Full Research Article

The use of latent variable models in policy: A road fraught with peril?

Danny Campbell*, Erlend Dancke Sandorf

University of Stirling, Stirling Management School, Economics Division

Abstract. This paper explores the potential usefulness and possible pitfalls of using integrated choice and latent variable models (hybrid choice models) on stated choice data to inform policy. Using a series of Monte-Carlo simulations, we consider how model selection depends on the strength of relationship between the latent variable and preferences and the strength of relationship between the latent variable and the indicator. Our findings show that integrated choice and latent variable models are difficult to estimate, even when the data generating process is known. Ultimately, we show that their use should be driven by the analyst's belief about the strength of correlations between preferences, the latent variable and indicator. We discuss the implications of our results for policy.

Keywords. Stated preferences, choice modelling, integrated choice and latent variables, hybrid choice model.

JEL codes. C25, H41, Q51.

1. Introduction

Many policies affect the natural environment: e.g. a new hydro-electric dam will provide clean renewable energy and jobs, but may cause damage to the local river; a new motorway will reduce travel time, but may be built in a vulnerable natural area; and, a new conservation area will protect a number of vulnerable species, but possibly displace existing and future industrial activity and development. Policy makers are routinely faced with these decisions and trade-offs, and in many countries they are required to undertake cost-benefit analyses or assessments. Problematically, many of these costs and benefits are not traded in markets and policy makers have no information on society's preferences for these non-market goods and services. Stated choice experiments, where people are asked to make a choice between competing policy alternatives, are a way to elicit people's preferences for non-market goods and services.

Economists have long recognized that people's choices are affected by a multitude of observable (e.g. gender, age and income) and unobservable (e.g. attitudes and beliefs)

---

*Corresponding author. E-mail: danny.campbell@stir.ac.uk

individual characteristics in addition to the characteristics of the options amongst which they choose. For example, when asked to choose whether to support a policy to protect a river from hydropower development, people's decision will likely depend on their income and where they live in relation to the river, but also their attitudes towards development, clean energy and conservation. Testing whether choices are different between high and low income people is trivial and straightforward, but how do we test for differences in attitudes and beliefs? How do we incorporate and consider them in our models? The most obvious, and perhaps most intuitive, way to test for the marginal effect of an attitude or belief is to use an interaction term the same way we would when exploring the marginal effects of age, gender or income. However, unlike age, gender and income, attitudes and beliefs are likely correlated with unobserved factors affecting choice (i.e. the error term) and indicators of attitudes and beliefs (e.g. Likert scale survey questions) are themselves imperfect measures of the true underlying attitude or belief. If either of these are true, then the model will be misspecified and the estimated parameters may be biased (endogeneity bias and measurement error) (Ben-Akiva et al., 2002; Hess, 2012).

Recently, the integrated choice and latent variable (ICLV), or hybrid choice model (Ben-Akiva et al., 2002; McFadden, 1986), popularized in transport (Bhat et al., 2015; Hess and Stathopoulos, 2013), has gained traction in environmental economics (Alemu and Olsen, 2019; Hoyos et al., 2015; Kassahun et al., 2016; Mariel and Meyerhoff, 2016; Taye et al., 2018; Zawojska et al., 2019). An ICLV model combines structural equation modelling with discrete choice modelling. In this modelling framework, we assume that (unobserved) character traits, such as pro-environmental attitudes, can be captured by one or more latent variables defined as functions of observable characteristics and measures intended to capture such attitudes, e.g. Likert scale questions. These latent variables can be included directly in our choice models to capture the effect of (latent) attitudes and beliefs on the probabilities of choice (Ben-Akiva et al., 2002). The popularity of ICLV models stems from claims that the inclusion of attitudes and beliefs through latent variables leads to improved forecasts (Vij and Walker, 2016; Yáñez et al., 2010), that it sheds more light on preference heterogeneity (Kassahun et al., 2016; Mariel and Meyerhoff, 2016), and that it allows for the inclusion of attitudinal variables and beliefs while avoiding issues with measurement error and possible endogeneity bias (Ben-Akiva et al., 2002; Guevara and Ben-Akiva, 2010).[1] The latter is only true under specific conditions (Vij and Walker, 2016).

Measurement error and endogeneity bias aside, the interpretability of the parameters in ICLV models remain a challenge, especially if we seek to use the model results to influence policy. In an ICLV model, indicators only affect choice indirectly through the latent variable. The latent variable is, by definition, unknown and has no direct interpretability. As such, the indicators can only be interpreted in relation to their directional impact on the latent variable and its directional impact on utility. For examples from environmental economics, see Kassahun et al. (2016) who study farmers' marginal willingness to pay (MWTP) to adopt irrigation methods, Taye et al. (2018) who study how people's environmental attitudes affect their MWTP for forest management options, Alemu and Olsen (2019) who try to understand how people's food choice motives affect their MWTP for

---

[1] For an overview of the historical development of hybrid discrete choice models, we refer the reader to (Bahamonde-Birke and Ortúzar, 2017).

insect based food products or Lundhede et al. (2015) who look at how perceived uncertainty about policy outcomes affect bird conservation under climate change. To aid interpretability of the latent variable and to gain a better understanding of what drives heterogeneity in welfare measures, Hoyos et al. (2015), Mariel and Meyerhoff (2016) and Mariel et al. (2018) argue in a series of papers, all in environmental economics, that practitioners should use exploratory factor analysis to identify which indicators are appropriate for each latent variable. This approach can also be helpful in model estimation, because more appropriate indicators should make estimation of the model easier. An alternative, or perhaps complement, to the exploratory analysis is to use already validated scales to elicit attitudes or personality traits (Alemu and Olsen, 2019; Boyce et al., 2019; Hoyos et al., 2015; Taye et al., 2018). That said, Vij and Walker (2016) show that a reduced form model without latent variables may fit the data at least as well as a latent variable model if the observable explanatory variables are good predictors of the latent variables, which is a specific case of the general result provided by (McFadden and Train, 2000). Chorus and Kroesen (2014) caution that using the results of an ICLV model to inform policies that seek to influence choice by targeting the latent variable is inappropriate given the cross-sectional nature of the data (i.e. only between-individual comparisons based on differences in the latent variable can be accommodated, rather than within-individual comparisons based on changes in the latent variable) and the possibly endogenous relationship between the latent variable and choice. It is also important to keep in mind that as the complexity of our models – and our ability to capture more heterogeneity – increase, we need to be careful that we do not tailor our model too close to the sample data. This may compromise our ability to generalize our model and results beyond the existing dataset and limit the usefulness to policy makers. While end users will often want to establish the relationship between the dependent variable(s) and a relatively small number of key independent variables, increasing model complexity is justified only if it produces reasonably more accurate results. While a familiar aphorism among econometricians is that "all models are wrong", some models are more wrong than others, and to be of practical use there is a need to ensure that our results are understandable and meaningful. That responsibility lies with us.

So what then, is the additional benefit of developing an ICLV model? We argue in this paper that while model fit is obviously important, it is not the be all and end all of model selection; and that while using hybrid models to suggest polices that target the latent variable itself is inappropriate (Chorus and Kroesen, 2014; Kroesen et al., 2017; Kroesen and Chorus, 2018), these models can provide rich insight into behaviour (Hess, 2012), help de-bias estimates (Vij and Walker, 2016), offer improvements in prediction in certain contexts (Vij and Walker, 2016) and reveal additional layers of heterogeneity (Hess, 2012; Mariel and Meyerhoff, 2016; Taye et al., 2018). However, we show that retrieving the true parameters of ICLV models can be challenging, and that the benefits of developing and using them are not always clear-cut.

This paper is a practical illustration of the points outlined above, and can work as a clarification for practitioners and policy makers alike. Using Monte Carlo simulations, we show the important role that correlation between the attributes, indicators and latent variables play in model selection and that the econometricians belief about the strength of this correlation is the main thing to consider when trying to decide whether an ICLV model is appropriate. Furthermore, we show that the bias of not accounting for these correlations in

parameters and MWTP is generally increasing with the strength of the correlations. The practical implication is that the strength of the endogeneity bias from including the indicator directly in the choice model is related to the strength of the correlation between the indicator and the latent variable. For low degrees of correlation, omitting the latent variable or using a reduced form model does not lead to substantial bias in MWTP, but for high degrees of correlation between the indicator and the latent variable, the bias in MWTP is less than the model without indicators or latent variables. As such, our results can be viewed as an illustration of Vij and Walker (2016) and Kroesen and Chorus (2018).

The rest of the paper is outlined as follows: Section 2 outlines our econometric approach, Section 3 details the Monte-Carlo data generation processes, Section 4 presents the results from the simulation study, and Section 5 discusses the implications of our results for the use of ICLV models for policy and concludes the paper.

2. Econometric approach

To illustrate our point and substantiate our conclusions, we use a straightforward stated choice data setup. We generate synthetic datasets and show through Monte-Carlo simulation how misspecification of the model can lead to bias and under which circumstances this may not be the case. In the following, we assume that the reader is somewhat familiar with discrete choice modelling. To introduce notation, and to save space, we start with a standard random parameters mixed logit model where the probability of observing the sequence of T_n choices y_n made by individual n is a K dimensional integral of the logit formula over all possible values of $\widehat{\boldsymbol{\beta}}_n$:[2]

$$\Pr(y_n \mid X_n, \widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\gamma}}, z_n, \widehat{\boldsymbol{\Sigma}}) = \underset{\breve{K}}{\int \int \dots \int} \prod_{t=1}^{T_n} \frac{\exp(\widehat{\boldsymbol{\beta}}_n x_{njt})}{\sum_{j=1}^J \exp(\widehat{\boldsymbol{\beta}}_n x_{njt})} \, f(\widehat{\boldsymbol{\beta}}_n \mid \widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\gamma}}, z_n, \widehat{\boldsymbol{\Sigma}}) d(\widehat{\boldsymbol{\beta}}_n), \tag{1}$$

where x_{njt} is a column vector of attribute levels and the joint density of the row vector $\widehat{\boldsymbol{\beta}}_n$ of marginal utilities is given by $f(\widehat{\boldsymbol{\beta}}_n \mid \cdot)$. A key consideration when specifying random parameters is the assumption regarding their distribution. In this paper, we express the individual marginal utility parameter for attribute k, $\hat{\beta}_{nk}$, as follows:

$$\hat{\beta}_{nk} = \hat{\mu}_k + \widehat{\boldsymbol{\gamma}}_k z_n + \epsilon_{nk}, \tag{2}$$

where $\hat{\mu}_k$ is the mean of the distribution for attribute k, z_n is a column vector of regressors relating to individual-specific characteristics, e.g. age, gender, attitudinal responses or latent variables, $\widehat{\boldsymbol{\gamma}}_k$ is a conformable row vector of estimated mean shifter parameters and ϵ_{nk} is a deviate from a multivariate normal distribution with zero mean and covariance $\widehat{\boldsymbol{\Sigma}}$. Introducing individual specific characteristics, e.g. responses to an attitudinal question, allows us to assess and interpret the marginal effect of the attitudinal response on margin-

