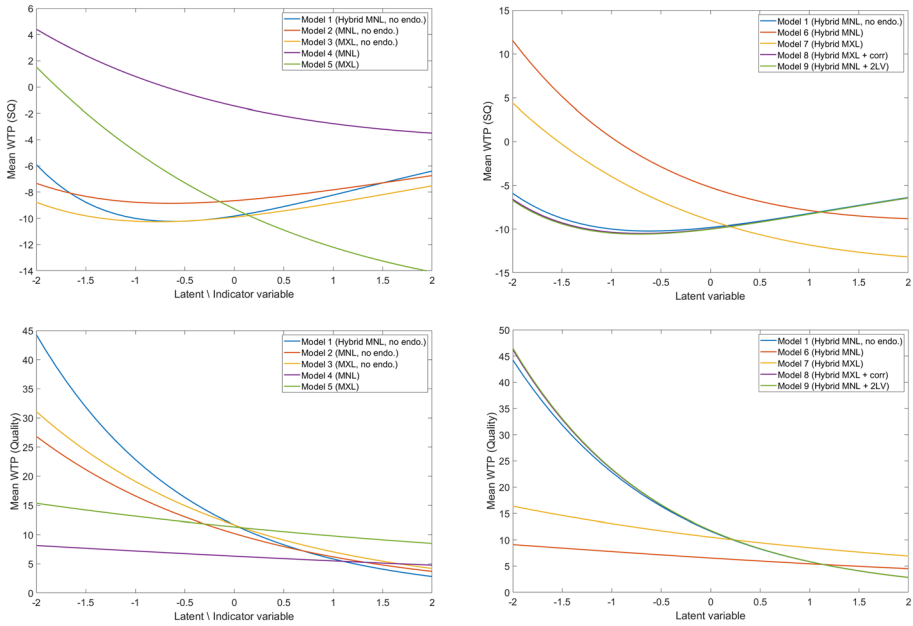


Fig. 1 Comparison of the effect of the latent variable on mean WTP under different specifications – mean estimate values in 1,000 simulations. The upper panels correspond to mean WTP for the *SQ* attribute, the lower panels correspond to mean WTP for the *Quality* attribute

heterogeneity, measurement errors, or LV-endogeneity. The results are presented in Fig. 1. We find that Models 2–7 result in substantially biased results relative to the true relationship represented by Model 1. Models suffering only from measurement error (Models 2 and 3 in the left panels in Fig. 1) recover a flatter relationship, which is probably caused by the lower estimates of coefficients for the interaction with the indicator variable, reported in Table 3. Nonetheless, these results are relatively close to the true relationship when compared with other models. Models 4–7 indicate decreasing relationship for the *SQ* attribute, even though it is actually U-shaped (as LV interactions with *SQ* and *Cost* have different signs). On the other hand, for the *Quality* attribute, the recovered relationship is much flatter. In contrast, Models 8 and 9 perform well, recovering the same relationship as assumed by the DGP (represented by the reference Model 1).²⁰ Overall, our results show that although not accounting for endogeneity had little effect on mean WTP, it substantially biased the relationships observed between LVs and choice attributes. This is a cause for concern, since observing such relationships is typically the main reason for using HC models.

²⁰ The deviations from the true relationship caused by model misspecification (not accounting for preference heterogeneity, measurement error or endogeneity) are even more severe for M-endogeneity, as illustrated in Appendix C.

5 Discussion and Conclusions

The hybrid choice framework is an approach that has quickly gained popularity. Vij and Walker (2016) analyze the possible advantages of employing the HC framework and identify a wide range of situations in which its use is justified. Most of the applications to date appear in the literature of environmental economics (e.g., Dekker et al. 2012; Hess and Beharry-Borg 2012; Hoyos et al. 2015; Czajkowski et al. 2017b, 2017c; Pakalniete et al. 2017) and transportation (e.g., Vredin Johansson et al. 2006; Daly et al. 2012; Daziano and Bolduc 2013). However, none of the existing studies explicitly account for the potential correlation between discrete choice and the other components of the model (for example, structural or measurement equations), which may arise when some variables are omitted from the model (for example, other attitudes).

