

# Consumer Choice Modeling: The Promises and the Cautions

Chandra Bhat

The University of Texas at Austin, Austin, TX, USA

and

The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

(Material for this chapter has been liberally drawn from earlier papers of the author)

The travel behavior field is entering into a new data landscape in which voluminous and diverse amounts of data are becoming available to analysts through a whole host of equipment that act as sensors — legacy roadway systems, smart phones and GPS systems, and smart cars themselves. The key issue is how to deal with such voluminous and diverse amounts of incoming data per unit of time, and translate them into usable information for near-real time operations purposes or for longer-term planning purposes. In addition, predictive analytics to translate data into information requires the ability to deal with data that may be from multiple sources, highly noisy, heterogeneous, and high-dimensional with complex interdependencies. On the last of these, the joint modeling of data with mixed (heterogeneous) types of dependent variables (including ordered-response or ordinal variables, unordered-response or nominal variables, count variables, and continuous variables) is a challenging problem. But, at the same time, new inference methods (and variants of these inference methods) have been proposed to address the increasing complexity of the models, so that it becomes feasible from a computational standpoint to estimate such models.

Figure 1 provides a diagrammatic representation of the temporal evolution of the types of dependent variables considered in the field (on the x-axis) and estimation methods (on the y-axis). This is discussed further below. In the ensuing discussion, we will focus solely on parametric models. We recognize that there have been some developments in non-parametric models too, but expanding to include such non-parametric models would be too wide a scope for a single paper.

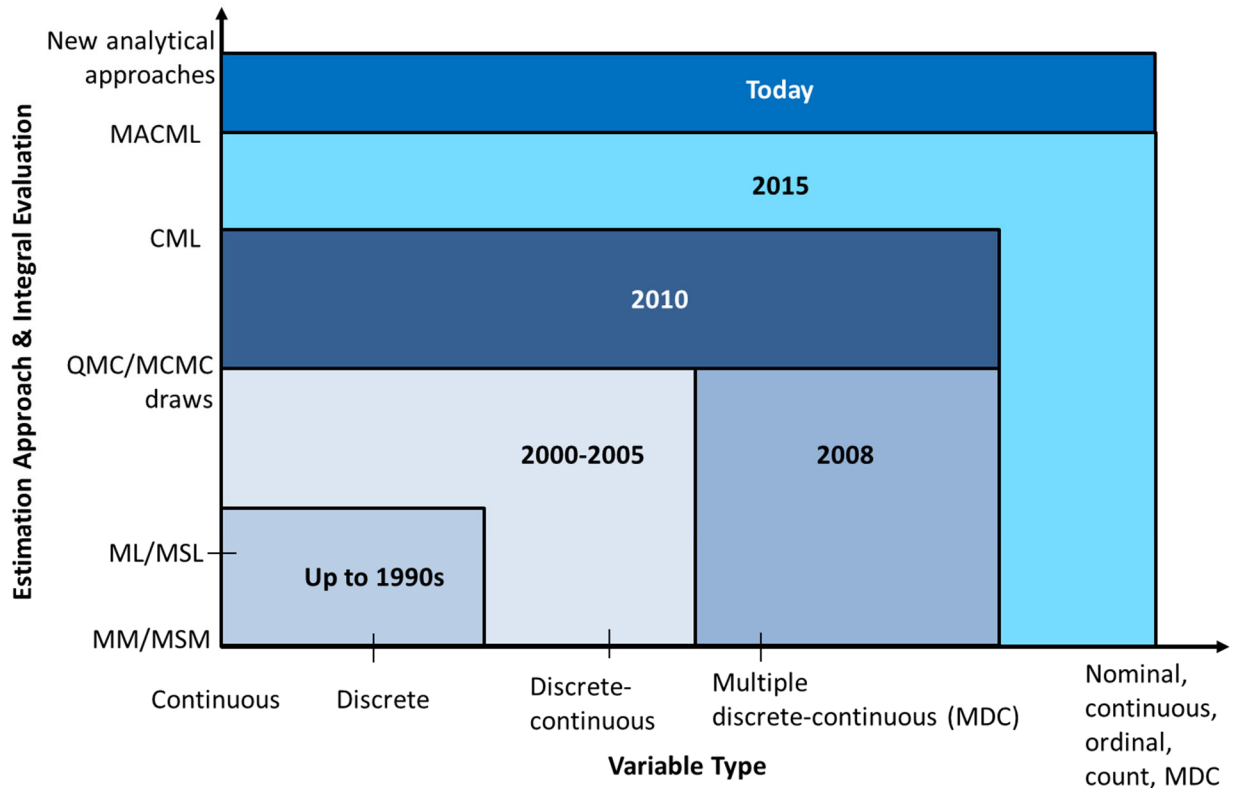


Figure 1. Evolution of dependent variable types and methods

### Period up to (and including) the 1990s

Much of the focus of travel behavior analysis (and applied econometrics in general) during this period was on continuous or single discrete variables. Among the single discrete variable models, the multinomial logit (Luce and Suppes, 1965; McFadden, 1974), the nested logit model (Williams, 1977; McFadden, 1978; Daly and Zachary, 1978), and multivariate extreme value models (McFadden, 1978; Ben-Akiva and Francois, 1983; Small, 1987; Chu, 1989; Vovsha, 1997; Bhat, 1998a; Koppelman and Wen, 2000) were the ones most applied in practice, primarily because of their closed-form probability expressions. During this period, on the inference method front, the predominant ones involved the method of moments or MOM (for continuous variables) or maximum likelihood or ML (for discrete variables). The MOM inference approach is based on the recognition that, in the population, the exogenous factors of the behavior being modeled and the model residuals are uncorrelated, and is implemented by making the residuals and exogenous “instrument” variables be uncorrelated in the sample. This MOM estimator is equivalent to the well-known least squares estimator for linear regression models when the exogenous “instrument” variables are the explanatory variables (see Train, 2009; page 240). The ML inference approach, used for discrete choice modeling, is based on estimating parameters by maximizing the joint probability (across all sample observations) of the observed individual choices (as a function of the parameters to be estimated), given the explanatory variables. While the ML technique, in the case of discrete choice models, is a special case of the MOM when the exogenous “instrument” variables in the MOM are replaced by the scores (the score refers to the derivative of the individual

log(probability) with respect to parameters), it is easier conceptualized as maximizing the joint probability of the observed sample.