---

[2] The ICLV model can be specified with other choice kernels as well, e.g. multinomial logit or latent class, but throughout this paper, whenever we refer to the ICLV model it is one specified with a random parameters mixed logit kernel.

al utility in the same way as we would for age, gender or income. However, as discussed above, by including attitudinal measures directly in the model, we assume that responses to these attitudinal questions are direct measures of attitudes, e.g. pro-environmental attitudes, and that they are exogenous, i.e. that the responses are uncorrelated with the error terms. If either assumption is violated, our model is misspecified and our parameters may be biased. To avoid some of the issues associated with measurement error and endogeneity bias, we can, for example, use a hybrid choice model. In this model, we assume that the responses to the attitudinal questions are mapped to a latent variable that is included directly in the marginal utility expression just like we would for any other individual characteristic. In our case, the latent variable is given by the following structural equation:

$$\hat{\mathcal{L}}_n = \hat{\omega}_n \text{SPACING}\hat{\omega} \sim N(0, \hat{\sigma}_{\mathcal{L}}^2), \tag{3}$$

where $\hat{\omega}_n$ is a normally distributed random disturbance with zero mean and standard deviation $\hat{\sigma}_{\mathcal{L}}$ to be estimated. Responses to our pro-environmental behaviour question are given on a three-point Likert scale, as explained below. Since the response is on an ordered scale, we need to use an ordered model for the measurement equations (Daly et al., 2012). Let us create an underlying continuous variable, $i^\star$, that determines the observed response to the indicator question. For individual n, we assume the following relationship with the latent variable:

$$i_n^* = \hat{\zeta} + \hat{\psi}\hat{\mathcal{L}}_n + \varepsilon_n, \tag{4}$$

where $\hat{\zeta}$ is a constant to estimate, $\hat{\psi}$ represents the variation of the underlying continuous variable for a unitary variation in the latent variable and ε_n is an idiosyncratic random disturbance term assumed to be a deviate from an identically and independently standard logistic distribution. Now, we can map the value of i_n^* to the observed cardinal response to the three-point indicator question. Specifically, with l denoting the index for the indicator response (i.e. $l \in \{1,2,3\}$), we have:

$$\hat{\text{i}}_n = \begin{cases} l = 1 \text{ if } -\infty < \hat{\zeta} + \hat{\psi}\hat{\mathcal{L}}_n \leq \hat{\tau}_1 \\ l = 2 \text{ if } \hat{\tau}_1 < \hat{\zeta} + \hat{\psi}\hat{\mathcal{L}}_n \leq \hat{\tau}_2 \\ l = 3 \text{ if } \hat{\tau}_2 < \hat{\zeta} + \hat{\psi}\hat{\mathcal{L}}_n \leq \infty \end{cases}, \tag{5}$$

where $\hat{\tau}_1$ and $\hat{\tau}_2$ are threshold parameters to be estimated. In order to preserve the positive signs of all of the probabilities and ensure that the support is over the entire real line, there is a strict ordering of threshold values that demarcate the observed ordinal levels of the indicator question, specifically $-\infty<\hat{\tau}_1<\hat{\tau}_2<\infty$, with $\tau_0=-\infty$ and $\tau_L=\infty$. With this in place, the probability for the response to the indicator question for individual n can be represented by the ordered logit model:

$$\Pr(i_n \mid \hat{\mathcal{L}}_n, \hat{\zeta}, \hat{\psi}, \hat{\boldsymbol{\tau}}) = \prod_{l=1}^L [\Lambda(\hat{\tau}_{l-1} + \hat{\tau}_l - \hat{\zeta} + \hat{\psi}\hat{\mathcal{L}}_n) - \hat{\zeta} + \hat{\psi}\Lambda(\hat{\tau}_{l-2} + \hat{\tau}_{l-1} - \hat{\mathcal{L}}_n)]^{\mathbb{I}_{n_l}}, \tag{6}$$

where $\Lambda(.)$ represents the standard logistic cumulative distribution function and $\mathbb{I}_{n_l}$ is a variable equal to one when the indicator level l is responded by individual n and zero otherwise.

To estimate the ICLV model, we need to maximize the joint likelihood of the observed sequence of choices and the observed responses to the Likert scale questions gauging pro-environmental behaviour. We can write the overall likelihood function as follows:

$$\Pr\big(y_n, i_n \mid X_n, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\Sigma}}, \hat{\sigma}_{\mathcal{L}}, \hat{\zeta}, \hat{\psi}, \hat{\boldsymbol{\tau}}\big) = \int \Pr\big(y_n \mid X_n, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\gamma}}, z_n, \hat{\boldsymbol{\Sigma}}\big) \Pr\big(i_n \mid \hat{\mathcal{L}}_n, \hat{\zeta}, \hat{\psi}, \hat{\boldsymbol{\tau}}\big)$$
$$\phi\big(\hat{\mathcal{L}}_n \mid 0, \hat{\sigma}_{\mathcal{L}}^2\big) d\big(\hat{\mathcal{L}}_n\big), \tag{7}$$

where $\phi\big(\hat{\mathcal{L}}_n \mid 0, \hat{\sigma}_{\mathcal{L}}^2\big)$ denotes the normal density with mean zero and variance $\hat{\sigma}_{\mathcal{L}}^2$. Note the probability now involves a $K{+}1$ dimensional integral.

3. Synthetic data generating process and approach

3.1 Data

We use Monte-Carlo experiments to generate synthetic datasets. This is particularly useful because we know the true parameters underlying the data generating process (DGP) and will enable us to judge model performance in terms of how close the model estimates are to the true values. For this demonstration, we construct a stated choice experiment characterized by three environmental attributes: "area" represents the protected area (in 1,000 km^2) with levels 2, 4, 6, 8, 10 and 12; "broadleaf" denotes the fraction of newly planted trees that are broad-leafed with levels 0.0, 0.2, 0.4, 0.6, 0.8, and 1.0; and, "recreation" is a zero-one indicator variable signifying if recreation opportunities are available. The "cost" attribute is specified as having six levels: €5, €10, €15, €20, €25 and €30. Next we generate a random experimental design consisting of 500 synthetic individuals completing six choice tasks comprising two alternatives.[3] For the indicator question we make use of a three-point Likert-scale indicating environmental tendency: anti-environmental tendencies, neutral-environmental tendencies and pro-environmental tendencies.

Our Monte-Carlo strategy involves 25 data generation processes. In all settings, the model specification used in the DGP is based on the ICLV model with a random parameters mixed logit kernel described above. Specifically, we assume:

$$\beta_{nk} = \mu_k + \gamma_k L_n + \sigma_k \upsilon_{nk}, \tag{8}$$

where υ_{nk} is an independent standard Normal deviate, meaning that σ_k can be interpreted as the standard deviation of the (underlying) Normal distribution.[4] The parameter vector $\boldsymbol{\gamma}$ determines the direction and strength of the relationship between the latent variable and the marginal utilities. To asses how findings are sensitive to different values of $\boldsymbol{\gamma}$ we consider different vectors. This goes from the case where the latent variable has no bearing on any of the marginal utility distributions (i.e. where $\gamma_k{=}0\forall K$) to one in which it plays a

---

[3] While this design ensures that all attribute levels can be estimated independently of each other, we recognise that a more efficient experimental design could have been used to minimise the variance of the parameters. However, in a Monte Carlo experiment with specified parameters it may be more appropriate to show that the results stand up in cases where the experimental design is not tailored too closely to the data-generating parameters. Indeed, this would be the case in a real-life empirical application.

[4] For the cost attribute, we specify $\beta_{nk}{=}{-}\exp(\mu_k+\gamma_k L_n+\sigma_k \upsilon_{nk})$ to ensure strictly negative values.

large role. Given our DGP of a positive correlation between the latent variable and environmental tendency, we achieve this by increasing the γ_k values for the non-cost attributes and decreasing it for the cost attribute. Furthermore, we consider different values of ψ to contrast the suitability of the indicator question as a manifestation of the underlying latent environmental tendency, respectively, from the case where the Likert responses are independent of environmental tendencies to one in which they are, for all intents and purposes, direct measures of environmental tendency. We make use of an orthogonal setup with five sets of parameters to control for the strength of relationship between the latent variable and the marginal utility parameters and five parameters to control the strength of relationship between the latent variable and the indicator response, thus producing 25 different DGPs enabling independent evaluation. The respective γ_k and ψ for each DGP is reported in Table 1. The other parameters remain constant across all DGPs. Respectively, for the cost, area, broadleaf and recreation attributes the values of μ_k are -1.0, 0.6, 2.5 and 1.4, and the values of σ_k are 0.4, 0.1, 0.8 and 0.4. For σ_L, ζ, τ_1 and τ_1 we use 1.2, -0.6, -1.0 and 0.1, respectively.

In practice, we generated a deviate for each synthetic individual from $N\sim(0,\sigma_L{}^2)$ to represent their specific latent variable, and independent deviates from $N\sim(0,\sigma_k{}^2)$ to obtain their specific marginal utility. Additionally, for each utility function and their underlying continuous variable relating to the indicator we retrieved deviates from independently and identically distributed type I extreme value distributions with variance $\pi^2/6$. The choices are produced by identifying the alternative associated with the largest utility value. The individual counterfactual response to the three-point indicator question is established by comparing the simulated indicator distribution against the demarcation thresholds. Since idiosyncratic results can arise from a single sample of individuals, we generate 100 replications for every simulation setting.

