

Chapter 5

Econometric Modelling: Basics



Abstract This chapter addresses basic topics related to choice data analysis. It starts by describing the coding of attribute levels and choosing the functional form of the attributes in the utility function. Next, it focuses on econometric models with special attention devoted to the random parameter mixed logit model. In this context, the chapter compares different coefficient distributions to be used, addresses specifics of the cost attribute coefficient and it pays attention to potential correlations between random coefficients. Finally, topics related to the estimation procedure such as assuring its convergence or random draws are discussed.

5.1 Coding of Attribute Levels: Effects, Dummy or Continuous

In the choice experiment literature, the two most common ways of coding attribute levels for modelling are dummy and effects coding. Most often the choice between dummy or effects coding arises when researchers consider how qualitatively (categorically) described attributes should enter the utility function, and when researchers want to relax assumptions about linearity of continuously coded attributes (see Sect. 5.2).

Consider an attribute with L levels. The (quantitative or qualitative) levels of this attribute are transformed into $L - 1$ dummy variables taking a value of one if a level is present in an alternative and equalling zero if it is not. The L -th level is omitted from analysis in order to avoid perfect collinearity. The utility of the L th level is per definition zero, and the $L - 1$ parameter estimates for the dummy variables capture the utility difference relative to the omitted baseline level. However, the utility of the status quo alternative is defined to be zero with respect to this attribute. In this case, the effect of the L th attribute level will be perfectly correlated with the constant term for the status quo alternative. This makes it impossible to independently identify the utility effect of the status quo alternative that is unrelated to the attributes characterising the non-status quo alternatives.

Effects coding implies that the effects of attribute levels are uncorrelated with a constant term for the status quo alternative. Again, $L - 1$ variables are created, which

receive a value of minus one if the L th attribute level is present, and a value of one for each of the $L - 1$ attribute levels. Importantly, the utility of the reference level is directly related to utility of the $L - 1$ attribute levels, defined as the negative sum of the $L - 1$ parameter estimates. Therefore, the estimates are likely uncorrelated with the constant for the status quo alternative. See Table 1 in Daly et al. (2016, p. 38) for an example of dummy and effects coding for a 4-level attribute.

The above differences, described by Bech and Gyrd-Hansen (2005), have quickly led to the widespread belief that effects coding would be “superior” to dummy coding and that dummy coding would imply confounding between base (reference) levels of dummy-coded attributes and dummy-coded alternative-specific constants (ASCs; typically associated with either the status quo or the non-status quo alternatives).

But does it really matter if effects coding or dummy coding is used? To start with, the concern about confounding of base (reference) levels of dummy-coded attributes and ASCs is only relevant for cases where the experimental design includes an opt-out or status quo alternative, and if any of the attributes relate exclusively to non-status quo alternatives. What characterises these cases is that none of the attribute levels of the experimental design will be shared with the status quo alternative. An example may be an attribute describing spatial location (e.g. an attribute describing where proposed changes would take place) in a WTP context. If there is no policy change in the status quo option, this attribute does actually not characterise the status quo. In supply contexts (e.g. studies aimed at estimating ecosystem service suppliers’ WTA to participate in contract schemes), attributes of conservation contracts (e.g. collaborative participation; contract length) do not apply for the status quo or opt-out alternative, which typically is a “no contract” alternative.

Furthermore, the estimation process of marginal values differs slightly. In a WTP context, effects coding requires taking the negative ratio of twice the utility parameter of interest, β_1 , plus the sum of the utility of utility parameters associated with remaining $L - 1$ levels and the cost parameter. For example, for a 2-level effects coded attribute, marginal $WTP = -2\beta_1/\beta$ and for 3-level effects coded attribute, marginal $WTP = -(2\beta_1 + \beta_2)/\beta$. Importantly, however, whether effects coding or dummy coding is used will not affect the log-likelihood value of multinomial logit models (i.e. the models are statistically equivalent), marginal WTP or WTA estimates, nor will it affect welfare estimates of CS (using the approach proposed by Hanemann 1984). This is the most important insight and is due to the fact that what matters are differences in utility between the individual levels of an attribute (or differences in utility between alternatives). Effects coding only uses a different normalisation of the reference level, while the difference in utility between levels remains the same for either form of coding.

Apart from requiring more attention when coding and estimating WTP, effects coding can further complicate the interpretation of utility effects. Daly et al. (2016) provide an example related to imposing an equality constraint on some of the attribute levels. Overall, therefore, the reasons above seem to be sufficient to discourage the use of effects coding as a “superior” model specification. We strongly recommend that researchers consult Daly et al. (2016) for a more detailed theoretical and empirical investigation of the impacts of coding before making any decisions on whether to

use effects and dummy coding. Beyond the arguments raised above, we also wish to note that the equivalence between dummy coding and effects coding disappears when random coefficients are applied to these variables (Burton 2018).

Given the above, in most circumstances effects coding offers no additional benefits over dummy coding and may lead to some undesirable complications. However, there may be a single reason for considering effects coding, which is traced back to an argument presented by Bech and Gyrd-Hansen (2005). This relates to the aim of giving the ASC parameter (associated with a status quo) a direct (behavioural) interpretation. If, for example, the aim is to interpret the ASC parameter as evidence for status quo bias, it may be tested if the coding of attributes has a strong impact on the ASC parameter estimate. However, in DCE applications that include a monetary attribute which enters the indirect utility function in a continuous fashion, the utility difference between zero (used only for status quo) and the lowest level of the monetary attribute in the non-status quo alternatives will still be captured by the ASC (see Appendix in Hess and Beharry-Borg 2012). Hence, in most applications the constant will in any case be confounded with attributes, even if effects coding has been applied, because the cost attribute is unlikely to be effects coded to enable estimation of WTP and WTA.

In summary, for multinomial choice models (DCE data) and the most typical applications, we recommend to use dummy coding; effects coding offers no advantages while making interpretation more difficult.

5.2 Functional Form of the Attributes in the Utility Function

Following random utility theory, the utility an individual derives from choosing an alternative comprises of a deterministic, observable component and a random, unobservable component (Sect. 1.2). The most common specification of the deterministic component is linear and additive in attributes. One of the first choice experiment applications in environmental economics that introduces non-linear terms into the utility specification is Adamowicz et al. (1998). They used quadratic terms for some of the continuously coded attributes, which helped identify threshold effects for these attributes: utility increases at a diminishing rate up to a threshold, beyond which utility decreases at an increasing rate. They find that the non-linear specification outperforms the linear one. This suggests that it can be important to test and assess if indeed there is non-linearity in sensitivity towards attributes.