The multinomial probit (MNP) model, though introduced early in the 1970s, did not see much use until the 1990s and well beyond, despite simulation-based developments by McFadden (1989) and Pakes and Pollard (1989). This perhaps was mainly because the MNP likelihood function involves a truncated multivariate integral (*i.e.*, the cumulative multivariate normal (MVN) function) that is generally not easy to evaluate using simulation methods. On the other hand, in the late 1990s, the mixed multinomial logit (MMNL) model that allows random coefficients, but retains an extreme value kernel error distribution, surged in popularity, primarily because of the ease of its conceptualization as well as ease in estimation through simulation draws over an untruncated multivariate distribution (see Revelt and Train, 1998 and Bhat, 1998b). The MMNL model is typically estimated using the maximum simulated likelihood (MSL) method that replaces the logarithm of the probability of each observation in the objective function (that is, the log-likelihood function) to be maximized by the logarithm of a simulated probability for that observation. Because the logarithm of the simulated probability is not unbiased for the logarithm of the true probability, even if the simulated probability is unbiased for the true probability, the MSL inference approach is technically a biased estimator (with the bias diminishing with increasing number of simulation draws). However, MSL is consistent if the number of draws rises with sample size (under standard regularity conditions), and becomes asymptotically normal and efficient if the number of draws rises faster than the square root of sample size. The problem, though, is that, while these are nice theoretical results, they provide little insight into the number of draws actually needed in a practical setting to make MSL consistent and asymptotically normal (CAN) and efficient. The Method of Simulated Moments (MSM) estimator, on the other hand, when applied to the MMNL (or any analytically intractable econometric model) has the appealing property that it is CAN because the probability expression enters linearly in the equation used to solve for the value of the parameters. However, typically, it is not as efficient as MSL.

### **The Period from 2000-2005**

This period saw a rapid growth in the application of the MMNL model for discrete choice, as well as expansion of the domain of the dependent variable to include a single discrete and its corresponding continuous quantity of consumption. But perhaps the single most important development in this period was the introduction of a new quasi-Monte Carlo (QMC) way of making simulation draws for MSL estimation of econometric models. The QMC approach, similar to the Monte Carlo approach, evaluates a multidimensional integral by replacing it with an average of values of the integrand computed at discrete points. However, rather than using pseudo-random sequences for the discrete points, the QMC approach uses “cleverly” crafted non-random and more uniformly distributed QMC sequences within the domain of integration. The underlying idea is that it is really inconsequential whether the discrete points are truly random; of primary importance is the even distribution (or maximal spread) of the points in the integration space.

Bhat (2000; 2001) proposed and introduced a simulation approach using QMC sequences for estimating discrete choice models with analytically intractable likelihood functions. In his approach, Bhat generates a multidimensional QMC sequence of length  $N*Q$ , then uses the first  $N$  points to compute the contribution of the first observation to the criterion function, the second  $N$

points to compute the contribution of the second observation, and so on. Each set of  $N$  points covers the domain of integration very well, resulting in more accurate probability evaluations for each sample observation. At the same time, each set of  $N$  points for each observation fills in the gaps left by the sets of  $N$  points used for previous observations. Consequently, the cancellation of simulation errors across observations is stronger when using QMC sequences than when using the ordinary pseudo Monte-Carlo (PMC) sequence. For example, Hensher (2001) found that the data fit and parameter values of the mixed logit model in his study remained about the same beyond 50 Halton draws and concludes that the QMC approach is “a phenomenal development in the estimation of complex choice models”. Many other studies during this period investigated the value of QMC draws, and examined different QMC variants (sequences) to use, including Bhat (2003) who suggests a scrambled QMC approach in high dimensions to reduce the correlation along high dimensions of a standard QMC sequence and shows that the scrambling improves the performance of the standard QMC sequence. A listing of the many applied studies of the MMNL model using the QMC approach during the 2000-2005 period is available in Bhat *et al.* (2008).

### **Period from 2005-2010**

This period saw the entry of multiple discrete-continuous choice models in the travel behavior literature. A detailed review of these models is provided in Bhat and Pinjari (2014), so we will only provide a brief overview of these kinds of models here.

Many choice situations are characterized by the choice of multiple alternatives at the same time (that is, multiple discreteness), as opposed to the choice of a single alternative. In addition, in such situations, the consumer usually also decides on a continuous dimension (or quantity) of consumption, which has prompted the label “multiple discrete-continuous” (MDC) choice (see Bhat, 2005; 2008). Specifically, an outcome is said to be of the MDC type if it exists in multiple states that can be jointly consumed to different continuous amounts.

The basic approach in a utility maximization framework for multiple discreteness hinges upon the use of a non-linear (but increasing and continuously differentiable) utility structure with decreasing marginal utility (or satiation). The origins of utility-maximizing MDC models may be traced back to the research of Wales and Woodland (1983) (see also Kim *et al.*, 2002; von Haefen and Phaneuf, 2003; Bhat, 2005). Bhat (2008) proposed a Box-Cox utility function form that is quite general and subsumes earlier utility specifications as special cases, and that is consistent with the notion of weak complementarity (see Mäler, 1974), which implies that the consumer receives no utility from a non-essential good’s attributes if she/he does not consume it. Then, using a multiplicative log-extreme value error term in the baseline preference for each alternative, Bhat (2005; 2008) proposed and formulated the multiple discrete-continuous extreme value (MDCEV) model, which has a closed-form probability expression and collapses to the MNL in the case that each (and every) decision-maker chooses only one alternative.