To determine how well the simulated data reflect the DGP, we report a number of Pearson correlation coefficients for each data generation setting in Table 1. Specifically, $\rho_{W_k L}$ and $\rho_{W_k i^*}$ denote the correlation between the MWTP for attribute k and the latent variable and the underlying continuous variable relating to the indicator, respectively. The correlation between the latent variable and the underlying continuous variable relating to the indicator is signified by ρ_{L,i^*}. We can see that the correlations reflect the DGP and, most importantly, that we separately control for differences in the influence of the latent variable on preferences and the indicator. It is also noticeable that the latent variable has a relatively stronger influence on the area attribute, followed by broadleaf and, lastly, recreation. This is a deliberate artefact of the parameters we used in DGP, since it allows us to compare the implications under a wider range of settings.

3.2 Analysis

For each dataset generated, we estimate six candidate models. This includes a random parameters mixed logit model (MXL), a random parameters mixed logit model with the indicators mapping directly to the marginal utilities (MXLIND) and a hybrid random parameters mixed logit model (LVMXL), that matches the DGP, where the latent variable enters both the marginal utility expressions and measurement equation relating to environmental tendency. It is widely acknowledged that models relying on the strict notion of

Table 1. Parameters used in the DGPs and mean correlations across the 100 Monte-Carlo simulations.

DGP	γ_{cost}	γ_{area}	$\gamma_{broadleaf}$	$\gamma_{recreation}$	ψ	$\rho W_{area,\mathcal{L}}$	$\rho W_{area,i*}$	$\rho W_{broadleaf,\mathcal{L}}$	$\rho W_{broadleaf,i*}$	$\rho W_{recreation,\mathcal{L}}$	$\rho W_{recreation,i*}$	$\rho \mathcal{L},i*$
1	0.000	0.000	0.000	0.000	0.000	-0.01 (-0.09 − 0.09)	0.00 (-0.07 − 0.09)	0.00 (-0.08 − 0.08)	0.01 (-0.09 − 0.10)	0.00 (-0.08 − 0.08)	0.01 (-0.07 − 0.09)	0.00 (-0.08 − 0.07)
2	0.000	0.000	0.000	0.000	0.700	0.01 (-0.09 − 0.10)	0.01 (-0.09 − 0.09)	0.01 (-0.07 − 0.09)	0.00 (-0.10 − 0.09)	0.01 (-0.08 − 0.09)	0.01 (-0.12 − 0.11)	0.54 (0.50 − 0.60)
3	0.000	0.000	0.000	0.000	1.400	0.00 (-0.08 − 0.08)	0.00 (-0.09 − 0.08)	0.00 (-0.08 − 0.08)	0.00 (-0.09 − 0.08)	0.00 (-0.08 − 0.09)	0.00 (-0.07 − 0.09)	0.79 (0.77 − 0.81)
4	0.000	0.000	0.000	0.000	2.100	0.00 (-0.09 − 0.08)	0.00 (-0.07 − 0.08)	0.00 (-0.08 − 0.08)	0.00 (-0.08 − 0.08)	-0.01 (-0.08 − 0.07)	-0.01 (-0.09 − 0.06)	0.89 (0.88 − 0.90)
5	0.000	0.000	0.000	0.000	2.800	0.00 (-0.08 − 0.09)	0.00 (-0.07 − 0.09)	0.00 (-0.08 − 0.09)	0.00 (-0.07 − 0.08)	0.00 (-0.09 − 0.10)	0.01 (-0.10 − 0.09)	0.93 (0.93 − 0.94)
6	-0.050	0.075	0.125	0.025	0.000	0.42 (0.36 − 0.47)	0.00 (-0.08 − 0.09)	0.22 (0.14 − 0.28)	0.00 (-0.08 − 0.07)	0.16 (0.08 − 0.24)	0.00 (-0.08 − 0.10)	0.00 (-0.08 − 0.09)
7	-0.050	0.075	0.125	0.025	0.700	0.41 (0.33 − 0.48)	0.23 (0.14 − 0.31)	0.22 (0.11 − 0.30)	0.12 (0.03 − 0.19)	0.16 (0.05 − 0.23)	0.09 (0.00 − 0.17)	0.54 (0.50 − 0.59)
8	-0.050	0.075	0.125	0.025	1.400	0.41 (0.35 − 0.47)	0.33 (0.27 − 0.39)	0.21 (0.15 − 0.28)	0.17 (0.09 − 0.24)	0.15 (0.07 − 0.23)	0.12 (0.05 − 0.19)	0.79 (0.77 − 0.81)
9	-0.050	0.075	0.125	0.025	2.100	0.41 (0.36 − 0.47)	0.37 (0.30 − 0.43)	0.21 (0.15 − 0.28)	0.19 (0.12 − 0.28)	0.15 (0.08 − 0.23)	0.13 (0.06 − 0.21)	0.89 (0.88 − 0.90)
10	-0.050	0.075	0.125	0.025	2.800	0.41 (0.36 − 0.48)	0.39 (0.33 − 0.45)	0.21 (0.13 − 0.30)	0.20 (0.12 − 0.29)	0.15 (0.07 − 0.25)	0.14 (0.06 − 0.24)	0.93 (0.93 − 0.94)
11	-0.100	0.150	0.250	0.050	0.000	0.64 (0.61 − 0.68)	-0.01 (-0.11 − 0.08)	0.40 (0.34 − 0.46)	0.00 (-0.09 − 0.10)	0.30 (0.23 − 0.36)	0.00 (-0.08 − 0.09)	-0.01 (-0.09 − 0.07)
12	-0.100	0.150	0.250	0.050	0.700	0.65 (0.61 − 0.68)	0.36 (0.30 − 0.41)	0.40 (0.33 − 0.46)	0.22 (0.16 − 0.29)	0.30 (0.22 − 0.37)	0.17 (0.09 − 0.26)	0.55 (0.50 − 0.61)
13	-0.100	0.150	0.250	0.050	1.400	0.64 (0.61 − 0.68)	0.51 (0.46 − 0.56)	0.40 (0.34 − 0.46)	0.32 (0.24 − 0.39)	0.30 (0.23 − 0.35)	0.23 (0.15 − 0.31)	0.80 (0.78 − 0.82)
14	-0.100	0.150	0.250	0.050	2.100	0.65 (0.60 − 0.69)	0.58 (0.52 − 0.62)	0.40 (0.33 − 0.46)	0.36 (0.29 − 0.42)	0.30 (0.22 − 0.37)	0.27 (0.18 − 0.33)	0.89 (0.88 − 0.90)
15	-0.100	0.150	0.250	0.050	2.800	0.64 (0.61 − 0.68)	0.60 (0.56 − 0.64)	0.40 (0.33 − 0.46)	0.38 (0.31 − 0.44)	0.29 (0.23 − 0.36)	0.28 (0.20 − 0.35)	0.93 (0.93 − 0.94)
16	-0.150	0.225	0.375	0.075	0.000	0.75 (0.71 − 0.78)	0.00 (-0.08 − 0.08)	0.53 (0.48 − 0.59)	0.00 (-0.07 − 0.09)	0.43 (0.37 − 0.49)	0.00 (-0.08 − 0.09)	0.00 (-0.08 − 0.08)
17	-0.150	0.225	0.375	0.075	0.700	0.75 (0.70 − 0.77)	0.41 (0.36 − 0.48)	0.53 (0.48 − 0.59)	0.29 (0.22 − 0.38)	0.42 (0.36 − 0.48)	0.23 (0.15 − 0.32)	0.55 (0.50 − 0.60)
18	-0.150	0.225	0.375	0.075	1.400	0.75 (0.72 − 0.77)	0.59 (0.55 − 0.64)	0.53 (0.48 − 0.57)	0.42 (0.35 − 0.50)	0.42 (0.35 − 0.47)	0.33 (0.27 − 0.39)	0.79 (0.77 − 0.82)
19	-0.150	0.225	0.375	0.075	2.100	0.75 (0.70 − 0.78)	0.67 (0.62 − 0.71)	0.54 (0.48 − 0.59)	0.48 (0.43 − 0.54)	0.42 (0.35 − 0.49)	0.38 (0.31 − 0.44)	0.89 (0.88 − 0.90)
20	-0.150	0.225	0.375	0.075	2.800	0.75 (0.71 − 0.77)	0.70 (0.65 − 0.73)	0.54 (0.48 − 0.58)	0.50 (0.44 − 0.55)	0.42 (0.34 − 0.47)	0.39 (0.32 − 0.46)	0.93 (0.93 − 0.94)
21	-0.200	0.300	0.500	0.100	0.000	0.79 (0.76 − 0.82)	0.00 (-0.07 − 0.10)	0.63 (0.59 − 0.66)	0.00 (-0.08 − 0.11)	0.51 (0.45 − 0.57)	0.00 (-0.07 − 0.08)	0.00 (-0.07 − 0.08)
22	-0.200	0.300	0.500	0.100	0.700	0.79 (0.75 − 0.82)	0.43 (0.37 − 0.49)	0.63 (0.57 − 0.67)	0.34 (0.26 − 0.41)	0.52 (0.47 − 0.56)	0.28 (0.20 − 0.35)	0.55 (0.50 − 0.60)
23	-0.200	0.300	0.500	0.100	1.400	0.79 (0.75 − 0.82)	0.63 (0.58 − 0.66)	0.62 (0.58 − 0.66)	0.49 (0.44 − 0.55)	0.51 (0.47 − 0.55)	0.41 (0.35 − 0.46)	0.79 (0.77 − 0.81)
24	-0.200	0.300	0.500	0.100	2.100	0.79 (0.75 − 0.82)	0.70 (0.66 − 0.74)	0.62 (0.58 − 0.66)	0.56 (0.51 − 0.60)	0.51 (0.47 − 0.55)	0.46 (0.39 − 0.51)	0.89 (0.88 − 0.90)
25	-0.200	0.300	0.500	0.100	2.800	0.79 (0.76 − 0.82)	0.74 (0.71 − 0.78)	0.62 (0.58 − 0.66)	0.58 (0.54 − 0.62)	0.51 (0.45 − 0.56)	0.48 (0.42 − 0.53)	0.93 (0.93 − 0.94)

Corresponding 2.5th and 97.5th percentiles are reported in the parentheses. Colour coding: -0.05 ☐ 0.95 (mean) Pearson correlation coefficient.

independent random parameters can be inferior to those that accommodate correlation (Mariel and Artabe, 2020; Mariel and Meyerhoff, 2018). While this correlation can stem from observable characteristics (e.g. gender, age and income), it may also be an artefact of unobserved latent variables. The importance of this latter point is often not fully appreciated. Indeed, a pertinent question is whether or not—and in what settings—allowing for correlation is an acceptable substitute for hybrid latent variable models and, conversely, if it is possible to say anything about the potential aptness of considering a hybrid latent variable model based on an inspection of the correlation structure of random parameters. To explore these issues we also estimate the corresponding models that allow for correlated random parameters (MXL-CORR, MXLIND-CORR and LVMXL-CORR, respectively). Estimating all six candidate models allows us to compare the effects under correctly specified and misspecified cases and to make inferences regarding the consequences of the naïve assumption(s). Combined, this leads to a total of 15,000 mixed logit models to estimate (i.e. 25 simulation treatments times 100 replications times six model specifications).

