J. Indian Inst. Sci.
A Multidisciplinary Reviews Journal
ISSN: 0970-4140 Coden-JIISAD

Check for updates

REVIEW ARTICLE

Discrete Choice Models with Alternate Kernel Error Distributions

Rajesh Paleti*

Abstract | The multinomial logit (MNL) and probit (MNP) models dominated the literature on consumer behavior analysis, particularly in the transportation planning context where the focus is on future travel demand prediction as an aggregated outcome of individual traveler choices. While Gumbel kernel errors in the MNL model are unbounded and positively skewed, normal kernel errors in the MNP model are symmetric and unbounded. However, choice models with alternative kernel errors (beyond Gumbel and normal distributions) have piqued the interest of choice modelers for behavioral and prediction accuracy reasons. In addition, researchers found evidence in support of these alternate kernel errors in a wide variety of empirical contexts. This paper compiles a synthesis of the past literature that developed choices models with flexible kernel errors, including both parametric and semi-parametric methods and concludes with possible avenues for further research.

Keywords: *Choice models, Utility functions, Logit, Probit, Kernel errors, Travel choices*

1 Introduction

Discrete choice models served as the modeling paradigm for analyzing scenarios where decision-makers choose one unique alternative from a set of competing options. While researchers explored several alternate decision rules, the random utility maximization (RUM) decision rule that assumes a rational decision-maker choosing the option with highest utility has dominated the choice modeling literature given its ability to approximate a wide gamut of choice rules[26]. Assuming additive separability, the utility $U_{d,i}$ associated with alternative to decision-maker d is written as: $U_{d,i} = V_{d,i} + \varepsilon_{d,i}$, where $V_{d,i}$ is the deterministic/observed/explained utility component and $\varepsilon_{d,i}$ is the stochastic/unobserved/un-explained utility component. Different assumptions about the error component $\varepsilon_{d,i}$ lead to different choice models.

Among these choice models, the multinomial logit (MNL) has seen wide applicability given its closed-form probability expression[31]. In the MNL model, the error components $\varepsilon_{d,i}$ are assumed to be independent and identically standard Gumbel random variables across choice alternatives. However, one of the main limitations of the MNL model is its independence from irrelevant alternatives (IIA) property, which implies that the relative odds of choosing an alternative A over another alternative B does not depend on other alternatives (excluding A and B). Equivalently, the cross elasticity defined as the percentage change in probability of alternative B for unit percentage change in attributes of alternative A does not depend on B, leading to proportional substitution effects in the MNL model. Later, generalized variants of the MNL model, including (a) the generalized extreme value (GEV) models with more flexible substitution patterns of choice probabilities across alternatives[27, 37, 40, 42], (b) the heteroscedastic logit model to allow different variances for utility functions across choice alternatives[4], and (c) the mixed logit (MMNL) model to handle random taste heterogeneity across decision makers and correlation of choice outcomes in panel data[32], were developed. In all these variants of the MNL model, the unobserved utility components

Discrete choices: Choice scenarios in which the decision-maker is faced with picking one unique alternative from the set of mutually exclusive and collectively exhaustive set of alternatives.

Random utility maximization (RUM) Theory: Stochastic extension of utility theory in which the decision-maker's utility function for each alternative is assumed to be composed of two parts—factors observed and unobserved by the analyst.

Utility: An alternative-specific representation of the decision-maker's net benefit from choosing an alternative.

Multinomial logit (MNL) model: The simplest and earliest discrete choice models that assumed that kernel errors are independent and identically distributed (*i.i.d.*) Gumbel or type-I extreme value random variables across choice alternatives.

Generalized extreme value (GEV) models: A group of choice models which relax the *i.i.d.* assumption of the MNL model; specifically, the kernel errors are assumed to multivariate type-I extreme value random variables.

[1] *Department of Civil and Environmental Engineering, The Pennsylvania State University, 226A Sackett Building, University Park, PA 16802, USA.*
**rpaleti@psu.edu*

Springer IISc Press

Multinomial probit (MNP) model: A discrete choice model that assumes the kernel errors to be multivariate normal random variables.

$\varepsilon_{d,i}$ are assumed to have a multivariate type-I extreme value or Gumbel distribution.

Alternatively, the multinomial probit (MNP) model with normally distributed utility functions is a one-stop solution to all the three limitations of the MNL model, including generic covariance structure for stochastic utility components, random taste heterogeneity, and correlation across decision-makers in panel data[15, 24, 38]. Unlike in the GEV models, the likelihood function in the MNP model does not have a closed form and entails evaluation of a multivariate integral of dimensions in the order of the number of choice alternatives. However, recent advances in the computational power coupled with the development of alternative inference methods, including both simulation and analytical approximations, have enabled faster estimation of MNP models[5, 21, 29, 30, 36]. The normal distributional assumption of the choice utilities in the MNP model implies a bell-shaped utility density curve that is perfectly symmetric. On the other hand, the Gumbel distribution in the MNL model is positively skewed with slightly fatter tails compared to a normal distribution. This property of Gumbel distribution is preferred in some choice contexts, given that it can allow more extreme behavior compared to a normal distribution[38]. However, the difference between the two distributions is empirically indistinguishable in many cases given their close semblance. In the MNP model, the deterministic utility component $V_{d,i}$ can be interpreted as the average utility over all the decision-makers (with same characteristics), out of which one decision-maker is chosen at random whereas the stochastic utility component $\varepsilon_{d,i}$ is the deviation from this average utility[24]. The error component $\varepsilon_{d,i}$ is comprised of several factors including omitted attributes describing the alternative and the decision-maker, imperfect information, and errors in the utility function, among others. So, the normal distributional assumption for the error term seems natural given that the average of several independent random variables converges to a normal distribution according to the central limit theorem. However, this is not always true particularly in cases when the random variables being aggregated are non-identical, correlated, and finite. In addition, wrong assumptions about the stochastic error part can lead to inconsistent parameter estimates resulting in inaccurate forecasts and policy implications[43]. Quite understandably, researchers explored alternate distributions for kernel errors in choice models.