The MDCEV model immediately was applied in a wide variety of travel behavior-related topics, including for time-use analysis, household vehicle fleet composition and use, household good expenditures, and vacation location choice (see Bhat and Eluru, 2010). The model continues to see substantial use in a variety of fields. Some recent examples include Yonezawa and Richards (2017) in the managerial economics field, Shin *et al.* (2015) in the technological and social change field,

and Wafa *et al.* (2015) in the regional science field. Of course, just as in the case of the traditional single choice models, advanced variants of the MDCEV such as the MDCGEV and random-coefficients MDCEV have also been introduced and applied (see, for example, Calastri *et al.*, 2017; Bernardo *et al.*, 2015; Pinjari, 2011) In addition, some studies have considered the replacement of the log-extreme value error term in the baseline preference with a log-normal error term, along with random-coefficients versions of the resulting MDC probit (MDCP) model (Bhat *et al.*, 2016a; Khan and Machemehl, 2017). Several other extensions of the MDC model have also been undertaken and there are yet many outstanding issues to be resolved (see Bhat and Pinjari, 2014; Bhat, 2018a).

From a model inference standpoint, the composite marginal likelihood (CML) approach found its way into travel behavior research during this period. The CML inference approach is a relatively simple approach that can be used when the full likelihood function is near impossible or plain infeasible to evaluate due to the underlying complex dependencies. For instance, Varin and Czado (2010) examined the headache pain intensity of patients over several consecutive days. In this study, a full information likelihood estimator would have entailed as many as 815 dimensions of integration to obtain individual-specific likelihood contributions, an infeasible proposition using the computer-intensive simulation techniques. As importantly, the accuracy of simulation techniques is known to degrade rapidly at medium-to-high dimensions, and the simulation noise increases substantially. This leads to convergence problems during estimation. In contrast, the CML method, which belongs to the more general class of composite likelihood function approaches (see Lindsay, 1988), is based on forming a surrogate likelihood function that compounds much easier-to-compute, lower-dimensional, marginal likelihoods. The CML approach can be applied using simple optimization software for likelihood estimation. It also represents a conceptually and pedagogically simpler simulation-free procedure relative to simulation techniques, and has the advantage of reproducibility of the results. Finally, as indicated by Varin and Vidoni (2009), it is possible that the “maximum CML estimator can be consistent when the ordinary full likelihood estimator is not”. This is because the CML procedures are typically more robust and can represent the underlying low-dimensional process of interest more accurately than the low dimensional process implied by an assumed (and imperfect) high-dimensional multivariate model.

The simplest CML approach is a pairwise likelihood function formed by the product of likelihood contributions of all or a selected subset of couplets (*i.e.*, pairs of variables or pairs of observations). Almost all earlier research efforts employing the CML technique have used the pairwise approach. Alternatively, the analyst can also consider larger subsets of observations, such as triplets or quadruplets or even higher dimensional subsets (see Caragea and Smith, 2007). In general, the issue of whether to use pairwise likelihoods or higher-dimensional likelihoods remains an open, and under-researched, area of research. However, it is generally agreed that the pairwise approach is a good balance between statistical and computation efficiency.

The properties of the CML estimator may be derived using the theory of estimating equations (see Cox and Reid, 2004; Bhat, 2014). Specifically, under usual regularity assumptions, the CML estimator is consistent and asymptotically normal distributed (this is because of the unbiasedness of the CML score function, which is a linear combination of proper score functions associated with the marginal event probabilities forming the composite likelihood). Of course, the maximum CML

estimator loses some asymptotic efficiency from a theoretical perspective relative to a full likelihood estimator (Zhao and Joe, 2005; Bhat, 2014). On the other hand, there is also a loss in asymptotic efficiency in the maximum simulated likelihood (MSL) estimator relative to a full likelihood estimator (see McFadden and Train, 2000).

Early applications of the CML approach in the transportation/econometric field (and specifically the pairwise likelihood approach) include Ferdous *et al.* (2010) in the context of cross-sectional multivariate ordered-response probit (CMOP) systems, and Bhat *et al.* (2010) to estimate their multivariate ordered-response probit system in the context of a spatially dependent ordered response outcome variable. Since then, the CML inference approach has been used widely to estimate multivariate econometric model systems and spatially-dependent model systems (see Bhat, 2014 for a detailed discussion). The conceptual and implementation simplicity of the CML approach makes it a promising and simple approach for analytically-intractable econometric models.

### **Period from 2010-2015**

This period saw the arrival of joint models of data with mixed types of dependent variables (including ordered-response or ordinal variables, unordered-response or nominal variables, count variables, continuous variables, and multiple discrete-continuous variables). Such mixed dependent variable models are of interest in several fields, including biology, developmental toxicology, finance, economics, epidemiology, social science, and transportation. In the transportation field, households that are not auto-oriented are likely to locate in transit- and pedestrian-friendly neighborhoods that are characterized by mixed and high-density land use; pedestrian-oriented design in such communities may also further structurally reduce motorized vehicle miles of travel. If that is the case, then it is likely that the choices of residential location (nominal variable), vehicle ownership (count), and vehicle miles of travel (continuous) are being made jointly as a bundle (see, for example, Bhat *et al.*, 2014).

The interest in mixed model systems has been spurred particularly by the recent availability of high-dimensional heterogeneous data with complex dependence structures, thanks to technology that allows the collection and archival of voluminous amounts of data (“big data”). Unlike standard correlated linear data that can be analyzed using traditional multivariate linear regression models, the presence of non-commensurate outcomes creates difficulty because of the absence of a convenient multivariate distribution to jointly (and directly) represent the relationship between discrete and continuous outcomes. Of course, one can choose to simply ignore the dependence and estimate separate models. However, such an approach is inefficient in estimating covariate effects for each outcome because it fails to borrow information on other outcomes, and is limiting in its ability to answer intrinsically multivariate questions such as the effect of a covariate on a multidimensional outcome (Teixeira-Pinto and Harezlak, 2013). But, more importantly, if some endogenous outcomes are used to explain other endogenous outcomes (such as examining the effect of density of residence on auto-ownership model), and if the outcomes are not modeled jointly to recognize the presence of unobserved exogenous variable effects, the result is inconsistent estimation of the effects of one endogenous outcome on another.