All models are coded and estimated using the maxLik library in R (see Henningsen and Toomet (2011) and R Core Team (2020) for further details). We used maximum simulated likelihood estimation using 500 quasi-random scrambled Sobol sequences for the simulation of the random parameters and latent variable. For all models, we started the estimation iterations using the parameters that were specified as part of the DGP.

4. Results

In Table 2, we show the mean difference in log-likelihood for all 25 DGPs over the 100 simulated Monte-Carlo datasets, and the corresponding 2.5th and 97.5th percentiles. Note that for the latent variable models we focus only on the fit of the choice model component, which we denote using LL*. First, and unsurprisingly, in accordance with Mariel and Meyerhoff (2018), we see that the models allowing for correlations between the random parameters fits the data better, i.e. produces higher log-likelihood values. This result holds for all three model specifications. However, we do note that the improvements in log-likelihood reported here do not penalise for the increased number of parameters. Second, including the indicator directly in the utility expression leads to better choice predictions. Referring back to the correlations between the indicator and latent variable for each DGP in Table 1, we see that this improvement in model fit is increasing in the degree of correlation between the two (i.e. $\rho_{W_k, i^\star}$). The most important take-away from this is that we find that the reduced form model without the latent variable fits the data equally well, which is consistent with Vij and Walker (2016). This fact really brings the question of what the additional benefit of a hybrid choice model is in many contexts to the forefront.

Moving beyond model fit is necessary to fully understand what is going on. While the reduced form models do "just as well" at predicting the chosen alternative, do they also retrieve unbiased and consistent estimates of the parameters and welfare measures? To explore this, in Table 3 and Table 4 we show the degree of bias in the parameters associated with the latent variable: specifically, with Table 3 and Table 4 comparing the mean error (i.e. the mean of all differences between the estimated values and the true value for each data generation setting) and the corresponding 2.5th and 97.5th percentiles for the models without and with correlations, respectively. We report the absolute bias, but rela-

Table 2. Mean improvement in log-likelihood (choice) over respective DGP baseline MXL model across the 100 Monte-Carlo simulations.

DGP	$LL_{MXL\text{-}corr} - LL_{MXL}$	$LL_{MXLIND} - LL_{MXL}$	$LL_{MXLIND\text{-}corr} - LL_{MXL}$	$LL^*_{LVMXL} - LL_{MXL}$	$LL^*_{LVMXL\text{-}corr} - LL_{MXL}$
1	2.61 (0.72 − 5.80)	3.94 (1.09 − 8.67)	6.49 (3.21 − 12.21)	0.80 (-0.48 − 3.18)	2.63 (0.40 − 5.61)
2	2.80 (0.83 − 6.01)	3.84 (1.19 − 8.16)	6.64 (2.61 − 11.60)	0.56 (-0.62 − 3.36)	2.80 (0.48 − 6.56)
3	3.17 (0.63 − 8.25)	3.86 (0.96 − 7.64)	6.96 (2.01 − 13.27)	1.23 (-0.51 − 6.15)	3.30 (0.27 − 8.91)
4	2.81 (0.73 − 6.60)	4.01 (0.81 − 8.27)	6.78 (2.71 − 12.01)	0.84 (-0.56 − 4.86)	2.96 (0.43 − 7.96)
5	2.70 (0.92 − 5.29)	4.22 (1.28 − 9.37)	6.93 (3.22 − 12.04)	0.71 (-0.49 − 3.82)	2.61 (0.48 − 5.66)
6	2.89 (0.82 − 6.70)	4.66 (1.71 − 9.66)	7.55 (3.52 − 13.17)	0.97 (-0.59 − 4.83)	2.92 (0.52 − 7.27)
7	3.07 (0.56 − 7.59)	7.76 (2.83 − 15.80)	10.75 (5.28 − 18.33)	0.68 (-1.07 − 5.66)	2.87 (0.07 − 6.94)
8	2.72 (0.86 − 5.88)	11.83 (4.58 − 21.67)	14.48 (6.66 − 24.49)	0.41 (-1.28 − 3.24)	2.59 (0.07 − 6.46)
9	3.05 (0.83 − 6.21)	13.24 (5.62 − 22.97)	16.23 (7.57 − 26.06)	0.75 (-1.35 − 4.56)	2.84 (0.38 − 6.20)
10	2.81 (0.65 − 7.12)	12.90 (6.00 − 20.92)	15.65 (8.04 − 23.95)	0.68 (-1.22 − 4.91)	2.55 (0.39 − 6.38)
11	3.66 (0.82 − 7.42)	3.93 (1.31 − 7.87)	7.61 (2.83 − 13.01)	1.63 (-0.64 − 5.93)	4.12 (0.55 − 8.51)
12	3.71 (0.84 − 7.82)	18.59 (8.54 − 32.76)	21.87 (11.68 − 35.90)	1.58 (-0.97 − 7.00)	3.61 (-0.10 − 8.02)
13	3.70 (1.21 − 8.33)	30.00 (18.04 − 42.02)	33.13 (21.65 − 45.63)	1.17 (-1.49 − 5.31)	3.31 (0.56 − 7.62)
14	3.40 (0.62 − 6.72)	34.15 (22.01 − 45.37)	36.93 (24.01 − 47.72)	1.04 (-1.60 − 3.83)	2.88 (-0.20 − 7.35)
15	3.78 (0.77 − 9.93)	35.66 (22.41 − 48.34)	38.77 (25.67 − 52.04)	1.23 (-1.80 − 5.07)	3.27 (-0.10 − 8.61)
16	5.85 (1.61 − 11.58)	4.27 (1.36 − 9.39)	10.13 (3.86 − 16.62)	2.73 (-0.48 − 8.77)	6.36 (1.70 − 13.09)
17	5.85 (2.69 − 12.59)	29.71 (16.42 − 43.96)	34.34 (22.39 − 49.87)	3.76 (-0.05 − 9.34)	5.91 (1.19 − 12.17)
18	6.27 (1.37 − 12.99)	51.78 (36.64 − 72.98)	55.72 (39.63 − 76.49)	3.84 (-1.36 − 9.92)	5.85 (-0.16 − 12.08)
19	6.00 (2.31 − 12.71)	60.54 (44.45 − 75.73)	63.91 (46.58 − 78.43)	3.77 (-0.62 − 9.16)	5.64 (1.16 − 12.07)
20	6.24 (1.97 − 13.13)	64.96 (44.41 − 84.90)	68.66 (46.50 − 89.99)	4.06 (-0.33 − 10.46)	5.91 (0.67 − 12.63)
21	10.28 (3.78 − 19.30)	3.74 (1.08 − 7.40)	14.08 (7.11 − 23.76)	6.53 (-0.51 − 16.35)	10.76 (3.75 − 17.82)
22	9.13 (3.71 − 15.91)	39.60 (24.32 − 55.60)	46.02 (30.55 − 63.27)	7.40 (0.71 − 14.11)	9.31 (3.69 − 16.53)
23	9.70 (3.56 − 17.87)	72.77 (51.73 − 94.77)	78.00 (56.74 − 100.87)	7.60 (1.09 − 15.96)	9.37 (2.90 − 17.13)
24	9.52 (3.49 − 16.60)	87.04 (68.13 − 107.79)	91.25 (72.57 − 111.79)	7.38 (1.42 − 14.40)	9.21 (2.54 − 16.81)
25	9.55 (3.25 − 18.01)	93.25 (71.70 − 116.24)	97.53 (76.31 − 119.89)	7.13 (-0.84 − 14.97)	8.92 (2.58 − 16.78)

Corresponding 2.5th and 97.5th percentiles are reported in the parentheses. Colour coding: 0 ▢▢▢▢▢ 100 (mean) improvement.

tive bias can be assessed by referring back to the true parameters associated with any given DGP in Table 1. Nonetheless, the tables are useful to compare the different DGPs and signing the bias.

First, looking at Table 3, the most striking result is that, in general, the standard deviation of the latent variable is underestimated (first column), the parameters for cost and recreation are overestimated while those for area and broadleaf are (for the most part) underestimated. We also remark that the latent variable interaction with the indicator shows a high degree of bias for all simulation settings. Indeed, in situations where $\psi=0$ we find that the interaction is underestimated, whereas for settings where $\psi>0$ we find that they are overestimated. Furthermore, while the extent of the bias is increasing in ψ, we see no such pattern for the standard deviation of the latent variable or the other estimated parameters. Recall that the DGP was based on the LVMXL model. While we can normally expect to see idiosyncratic bias because the integrals are simulated and the data randomly generated, the fact that we observe systematic bias is a cause for concern and should make any practitioner think twice about using hybrid choice models. The inablity to recover the true parameters—even when the DGP is known and we start the estimation at the true parameters—is disconcerting and underlines the point that these models are difficult to estimate even under "perfect" conditions.[5] So what does this say about our ability to

---

[5] We recognise that 500 quasi-random draws may not have been sufficient and that increasing the number of simulation draws may have led to a more stable set of parameter estimates. We justify this on the grounds that,

Table 3. Bias for parameters connected with the latent variable in the LVMXL model.