Perhaps the simplest and most effective way of testing for non-linearity is to use dummy coding for $L - 1$ attribute levels. The resulting estimates for each level can be directly inspected or displayed graphically to investigate if utility changes linearly to (proportionally with) changes in the attribute. Figure 5.1 presents example of two different effects the attributes can have on the utility. The effect of the first attribute (Attr. 1) is approximately linear over attribute levels and the effect of the second

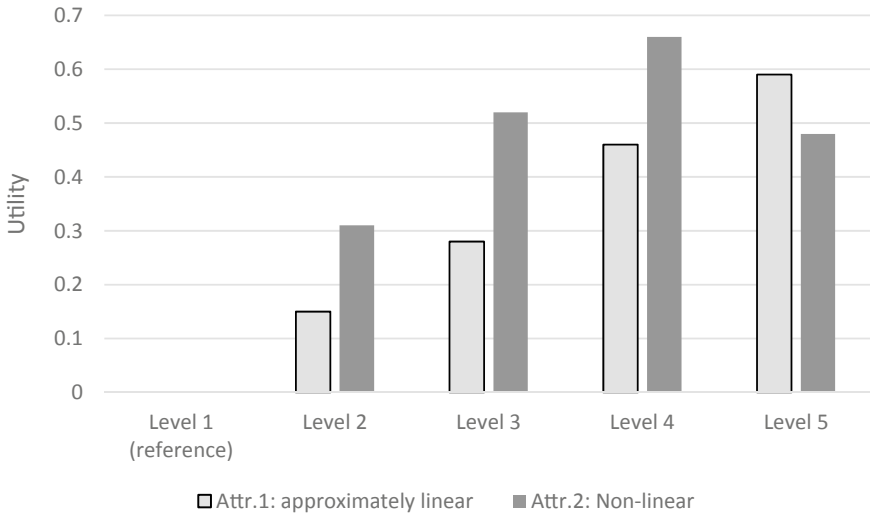


Fig. 5.1 Example of plotting utility of dummy-coded attribute levels of two 5-level attributes

attribute (Attr 2.) is non-linear. Generally, likelihood-ratio tests can be used to test if a non-linear specification outperforms a linear one. To test if utility increases proportionally over two dummy-coded attribute levels, a Wald test may be used to test if $\beta_1 - \beta_2 \left(\frac{X_2}{X_1} \right) = 0$, where X_1 and X_2 refer to quantities associated with two attribute levels and β_1 and β_2 are correspondings parameters. When investigating or plotting utility parameters of dummy or effects coded levels, care should be taken if the attribute levels are not evenly spaced so that utility differences between attribute levels relate to varying differences in quantities. If there is an indication of a non-linear utility surface across the range of attribute levels, the dummy specification can be retained, or a number of alternative options can be considered for introducing non-linearity in an attribute's utility surface.

The first is a non-linear transformation of the independent variables (attributes and/or socio-demographic variables if included). For example, this may be done through the use of polynomials (quadratic, cubic, etc.) as described above, or a logarithmic transformation. Note that the resulting specification is still linear and additive in parameters—merely the utility of an attribute will be described by a non-linear function. Non-linear transformation may also include Box-Cox, Box-Tukey, or alternative power transformations. Examples of the use of such applications include Farsi (2010), Glenk and Colombo (2013) and Tuhkanen et. al. (2016) (see also Stathopoulos and Hess 2011 for a transportation application). Non-linear transformation does not make sense for 2-level attributes, and the advantages of a non-linear transformation over simple dummy coding may be questioned for 3-level attributes. As an aside, of course any of the above non-linear transformations may also be applied to socio-demographic variables included in the model (as interactions with attributes,

as independent variables explaining class membership in latent class models or as part of structural equations in hybrid choice models).

A second approach is the use of a piecewise specification (splines). Assume a continuous attribute with five attribute levels ranging from 10 to 100 (10, 30, 60, 80, 100). The “dummy specification test” described above suggests marginal utility is not significantly different between 10 and 30, then increases but is not different between 30 and 60, then increases again but is constant between 60 and 100. In this case, the following three categories and corresponding dummy variables could be created (as opposed to five dummy variables, of which four enter the utility function for the “dummy specification test”): X_1 : $[10-30[= 1$, otherwise 0; X_2 : $[30-60[= 1$, otherwise 0; $[60-100] = 1$, otherwise 0. With one acting as a reference category, two of the above dummies can be entered into the utility function. This will produce two utility parameters corresponding to the attribute-level intervals defined above. However, this approach has the drawback that marginal utility is zero within the intervals, and the resulting function is discontinuous (there are discrete steps or “jumps”, see Fig. 5.2). One alternative is to use a piecewise linear specification. Marginal utility is allowed to vary within intervals and a continuous function is enforced (i.e. no “jumps”, see Fig. 5.2)—in the example above this may be relevant if marginal utility is found not to be zero within intervals but is, for example, increasing, but at different rates. For the example above, three variables are specified as follows: $X_1 = \min(x, 30)$; $X_2 = \max(0, \min(x - 30, 30))$; $X_3 = \max(0, \min(x - 60, 40))$, where x describes the value of the initial attribute level. A piecewise linear specification can be useful in choice experiment applications that investigate some form of

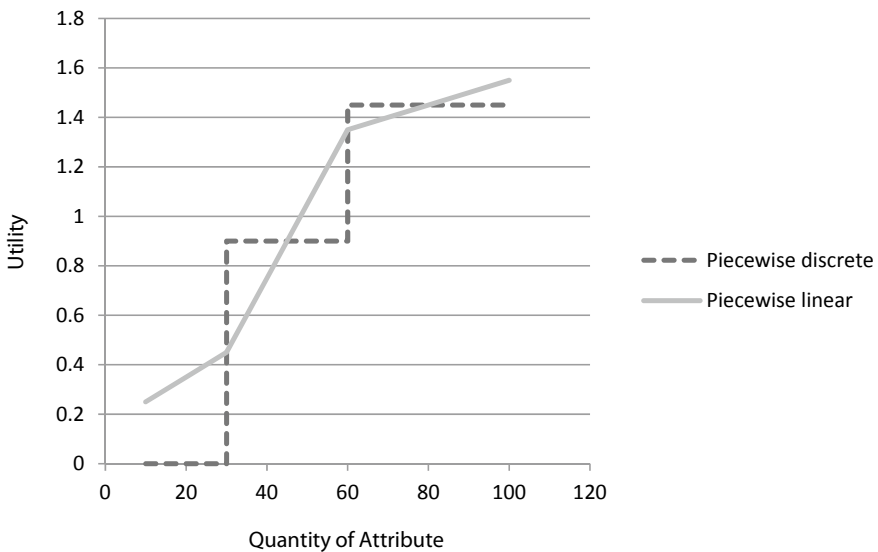


Fig. 5.2 Example of piecewise discrete and piecewise linear functional forms

reference dependency (Lanz et al. 2010; Ahtaiainen et al. 2015). Combining the polynomial approach with the piecewise approach can result in a piecewise non-linear specification (Glenk 2011).

The use of any type of non-linear specification for the utility of attributes implies that marginal WTP/WTA estimates depend on the attribute value, i.e. there is no unique estimate, but values differ for different levels of provision of an attribute. Sensitivity to cardinally scaled attributes (e.g., distance in km, area in km², frequency, percentage change, change in objects that can be counted) may not change at a constant rate. Such non-linear utility effects can sometimes be motivated by economic theory, for example, through the principle of diminishing marginal utility. See Sagebiel et al. (2017) for an example related to changes in forest cover, or Glenk et al. (2011) for an example related to water quality improvements.

It can be easily assessed if non-linearity in the utility surface is present by using dummy coding of attribute levels. If there is an indication of a non-linearity, the dummy specification can be retained, or a number of alternative options for specifying the observed part of utility for an attribute can be considered. Such non-linear transformations might be more convenient compared to a dummy specification if the attribute entails at least three, but is better if there are more than three levels. When reporting marginal WTP/WTA estimates for attributes based on non-linear specifications, researchers should clarify the corresponding level of provision that underpins the estimate. With reference to Sect. 1.1, one should be reluctant to introduce any form of non-linearity on the cost coefficient to maintain valid welfare estimates.