It is commonly assumed that the HC framework addresses the endogeneity and measurement problems associated with incorporating indicator variables into the choice model. We show that although this is true for the latter, resolving the former requires a specific formulation of the model. Using a Monte Carlo simulation, in which we can control the DGP and induce endogeneity, we are able to study the performance of different specifications of choice models, in terms of the resulting bias of model parameters and implied WTPs. We show how endogeneity can be controlled for by explicitly allowing for correlation between structural and discrete choice component error terms (or with random parameters in the utility function), or by introducing an additional latent variable. The latter approach is probably easier to implement with the existing software, although it requires putting auxiliary latent factors into the same measurement equations as a latent factor which a researcher suspects of endogeneity. This may render interpretation of the results difficult. We demonstrate that these approaches work as expected, and they successfully recover the true values of all parameters. Although the practical usefulness of these approaches is yet to be confirmed,²¹ they demonstrate that endogeneity should and can be controlled for.

Our results demonstrate that failure to account for unobserved preference heterogeneity, measurement error, or endogeneity leads to biased results, in terms of both utility function parameters and WTP. Interestingly, we find that controlling for endogeneity is of secondary importance (relative to measurement error and unobserved preference heterogeneity) for mean WTP estimates. However, it is necessary because it facilitates the observation of the correct functional relationships between latent constructs and preferences associated with choice attributes. We believe this is an important reason to control for endogeneity whenever possible, as observing the relationship between WTP and latent factors is often the main reason for using HC models.

In Appendix C we consider M-endogeneity, in which a missing variable enters measurement equations, rather than the structural equation. We find that it generally mimics the results that we observed for LV-endogeneity. The main difference is that controlling for the correlation between the structural equation and choice model (Model 8) in relation to M-endogeneity no longer allows us to recover the true values of the coefficients.

As HC models combine different types of data (e.g., discrete choice and Likert scale) and utilize several random components to account for preference heterogeneity (e.g., random parameters and latent variables), they can quickly become very complex. For example,

²¹ For example, because of the possible non-identification of the model with more latent variables given available data.

we believe that Models 8 and 9 may be difficult to estimate, and therefore may require relatively large sample sizes. We conducted auxiliary simulations with varying sample sizes to test what minimum sample sizes are sufficient in order to accurately recover the DGP parameters. We considered sample sizes of 200, 500, and 700 respondents (with 6 choice tasks per respondent, as in our main simulation). We found that Model 1, which is relatively simple (no random parameters), recovered the true parameter values with a sample of 500, whereas with a sample of 200 some parameters were biased. Due to their complexity Models 8 and 9 required a sample of 700. We note, however, that our DGP is relatively simple, with only 3 attributes in the choice component of the model. If one were to utilize a more complex design with more attributes and random parameters, a larger sample than 700 would probably be needed. In the context of structural equation models (SEMs), Marsh et al. (1998) show that by having more indicator variables per latent factor, SEM can be estimated with a smaller sample size. This is also likely true for HC models. In future research, it would be useful to establish some rule-of-thumb guidelines for the necessary sample size for HC models, similar to those proposed by Westland (2010) for SEMs.

We believe it would be of interest to consider other methods that could address the endogeneity issues associated with indicator variables. The use of proxies (Guevara 2015) is probably the most straightforward way, although this is highly dependent on whether a researcher can identify a good variable that could serve as a proxy. In our simulation setting, X_i^{SD} would be a straightforward choice as it affects preferences solely through its effect on the latent variable. In practice, selecting such a variable may not be obvious or data may not be available. Other methods such as the multiple indicator solution (Guevara and Polanco 2016) may not be suitable, as in the presence of LV-endogeneity the other indicator needed will also be correlated with the error term of the choice model. Nonetheless, in future research, it would be relevant to establish how methods which are less computationally intensive can mitigate the bias caused by LV-endogeneity.

Acknowledging the limitations of our study, we note a possible taxonomy confusion, which could, to some extent, explain the overall belief in the ability of a hybrid choice model to mitigate endogeneity bias caused by indicator variables.²² In the hybrid choice framework, indicator variables do not enter the choice model directly, but rather they are treated as dependent variables. As such, they cannot cause endogeneity as it only arises when independent variables are correlated with an error term. However, in the HC models, the latent factors are independent variables, so they can still induce endogeneity if they are correlated with an error term in the choice model. Furthermore, since measurement error can be considered a special case of endogeneity (Walker et al. 2010), by addressing measurement bias hybrid choice models can resolve this source of endogeneity. Nevertheless, as illustrated by our study, the typical specification of the hybrid choice model does not address endogeneity arising from other sources.