DGP	$\hat{\sigma}_{\mathscr{L}} - \sigma_{\mathscr{L}}$	$\hat{\gamma}_{cost} - \gamma_{cost}$	$\hat{\gamma}_{area} - \gamma_{area}$	$\hat{\gamma}_{broadleaf} - \gamma_{broadleaf}$	$\hat{\gamma}_{recreation} - \gamma_{recreation}$	$\hat{\psi} - \psi$
1	-0.79 (-1.20 – 1.10)	1.72 (0.09 – 3.88)	-0.17 (-0.63 – 0.44)	0.00 (-0.35 – 0.27)	-0.44 (-9.47 – 2.06)	-1.33 (-15.89 – 8.56)
2	-0.94 (-1.20 – 1.35)	2.45 (0.23 – 10.96)	0.02 (-0.32 – 0.58)	0.02 (-0.14 – 0.25)	-0.11 (-0.91 – 0.62)	0.96 (-0.64 – 4.79)
3	-0.93 (-1.20 – 0.06)	1.81 (0.32 – 6.72)	-0.02 (-1.01 – 0.52)	0.00 (-0.19 – 0.18)	0.00 (-1.14 – 1.39)	0.17 (-1.39 – 5.40)
4	-0.85 (-1.20 – 1.20)	1.42 (-0.02 – 7.48)	0.05 (-0.44 – 0.75)	0.04 (-0.18 – 0.63)	0.39 (-1.52 – 7.10)	0.16 (-2.18 – 10.25)
5	-0.89 (-1.20 – 0.23)	1.70 (0.41 – 8.64)	-0.03 (-0.46 – 0.46)	-0.01 (-0.25 – 0.18)	-0.09 (-1.99 – 1.50)	1.81 (-2.69 – 17.37)
6	-0.48 (-1.20 – 3.45)	1.27 (-0.16 – 2.97)	-0.24 (-1.85 – 1.37)	-0.07 (-0.49 – 0.90)	0.86 (-1.80 – 6.15)	-0.69 (-17.27 – 14.39)
7	-1.08 (-1.20 – -0.78)	2.04 (0.89 – 4.89)	-0.14 (-0.28 – -0.01)	-0.08 (-0.15 – 0.01)	0.08 (-0.30 – 0.66)	1.61 (-0.49 – 3.11)
8	-1.07 (-1.19 – -0.79)	1.81 (0.90 – 4.61)	-0.14 (-0.38 – -0.01)	-0.05 (-0.13 – 0.05)	0.11 (-0.33 – 0.75)	1.72 (-1.01 – 11.07)
9	-1.06 (-1.19 – -0.71)	1.79 (0.78 – 9.70)	-0.18 (-0.59 – -0.05)	-0.05 (-0.12 – 0.04)	0.10 (-0.46 – 0.65)	1.62 (-1.48 – 13.00)
10	-1.02 (-1.20 – -0.56)	1.71 (0.76 – 9.38)	-0.19 (-0.61 – -0.03)	-0.04 (-0.14 – 0.09)	0.06 (-0.72 – 0.85)	0.93 (-2.22 – 12.37)
11	-0.03 (-1.19 – 8.87)	1.31 (-0.07 – 4.69)	-0.36 (-6.72 – 1.43)	-0.01 (-1.71 – 5.00)	0.43 (-8.26 – 10.50)	-1.11 (-9.53 – 6.77)
12	-1.10 (-1.19 – -0.78)	2.21 (1.07 – 7.98)	-0.23 (-0.46 – -0.14)	-0.16 (-0.24 – -0.03)	0.09 (-0.28 – 0.55)	2.72 (-0.31 – 12.13)
13	-1.11 (-1.20 – -0.93)	1.84 (1.06 – 3.92)	-0.24 (-0.41 – -0.14)	-0.13 (-0.20 – 0.00)	0.15 (-0.20 – 0.65)	3.54 (-0.55 – 32.94)
14	-1.09 (-1.19 – -0.93)	1.60 (1.00 – 3.59)	-0.25 (-0.36 – -0.16)	-0.11 (-0.20 – 0.00)	0.19 (-0.17 – 0.50)	3.18 (-0.93 – 24.82)
15	-1.07 (-1.20 – -0.81)	1.85 (0.86 – 7.87)	-0.26 (-0.45 – -0.15)	-0.10 (-0.23 – 0.01)	0.22 (-0.12 – 0.71)	4.93 (-1.37 – 40.58)
16	0.37 (-1.19 – 6.28)	1.12 (-0.39 – 4.17)	-0.39 (-7.00 – 4.43)	-0.09 (-7.46 – 6.36)	0.62 (-17.72 – 14.88)	0.13 (-7.86 – 11.38)
17	-1.11 (-1.20 – -0.96)	2.13 (1.44 – 3.77)	-0.33 (-0.44 – -0.23)	-0.24 (-0.33 – -0.15)	0.14 (-0.25 – 0.50)	1.66 (-0.26 – 11.99)
18	-1.09 (-1.20 – -0.86)	1.60 (1.04 – 2.59)	-0.36 (-0.52 – -0.27)	-0.18 (-0.29 – -0.06)	0.26 (-0.10 – 0.80)	1.72 (-0.32 – 12.87)
19	-1.08 (-1.19 – -0.91)	1.53 (1.09 – 2.40)	-0.38 (-0.50 – -0.27)	-0.17 (-0.26 – -0.06)	0.27 (-0.10 – 0.64)	4.19 (-0.54 – 32.10)
20	-1.09 (-1.20 – -0.89)	1.62 (0.99 – 5.40)	-0.39 (-0.59 – -0.26)	-0.14 (-0.32 – -0.02)	0.30 (-0.02 – 0.70)	3.78 (-0.88 – 21.76)
21	0.82 (-1.20 – 8.75)	1.27 (-0.01 – 4.31)	-1.32 (-9.88 – 4.62)	0.90 (-8.21 – 8.95)	2.87 (-15.59 – 16.90)	-0.81 (-9.48 – 4.24)
22	-1.11 (-1.20 – -0.95)	2.04 (1.56 – 2.49)	-0.44 (-0.52 – -0.37)	-0.31 (-0.38 – -0.20)	0.23 (-0.06 – 0.45)	0.19 (-0.25 – 1.59)
23	-1.08 (-1.19 – -0.87)	1.80 (1.08 – 3.42)	-0.48 (-0.63 – -0.36)	-0.25 (-0.40 – -0.08)	0.32 (-0.04 – 0.73)	2.88 (-0.17 – 15.75)
24	-1.07 (-1.20 – -0.90)	1.58 (1.18 – 3.59)	-0.51 (-0.62 – -0.36)	-0.21 (-0.40 – -0.07)	0.36 (-0.09 – 0.83)	3.43 (-0.14 – 18.75)
25	-1.07 (-1.19 – -0.89)	1.72 (1.14 – 4.66)	-0.49 (-0.63 – -0.34)	-0.21 (-0.42 – -0.10)	0.33 (-0.09 – 0.73)	6.15 (-0.25 – 42.25)

Corresponding 2.5th and 97.5th percentiles are reported in the parentheses. Colour coding: -9.3 ▬ 8.8 (mean) bias.

Table 4. Bias for parameters connected with the latent variable in the LVMXL-CORR model.

DGP	$\hat{\sigma}_{\mathscr{L}} - \sigma_{\mathscr{L}}$	$\hat{\gamma}_{cost} - \gamma_{cost}$	$\hat{\gamma}_{area} - \gamma_{area}$	$\hat{\gamma}_{broadleaf} - \gamma_{broadleaf}$	$\hat{\gamma}_{recreation} - \gamma_{recreation}$	$\hat{\psi} - \psi$
1	5.28 (-1.17 – 53.81)	-0.03 (-0.79 – 1.07)	-0.04 (-0.31 – 0.25)	-0.61 (-7.76 – 0.71)	-0.08 (-1.45 – 0.51)	-7.97 (-56.61 – -1.42)
2	3.62 (-1.18 – 40.18)	0.05 (-0.27 – 0.68)	-0.01 (-0.20 – 0.15)	-0.08 (-0.88 – 1.05)	-0.02 (-0.83 – 0.59)	-6.92 (-43.68 – -1.96)
3	1.60 (-1.16 – 23.06)	-0.02 (-1.03 – 0.50)	0.00 (-0.32 – 0.38)	0.05 (-1.66 – 2.00)	0.04 (-1.00 – 1.02)	-5.37 (-27.26 – -2.52)
4	1.51 (-1.16 – 22.59)	-0.13 (-1.79 – 0.61)	-0.03 (-0.57 – 0.22)	0.20 (-2.75 – 6.69)	0.02 (-1.40 – 1.74)	-5.82 (-27.49 – -3.13)
5	2.27 (-1.13 – 20.64)	-0.03 (-0.85 – 0.28)	-0.01 (-0.11 – 0.14)	-0.06 (-1.81 – 1.07)	-0.04 (-0.81 – 0.51)	-7.34 (-26.24 – -3.82)
6	3.76 (-1.18 – 50.47)	-0.21 (-1.12 – 1.04)	-0.10 (-0.96 – 0.35)	0.46 (-2.49 – 5.91)	-0.29 (-5.01 – 2.38)	-6.44 (-53.27 – -1.42)
7	5.15 (-1.16 – 44.43)	0.03 (-0.11 – 0.14)	-0.03 (-0.08 – 0.07)	-0.05 (-0.38 – 0.35)	0.00 (-0.23 – 0.24)	-8.53 (-47.93 – -2.02)
8	2.16 (-1.18 – 28.30)	-0.02 (-0.30 – 0.10)	-0.01 (-0.08 – 0.12)	-0.01 (-0.49 – 0.50)	-0.02 (-0.47 – 0.24)	-6.08 (-32.50 – -2.56)
9	2.55 (-1.17 – 25.35)	-0.04 (-0.32 – 0.08)	0.00 (-0.08 – 0.11)	-0.02 (-0.57 – 0.55)	0.00 (-0.33 – 0.40)	-7.06 (-30.25 – -3.18)
10	2.11 (-1.12 – 21.01)	-0.04 (-0.31 – 0.11)	0.00 (-0.07 – 0.15)	-0.05 (-0.63 – 0.53)	-0.02 (-0.40 – 0.41)	-7.16 (-26.61 – -3.83)
11	1.85 (-1.16 – 42.68)	-0.14 (-6.11 – 3.48)	-0.04 (-1.84 – 4.14)	0.18 (-10.73 – 13.76)	0.16 (-2.60 – 6.40)	-4.49 (-45.48 – -1.42)
12	6.38 (-1.19 – 45.15)	0.03 (-0.19 – 0.11)	-0.07 (-0.14 – 0.05)	-0.14 (-0.50 – 0.22)	-0.02 (-0.27 – 0.21)	-9.78 (-48.65 – -2.05)
13	3.58 (-1.17 – 32.89)	0.01 (-0.14 – 0.10)	-0.04 (-0.13 – 0.10)	-0.06 (-0.43 – 0.49)	-0.04 (-0.22 – 0.16)	-7.63 (-37.09 – -2.70)
14	2.99 (-1.18 – 23.36)	0.01 (-0.14 – 0.09)	-0.03 (-0.13 – 0.09)	-0.06 (-0.39 – 0.29)	-0.01 (-0.26 – 0.25)	-7.72 (-28.26 – -3.28)
15	3.47 (-1.14 – 20.67)	0.00 (-0.16 – 0.09)	-0.02 (-0.13 – 0.12)	-0.02 (-0.37 – 0.45)	-0.03 (-0.32 – 0.31)	-8.79 (-26.27 – -3.92)
16	3.30 (-1.17 – 50.50)	0.21 (-12.23 – 4.72)	0.08 (-3.78 – 5.22)	0.22 (-17.39 – 13.41)	-0.21 (-4.24 – 4.12)	-5.91 (-53.30 – -1.42)
17	4.30 (-1.18 – 44.74)	0.06 (-0.17 – 0.14)	-0.11 (-0.20 – -0.01)	-0.22 (-0.55 – 0.11)	-0.04 (-0.21 – 0.16)	-7.70 (-48.24 – -2.07)
18	2.30 (-1.17 – 27.76)	0.03 (-0.15 – 0.13)	-0.06 (-0.18 – 0.08)	-0.11 (-0.45 – 0.32)	-0.03 (-0.24 – 0.24)	-6.40 (-31.96 – -2.72)
19	2.58 (-1.19 – 23.51)	0.02 (-0.15 – 0.12)	-0.04 (-0.18 – 0.13)	-0.07 (-0.37 – 0.40)	-0.01 (-0.20 – 0.22)	-7.35 (-28.41 – -3.35)
20	2.67 (-1.18 – 20.77)	0.00 (-0.19 – 0.13)	-0.02 (-0.18 – 0.14)	-0.05 (-0.35 – 0.45)	-0.02 (-0.21 – 0.23)	-8.06 (-26.37 – -3.95)
21	2.38 (-1.17 – 42.52)	-1.40 (-9.36 – 8.25)	0.98 (-2.60 – 12.66)	1.56 (-8.79 – 16.31)	0.87 (-4.04 – 13.92)	-4.98 (-45.32 – -1.42)
22	2.29 (-1.19 – 41.54)	0.07 (-0.05 – 0.18)	-0.13 (-0.21 – 0.00)	-0.23 (-0.53 – 0.14)	-0.07 (-0.31 – 0.16)	-5.65 (-45.04 – -2.06)
23	4.80 (-1.19 – 34.98)	0.06 (-0.10 – 0.17)	-0.09 (-0.26 – 0.11)	-0.15 (-0.47 – 0.26)	-0.04 (-0.29 – 0.22)	-8.91 (-39.18 – -2.70)
24	2.28 (-1.17 – 22.37)	0.00 (-0.21 – 0.16)	-0.03 (-0.26 – 0.15)	-0.10 (-0.51 – 0.37)	-0.02 (-0.28 – 0.19)	-7.01 (-27.27 – -3.35)
25	3.67 (-1.19 – 22.58)	0.03 (-0.16 – 0.16)	-0.05 (-0.23 – 0.12)	-0.14 (-0.50 – 0.41)	-0.02 (-0.30 – 0.19)	-9.12 (-28.18 – -4.03)