In the economics and transportation fields, the outcomes of mixed models are tied based on their latent variable representations (for non-continuous outcome variables) or observed continuous variable representations. The models typically assume a multivariate normal (MVN) distribution over the entire set of latent variables and observed continuous variables characterizing the many types of outcomes. Models with nominal variables are based on latent variable representations of the nominal variables (these latent variable representations take the form of the utilities of each alternative for each nominal variable; for example, see Paleti *et al.*, 2013 and Bhat *et al.*, 2014). A variant of this methodology uses a Gaussian copula function to tie the latent and observed continuous variables if the variables have different marginal distributions, though this approach has been confined to scenarios without a nominal variable (see, for example, Wu *et al.*, 2013). Another variant introduces random error terms linearly in the latent and observed continuous variable equations associated with the discrete outcomes and continuous outcomes, respectively. The underlying continuous variables are considered to be independent, conditional on these random error terms. Then, if these random error terms are common or correlated, the result is an association structure among the mixed outcomes. Such a specification falls under the label of a multivariate generalized linear latent and mixed model (GLLAMM), and is particularly helpful when considering clustering effects (due to multiple observations from the same person or due to spatial dependency) in addition to correlation across mixed outcomes (see, for example, Faes *et al.*, 2009 and Bhat *et al.*, 2014).

Another approach, originating from the social sciences, implicitly generates dependence among mixed outcomes by writing the latent and observed continuous variables as a function of unobserved psychological constructs. These relationships are characterized as measurement equations, in that the psychological constructs are manifested in the larger combination of mixed outcomes. The constructs themselves are related to exogenous variables and may be correlated with one another in a structural relationship. In this approach, the unobserved psychological constructs serve as latent factors that provide a structure to the dependence among the many mixed indicator variables. Seen from this perspective, the approach can also be viewed as a parsimonious attempt to explain the covariance relationship among a large set of mixed outcomes through a much smaller number of unobservable latent factors. Sometimes referred to as *factor analysis*, the approach represents a powerful dimension-reduction technique to analyze high-dimensional heterogeneous outcome data by representing the covariance relationship among the data through a smaller number of unobservable latent factors. An entire field of structural equations modeling (SEM) has been developed around this psychological construct-based dependence modeling, originating in some of the early works of Jöreskog (1977). However, the SEM field has focused almost exclusively on non-nominal variable analysis (see Gates *et al.*, 2011 and Hoshino and Bentler, 2013). Indeed, traditional SEM software (such as LISREL, MPLUS, and EQS) is either not capable of handling nominal variables or at least are not readily suited to handle nominal variables (see Temme *et al.*, 2008). But when this approach is extended to include a nominal variable, it essentially takes the form of an integrated choice and latent variable (ICLV) model (Ben-Akiva *et al.*, 2002; Bolduc *et al.*, 2005). But, while traditional SEM techniques typically adopt normally distributed latent factors along with normally distributed measurement error terms (leading to probit models in the presence of binary/ordered outcomes), ICLV models tend to use normally distributed latent factors mixed with logistically distributed errors in the measurement equations for ordinal variables and type-1 extreme value errors in the nominal outcome utility functions (leading to a probability expression that involves a multivariate integral over the product

of logit-type probabilities for the outcomes). In both the SEM and ICLV cases, the standard estimation methodology is the method of maximum likelihood estimation. When there are many binary/ordered-response outcomes (indicators) and/or a nominal variable, the integrals in the overall probability expression are computed using simulation techniques. As indicated by Hoshino and Bentler (2013), this can “be difficult to impossible when the model is complex or the number of variables is large.” This is particularly the case with the mixture formulation of ICLV models in general, and particularly when there are several latent factors (see Daziano and Bolduc, 2013).

Bhat and Dubey (2014) proposed a different way of formulating ICLV models, in which they use a SEM-like probit approach while also accommodating a single nominal variable. Essentially, this approach combines the power and parsimony of the dimension-reduction factor analysis structure of SEMs (as just discussed above) with an approach that uses a symmetric, latent continuous variable representation for all non-continuous outcomes (as in Paleti *et al.*, 2013 and Bhat *et al.*, 2014). Bhat (2015a) and Bhat *et al.* (2016b) later generalized this approach to the case of multiple nominal outcomes, multiple ordinal variables, multiple count variables, multiple continuous variables, and a multiple discrete-continuous variable. The resulting model, which is labeled simply as the *generalized heterogeneous data model* (GHDM), is general enough to accommodate other models in the literature as special cases. Bhat and colleagues propose a new maximum approximate CML (or MACML) inference approach for their probit-based ICLV type model as well as for their more general GHDM model, as discussed further below. Straightforward extensions of the GHDM model are available to accommodate longitudinal and spatial clustering.

The reason for the development of the MACML approach is that the presence of a nominal outcome variable or a multiple discrete-continuous variable in mixed variable modeling still can create a computational challenge, even after one applies the traditional CML technique. For instance, in the presence of even just a single nominal variable with  $I$  choice alternatives per individual (assume for ease in presentation that all individuals have all  $I$  choice alternatives), and an additional ordered-response variable, the CML approach involves compounding the likelihood of the joint probability of the two observed outcomes (variables) for each individual. However, this joint probability itself entails the integration of a multivariate normal cumulative distribution (MVNCD) function of dimension equal to  $(I - 1) + 1 = I$ . The evaluation of such an integration function cannot be pursued using quadrature techniques due to the curse of dimensionality when the dimension of integration exceeds two (see Bhat, 2003). In this case, the MVNCD function evaluation for each agent has to be evaluated using simulation or other analytic approximation techniques. Typically, the MVNCD function is approximated using simulation techniques through the use of the Geweke-Hajivassiliou-Keane (GHK) simulator, which is perhaps the most effective simulator available for evaluating the MVNCD function. In addition, Bayesian simulation using Markov Chain Monte Carlo (MCMC) techniques (instead of MSL techniques) have been used in the literature (see Albert and Chib, 1993; McCulloch and Rossi, 2000; and Train, 2009). However, all these MSL and Bayesian techniques require extensive simulation, are time-consuming, are not very straightforward to implement, and create convergence assessment problems as the number of dimensions of integration increases. Besides, they do not possess the simulation-free appeal of the CML function in the first place.