Finally, a practical challenge with the testing of different functional forms of attributes, relates to the design and how the attribute was specified there (see Chap. 3). If, for example, a D-efficient design is generated assuming a continuous attribute, and we want to code it using dummy levels, the results obtained may not reflect the underlying preferences. Thus, care should be taken when the functional form deviates from the design (see Sect. 3.3).

5.3 Econometric Models

5.3.1 *Multinomial (Conditional) Logit*

The econometrics literature makes a distinction between the multinomial logit and the conditional logit model. In practice, the two labels are often used interchangeably. The term multinomial refers to a situation where more than two alternatives are available to consumers (or respondents). Both the multinomial and the conditional logit model aim at explaining the observed choices. The distinction between the multinomial and the conditional model arises in the use of explanatory variables.

The multinomial logit model explains the observed choices only by means of characteristics of the individuals (e.g. gender and age). The conditional logit model, as introduced by McFadden (1974), allows choices to be explained by means of

variables describing the characteristics of the available alternatives (e.g. quality and cost). The latter are henceforth denoted as (product) attributes. It is good practice to use a combination of individual and attribute specific variables in discrete choice models. As a result, a more general model combining the multinomial and conditional logit formulations is commonly used, hence explaining the inconsistent use of terminology. To avoid confusion, this document will refer to the above general model as the MNL model. Moreover, the two models are identical from a mathematical perspective (see also Greene 2017; Long 1997).

Chapter 2 in Train (2009) makes the connection between the MNL model both as a behavioural and as an econometric model. It should be noted that the econometric flexibility offered by the MNL model (and more advanced choice models) is not always in line with the behavioural restrictions imposed by, for example, economic theory (e.g. Batley and Ibañez-Rivas 2013). Section 2.5 in Train (2009) is particularly important as it highlights a number of identification issues related to the MNL model with significant implications for the use of constants, sociodemographic variables and scale parameters in the MNL model.

5.3.2 *Mixed Logit Models—Random Parameter, Error Component and Latent Class Models*

Mixed logit models (MXL), including the random parameters MXL (RP-MXL), error components MXL (EC-MXL) and latent class models (LCM), extend the MNL model by allowing for unobserved heterogeneity in the estimated parameters. While in the MNL model, the estimated preference parameters are fixed, MXL models allow preferences to vary across choices (Brownstone and Train 1998—also known as cross-sectional modelling), individuals (Revelt and Train 1998—also known as panel modelling) or both (Hess and Rose 2009—also known as inter-and-intra respondent heterogeneity).

The MXL assumes the unobserved heterogeneity follows a continuous or discrete distribution across the population. In the early years of its application, the discrete distributions were represented by LCM and the continuous distributions were restricted usually to univariate normal densities.

However, the unrestricted domain (negative and positive values between minus and plus infinity) of the normal density goes against many behavioural predictions, such as negative cost sensitivities. We often know whether an attribute has a positive or negative influence on utility and this has resulted in the implementation of alternative distributional forms, such as the log-normal density (e.g. Train and Sonnier 2005). Also, correlations between preferences across distributions have been implemented to make the models more flexible (see again Train and Sonnier 2005). Another important extension has been the introduction of willingness-to-pay space models, as will be shown in Sect. 6.1 (Train and Weeks 2005; Daly et al. 2012a).

The extensions of the MNL and MXL, particularly beyond univariate normal densities and working in willingness-to-pay space, have significant implications on the data and estimation requirements. First, a significant number of draws need to be taken to accurately approximate the integral in the likelihood function (Train 2009, Chap. 6) and alternative routines (e.g. Halton, MLHS and Sobol) to draw from the specified densities might work better than others (Bhat 2003; Hess et al. 2006; Czajkowski and Budziński 2019). Irrespective of the number of draws, the standard classical maximum simulated likelihood method struggles with the complexity of the MXL, especially when multivariate distributions are included in the model specification. Moreover, robustness across starting values and the number of draws needs to be tested, but alternative estimation strategies, such as Bayesian analysis (Huber and Train 2001), Expectation Maximisation (see Train 2009, Chap. 14) or Maximum Approximate Composite Marginal Likelihood (Bhat and Sidharthan 2011) have been proposed.

A special case of the MXL is the EC-MXL (Scarpa et al. 2005). The error component acts as an additional error term in the utility function and is either implemented at the cross-sectional or panel level. Most applications now adopt a panel specification, that is more than one choice occasion per respondent. The benefit of the additional error term is that it allows for different correlation patterns across alternatives and thereby alleviates the Independence of Irrelevant Alternatives assumption underlying the MNL model. A common application is that of an additional error component on the status quo and none on the hypothetical alternatives in the choice set. The error component is modelled as a zero mean continuously (normally) distributed variable such that only the variance term needs to be estimated. That can easily lead to identification issues as can be the case of error components and random alternative specific constants. Important identification and normalisation strategies are discussed in Walker et al. (2007).

Discrete mixed logit (DM-MXL) and LCMs move away from the continuously distributed random parameters specification. Instead, they assume the random parameter(s) can only take a limited (discrete) number of values and each mass point is associated with a probability (Hess 2014). Hence, for each discrete random parameter K location and $(K - 1)$ probabilities need to be estimated. Here, K refers to the number of mass points and since the probabilities need to sum to one, one probability parameter cannot be estimated. In estimation, the number of combinations between mass points and the total number of parameters rapidly increases with the number of random parameters included. For this reason, the DM-MXL model is hardly applied in practice. Instead, the LCM has gained much more in popularity (Greene and Hensher 2003). Like the DM-MXL model, the LCM assumes a finite number of potential values for the random parameters, but assumes there exists correlation across some of the random parameters in the model—preference parameters are random (discrete), but assumed the same for several “classes” of preferences. In addition, a probability of each respondents’ membership in each class is estimated (and can be made a function of their socio-demographic characteristics). As a result, there is a probability of belonging to a class. More recent versions of the LCM allow

for random parameter heterogeneity within classes (e.g. Greene and Hensher 2013; Campbell et al. 2014; Karlõševa et al. 2016).

The LCM has often been misinterpreted due to its behavioural appeal. Many applications forget that class membership is only known up to a probability and each respondent has a finite probability of belonging to each class. Mean probability of belonging to each class is not the same as a share of respondents belonging to each class. Moreover, LCMs are even more subject to local maxima and convergence issues than the conventional MXL. The EM-algorithm is very effective in estimating LCMs, but even then, robustness checks using a wide range of starting values are of the utmost importance. Recent improvements in the estimation of LCMs with a large number of mass points and flexible mixing densities are reported in Train (2016).

5.3.3 *G-MXL Model*

The generalised mixed logit (G-MXL) model (Fiebig et al. 2010; Greene and Hensher 2010) received significant attention for its acclaimed ability to separate preference and scale heterogeneity. As argued by Hess and Rose (2012), this claim is incorrect, because preferences and scale are confounded and the interpretation of the estimated parameters as preferences or scale is always the arbitrary decision of the researcher. Some forms of the G-MXL model can be seen as a middle-ground between the RP-MXL model without correlated parameters and the RP-MXL model with fully correlated parameters, as G-MXL introduces a single additional parameter capturing simultaneous correlation between all parameters, rather than modelling all correlations separately, which requires many more parameters. Overall, the model is now rarely used, because of the problems with the interpretation of the estimated correlation coefficient, which cannot be definitely attributed to scale heterogeneity. However, its specific versions can still be useful, provided they are correctly interpreted (e.g. Keane and Wasi 2013; Hess and Train 2017).