It must also be noted that our study deals with the endogeneity of indicator variables and latent factors, and not with the endogeneity of the observed attributes (e.g., the travel cost in revealed preference studies). The latter case has attracted some attention in the literature, with proposed solutions that include imputing the missing attributes as latent

²² HC models and indicator variables are often used to resolve the endogeneity problem (Guevara and Ben-Akiva 2010), whereas in the case considered here, indicator variables are a source of endogeneity. Therefore, the results of the current study are of importance for cases in which the researcher is interested in the influence of certain attitudes on preferences, e.g. the effects of consequentiality on the willingness to pay for some environmental good (Zawojnska et al. 2019).

variables using several indicator variables as measurement equations (Guevara and Ben-Akiva 2010), the BLP method (Berry et al. 1995), using a control function (Rivers and Vuong 1988; Train 2009), and Multiple Indicator Solution (Guevara and Polanco 2016). For the comparison of the performance of different methods used to account for the endogeneity of attributes see Guevara (2015).

It is no easy matter to assess the relative importance of the LV-endogeneity when compared with the endogeneity of observed attributes. In stated preference studies, endogeneity of observed attributes is usually a lesser concern, as their levels are varied exogenously by the researcher. Furthermore, if sufficient care is taken when designing a study then all relevant attributes should be accounted for. If the estimation of mean WTP is an objective of the study, then LV-endogeneity can usually be avoided by simply not including any latent factors or indicator variables in the model, and letting the variation in preferences be captured by the random parameters. The exception is when identification of the effects of the latent variables is necessary for the aim of the study or the proper interpretation of the model. An example of the former is described by Buckell et al. (2021) where the aim of the study was to identify the effects of the intensity of nicotine addiction, which is measured through several indicator variables. An example of the latter is consequentiality (Zawojcka et al. 2019), where the parameters of the choice model cannot be interpreted as true marginal utilities for individuals who do not perceive a given survey as consequential. On the other hand, in revealed preference studies, endogeneity of attributes is much more likely as researchers usually do not observe a full context of the observed choices. As a result, even without any latent factors, it may be difficult to obtain correct estimates of mean WTP. We note, however, that in the revealed preference context, it is also likely that what matters for the decision process are perceptions of certain attributes rather than their observed measures. For example, when valuing water quality improvements, there could be a systemic difference between the objective measures and respondents' perceptions (Artell et al. 2013). If this is the case, then one could account for individuals' perceptions by using a hybrid choice framework, which would then make the LV-endogeneity discussion relevant also for revealed preference studies.

Finally, we acknowledge that our investigation relates to individual-specific latent variables, or as Bahamonde-Birke et al. (2015) refer to them, “*non-alternative related attitudes*,” in contrast to “*alternative related attitudes*” and “*perceptions*.” We believe that our results are general, although we note that addressing endogeneity in relation to alternative related attitudes and perceptions would likely be much more difficult to deal with from the modeling perspective.

In summary, our study shows that while typical hybrid choice models mitigate measurement error, they can still suffer from endogeneity bias, for example, caused by omitted variables that affect both choice and indicator variables. We highlight the potential problem, provide a thorough analysis of its potential causes and effects, and propose a method of classification for different types of endogeneity in HC models. We use a Monte Carlo experiment to demonstrate the existence and extent of endogeneity bias, propose two ways of addressing it, and verify that they are effective. Overall, we hope that our study stimulates further research in this area and is useful to applied researchers who deal with the endogeneity of indicator variables.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10640-022-00702-0>.

Acknowledgements The authors wish to thank the participants of the 5th International Choice Modeling Conference, Cape Town, 2017, and the participants of the 23rd Annual Conference of the European

Association of Environmental and Resource Economists, Athens, 2017, who facilitated this study with helpful comments on earlier versions of this paper. WB gratefully acknowledges the support of the National Science Centre of Poland (project 2016/21/N/HS4/02094) and the support of the Foundation for Polish Science (FNP). MC gratefully acknowledges the support of the National Science Centre of Poland (project 2017/25/B/HS4/01076). The funding sources had no involvement in study design; in the analysis and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