Corresponding 2.5th and 97.5th percentiles are reported in the parentheses. Colour coding: -9.3 ▬ 8.8 (mean) bias.

retrieve unbiased parameters in empirical settings when the DGP and its parameters are unknown?

---

in total, we estimated 15,000 mixed logit models. Increasing the number of draws would have entailed considerably more estimation time.

Turning our attention to the model with correlation reported in Table 4, we see some stark differences compared to the models without correlation. The most notable change is the switching signs and larger magnitude of the bias in the standard deviation of the latent variable and ψ. This shows overwhelming evidence that allowing for correlation in the random parameters when this was not part of the DGP leads to severe bias in the parameters associated with the latent variable when the latent variable is the only source of correlation in the data. Intuitively, this makes sense. We now have a whole correlation structure, in addition to the latent variable, trying to describe the influence of the latent variable. Crucially, the magnitude of the bias of ψ is important because it can lead to an entirely misleading interpretation of the latent variable. Note that given the true parameters of ψ in Table 1, the magnitude of the bias implies that the estimated value of the impact of the latent variable will be negative. Consequently, *ceteris paribus*, we would wrongly conclude that an increase in the latent variable is associated with an increase in the MWTP for the environmental attributes and a decrease in the tendency to report pro-environmental attitudes on our three-point Likert scale question. If we look at the bias in the γ parameters, we see that this is much smaller compared to the LVMXL model. While the bias for the standard deviation of the latent variable and the latent variable indicator interactions switch signs and are considerably larger, the bias for γ is much smaller, which makes it difficult to ascertain the net effect on welfare estimates. What it does highlight, and we cannot stress this enough, is that there appears to be a dilemma and a set of unforeseen trade-offs when it comes to hybrid choice model selection. The extent to which this is just an artefact of our DGP parameters and assumptions remain unclear, as this would require further simulation work under a broader range of settings. Nonetheless, it does show that model selection comes down to the analyst's belief about correlations and that model selection and the use of these models truly is "a road fraught with peril".

To determine how the above results affect MWTP, we compare the overlapped estimated area of the actual MWTP kernel density estimates to that of the kernel density of the distribution of the means of the individual-specific posterior MWTP. This is an easy way to quantify the similarities or differences between the actual and predicted MWTP distributions. To make the comparison more intuitive, we consider the difference in the percentage overlap of each model to the basic MXL model, which represents the most naïve assumptions about the DGP. To illustrate how this difference is sensitive to the correlation between the DGP MWTP and the latent variable, as well as the indicator, we plot the differences against $\rho_{W_k L}$ and $\rho_{W_k i}$, respectively. We sort the corresponding points by the correlation measure and graph this using a technique known as locally estimated scatterplot smoothing (LOESS).[6] We show these in Figure 1, Figure 2 and Figure 3 (and their associated 95 percent confidence level) for the area, broadleaf and recreation attributes, respectively. Specifically, the locally regressed and smoothed percentage point differences in overlap of each candidate model relative to the MXL model are plotted against: (i) the correlation between the actual MWTP and the latent variable in the left panel; and, (ii) the correlation between the actual MWTP and the underlying continuous variable relat-

---

[6] The LOESS method is a non-parametric approach where fitting is done locally (in our case with a neighbourhood proportion of 0.4). The result is a smooth curve, which makes it easier to detect trends. This was achieved using the stats library in R .

Figure 1. Percentage point difference in overlap of MWTP distributions relative to the true MWTP distribution for area.
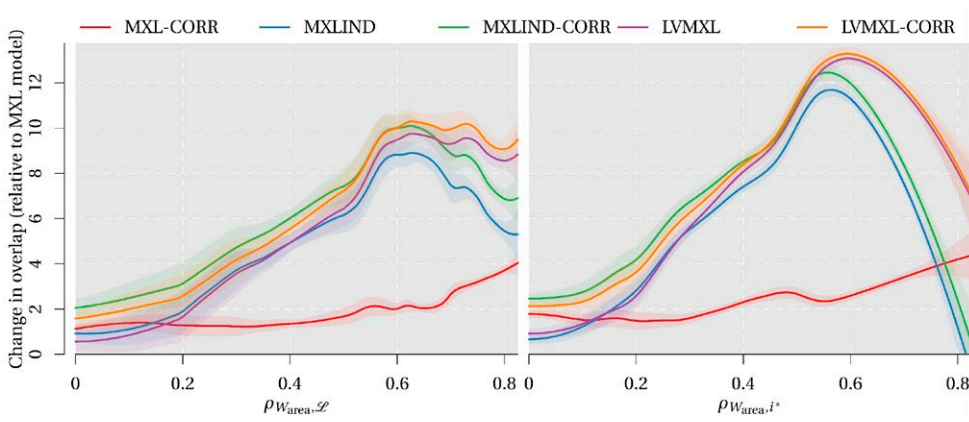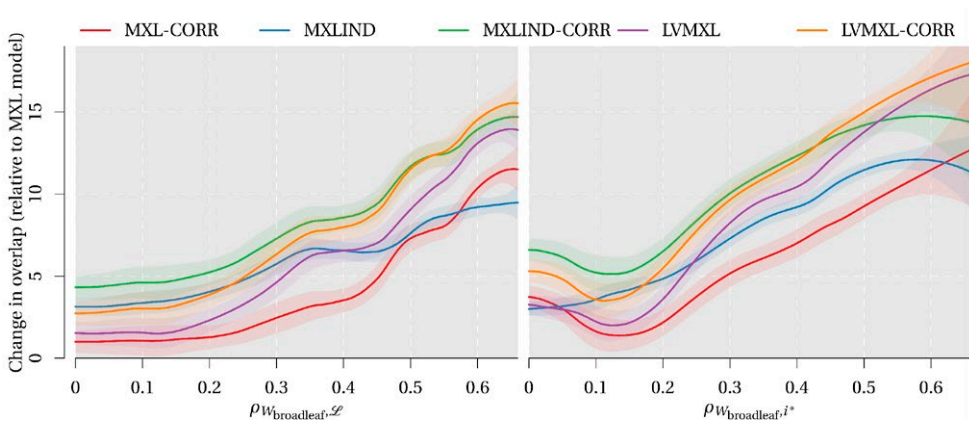

Figure 2. Percentage point difference in overlap of MWTP distributions relative to the true MWTP distribution for broadleaf.

ing to the indicator in the right panel. As we move from the origin to the right, the degree of correlation increases. The vertical axis shows the percentage point difference in overlap relative to the MXL model, meaning that a move up this axis signifies that the candidate model does better at predicting the true MWTP distribution relative to the MXL model.