5.3.4 *Hybrid Choice Models*

Besides the use of socio-economic characteristics (such as age and gender), it is often useful to explain the observed choices by means of attitudes or other not directly observable constructs (latent behavioural traits). For example, assume one wishes to explain the decision to buy an electric car by the extent to which the respondent is “environmentally friendly”. Attitudinal statements included in surveys measure the degree of environmental friendliness of respondents. However, these responses are associated with measurement error and introducing them directly in the choice model may result in endogeneity issues. Therefore, we instead try to explain the observed choices (the choice model) and responses to attitudinal statements (the measurement

model) by means of the same underlying latent variable. In addition, the structural component can be added, which explains how latent environmental friendliness varies across respondents with respect to their socio-demographic characteristics. While the hybrid choice model (HCM, sometimes called integrated choice and latent variable—ICLV model) addresses the measurement error of indicator variables, to date there has only been limited attention to the extent that the HCM addresses endogeneity as opposed to more traditional instrumental variable approaches or alternatives (Hoyos et al. 2015; Budziński and Czajkowski 2018; Mariel et al. 2018; Guevara et al. 2018).

Hybrid choice models can be estimated simultaneously or sequentially. The former is more efficient, but makes the estimation more complex and slower due to the use of additional integrals in the likelihood function. Estimation times can be significantly reduced by means of Bayesian estimation (e.g. Dekker et al. 2016). Mariel and Meyerhoff (2016) show that the use of hybrid models is justified when one is primarily interested in learning about preference heterogeneity, but not in predictive power.

Chorus and Kroessen (2014) represent the main source of criticism of the HCM when commenting on its use in transportation. Their first point of criticism is that HCMs do not support the derivation of travel demand policies that aim to change travel behaviour through changes in a latent variable, because of the non-trivial endogeneity of the latent variable regarding travel choice. Their second point is that the cross-sectional nature of the latent variable does not allow for claims concerning changes in the variable at the individual level.

Vij and Walker (2016) present a comprehensive comparison between MXL and HCM concluding that HCM can lead to an improvement in the analyst's ability to predict outcomes in the choice data, and it allows for the identification of structural relationships between variables that could not otherwise be identified by a choice model without latent variables. Moreover, HCM can help correct for bias arising from omitted variables and measurement error, it can in some cases reduce the variance of the parameter estimates, and, finally, it can quantify the impact of latent constructs on observable behaviour, preferences and WTP and demand elasticities (e.g. Boyce et al. 2019; Czajkowski et al. 2017a, b; Zawojska et al. 2019).

The latest discussion in the HCM literature points to perceptions and their use in HCM. For example, Bahamonde-Birke et al. (2017) discuss the differences between attitudes and perceptions and Borriello and Rose (2019) highlight the difference between global and localised attitudinal responses. The potential of perceptions within HCMs has thus far been largely overlooked in environmental valuation.

5.4 Coefficient Distribution in RP-MXL

Choosing the correct distribution to capture the heterogeneity in underlying population preferences has been one of the major research interests in DCE in recent years, but it is still a central problem. Extensive literature exists on this subject (Daly et al. 2012b; Dekker 2016; Fosgerau and Bierlaire 2007; Fosgerau and Mabit 2013; Scarpa

et al. 2008a; Train 2016), but the question of choosing the correct distribution still seems to be open.

Before we analyse the type of distribution to be used for our coefficients, we could test whether they should be considered random. The z-statistics of the estimated deviations of the random parameters are usually used to test this. An alternative and more sophisticated way to test the randomness of a coefficient is the Lagrange Multiplier (LM) test proposed by McFadden and Train (2000). Mariel et al. (2013), however, show that, on the one hand, the z-statistic test usually has high power achieved at the expense of a distorted empirical size and that, on the other hand, the LM test has very low power.

A model in which all parameters are assumed to be random is more flexible than a model in which some of the parameters are assumed to be non-random. Obviously, this flexibility is achieved at the expense of the loss of degrees of freedom. Nevertheless, it is better to work with a more flexible model by assuming that all parameters are random than imposing incorrect constraints on the randomness of the parameters.

If a parameter is assumed to be random, the distributions of monetary and non-monetary attributes are often set to normal. But that distribution choice can cause serious problems in the WTP calculation. The distribution of WTP for an attribute is generally derived from the distribution of the ratio of coefficients corresponding to the non-monetary and monetary attribute. Since the monetary coefficient enters the denominator, its distribution is crucial for the distribution of WTP. Daly et al. (2012b) show that some popular distributions used for the monetary coefficient in random coefficient models, including normal, truncated normal, uniform and triangular, can imply infinite moments for the distribution of WTP, even if truncated or bounded at zero. Therefore, their mechanical application to the monetary coefficient may lead to undefined moments of the distribution (e.g. mean). To avoid that problem, the analyst can interpret quantiles of the distribution (e.g. median), adopt other distributions for the monetary coefficient (e.g. log-normal) or re-parameterise the model in WTP space (Sect. 6.1). It is important to note that if a log-normal distribution is assumed, then the interpretation of the estimated parameters (β) and standard deviation (σ) is not straightforward, as the median, mean and standard deviation of the distribution are given by $\exp(\beta)$, $\exp(\beta + \sigma^2/2)$ and $\exp(\beta + \sigma^2/2) \cdot \sqrt{\exp(\sigma^2) - 1}$ respectively.

Independently of the type of distribution used, a wide spread of the preferences often means extreme preferences for a non-marginal share of respondents and in this case, a pre-analysis of the data is needed. It can indicate lexicographic, non-trading or protest responses that can require a specific treatment (Hess et al. 2010).

The analyst can also properly test the coefficients' distribution by the use of the Fosgerau and Bierlaire (2007) semi-nonparametric test for mixing distributions in discrete choice models. It tests if a random parameter follows an a priori postulated distribution. Unfortunately, it is not included in standard software packages with the exception of Biogeme (Bierlaire 2020). A less sophisticated but simpler procedure is proposed by Hensher and Greene (2003). They suggest plotting the contributions (incremental marginal utility) of all individuals to the overall sample mean parameter

estimate and hence the profile of individual preference heterogeneity. It is a procedure which is easy to perform with any software package and although it is not a proper statistical test, it can give an idea of the underlying distributions of the parameters. Alternatively, Fosgerau and Mabit (2013) propose a method to generate flexible mixture distributions. Their proposal takes draws from a distribution and transforms them using a power series. It allows for a use of a flexible and non-standard distribution without restrictive assumptions. In general, there is still limited information for practitioners regarding the performance of the above-described tests and approaches.

If the underlying theory implies a specific sign of the coefficient (e.g. negative sign for cost coefficient) a log-normal (with changed or unchanged sign) can be used. In other cases, the normal distribution can be appropriate bearing in mind that wide standard deviations indicate inappropriateness of the assumed distribution. If the focus of the analysis is on WTP, the model can be estimated in WTP space (Sect. 6.1). If there is a clear indication that the distribution of the parameters is non-standard (non-symmetric, bi-modal) some of the flexible, but technically complex, parametric (e.g. LCM) or non-parametric approaches can be applied (Train 2016).