Code availability The models estimated herein used the DCE package developed in Matlab and are available at <https://github.com/czaj/DCE>. The code and data for estimating the specific models presented in this study are available from <http://czaj.org/research/supplementary-materials>.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ariely D, Loewenstein G, Prelec D (2003) "Coherent Arbitrariness": Stable Demand Curves Without Stable Preferences. *Quart J Econ* 118:73–105
- Artell J, Ahtiainen H, Pouta E (2013) Subjective vs. objective measures in the valuation of water quality. *J Environ Manage* 130:288–296
- Bahamonde-Birke, F. J., U. Kunert, H. Link, and d. J. D. Ortúzar. 2015. About attitudes and perceptions: finding the proper way to consider latent variables in discrete choice models. *Transportation*:1–19.
- Bartczak A, Mariel P, Chilton S, Meyerhoff J (2016) The impact of latent risk preferences on valuing the preservation of threatened lynx populations in Poland. *Australian Journal of Agricultural and Resource Economics* 60:284–306
- Bello, M., and A. Abdulai. 2015. Measuring heterogeneity, survey engagement and response quality in preferences for organic products in Nigeria. *Applied Economics*:1–13.
- Ben-Akiva, M., J. Walker, A. T. Bernardino, D. A. Gopinath, T. Morikawa, and A. Polydoropoulou. 2002. Integration of choice and latent variable models. Perpetual motion: Travel behaviour research opportunities and application challenges:431–470.
- Berry S, Levinsohn J, Pakes A (1995) Automobile Prices in Market Equilibrium. *Econometrica* 63:841–890
- Bollen KA, Davis WR (2009) Two Rules of Identification for Structural Equation Models. *Struct Equ Modelling* 16:523–536
- Boyce C, Czajkowski M, Hanley N (2019) Personality and economic choices. *J Environ Econ Manag* 94:82–100
- Buckell J, Hensher DA, Hess S (2021) Kicking the habit is hard: A hybrid choice model investigation into the role of addiction in smoking behavior. *Health Econ* 30:3–19
- Campbell D, Sandorf ED (2020) The use of latent variable models in policy: A road fraught with peril? *Bio-Based and Applied Economics* 9:305–324
- Chorus CG, Kroesen M (2014) On the (im-)possibility of deriving transport policy implications from hybrid choice models. *Transp Policy* 36:217–222
- Czajkowski M, Ahtiainen H, Artell J, Meyerhoff J (2017a) Choosing a Functional Form for an International Benefit Transfer: Evidence from a Nine-country Valuation Experiment. *Ecol Econ* 134:104–113
- Czajkowski M, Budziński W (2019) Simulation error in maximum likelihood estimation of discrete choice models. *Journal of Choice Modelling* 31:73–85
- Czajkowski M, Hanley N, Nyborg K (2017b) Social Norms, Morals and Self-interest as Determinants of Pro-environment Behaviours: The Case of Household Recycling. *Environ Resource Econ* 66:647–670
- Czajkowski M, Ščasný M (2010) Study on benefit transfer in an international setting. How to improve welfare estimates in the case of the countries' income heterogeneity? *Ecol Econ* 69:2409–2416
- Czajkowski M, Vossler CA, Budziński W, Wiśniewska A, Zawojńska E (2017c) Addressing empirical challenges related to the incentive compatibility of stated preference methods. *J Econ Behav Organ* 142:47–63