To accommodate the situation when the CML function itself may involve the evaluation of MVNCD functions, Bhat (2011) proposed a combination of an *analytic approximation* method to



evaluate the MVNCD function with the CML function, and labeled this as the MACML approach. While several analytic approximations have been reported in the literature for MVNCD functions, the one Bhat proposed for his MACML approach is based on decomposition into a product of conditional probabilities. A recent paper (Batram and Bauer, 2019) discusses the properties of the MACML estimator. Similar to the CML approach that decomposes a large multidimensional problem into lower level dimensional components, the analytic approximation method also decomposes the MVNCD function to involve only the evaluation of lower dimensional univariate and bivariate normal cumulative distribution functions. Thus, there is a type of conceptual consistency in Bhat's proposal of combining the CML method with the MVNCD analytic approximation. The net result is that the approximation approach is fast and lends itself nicely to combination with the CML approach. Further, unlike Monte-Carlo simulation approaches, even two to three decimal places of accuracy in the analytic approximation is generally adequate to accurately and precisely recover the parameters and their covariance matrix estimates because of the smooth nature of the first and second derivatives of the approximated analytic log-likelihood function. Indeed, Patil *et al.* (2017) have recently shown in the case of a single nominal variable, and based on three metrics (accuracy, precision of parameter recovery and estimation time), that the MACML approach provided, by far, the best performance for the MNP-based data generation settings examined in the study (relative to traditional MSL-based approaches).

## **2015-Present**

In many ways, the field seems to have come full circle. While the multinomial probit (MNP) model (Daganzo, 1979) has been known for a long time as a basis for a single nominal variable, it was literally spurned in favor of the MMNL model in the late 1990s and early 2000s due to the simulation ease of the latter compared to the former. However, new estimation techniques, including the MACML approach, make the estimation of the MNP much easier than for MMNL models, creating a resurgence in interest in MNP-based models. Besides, the MNP model can indeed be more parsimonious (computationally) than the MMNL in many situations, such as when the number of random coefficients is much more than the number of alternatives (and when the random coefficients are normally distributed). This is because the MNP likelihood function can be expressed as an integral whose dimensionality does not depend on the number of random coefficients in the specification. Also, in some contexts, the MVN distributional assumption of the MNP may carry better appeal than the extreme value (or multivariate extreme value) distribution used in logit-based models. For example, in social or spatial interaction models, it is much easier to specify parsimonious correlation structures using the MNP kernel than the logit kernel, primarily because of the conjugate nature of the multivariate normal distribution under affine transformations. This is reflected in the almost exclusive use of the MNP kernel for discrete choice models with spatial/social dependence (see a review in Bhat, 2015b). Similarly, it is much easier to work with a probit kernel for a nominal variable in situations with multi-dimensional joint mixed variables, again because of the ability to use the CML method to break down the multidimensionality, given that multivariate marginals of the multivariate normal distribution are themselves multivariate normally distributed.

Similar to the resurgence of the popularity of MNP-based models for nominal variables, there is also a resurgence in interest in analytic approximations to compute the MVNCD function (of course, the resurgence in MNP use and in the interest in analytic approximations for the MVNCD

function is not without coincidence). Among the analytic approximation techniques, one of the first approaches was that proposed by Clark (1961). Unfortunately, this approximation does not perform well for MVNCD evaluations when the random variables are highly correlated or have different variances. In another analytic approximation study, Mendell and Elston (1974) (ME) use the same univariate conditioning approach that formed the basis later for the GHK, except they replace draws from the truncated normal at each conditioning step with approximations of the first two moments of the truncated variables at earlier conditioning steps. This method has also been used by many other authors since, including Rice *et al.* (1979), Kamakura (1989), and Hutmacher and French (2011). Yet another MVNCD analytic approximation was first proposed by Solow (1990) based on Switzer (1977), and then refined by Joe (1995). This procedure entails the decomposition of the multivariate integral into a product of conditional probabilities, and was used by Bhat (2011) in developing his MACML inference approach. However, the early analytic approximations for the MVNCD evaluations were spurned in favor of simulation-based approaches, only to see a move back today toward the consideration of analytic approximations.

In a recent paper, Bhat (2018b) develops new analytic ways to evaluate the multivariate normal cumulative distribution (MVNCD) function. Unlike traditional simulation-based methods to MVNCD evaluation for econometric models, which can be saddled with convergence and computational cost problems, these analytic approximation techniques for MVNCD evaluation are known to provide likelihood surfaces (and the derivatives and Hessians of these surfaces with respect to model parameters) that are more smooth, reducing convergence and covariance matrix computation problems that can occur routinely in the maximum likelihood estimation of consumer choice models with analytically intractable likelihood functions (see Bhat and Sidharthan, 2011). Bhat proposes a streamlined and matrix-based version of the ME method, which relies on a single-sided truncation of a multivariate normal distribution in which some variables are truncated while others are not. Further, his new matrix-based implementation for the ME algorithm allows us to write, in a streamlined manner, the analytic matrix-based gradients of the approximated MVNCD function with respect to the abscissae and correlation parameters, an issue that is important in model estimation. In addition, he proposed four new methods for approximating the MVNCD function, based on recognizing that, when untruncated variables are normally distributed, the marginal distribution of one of the untruncated variates given that other variables are truncated (or screened) is skew-normally distributed and not normally distributed. His results suggest that, for parameter estimation, it is better (at least in the context of analytic approximations) to evaluate individual choice probabilities (and, thus, individual log-likelihoods) accurately rather than relying on any ordering noise cancellation across observations.

## Looking Forward

The field of consumer choice modeling has substantially matured over the course of the past few years, especially in the context of heterogeneous data modeling. This has been made possible through the introduction of new inference approaches as well as new analytic approximations of the MVNCD function. Of course, one still needs to recognize that the first order of business in any model should be on developing the best systematic specification possible (that is, introduce a rich set of explanatory variables). For policy analysis and forecasting, it is critical that the appropriate explanatory variables be present in the model. The recognition of jointness among dependent variables (or covariations and heteroscedasticity in the utilities of alternatives in nominal variable