5.5 Specifics for the Cost Attribute

The cost coefficient is one of the most important elements of the choice model. Indeed, when one treats a choice model as an econometric model, the cost coefficient is just like any other explanatory variable in the model. Similarly, when the choice model is a (non-economic) behavioural model, then prices and their corresponding coefficients simply describe how the attractiveness of an alternative varies as a function of its price. However, we often treat the choice model as an economic model and wish to use it to obtain welfare measures expressed in terms of WTP or WTA. By working in a RUM context, we implicitly assume the respondent maximises his/her utility subject to a budget constraint (Small and Rosen 1981). For the indirect utility function, which we estimate and specify in the choice model, to be consistent with economic theory the utility function needs to be linear in price when conditional demand is restricted to unity (Batley and Ibáñez Rivas 2013). If that is not the case, any estimated welfare measure is invalid. When the utility function is linear for prices, then the negative of the cost coefficient measures the marginal utility of income and can be used to translate utilities into monetary terms. In effect, this is what the WTP-space model (see Sect. 6.1) does directly (Train and Weeks 2005). For this reason, we also often use a negative log-normal density to describe any preference heterogeneity in the cost coefficient (Daly et al. 2012b).

While for simplicity of the calculation of WTP some applications assume the monetary attribute parameter to be fixed, this is discouraged, as heterogeneity with respect to the monetary variable is almost always substantial, and constraining it introduces significant bias to the model and the estimated WTP. This point is analysed further in Sect. 6.1.

5.6 Correlation Between Random Coefficients

McFadden and Train (2000) showed that any choice model, with any distribution of preferences, can be approximated to any degree of accuracy by a mixed logit, demonstrating that this is a highly flexible model. Nevertheless, Train and Weeks (2005) show that widely applied models with uncorrelated utility coefficients imply that scale is constant across all utilities and implies that corresponding WTP values are correlated in a very particular way. Similarly, a model specified in WTP space with uncorrelated parameters implies a specific pattern of correlation in utility coefficients (Carson and Czajkowski 2019).

The assumed distributions of random coefficients and their possible correlations impose specific restrictions, which should be in line with the actual respondents' behaviour. Correlations among utility coefficients can arise for many reasons and have been observed in different fields (Revelt and Train 1998; Scarpa et al. 2008b; Hess and Train 2017). The most general form of a RP-MXL allows all utility coefficients to be randomly distributed and correlated. It allows for the type of correlation that would result from behavioural sources as well as scale heterogeneity. It is not possible to empirically determine what portion is due to behavioural sources and what portion of an estimated correlation is due to scale heterogeneity. In the case of behavioural source, people who support one attribute can also be supportive of the second attribute, creating a positive correlation between the coefficients of the two attributes (Mariel and Meyerhoff 2018).

In the case of scale heterogeneity, the correlation appears to be due to the fact that a respondent's choice can be more random (with all of the coefficients being smaller) or more deterministic (with all of the coefficients being larger in magnitude). The scale of utility, that is the magnitude of all utility coefficients, differs across individuals. According to Hess and Train (2017), it is impossible to empirically distinguish various sources of heterogeneity. A RP-MXL with full covariance among coefficients includes not only correlation induced by scale heterogeneity but also any other source of correlation. They recommend estimating a RP-MXL with full covariance if the aim is to allow for all forms of correlation among utility coefficients. The estimation of such a model is more complex than estimating a model with uncorrelated coefficients, given the high number of parameters and possible non-concave areas of the log-likelihood function, but software packages which do this are widely available.

In summary, if enough data is available to warrant identification and convergence of the model, allowing for correlations of random parameters is highly recommended. This is especially important for utility functions with dummy-coded attribute levels and error component models.

5.7 Assuring Convergence

There are many papers in the literature presenting estimations of different discrete choice models without giving any details regarding the estimation procedure. This estimation procedure is usually a maximum likelihood estimator or a maximum simulated likelihood estimator. If we focus on the most applied models, the maximised likelihood function is globally concave only for a multinomial logit model, whereas the likelihood function or simulated likelihood function of MXL models or HCMs is, generally, not globally concave and may feature several local maxima. The selection of starting values is a crucial issue to avoid an inferior local maximum. Empirical studies rarely describe the choice of starting values used and there are not many studies devoted to this topic, and, unfortunately, little guidance to solve this issue exists.

The work of Hole and Yoo (2017) is an exception in this regard, as it proposes an estimation strategy based on the joint use of heuristic optimisation algorithms and the usual gradient-based algorithms to obtain the estimates from different models through a simulated maximum likelihood method. The central idea is to use heuristic algorithms to locate a starting point which is likely to be close to the global maximum, and then to use gradient-based algorithms to refine this point further. Their results based on simulation studies indicate that repeatedly finding a particular maximum from several starting points is not reliable proof that this is in fact the global maximum and their strategy generally results in higher maxima than more conventional estimation strategies.

Optimisation procedures like Newton–Raphson, BHHH, BFGS or Nelder–Mead, usually implemented in software packages, can end up easily in local maxima in relatively simple MXL models. It is important to bear in mind that convergence to the global maximum of the likelihood function or simulated likelihood function is not guaranteed by any of these optimisation algorithms. Their performance is case specific and it is difficult to give a general recommendation. Nevertheless, it is strongly recommended to always estimate more complex models a couple of times with different starting values, different optimisation procedures and, if possible, use a mixture of heuristic procedures and the usual gradient-based algorithms as indicated by Hole and Yoo (2017).

Mebane and Sekhon (2011) is another example of a similar approach that combines evolutionary search algorithms with derivative-based (Newton or quasi-Newton) methods to solve difficult optimisation problems. Their procedure finds a global maximum of functions that are not globally concave and they may even have irregularities such as saddle points or discontinuities. The generally implemented optimisation methods such as Newton–Raphson, BHHH, BFGS or Nelder–Mead that rely on derivatives of the objective function may be in these cases unable to find any optimum at all. The only drawback of heuristic search methods is that the computation time is usually much higher than that necessary for gradient-based optimisation.

As a general recommendation, the researcher can always use different sets of random starting values and compare the final value of the log-likelihood function. On the other hand, MNL estimates should be reasonable starting values for an RP-MXL model without correlations, and RP-MXL estimates without correlations should provide reasonable starting values for RP-MXL with correlations. If it is a RP-MXL model with normally distributed coefficients, the means can be set to the MNL estimates, standard deviation to a small positive constant (e.g. 0.5) and the remaining coefficients related to the covariances of the random coefficient can be set to zero. It is important to highlight that very often the LCM ends up in flat regions of its log-likelihood function. The more dummy variables are included in a model, the more the severe problem of a flat region is likely to appear. If we do not have any a priori hypothesis about the coefficients in different classes of our LCM that could function as starting values in each class, the model should be estimated repeatedly for randomly generated starting values.