- Daly A, Hess S, Patruni B, Potoglou D, Rohr C (2012) Using ordered attitudinal indicators in a latent variable choice model: a study of the impact of security on rail travel behaviour. *Transportation* 39:267–297
- Daziano RA, Bolduc D (2013) Incorporating pro-environmental preferences towards green automobile technologies through a Bayesian hybrid choice model. *Transportmetrica a: Transport Science* 9:74–106
- Dekker, T., S. Hess, R. Brouwer, and M. Hofkes. 2012. Implicitly or explicitly uncertain? International Choice Modelling Conference, Sydney, Australia.
- Guevara CA (2015) Critical assessment of five methods to correct for endogeneity in discrete-choice models. *Transportation Research Part a: Policy and Practice* 82:240–254
- Guevara, C. A., and M. Ben-Akiva. 2010. Addressing endogeneity in discrete choice models: Assessing control-function and latent-variable methods. *Choice Modelling: The State-of-the Art and the State-of-practice*:353–371.
- Guevara CA, Polanco D (2016) Correcting for endogeneity due to omitted attributes in discrete-choice models: the multiple indicator solution. *Transportmetrica a: Transport Science* 12:458–478
- Hauck WW, Anderson S (1984) A new statistical procedure for testing equivalence in two-group comparative bioavailability trials. *J Pharmacokinetic Biopharm* 12:83–91
- Hess S, Beharry-Borg N (2012) Accounting for Latent Attitudes in Willingness-to-Pay Studies: The Case of Coastal Water Quality Improvements in Tobago. *Environ Resource Econ* 52:109–131
- Hess S, Shires J, Jopson A (2013) Accommodating underlying pro-environmental attitudes in a rail travel context: Application of a latent variable latent class specification. *Transp Res Part D: Transp Environ* 25:42–48
- Hess S, Stathopoulos A (2013) A mixed random utility — Random regret model linking the choice of decision rule to latent character traits. *Journal of Choice Modelling* 9:27–38
- Hoyos D, Mariel P, Hess S (2015) Incorporating environmental attitudes in discrete choice models: An exploration of the utility of the awareness of consequences scale. *Sci Total Environ* 505:1100–1111
- Jöreskog KG, Goldberger AS (1975) Estimation of a Model with Multiple Indicators and Multiple Causes of a Single Latent Variable. *J Am Stat Assoc* 70:631–639
- Kim J, Rasouli S, Timmermans H (2017) Satisfaction and uncertainty in car-sharing decisions: An integration of hybrid choice and random regret-based models. *Transportation Research Part a: Policy and Practice* 95:13–33
- Kløjgaard ME, Hess S (2014) Understanding the formation and influence of attitudes in patients' treatment choices for lower back pain: Testing the benefits of a hybrid choice model approach. *Soc Sci Med* 114:138–150
- Kristofersson D, Navrud S (2005) Validity Tests of Benefit Transfer – Are We Performing the Wrong Tests? *Environ Resource Econ* 30:279–286
- Lundhede T, Jacobsen JB, Hanley N, Strange N, Thorsen BJ (2015) Incorporating outcome uncertainty and prior outcome beliefs in stated preferences. *Land Econ* 91:296–316
- Mariel P, Meyerhoff J, Hess S (2015) Heterogeneous preferences toward landscape externalities of wind turbines – combining choices and attitudes in a hybrid model. *Renew Sustain Energy Rev* 41:647–657
- Marsh HW, Hau K-T, Balla JR, Grayson D (1998) Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivar Behav Res* 33:181–220
- Pakalnite K, Aigars J, Czajkowski M, Strake S, Zawojcka E, Hanley N (2017) Understanding the distribution of economic benefits from improving coastal and marine ecosystems. *Sci Total Environ* 584–585:29–40
- Palma D, J. d. D. Ortúzar, L. I. Rizzi, C. A. Guevara, G. Casaubon, and H. Ma. (2016) Modelling choice when price is a cue for quality: a case study with Chinese consumers. *Journal of Choice Modelling* 19:24–39
- Raveau S, Yáñez MF, J. d. D. Ortúzar. (2012) Practical and empirical identifiability of hybrid discrete choice models. *Transportation Research Part B: Methodological* 46:1374–1383
- Revelt D, Train K (1998) Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level. *Rev Econ Stat* 80:647–657
- Rivers D, Vuong QH (1988) Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of Econometrics* 39:347–366
- Train KE (2009) *Discrete Choice Methods with Simulation*, 2nd edn. Cambridge University Press, New York
- Vij A, Walker JL (2016) How, when and why integrated choice and latent variable models are latently useful. *Transportation Research Part B: Methodological* 90:192–217
- Vredin Johansson M, Heldt T, Johansson P (2006) The effects of attitudes and personality traits on mode choice. *Transportation Research Part a: Policy and Practice* 40:507–525
- Walker J, Li J, Srinivasan S, Bolduc D (2010) Travel demand models in the developing world: correcting for measurement errors. *Transportation Letters* 2:231–243

- Westland JC (2010) Lower bounds on sample size in structural equation modeling. *Electron Commer Res Appl* 9:476–487
- Zawojcka E, Bartczak A, Czajkowski M (2019) Disentangling the effects of policy and payment consequentiality and risk attitudes on stated preferences. *J Environ Econ Manag* 93:63–84

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.