As might be expected, a visual inspection of Figures 1-3 reveals that all models generally retrieve the same MWTP distributions when the correlation between MWTP and either the latent variable or indicator is low. But, as the degree of correlation increases, we can see that the models that accommodate correlated random parameters and/or environmental tendency (either directly or indirectly) are better at explaining the true MWTP distributions. Recall the discussion relating to the switching signs for the bias in the standard deviation of the latent variable and the interaction of the latent variable;

Figure 3. Percentage point difference in overlap of MWTP distributions relative to the true MWTP distribution for recreation.
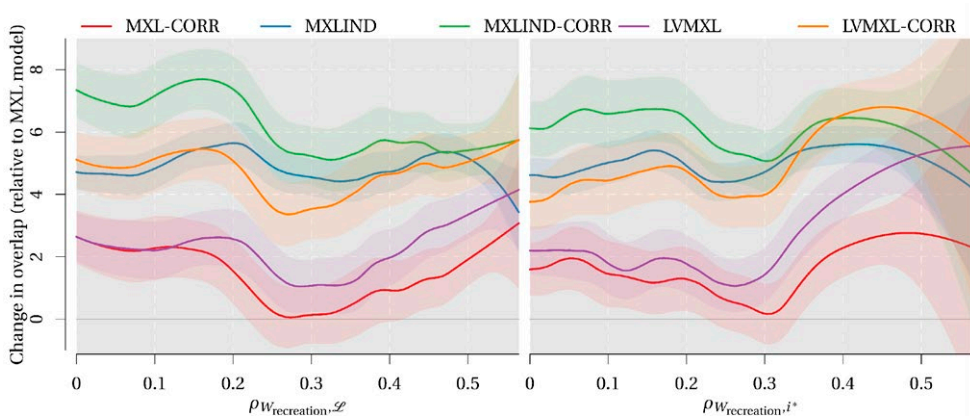
this switch does not appear to affect the estimation of MWTP. Relatively speaking, for the most part, the LVMXL and LVMXL-CORR curves are closely aligned.

Focusing on Figure 1, we see that as the correlation with the latent variable (left panel) increases beyond 0.2, the models that directly or indirectly include the indicator outperform the MXL and MXL-CORR models. This is an important finding, since it suggests that simply allowing for correlation does not, in itself, allow us to recover the correct MWTP distribution. However, it must be noted that this result is strongest in cases where the correlation with the latent variable is moderate. As the strength of the relationship gets very high ($\rho > 0.6$) there is a clear turning point, indicating that the relative importance of directly or indirectly including the indicator lessens. But this same finding is not observed for the MXL-CORR model, to the extent that just allowing for correlated random parameters does all most just as well at retrieving the correct MWTP to pay distribution. Importantly, this suggests that if the analyst believes that most, if not all, of the correlation between the random parameters are caused by a single unobserved latent variable, and that the effect of this variable is sufficiently strong, then simply estimating a standard mixed logit model with a full correlation structure may be sufficient if MWTP is the key measure of interest. Though, of course, this comes at the expense of not knowing the underlying source of heterogeneity, which may, or may not, be of interest. We also observe that the MXLIND and MXLIND-CORR are better able to uncover the true MWTP distribution compared to the LVMXL and LVMXL-CORR, respectively, when the strength of relationship between the latent variable and MWTP is weak or moderate. As the strength of relationship increases, however, we remark that this no longer holds. This additional insight implies that the relative advantage of ICLV models over simpler models to retrieve the correct MWTP distribution is dependent on the strength of the role that the latent variable plays on the distribution. While not a surprising finding, it reinforces the need to think twice about using hybrid choice models in situations where it is believed that the latent variable is weakly related. These findings are perhaps better illustrated when the change in overlap is plotted against the correlation with the indicator. The downward turn

towards the MXL baseline is even more pronounced at higher levels of correlation for all but the MXL-CORR model and especially so for the models that directly include the indicator responses in the utility function. At this point, recall that the "area" attribute is also the one that is linked the strongest to the latent variable and the indicator. This explains why the difference between the models are so stark and why we see that this result is mitigated as the relationship between MWTP and the latent variable and indicator becomes weaker. For example, looking at the Figures 2 and 3, where the strengths of association are lower, we see that the predicted curves are more closely aligned and do not exhibit an inverted U-shape. This implies that, for these attributes, models that include the indicator (either directly or indirectly) do not produce markedly better predictions of the MWTP distribution compared to the MXL-CORR model, and this holds irrespective of the correlation between MWTP and either the latent variable or indicator. For the recreation attribute (Figure 3), which had the lowest association with the latent variable and indicator response, the predicted curves are relatively flat, suggesting that the prediction of the MWTP distribution is less sensitive to which of candidate models is used. While these are also obvious findings, the fact that we are able to retrieve, show and prove them through our simulation is reassuring.

In generating the results illustrated in Figures 1-3 we took account of all synthetic individuals per DGP. However, this may mask the relative performance of each candidate model to correctly predict MWTP for individuals who hold a particular environmental tendency. Indeed, one of the often-purported advantages of ICLV models is their ability to provide additional insight on preference heterogeneity, particular among those with different latent attitudes. While, as stated earlier, we should be prudent about making policy recommendations on the basis of a latent variable – as well as the, obvious, impracticality, and futility, of targeting policy on the basis of an indicator response – policy makers may still be interested in knowing how members of society with different environmental tendencies judge their policies. For this reason, in Figures 4-5, we plot the locally regressed and smoothed mean bias in MWTP for each attribute against the correlation between the attribute and the latent variable broken down by whether an individual holds anti-, neutral or pro-environmental tendencies, depicted on the first, second and third panel, respectively.[7] For comparison, in the fourth panel, we also present this for all individuals. Looking firstly at this fourth panel, we see that the curves essentially overlap and are not significantly different from zero when the correlation between MWTP and the latent variable is weak or moderate. In these cases, the ability to retrieve the mean MWTP (across all individuals) does not appear to be affected by the degree of correlation with the latent variable nor by which of the candidate models we use. As can be seen in Figures 4-5, however, these curves begin to diverge as the degree of correlation increases ($\rho > 0.5$) to the extent that some are significantly different from zero. This insight suggests that if the main interest is on describing the means of the posterior MWTP distributions at the sample level, model choice is perhaps only consequential when the MWTP distribution

---

[7] For this, we subtract the actual individual-specific MWTP from the mean of the predicted individual-specific posterior MWTP and take the arithmetic mean for each data generation setting and model and, again, apply the LOESS method with a smoothing parameter of 0.4. The results are qualitatively similar for correlations between the attribute and the indicator and is omitted from the paper for brevity, but are available from the corresponding author upon request.

Figure 4. Mean bias in MWTP broken down by anti-, neutral and pro-environmental tendencies for area.
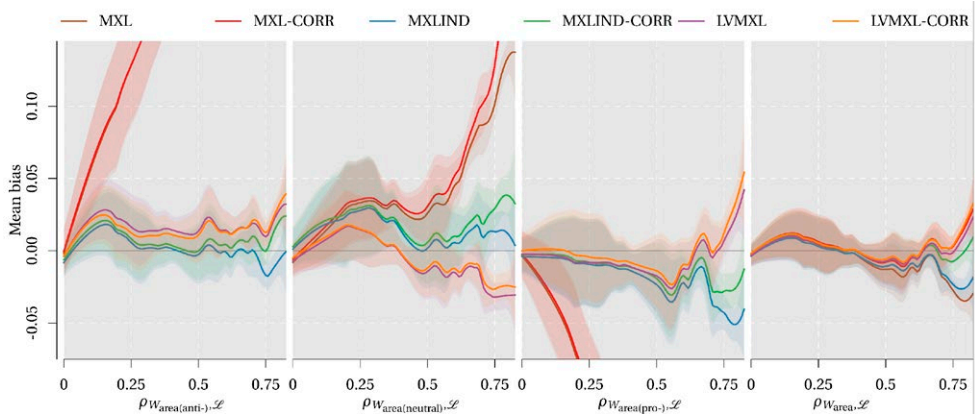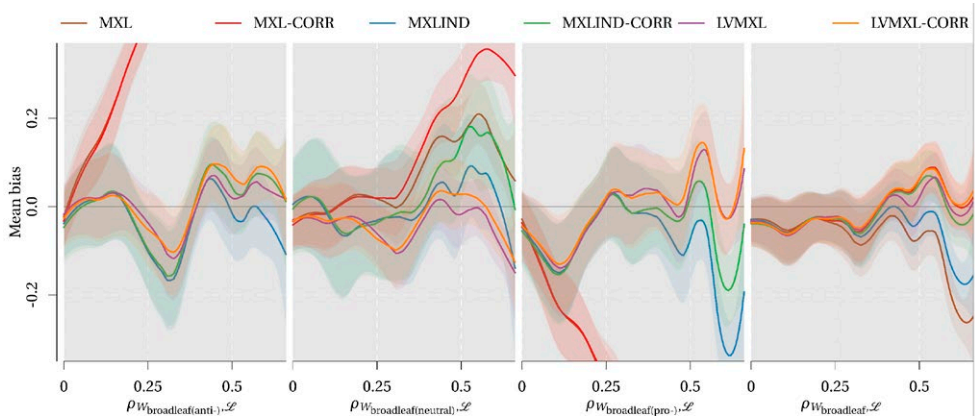

Figure 5. Mean bias in MWTP broken down by anti-, neutral and pro-environmental tendencies for broadleaf.

is believed to be strongly correlated with the latent variable. This is expected given the results above and that more flexible models are preferred if you suspect high degrees of correlation between MWTP and the latent variable. However, the corresponding curves produced for individuals who hold anti-, neutral or pro-environmental tendencies tell a somewhat different story. Only when the MWTP and latent variable distributions are uncorrelated do we find that all models produce relatively unbiased estimates of MWTP irrespective of environmental tendency. However, with any degree of correlation we see that the MXL and MXL-CORR models produce biased MWTP estimates for each sub-group. Specifically, these models overestimate individual-specific MWTP for individuals who hold anti- and (albeit to a lesser extent) neutral environmental tendencies, whereas they underestimate for the subgroup with pro-environmental tendencies. An important

finding for analysts who make use of individual-specific posterior MWTP estimates is that the extent of these biases increase with the degree of correlation. While this trend of overestimating MWTP for anti- as well as neutral environmental tendencies and under-estimating for pro-environmental tendencies still largely holds for the other candidate models, it is less evident and we observe it to be less sensitive to the degree of correlation. Nonetheless, there appears to be systematic differences between the models where we have included the indicators directly and their analogous latent variable models. For example, in Figures 4-5, relative to the MXLIND and MXLIND-CORR models, the LVMXL and LVMXL-CORR models, respectively, produce higher MWTP estimates for the anti- and pro-environmental tendency subgroups, but lower estimates for the neutral subgroup. Furthermore, these differences become more apparent as the degree of correlation between the MWTP and latent variable distributions increase. While the extent to which this finding can be generalized beyond our data generation settings is unclear, it does, nonetheless, further emphasise the difficultly associated with model selection when latent attitudes are believed to play an important role on MWTP.