or MDC variable modeling) is simply to recognize the inevitable presence of factors that we will never observe and/or capture in our data collections and that impact the dependent variables. This also brings up the issue of conceptual and theoretical thinking driving model specifications. Doing so is important for two reasons. The first is that, in most consumer choice modeling, we want *Explainability* -- the attribute of being able to explain why specific factors impact specific outcomes. This goes well beyond predictability. For example, a purely “black-box” model may be able to predict things well in the very short term or in response to small perturbations in the system. But it does not explain why a specific combination of circumstances leads to a specific outcome, or what the factors are that most impact the outcome at hand. On the other hand, a more “explainable” model may not be able to predict things as well in the very short term, but may be able to explain the outcome process much better than a “black-box” model. A more “explainable” model is likely to do better in long-term prediction or in response to large perturbations in a system. The essential question is how much do we want causality and explainability in our models and how much do we want a pure prediction accuracy model. For most consumer choice modeling, the emphasis is likely to be more on policy analysis in response to relatively large perturbations or for forecasting over a long planning horizon, and so explainability must never be dismissed or forgotten in our excitement of our new found prowess to estimate joint models accommodating error covariances. The second reason for keeping conceptual and theoretical thinking at the foundation of modeling building is more pragmatic. This is indeed the basis of the GHDM model, where latent psychological variables are used to engender jointness among multiple endogenous variables. Doing so immediately allows for a parsimonious covariance matrix in the joint model. Of course, in these kinds of models, investigating the literature and determining what kinds of latent constructs are appropriate and how these latent constructs may impact the endogenous variables of interest is critical. On the other hand, allowing for a general covariance matrix is likely to result in too many parameters to be estimated, and can lead to convergence and estimation problems.

Another important issue to keep in mind is that of econometric identification. As we get into the realm of increasingly heterogeneous data modeling, it is important that we understand what parameters are identifiable and what are not. In some instances, even if parameters may be theoretically identifiable, they may not be empirically identifiable. A classic case of the latter is in MDC models, where, as Bhat (2008) explains, two parameters which result in satiation behavior and operate in different ways are theoretically identifiable, but provide utility curves that are essentially indifferntiable. This results in a situation of empirical unidentifiability of the two separate parameters. More broadly, the necessary identification conditions get tricky as we enter into more complex heterogeneous data modeling, and the analyst typically resorts to specifications that meet sufficiency conditions. In any case, there is a need to “dig in” with investigative curiosity when estimating heterogeneous consumer data models, because the **identifiability is in the details**.

As a final word, we are in an exciting time in consumer choice modeling with endless possibilities, even if there are still new challenges to address. But our new found ability to estimate even multi-dimensional models with relative ease should not lull us into being complacent in our responsibility to have conceptual and theoretical thinking drive our models.

## Acknowledgements

This research was partially supported by the U.S. Department of Transportation through the Data-Supported Transportation Operations and Planning (D-STOP) Tier 1 University Transportation Center. The author is grateful to Aarti Bhat and Lisa Macias for their help in formatting this document.

## References

- Albert, J.H. and S. Chib (1993) Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88(422), 669-679.
- Batram, M. and D. Bauer (2019) On consistency of the MACML approach to discrete choice modelling. *Journal of Choice Modelling*, 30, 1-16.
- Ben-Akiva, M. and B. Francois (1983) Mu-homogenous generalized extreme value model. Working Paper, Department of Civil Engineering, MIT.
- Ben-Akiva, M., D. McFadden, K. Train, J. Walker, C. Bhat, M. Bierlaire, M., D. Bolduc, A. Boersch-Supan, D. Brownstone, D.S. Bunch, A. Daly, A. de Palma, D. Gopinath, A. Karlstrom, and M.A. Munizaga (2002) Hybrid choice models: Progress and challenges. *Marketing Letters*, 13(3), 163-175.
- Bernardo, C., R. Paleti, M. Hoklas, and C.R. Bhat (2015) An empirical investigation into the time-use and activity patterns of dual-earner couples with and without young children. *Transportation Research Part A*, 76, 71-91.
- Bhat, C.R. (1998a) An analysis of travel mode and departure time choice for urban shopping trips. *Transportation Research Part B*, 32(6), 361-371.
- Bhat, C.R. (1998b) Accommodating variations in responsiveness to level-of-service variables in travel mode choice modeling. *Transportation Research Part A*, 32(7), 495-507.
- Bhat, C.R. (2000) A multi-level cross-classified model for discrete response variables. *Transportation Research Part B*, 34(7), 567-582.
- Bhat, C.R. (2001) Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research Part B*, 35(7), 677-693.
- Bhat, C.R. (2003) Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences. *Transportation Research Part B*, 37(9), 837-855.
- Bhat, C.R. (2005) A multiple discrete-continuous extreme value model: Formulation and application to discretionary time-use decisions. *Transportation Research Part B*, 39(8), 679-707.
- Bhat, C.R. (2008) The multiple discrete-continuous extreme value (MDCEV) model: Role of utility function parameters, identification considerations, and model extensions. *Transportation Research Part B*, 42(3), 274-303.
- Bhat, C.R. (2011) The maximum approximate composite marginal likelihood (MACML) estimation of multinomial probit-based unordered response choice models. *Transportation Research Part B*, 45(7), 923-939.
- Bhat, C.R. (2014) The composite marginal likelihood (CML) inference approach with applications to discrete and mixed dependent variable models. *Foundations and Trends in Econometrics*, 7(1), 1-117.
- Bhat, C.R. (2015a) A new generalized heterogeneous data model (GHDM) to jointly model mixed types of dependent variables. *Transportation Research Part B*, 79, 50-77.