There are other ways to reduce the problems with the optimisation procedures. The magnitude of a coefficient multiplying an explanatory variable (usually an attribute) in a discrete choice model is related to the scale of this variable. If the scale of the variable is thousands of units, the corresponding coefficient is expected to be in order of thousandths. The higher the differences in orders of the coefficients, the more difficult it will be for the optimiser to reach a global maximum. It is, therefore, a good practice to re-scale the attribute so that the order of all coefficients is similar (ideally between 0.1 and 1). This simple procedure will ease the convergence of the estimation procedure based on numerical optimisation. In order to recover the original units, the estimated coefficients should be multiplied or divided accordingly. For example, the values of an attribute x_1 are between 10 and 100, and its contribution to the utility is $\Phi_{1njt} = \beta_1 \cdot x_{1njt}$. If the values of this attribute are divided by 100 ($x_{1njt}^* = x_{1njt}/100$) and, therefore, values between 0.1 and 1.0 are used in the estimation process, the contribution to the utility remains the same: $\Phi_{1njt} = \beta_1 \cdot x_{1njt} = (\beta_1 \cdot 100) \cdot (x_{1njt}/100) = \beta_1^* \cdot x_{1njt}^*$. The corresponding estimated parameter $\hat{\beta}_1^* = \widehat{(\beta_1 \cdot 100)}$ should then be divided by 100 to obtain the estimation of the original β_1 .

McCullough and Vinod (2003) present a four-step process for the verification of the solution obtained. The procedure is based on the detailed analysis of the final gradient and trace, hessian and quadratic approximation of the likelihood function in the proximity of the obtained solution. Practitioners should always be cautious about solutions obtained from statistical packages, no matter how convenient and expected the outcomes are, and apply at least some of the checks described above.

5.8 Random Draws in RP-MXL

Simulated maximum likelihood is the preferred estimator of most researchers dealing with discrete choice models, as it is relatively straightforward and readily implemented in most statistical software packages. Simulating the value of the log-likelihood function is necessarily associated with the simulation error that depends on the number and type of draws used. By using a different set of draws or even changing the order of explanatory variables a researcher will arrive at somewhat different estimation results, in terms of the value of the log-likelihood function, parameter estimates and their estimated standard errors (and hence the associated *z*-statistics).

Several studies have demonstrated the advantages of using Quasi Monte Carlo (QMC) methods in terms of reducing simulation-driven variation of the results (e.g. using Halton draws, Modified Latin Hypercube Sampling or Sobol draws rather than pseudo-random draws) and this has led to their wide proliferation. Unfortunately, examples of 100 Halton draws leading to a smaller bias than 1,000 pseudo-random draws (e.g. Bhat 2001) have in fact led some to use very few draws for simulations, when in fact not much is known about the extent of the possible bias resulting from using different numbers of different types of draws in various conditions (datasets).

One problem with using the popular Halton sequence is its poor performance in higher dimensions (i.e. generating Halton sequences for a high number of coefficients), because the sequences generated using high prime numbers as bases tend to be highly correlated. To address this problem, the use of scrambling or shuffling the sequence (or other QMC methods) has been suggested. Wang and Kockelman (2006) compared scrambled and shuffled Halton sequences, concluding that although scrambling seems to perform better, the difference is relatively small.

Using more draws is always better than using fewer—not only will the estimates become more precise (lower simulation error) but this can also lead to the detection of identification problems (Chiou and Walker 2007). Using very few draws (e.g. below 100) can therefore hinder estimation of even preliminary models, which are used by researchers to guide their subsequent analyses, i.e. the choice of the final preferred and published model. It is important to keep in mind that the estimation by maximisation of a simulated likelihood function is based on an approximation of an integral. A low number of draws lead to a poor approximation of this integral. That can lead to a situation in which the log-likelihood of the estimated model with, e.g. 100 draws is higher than the log-likelihood with 500 draws. It does not mean that we should use 100 draws, it just reflects a poor approximation of the integral. For a low number of draws, the log-likelihood of the estimated model can differ markedly, but the increase of draws usually leads to a stabilisation of the log-likelihood. We can even find models that converge with a low number of draws but do not converge with a high number of draws. This indicates an identification problem of the model that must be solved and that we cannot trust the results with a low number of draws. In the specification search stage of our research, we usually estimate different variations of our model and to lower the estimation cost of numerous preliminary models we

set a low number of draws. This is also an incorrect approach as a low number of draws can lead to an incorrect specification decision. All the models we estimate in our specification search must be estimated with a sufficiently high number of draws.

Czajkowski and Budziński (2019) provide a systematic comparison of pseudo-random, Halton, Sobol and Modified Latin Hypercube Sampling draws under a wide set of experimental conditions in terms of experimental designs, the number of individuals (400–1,200) and the number of choice tasks per individual (4–12). Based on a Monte Carlo simulation, they demonstrate the extent of the simulation error resulting from using 100 up to 1,000,000 draws. They show that a scrambled Sobol sequence results in the lowest simulation error of all the simulation methods compared, irrespective of the experimental conditions, with Halton draws a close second.

Czajkowski and Budziński (2019) propose guidelines regarding how many draws are “enough” for the required precision level. Their measure is based on 95% confidence that log-likelihoods do not lead to simulation-driven erroneous inference and that parameter estimates are within 5% of their true values for all experimental settings considered. They find that as the number of observations increases, so do the absolute levels of log-likelihood, and the minimum number of draws for the required precision level. Conversely, in the case of parameter estimates, the reverse relationship is observed—increasing the number of observations reduces the number of draws required for a given precision level.

Overall, Czajkowski and Budziński (2019) show that satisfying these precision criteria depends on the number of observations and may require using over 2,000 *Sobol* draws in the case of 5-attribute designs and over 25,000 *Sobol* draws in the case of 10-attribute designs. It is also recommended to verify if the results are stable with respect to an increase in the number of draws.

References

- Adamowicz W, Boxall P, Williams M, Louviere J (1998) Stated preference approaches for measuring passive use values: choice experiments and contingent valuation. *Am J Agr Econ* 80:64–75. <https://doi.org/10.2307/3180269>
- Ahtiainen H, Pouta E, Artell J (2015) Modelling asymmetric preferences for water quality in choice experiments with individual-specific status quo alternatives. *Water Resour Econ* 12:1–13. <https://doi.org/10.1016/j.wre.2015.10.003>
- Bahamonde-Birke FJ, Kunert U, Link H, Ortúzar J de D (2017) About attitudes and perceptions: finding the proper way to consider latent variables in discrete choice models. *Transportation* 44:475–493. <https://doi.org/10.1007/s11116-015-9663-5>
- Batley R, Ibáñez Rivas JN (2013) Applied welfare economics with discrete choice models: implications of theory for empirical specification. In: Hess S, Daly A (eds) *Choice Modelling*. Edward Elgar Publishing, pp 144–171
- Bech M, Gyrd-Hansen D (2005) Effects coding in discrete choice experiments. *Health Econ* 14:1079–1083. <https://doi.org/10.1002/hec.984>