5. Discussion and concluding remarks

In this paper, we generate a series of Monte-Carlo simulations that separately control for the strength of relationship between the latent variable and preferences and the strength of relationship between the latent variable and the indicator. In the real world, structural equations usually comprise standard socio-demographic characteristics and are often weak. To mimic this in the present paper, without complicating the DGP more than necessary, we treat the latent variable as normally distributed with zero mean and estimated standard deviation. This is exactly identical to a structural equation containing only an error-term. This also means that our reduced form model is a mixed logit model with an additional random error component (Vij and Walker, 2016). In the present paper, we used a simple three-point Likert scale question as an indicator of environmental tendency. This indicator was included in an ordered logit measurement equation. For each dataset generated, we estimate a random parameters mixed logit model, a random parameters mixed logit model with the indicators mapping directly to the marginal utilities and an ICLV random parameters mixed logit model, each with and without allowing for correlation among the random parameters.

From our results, it is clear that if you are only interested in choice prediction, then a mixed logit model with correlation may perform equally well to a hybrid choice model. While this is consistent with the general result of Vij and Walker (2016), who suggest that a reduced form model will fit the data at least as well, this is not in and of itself a reason to not use ICLV models. As we, and others, have shown, such models can offer greater insight into underling behavioural phenomena and contribute to decomposing marginal effects of the latent attitude on welfare estimates. But whether or not these additional behavioural insights outweigh the costs of estimating them remains an empirical question and will be entirely context dependent. In our simulations, we show that if the structural and measurement equations are weak (i.e. if observable characteristics are poor predictors of the latent variables, if appropriate indicators are not available and/or if the correlation between preferences, the latent construct and the indicator is weak), then

the model's ability to separately identify the marginal effects are likely limited and the benefits of developing and using an ICLV model are less clear-cut. In the cases where we do have weak structural equations, the use of measurement equations can help explain the latent variable and improve the fit of our choice model. Unfortunately, in real world applications we do not know *a priori* whether an indicator is good or bad, nor is there much guidance on the strengths of correlation. But there are ways to identify *better* indicators using, for example, exploratory factor analysis (Hoyos et al., 2015; Mariel et al., 2018; Mariel and Meyerhoff, 2016). Ultimately, however, we show that model selection should be driven by the analyst's belief about the strength of correlations between preferences, the latent variable and indicator. In any case, we need to be careful and mindful of the criticisms of Chorus and Kroesen (2014) and Kroesen and Chorus (2018): given the potential endogenous relationship between the latent variable and choice and the cross-sectional nature of the data, it is impossible to ascertain a causal relationship between attitudes and behaviour meaning that we should be very careful recommending policies that target the latent variable itself.

While we have not spent significant time talking about prediction in the present paper, we do feel it is prudent to reiterate that hybrid choice models can lead to improved predictions, but that any improvements are only likely in the case where it would be possible to predict the future state of the latent variable itself (Vij and Walker, 2016; Yáñez et al., 2010). More likely than not, this type of data will not be available. This is also possibly why we see that our models fit the data equally well, i.e. in terms of explaining the sequence of choices made by individuals. That said, the conclusions in this paper echo those of many others (Chorus and Kroesen, 2014; Kroesen and Chorus, 2018; Vij and Walker, 2016), that we need to take better heed of the quality of our data and recognize the limitations of it. The usefulness of ICLV models hinge on the quality of the data, and an ICLV model applied to poor data may add nothing to explanatory power and even less to policy. Furthermore, it is clear from our simulation work that even under "perfect" conditions, we struggled to retrieve the true parameters of the model, and the appropriateness of the model itself came down to the degree of correlation between our attributes, latent variable and indicator. Taken together, this makes the use of latent variable models, perhaps especially to inform policy, a road fraught with peril.

Acknowledgements

References

Alemu, M.H., Olsen, S.B., 2019. Linking Consumers' Food Choice Motives to their Preferences for Insect-based Food Products: An Application of Integrated Choice and Latent Variable Model in an African Context. *Journal of Agricultural Economics* 70, 241–258. https://doi.org/10.1111/1477-9552.12285

Bahamonde-Birke, F.J., Ortúzar, J. de D., 2017. Analyzing the continuity of attitudinal and perceptual indicators in hybrid choice models. Journal of Choice Modelling, IATBR 2015 - 14th International Conference on Travel Behaviour Research (IATBR) 25, 28–39. https://doi.org/10.1016/j.jocm.2017.01.003

Ben-Akiva, M., Mcfadden, D., Train, K., Walker, J., Bhat, C., Bierlaire, M., Bolduc, D., Boersch-Supan, A., Brownstone, D., Bunch, D.S., Daly, A., De Palma, A., Gopinath, D., Karlstrom, A., Munizaga, M.A., 2002. Hybrid Choice Models: Progress and Challenges. Marketing Letters 13, 163–175. https://doi.org/10.1023/A:1020254301302

Bhat, C.R., Dubey, S.K., Nagel, K., 2015. Introducing non-normality of latent psychological constructs in choice modeling with an application to bicyclist route choice. *Transportation Research Part B: Methodological* 78, 341–363. https://doi.org/10.1016/j.trb.2015.04.005

Boyce, C., Czajkowski, M., Hanley, N., 2019. Personality and economic choices. *Journal of Environmental Economics and Management* 94, 82–100. https://doi.org/10.1016/j.jeem.2018.12.004

Chorus, C.G., Kroesen, M., 2014. On the (im-)possibility of deriving transport policy implications from hybrid choice models. *Transport Policy* 36, 217–222. https://doi.org/10.1016/j.tranpol.2014.09.001

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. https://doi.org/10.1007/s11116-011-9351-z

Guevara, C.A., Ben-Akiva, M., 2010. Addressing Endogeneity in Discrete Choice Models: Assessing Control-Function and Latent-Variable Methods, in: Hess, S., Daly, A. (Eds.), Choice Modelling: The State-of-the-Art and The State-of-Practice. Emerald Group Publishing Limited, pp. 353–370. https://doi.org/10.1108/9781849507738-016

Henningsen, A., Toomet, O., 2011. maxLik: A package for maximum likelihood estimation in R. *Computational Statistics, Computational* S 26, 443–458. https://doi.org/10.1007/s00180-010-0217-1

Hess, S., 2012. Rethinking heterogeneity: the role of attitudes, decision rules and information processing strategies. *Transportation Letters* 4, 105–113. https://doi.org/10.3328/TL.2012.04.02.105-113

Hess, S., Stathopoulos, A., 2013. Linking response quality to survey engagement: A combined random scale and latent variable approach. *Journal of Choice Modelling* 7, 1–12. https://doi.org/10.1016/j.jocm.2013.03.005

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. *Science of The Total Environment* 505, 1100–1111. https://doi.org/10.1016/j.scitotenv.2014.10.066

Kassahun, H.T., Nicholson, C.F., Jacobsen, J.B., Steenhuis, T.S., 2016. Accounting for user expectations in the valuation of reliable irrigation water access in the Ethiopian highlands. *Agricultural Water Management* 168, 45–55. https://doi.org/10.1016/j.agwat.2016.01.017

Kroesen, M., Chorus, C., 2018. The role of general and specific attitudes in predicting travel behavior – A fatal dilemma? *Travel Behaviour and Society* 10, 33–41. https://doi.org/10.1016/j.tbs.2017.09.004

Kroesen, M., Handy, S., Chorus, C., 2017. Do attitudes cause behavior or vice versa? An alternative conceptualization of the attitude-behavior relationship in travel behavior modeling. *Transportation Research Part A: Policy and Practice* 101, 190–202. https://doi.org/10.1016/j.tra.2017.05.013

Lundhede, T., Jacobsen, J.B., Hanley, N., Strange, N., Thorsen, B.J., 2015. Incorporating Outcome Uncertainty and Prior Outcome Beliefs in Stated Preferences. *Land Economics* 91, 296–316. https://doi.org/10.3368/le.91.2.296

Mariel, P., Artabe, A., 2020. Interpreting correlated random parameters in choice experiments. *Journal of Environmental Economics and Management* 103, 102363. https://doi.org/10.1016/j.jeem.2020.102363

Mariel, P., Hoyos, D., Artabe, A., Guevara, C.A., 2018. A multiple indicator solution approach to endogeneity in discrete-choice models for environmental valuation. *Science of The Total Environment* 633, 967–980. https://doi.org/10.1016/j.scitotenv.2018.03.254

Mariel, P., Meyerhoff, J., 2018. A More Flexible Model or Simply More Effort? On the Use of Correlated Random Parameters in Applied Choice Studies. *Ecological Economics* 154, 419–429. https://doi.org/10.1016/j.ecolecon.2018.08.020

Mariel, P., Meyerhoff, J., 2016. Hybrid discrete choice models: Gained insights versus increasing effort. *Science of The Total Environment* 568, 433–443. https://doi.org/10.1016/j.scitotenv.2016.06.019

McFadden, D., 1986. The Choice Theory Approach to Market Research. *Marketing Science* 5, 275–297. https://doi.org/10.1287/mksc.5.4.275

McFadden, D., Train, K., 2000. Mixed MNL models for discrete response. *Journal of Applied Econometrics* 15, 447–470. https://doi.org/10.1002/1099-1255(200009/10)15:5<447::AID-JAE570>3.0.CO;2-1

R Core Team, 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Taye, F.A., Vedel, S.E., Jacobsen, J.B., 2018. Accounting for environmental attitude to explain variations in willingness to pay for forest ecosystem services using the new environmental paradigm. *Journal of Environmental Economics and Policy* 7, 420–440. https://doi.org/10.1080/21606544.2018.1467346

Vij, A., Walker, J.L., 2016. How, when and why integrated choice and latent variable models are latently useful. *Transportation Research Part B: Methodological* 90, 192–217. https://doi.org/10.1016/j.trb.2016.04.021

Yáñez, M.F., Raveau, S., Ortúzar, J. de D., 2010. Inclusion of latent variables in Mixed Logit models: Modelling and forecasting. *Transportation Research Part A: Policy and Practice* 44, 744–753. https://doi.org/10.1016/j.tra.2010.07.007

Zawojska, E., Bartczak, A., Czajkowski, M., 2019. Disentangling the effects of policy and payment consequentiality and risk attitudes on stated preferences. *Journal of Environmental Economics and Management* 93, 63–84. https://doi.org/10.1016/j.jeem.2018.11.007