- Bhat, C.R. (2015b) A new spatial (social) interaction discrete choice model accommodating for unobserved effects due to endogenous network formation. *Transportation*, 42(5), 879-914.
- Bhat, C.R. (2018a) A new flexible multiple discrete-continuous extreme value (MDCEV) choice model. *Transportation Research Part B*, 110, 261-279.
- Bhat, C.R. (2018b) New matrix-based methods for the analytic evaluation of the multivariate cumulative normal distribution function. *Transportation Research Part B*, 109, 238-256.
- Bhat, C.R., and S.K. Dubey (2014) A new estimation approach to integrate latent psychological constructs in choice modeling. *Transportation Research Part B*, 67, 68-85.
- Bhat, C.R., and N. Eluru (2010) The multiple discrete-continuous extreme value (MDCEV) model: Formulation and applications. In: S. Hess and A. Daly (Eds.), *Choice Modelling: The State-of-the-Art and the State-of-Practice, Proceedings from the inaugural International Choice Modelling Conference*, Ch. 4, pp. 71-100, Emerald Group Publishing Limited.
- Bhat, C.R., and A. Pinjari (2014) Multiple discrete-continuous choice models: A reflective analysis and a prospective view. In: S. Hess and A. Daly (Eds.), *Handbook of Choice Modelling*, Ch. 19, pp. 427-453, Edward Elgar Publishing Ltd.
- Bhat, C. R., and R. Sidharthan (2011) A simulation evaluation of the maximum approximate composite marginal likelihood (MACML) estimator for mixed multinomial probit models. *Transportation Research Part B*, 45(7), 940-953.
- Bhat, C.R., N. Eluru and R.B. Copperman (2008) Flexible model structures for discrete choice analysis. In: D.A. Hensher and K. J. Button (Eds.), *Handbook of Transport Modelling, 2nd edition*, Ch. 5, pp. 75-104, Elsevier Science.
- Bhat, C.R., I.N. Sener and N. Eluru (2010) A flexible spatially dependent discrete choice model: Formulation and application to teenagers' weekday recreational activity participation. *Transportation Research Part B*, 44(8-9), 903-921.
- Bhat, C.R., S. Astroza, R. Sidharthan, M. Jobair Bin Alam, and W.H. Khushefati (2014) A joint count-continuous model of travel behavior with selection based on a multinomial probit residential density choice model. *Transportation Research Part B*, 68, 31-51.
- Bhat, C.R., S. Astroza, and A.C. Bhat (2016a) On allowing a general form for unobserved heterogeneity in the multiple discrete-continuous probit model: Formulation and application to tourism travel. *Transportation Research Part B*, 86, 223-249.
- Bhat, C.R., S. Astroza, A.C. Bhat, and K. Nagel (2016b) Incorporating a multiple discrete-continuous outcome in the generalized heterogeneous data model: Application to residential self-selection effects analysis in an activity time-use behavior model. *Transportation Research Part B*, 91, 52-76.
- Bolduc, D., M. Ben-Akiva, J. Walker, A. Michaud (2005) Hybrid choice models with logit kernel: Applicability to large scale models. In: M. Lee-Gosselin and S. Doherty (Eds.), *Integrated Land-Use and Transportation Models: Behavioral Foundations*, pp. 275-302, Elsevier, Oxford.
- Calastri, C., S. Hess, A. Daly, and J.A. Carrasco (2017) Does the social context help with understanding and predicting the choice of activity type and duration? An application of the Multiple Discrete-Continuous Nested Extreme Value model to activity diary data. *Transportation Research Part A*, 104, 1-20.
- Caragea, P.C. and R.L. Smith (2007) Asymptotic properties of computationally efficient alternative estimators for a class of multivariate normal models. *Journal of Multivariate Analysis*, 98(7), 1417-1440.

- Chu, C. (1989) A paired combinational logit model for travel demand analysis. In: *Transport Policy, Management and Technology Towards 2001: Selected Proceedings of the Fifth World Conference on Transport Research*, 4, pp. 295-309, Western Periodicals, Ventura, CA.
- Clark, C. (1961) The greatest of a finite set of random variables. *Operations Research*, 9(2), 145-162.
- Cox, D.R. and N. Reid (2004) A note on pseudolikelihood constructed from marginal densities. *Biometrika*, 91(3), 729-737.
- Daganzo, C. (1979) *Multinomial Probit: The Theory and its Application to Demand Forecasting*. Academic Press, New York.
- Daly, A. and S. Zachary (1978) Improved multiple choice models. In D.A. Hensher and M.Q. Dalvi (Eds.), *Determinants of Travel Choice*, pp. 335-357, Saxon House, Westmead.
- Daziano, R.A., and D. Bolduc (2013) Incorporating pro-environmental preferences towards green automobile technologies through a Bayesian hybrid choice model. *Transportmetrica A: Transport Science*, 9(1), 74-106.
- Faes, C., H. Geys and P. Catalano (2009) Joint models for continuous and discrete longitudinal data. In: G. Fitzmaurice, M. Davidian, G. Verbeke, G. Molenberghs (Eds.), *Longitudinal Data Analysis*, pp. 327-348, Chapman and Hall/CRC, Taylor & Francis Group, Boca Raton, FL.
- Ferdous, N., N. Eluru, C.R. Bhat and I. Meloni (2010) A multivariate ordered-response model system for adults' weekday activity episode generation by activity purpose and social context. *Transportation Research Part B*, 44(8-9), 922-943.
- Gates, K.M., P. Molenaar, F.G. Hillary and S. Slobounov (2011) Extended unified SEM approach for modeling event-related fMRI data. *NeuroImage*, 54(2), 1151-1158.
- Hensher, D.A. (2001) Measurement of the valuation of travel time savings. *Journal of Transport Economics and Policy*, 35(1) 71-98.
- Hoshino, T. and P.M. Bentler (2013) Bias in factor score regression and a simple solution. In: A.R. De Leon and K.C. Chough (Eds.), *Analysis of Mixed Data: Methods & Applications*, pp. 43-61, CRC Press, Taylor & Francis Group, Boca Raton, FL.
- Hutmacher, M.M. and J.L. French (2011) Extending the latent variable model for extra correlated longitudinal dichotomous responses. *Journal of Pharmacokinetics and Pharmacodynamics*, 38(6), 833-859.
- Joe, H. (1995) Approximations to multivariate normal rectangle probabilities based on conditional expectations. *Journal of the American Statistical Association*, 90(431), 957-964.
- Jöreskog, K.G. (1977) Factor analysis by least squares and maximum likelihood methods. In: K. Enslein, A. Ralston, and H.S. Wilf (Eds.), *Statistical Methods for Digital Computers*, John Wiley & Sons, New York.
- Kamakura, W.A. (1989) The estimation of multinomial probit models: A new calibration algorithm. *Transportation Science*, 23(4), 253-265.
- Khan, M. and R. Machemehl (2017) Commercial vehicles time of day choice behavior in urban areas. *Transportation Research Part A*, 102, 68-83.
- Kim, J., G.M. Allenby, and P.E. Rossi (2002) Modeling consumer demand for variety. *Marketing Science*, 21, 229-250.
- Koppelman, F. and C. Wen (2000) The paired combinatorial logit model: Properties, estimation and application. *Transportation Research Part B*, 34(2) 75-89.
- Lindsay, B.G., 1988. Composite likelihood methods. *Contemporary Mathematics*, 80, 221-239.
- Luce, R. and P. Suppes (1965) Preference, utility and subjective probability. In: R. Luce, R. Bush, and E. Galanter (Eds.), *Handbook of Mathematical Probability*, Vol. 3, Wiley, New York.