- Bhat CR (2003) Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences. *Transp Res Part B Methodol* 37:837–855. [https://doi.org/10.1016/S0191-2615\(02\)00090-5](https://doi.org/10.1016/S0191-2615(02)00090-5)
- Bhat CR (2001) Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transp Res Part B Methodol* 35:677–693. [https://doi.org/10.1016/S0191-2615\(00\)00014-X](https://doi.org/10.1016/S0191-2615(00)00014-X)
- Bhat CR, Sidharthan R (2011) A simulation evaluation of the maximum approximate composite marginal likelihood (MACML) estimator for mixed multinomial probit models. *Transp Res Part B Methodol* 45:940–953. <https://doi.org/10.1016/j.trb.2011.04.006>
- Bierlaire M (2020) Biogeme. <https://biogeme.epfl.ch/>. Accessed 21 May 2020
- Borriello A, Rose JM (2019) Global versus localised attitudinal responses in discrete choice. *Transportation*. <https://doi.org/10.1007/s11116-019-10045-3>
- Boyce C, Czajkowski M, Hanley N (2019) Personality and economic choices. *J Environ Econ Manag* 94:82–100. <https://doi.org/10.1016/j.jeem.2018.12.004>
- Brownstone D, Train K (1998) Forecasting new product penetration with flexible substitution patterns. *J Econometrics* 89:109–129
- Budziński W, Czajkowski M (2018) Hybrid choice models vs. endogeneity of indicator variables: a Monte Carlo investigation. Faculty of Economic Sciences, University of Warsaw
- Burton M (2018) Model invariance when estimating random parameters with categorical variables. Working Paper 1804, Agricultural and Resource Economics, The University of Western Australia, Crawley, Australia.
- Campbell D, Hensher DA, Scarpa R (2014) Bounding WTP distributions to reflect the “actual” consideration set. *J Choice Model* 11:4–15. <https://doi.org/10.1016/j.jocm.2014.02.004>
- Carson RT, Czajkowski M (2019) A new baseline model for estimating willingness to pay from discrete choice models. *J Environ Econ Manage* 95:57–61. <https://doi.org/10.1016/j.jeem.2019.03.003>
- Chiou L, Walker JL (2007) Masking identification of discrete choice models under simulation methods. *J Econometrics* 141:683–703. <https://doi.org/10.1016/j.jeconom.2006.10.012>
- Chorus CG, Kroesen M (2014) On the (im-)possibility of deriving transport policy implications from hybrid choice models. *Transp Policy* 36:217–222. <https://doi.org/10.1016/j.tranpol.2014.09.001>
- Czajkowski M, Budziński W (2019) Simulation error in maximum likelihood estimation of discrete choice models. *J Choice Model* 31:73–85. <https://doi.org/10.1016/j.jocm.2019.04.003>
- Czajkowski M, Hanley N, Nyborg K (2017) Social norms, morals and self-interest as determinants of pro-environment behaviours: the case of household recycling. *Environ Resource Econ* 66:647–670. <https://doi.org/10.1007/s10640-015-9964-3>
- Czajkowski M, Vossler CA, Budziński W et al (2017) Addressing empirical challenges related to the incentive compatibility of stated preferences methods. *J Econ Behav Organ* 142:47–63. <https://doi.org/10.1016/j.jebo.2017.07.023>
- Daly A, Dekker T, Hess S (2016) Dummy coding vs effects coding for categorical variables: clarifications and extensions. *J Choice Model*. <https://doi.org/10.1016/j.jocm.2016.09.005>
- Daly A, Hess S, de Jong G (2012) Calculating errors for measures derived from choice modelling estimates. *Transp Res Part B Methodol* 46:333–341. <https://doi.org/10.1016/j.trb.2011.10.008>
- Daly A, Hess S, Train K (2012) Assuring finite moments for willingness to pay in random coefficient models. *Transportation* 39:19–31. <https://doi.org/10.1007/s11116-011-9331-3>
- Dekker T (2016) Asymmetric triangular mixing densities for mixed logit models. *J Choice Model* 21:48–55. <https://doi.org/10.1016/j.jocm.2016.09.006>
- Dekker T, Hess S, Brouwer R, Hofkes M (2016) Decision uncertainty in multi-attribute stated preference studies. *Resource Energy Econ* 43:57–73. <https://doi.org/10.1016/j.reseneeco.2015.11.002>
- Farsi M (2010) Risk aversion and willingness to pay for energy efficient systems in rental apartments. *Energy Policy* 38:3078–3088. <https://doi.org/10.1016/j.enpol.2010.01.048>

- Fiebig DG, Keane MP, Louviere J, Wasi N (2010) The generalized multinomial logit model: accounting for scale and coefficient heterogeneity. *Market Sci* 29:393–421. <https://doi.org/10.1287/mksc.1090.0508>
- Fosgerau M, Bierlaire M (2007) A practical test for the choice of mixing distribution in discrete choice models. *Transp Res B Methodol* 41:784–794. <https://doi.org/10.1016/j.trb.2007.01.002>
- Fosgerau M, Mabit SL (2013) Easy and flexible mixture distributions. *Econ Lett* 120:206–210. <https://doi.org/10.1016/j.econlet.2013.03.050>
- Glenk K (2011) Using local knowledge to model asymmetric preference formation in willingness to pay for environmental services. *J Environ Manage* 92:531–541. <https://doi.org/10.1016/j.jenvman.2010.09.003>
- Glenk K, Colombo S (2013) Modelling outcome-related risk in choice experiments. *Aust J Agric Resource Econ* 57:559–578. <https://doi.org/10.1111/1467-8489.12012>
- Glenk K, Lago M, Moran D (2011) Public preferences for water quality improvements: implications for the implementation of the EC Water Framework Directive in Scotland. *Water Policy* 13:645–662. <https://doi.org/10.2166/wp.2011.060>
- Greene WH (2017) *Econometric analysis*, 8th edn. Pearson, New York, NY
- Greene WH, Hensher DA (2003) A latent class model for discrete choice analysis: contrasts with mixed logit. *Transp Res Part B Methodol* 37:681–698. [https://doi.org/10.1016/S0191-2615\(02\)00046-2](https://doi.org/10.1016/S0191-2615(02)00046-2)
- Greene WH, Hensher DA (2013) Revealing additional dimensions of preference heterogeneity in a latent class mixed multinomial logit model. *Appl Econ* 45:1897–1902. <https://doi.org/10.1080/00036846.2011.650325>
- Greene WH, Hensher DA (2010) Does scale heterogeneity across individuals matter? An empirical assessment of alternative logit models. *Transportation* 37:413–428. <https://doi.org/10.1007/s1116-010-9259-z>
- Guevara CA, Tirachini A, Hurtubia R, Dekker T (2018) Correcting for endogeneity due to omitted crowding in public transport choice using the Multiple Indicator Solution (MIS) method. *Transp Res Part A: Policy Pract.* <https://doi.org/10.1016/j.tra.2018.10.030>
- Hanemann WM (1984) Discrete/continuous models of consumer demand. *Econometrica* 52:541–561
- Hensher DA, Greene WH (2003) The mixed logit model: the state of practice. *Transportation* 30:133–176. <https://doi.org/10.1023/A:1022558715350>
- Hess S (2014) Latent class structures: taste heterogeneity and beyond. In: Hess S, Daly A (eds) *Handbook of choice modelling*. Edward Elgar Publishing, Cheltenham, UK, pp 311–329
- 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. <https://doi.org/10.1007/s10640-011-9522-6>
- Hess S, Rose JM (2009) Allowing for intra-respondent variations in coefficients estimated on repeated choice data. *Transp Res B Methodol* 43:708–719. <https://doi.org/10.1016/j.trb.2009.01.007>
- Hess S, Rose JM (2012) Can scale and coefficient heterogeneity be separated in random coefficients models? *Transportation* 39:1225–1239. <https://doi.org/10.1007/s11116-012-9394-9>
- Hess S, Rose JM, Polak J (2010) Non-trading, lexicographic and inconsistent behaviour in stated choice data. *Transp Res D Transp Environ* 15:405–417. <https://doi.org/10.1016/j.trd.2010.04.008>
- Hess S, Train K (2017) Correlation and scale in mixed logit models. *J Choice Model* 23:1–8. <https://doi.org/10.1016/j.jocm.2017.03.001>
- Hess S, Train KE, Polak JW (2006) On the use of a Modified Latin Hypercube Sampling (MLHS) method in the estimation of a Mixed Logit Model for vehicle choice. *Transp Res B Methodol* 40:147–163. <https://doi.org/10.1016/j.trb.2004.10.005>
- Hole AR, Yoo HI (2017) The use of heuristic optimization algorithms to facilitate maximum simulated likelihood estimation of random parameter logit models. *J Roy Stat Soc: Ser C (Appl Stat)* 66:997–1013. <https://doi.org/10.1111/rssc.12209>