- Mäler, K-G. (1974) *Environmental Economics: A theoretical inquiry*. The Johns Hopkins University Press for Resources for the Future, Baltimore, MD.
- McCulloch, R.E. and P.E. Rossi (2000) Bayesian analysis of the multinomial probit model. In: R. Mariano, T. Schuermann, and M.J. Weeks (Eds.), *Simulation-Based Inference in Econometrics*, pp. 158-178, Cambridge University Press, New York.
- McFadden, D. (1974) Conditional logit analysis of qualitative choice behavior. In: P. Zarembka (Ed.), *Frontiers in Econometrics*, pp. 105-142, Academic Press, New York.
- McFadden, D. (1978) Modelling the choice of residential location. In: A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull (Eds.), *Spatial Interaction Theory and Planning Models*, pp. 75-96, North-Holland, Amsterdam.
- McFadden, D. (1989) A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica*, 57(5), 995-1026.
- McFadden, D. and K. Train (2000) Mixed MNL models of discrete response. *Journal of Applied Econometrics*, 15(5), 447-470.
- Mendell, N.R. and R.C. Elston (1974) Multifactorial qualitative traits: Genetic analysis and prediction of recurrence risks. *Biometrics*, 30(1), 41-57.
- Pakes, A and D. Pollard (1989) Simulation and the asymptotics of optimization estimators. *Econometrica*, 57(5), 1027-1057.
- Paleti, R., C.R. Bhat and R.M. Pendyala (2013) Integrated model of residential location, work location, vehicle ownership, and commute tour characteristics. *Transportation Research Record: Journal of the Transportation Research Board*, 2382, 162-172.
- Patil, P.N., S.K. Dubey, A.R. Pinjari, E. Cherchi, R. Daziano and C.R. Bhat (2017) Simulation evaluation of emerging estimation techniques for multinomial probit models. *Journal of Choice Modelling*, 23, 9-20.
- Pinjari, A.R. (2011) Generalized extreme value (GEV)-based error structures for multiple discrete-continuous choice models. *Transportation Research Part B*, 45(3), 474-489.
- Revelt, D. and K. Train (1998) Mixed logit with repeated choices: households' choices of appliance efficiency level. *Review of Economics and Statistics*, 80(4) 647-657.
- Rice, J., T. Reich, C.R. Cloninger and R. Wette (1979) An approximation to the multivariate normal integral: Its application to multifactorial qualitative traits. *Biometrics*, 35(2), 451-459.
- Shin, J., J. Ahn, and D. Lee (2015) The impact of diversification with and without commodity bundling on corporate performance: An empirical analysis in Korea telecommunication markets. *Technological Forecasting and Social Change*, 101, 314-319.
- Small, K. (1987) A discrete choice model for ordered alternatives. *Econometrica*, 55(2), 409-424.
- Solow, A.R. (1990) A method for approximating multivariate normal orthant probabilities. *Journal of Statistical Computation and Simulation*, 37(3-4), 225-229.
- Switzer, P. (1977) Estimation of spatial distributions from point sources with application to air pollution measurement. *Bulletin of the International Statistical Institute*, 47, 123-137.
- Teixeira-Pinto, A., and J. Harezlak (2013) Factorization and latent variable models for joint analysis of binary and continuous outcomes. In: A.R. De Leon and K.C. Chough (Eds.), *Analysis of Mixed Data: Methods & Applications*, pp. 81-91, CRC Press, Taylor & Francis Group, Boca Raton, FL.
- Temme, D., M. Paulssen and T. Dannewald (2008) Incorporating latent variables into discrete choice models-A simultaneous estimation approach using SEM software. *Business Research*, 1(2), 220-237.

- Train, K.E. (2009) *Discrete Choice Methods with Simulation*, 2nd ed. Cambridge University Press, Cambridge.
- Varin, C. and C. Czado (2010) A mixed autoregressive probit model for ordinal longitudinal data. *Biostatistics*, 11(1), 127-138.
- Varin, C. and P. Vidoni (2009) Pairwise likelihood inference for general state space models. *Econometric Reviews*, 28(1-3), 170-185.
- von Haefen, R.H. and D.J. Phaneuf (2003) Estimating preferences for outdoor recreation: A comparison of continuous and count data demand system frameworks. *Journal of Environmental Economics and Management*, 45(3), 612-630.
- Vovsha, P. (1997) Cross-nested logit model: An application to mode choice in the Tel-Aviv metropolitan area. *Transportation Research Record: Journal of the Transportation Research Board*, 1607, 6-15.
- Wafa, Z., C.R. Bhat, R.M. Pendyala, and V.M. Garikapati (2015) Latent-segmentation-based approach to investigating spatial transferability of activity-travel models. *Transportation Research Record: Journal of the Transportation Research Board*, 2493, 136-144.
- Wales, T.J. and A.D. Woodland (1983) Estimation of consumer demand systems with binding non-negativity constraints. *Journal of Econometrics*, 21(3), 263-85.
- Williams, H. (1977) On the formation of travel demand models and economic evaluation measures of user benefits. *Environment and Planning A*, 9, 285-344.
- Wu, B., A. de Leon and N. Withanage (2013) Joint analysis of mixed discrete and continuous outcomes via copulas. In: A.R. De Leon and K.C. Chough (Eds.), *Analysis of Mixed Data: Methods & Applications*, pp. 139-156, CRC Press, Taylor & Francis Group, Boca Raton, FL.
- Yonezawa, K. and T.J. Richards (2017) Consumer risk-reduction behavior and new product purchases. *Managerial and Decision Economics*, 38(7), 1003-1016.
- Zhao, Y. and J.H. Joe (2005) Composite likelihood estimation in multivariate data analysis. *The Canadian Journal of Statistics*, 33(3), 335-356.