- 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. <https://doi.org/10.1016/j.scitotenv.2014.10.066>
- Huber J, Train K (2001) On the similarity of classical and Bayesian estimates of individual mean partworths. *Market Lett* 12:259–269. <https://doi.org/10.1023/A:1011120928698>
- Karlöševa A, Nömmann S, Nömmann T et al (2016) Marine trade-offs: comparing the benefits of off-shore wind farms and marine protected areas. *Energy Econ* 55:127–134. <https://doi.org/10.1016/j.eneco.2015.12.022>
- Keane M, Wasi N (2013) Comparing alternative models of heterogeneity in consumer choice behavior. *J Appl Econometrics* 28:1018–1045. <https://doi.org/10.1002/jae.2304>
- Lanz B, Provins A, Bateman IJ et al (2010) Investigating willingness to pay–willingness to accept asymmetry in Choice experiments. In: Hess S, Daly A (eds) *Choice modelling: the state-of-the-art and the state-of-practice*, pp 517–541
- Long JS (1997) *Regression models for categorical and limited dependent variables*, 1st edn. Sage, Thousand Oaks
- Mariel P, De Ayala A, Hoyos D, Abdullah S (2013) Selecting random parameters in discrete choice experiment for environmental valuation: a simulation experiment. *J Choice Model* 7:44–57. <https://doi.org/10.1016/j.jocm.2013.04.008>
- Mariel P, Hoyos D, Artabe A, Guevara CA (2018) A multiple indicator solution approach to endogeneity in discrete-choice models for environmental valuation. *Sci Total Environ* 633:967–980. <https://doi.org/10.1016/j.scitotenv.2018.03.254>
- Mariel P, Meyerhoff J (2016) Hybrid discrete choice models: gained insights versus increasing effort. *Sci Total Environ* 568:433–443. <https://doi.org/10.1016/j.scitotenv.2016.06.019>
- Mariel P, Meyerhoff J (2018) A more flexible model or simply more effort? On the use of correlated random parameters in applied choice studies. *Ecol Econ* 154:419–429. <https://doi.org/10.1016/j.ecolecon.2018.08.020>
- McCullough BD, Vinod HD (2003) Verifying the solution from a nonlinear solver: a case study. *Am Econ Rev* 93:873–892. <https://doi.org/10.1257/000282803322157133>
- McFadden D (1974) Conditional logit analysis of qualitative choice behaviour. In: Zarembka P (ed) *Academic Press, New York*, pp 105–142
- McFadden D, Train K (2000) Mixed MNL models for discrete response. *J Appl Econometrics* 15:447–470. [https://doi.org/10.1002/1099-1255\(200009/10\)15:5%3c447::AID-JAE570%3e3.0.CO;2-1](https://doi.org/10.1002/1099-1255(200009/10)15:5%3c447::AID-JAE570%3e3.0.CO;2-1)
- Mebane WR, Sekhon JS (2011) Genetic Optimization Using Derivatives: the rgenoud Package for R. *J Stat Softw* 42:1–26. <https://doi.org/10.18637/jss.v042.i11>
- Revelt D, Train K (1998) Mixed Logit with repeated choices: households' choices of appliance efficiency level. *Rev Econ Stat* 80:647–657. <https://doi.org/10.1162/003465398557735>
- Sagebiel J, Glenk K, Meyerhoff J (2017) Spatially explicit demand for afforestation. *Forest Policy Econ* 78:190–199. <https://doi.org/10.1016/j.forpol.2017.01.021>
- Scarpa R, Ferrini S, Willis K (2005) Performance of error component models for status-quo effects in choice experiments. In: Scarpa R, Alberini A (eds) *Applications of simulation methods in environmental and resource economics*. Springer, Netherlands, Dordrecht, pp 247–273
- Scarpa R, Thiene M, Marangon F (2008a) Using flexible taste distributions to value collective reputation for environmentally friendly production methods. *Can J Agric Econ* 56:145–162. <https://doi.org/10.1111/j.1744-7976.2008.00122.x>
- Scarpa R, Thiene M, Train K (2008b) Utility in willingness to pay space: a tool to address confounding random scale effects in destination choice to the Alps. *Am J Agr Econ* 90:994–1010. <https://doi.org/10.1111/j.1467-8276.2008.01155.x>
- Small K, Rosen H (1981) applied welfare economics with discrete choice models. *Econometrica* 49:105–130
- Stathopoulos A, Hess S (2011) Referencing, gains-losses asymmetry and non-linear sensitivities in commuter decisions: one size does not fit all! Working Papers 0511. CREI Università degli Studi Roma Tre

- Train K (2016) Mixed logit with a flexible mixing distribution. *J Choice Model* 19:40–53. <https://doi.org/10.1016/j.jocm.2016.07.004>
- Train K (2009) *Discrete choice methods with simulation*, 2nd edn. Cambridge University Press, New York
- Train K, Sonnier G (2005) Mixed logit with bounded distributions of correlated partworths. In: Scarpa R, Alberini A (eds) Springer. The Netherlands, Dordrecht, pp 1–16
- Train K, Weeks M (2005) Discrete choice models in preference space and willingness-to-pay space. In: Scarpa R, Alberini A (eds) Springer. The Netherlands, Dordrecht, pp 1–16
- Tuhkanen H, Piirsalu E, Nömmann T et al (2016) Valuing the benefits of improved marine environmental quality under multiple stressors. *Sci Total Environ* 551–552:367–375. <https://doi.org/10.1016/j.scitotenv.2016.02.011>
- Vij A, Walker JL (2016) How, when and why integrated choice and latent variable models are latently useful. *Transp Res B Methodol* 90:192–217. <https://doi.org/10.1016/j.trb.2016.04.021>
- Walker JL, Ben-Akiva M, Bolduc D (2007) Identification of parameters in normal error component logit-mixture (NECLM) models. *J Appl Econ* 22:1095–1125. <https://doi.org/10.1002/jae.971>
- Wang X, Kockelman KM (2006) Tracking Land cover change in mixed logit model: recognizing temporal and spatial effects. *Transp Res Rec* 1977:112–120. <https://doi.org/10.1177/0361198106197700114>
- 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. <https://doi.org/10.1016/j.jeem.2018.11.007>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), 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 license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license 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.