Kernel Errors: The part of the utility function that is composed of factors unknown to the analyst which is typically assumed to a random variable with pre-specified parametric distribution.

2 Review Scope and Focus

There is considerable literature that developed choice models with non-normally distributed random parameters. For instance, researchers used non-normal distributions including lognormal, triangular, and uniform distributions for random parameters within the mixed logit framework[19]. Along similar lines, Bhat and his colleagues developed MNP model with skew-normal random parameters estimated using an alternate inference method that combines analytical approximation of multivariate normal cumulative distribution function and composite marginal likelihood[6, 9, 10]. More recently, MNP model with truncated normal random parameters was developed[33]. These studies are mainly motivated by the need to constrain the domain of cost/time parameters to reflect strictly negative (or positive) behavioral sensitivities. While these studies introduce skewness in the effective error term (i.e., combination of random parameters and kernel error), they retain either the Gumbel or normal distributional assumption for the kernel errors. In this paper, we focus exclusively on past research that explored alternate distributions for the kernel errors (and not taste sensitivity parameters).

Researchers developed ordered response models for modeling categorical data in which the categories are intrinsically ordered (e.g., attitudes measured on Likert scale, or intensity of events). These ordinal responses are interpreted as outcomes of a data generation process in which a single latent propensity y_d^* is mapped into different ordinal categories by latent threshold parameters $\psi_{d,j}$ $(j = 1, 2, \ldots, J-1)$ that partition the propensity space into as many intervals as the number of ordinal categories J. Most earlier applications of OR models assume (1) additive separability of latent propensity into two components—V_d (observed) and ε_d (unobserved), and (2) fixed thresholds that do not vary across observations, i.e. $\psi_{d,j} = c_j$. Unlike in unordered response models, exploring alternate distributions for ε_d in propensity is much easier given that the implied choice probabilities can be written as the difference of the univariate cumulative distribution function of ε_d evaluated at two different points. Among different parametric distributions, the normal (leading to ordered probit model) and logistic (ordered logit model) distributions are frequently used in the literature. Other parametric distributions including complementary log–log and Gompertz distributions were also used but do not have any clear underlying motivation[23]. Recently, ordered response model with skew-normal error term was developed to explore skewness and asymmetry in unobserved factors affecting bicycling frequency[7]. Also, generalized ordered response (GOR) models with both systematic and stochastic heterogeneity in thresholds (as opposed to treating them as fixed) were developed using random parameters specification within the thresholds[18, 22]. More recently, a new class of ordered response models with intrinsically stochastic thresholds that are truncated realizations of Gumbel and normal random variables were developed[35]. Ordinal data can also be modeled within the unordered response framework but with flexible kernel error dependency structure. For instance, ordered GEV (OGEV) model which retains the Gumbel marginal distribution for kernel error terms but allows correlation between two ordinal alternatives A and B to be a function of the gap or difference between the alternatives, i.e. $|A-B|$ was developed[37]. In this paper, we limit the literature synthesis to unordered choice models with alternate kernel error distributions beyond the Gumbel and normal distributions used in the GEV and MNP models, respectively.

3 Relevant Studies

Table 1 presents key details of different studies in the literature that explored alternate kernel error distributions. We now proceed to discuss each of these studies in greater detail. As mentioned earlier, the restrictive IIA property of the MNL model has led researchers to explore alternate kernel error distributions that allow flexible substitution patterns but have closed-form probability expression. For instance, Daganzo introduced a choice model with negative exponential distribution (NED) for the kernel errors but did not explore its properties in complete detail[15]. More recently, this NED model re-surfaced again as the exponomial choice (EC) model in which the utility function is bounded from above and negatively skewed[1, 15]. In the EC model, the stochastic part of the utility function $\varepsilon_{d,i}$ is assumed to have exponential distribution as follows: $U_{d,i} = V_{d,i} - \varepsilon_{d,i}$. It is important to note that the error component in the EC model appears with a negative sign making the effective error term $-\varepsilon_{d,i}$ that has additive inverse of exponential distribution. In the EC model, the deterministic part of the utility function $V_{d,i}$ can be interpreted as maximum attractiveness or ideal utility associated with an alternative and the error component captures heterogeneity across decision-makers. In choice contexts where decision-makers are more knowledgeable about the choice market (i.e. competing alternatives and their prices), the utility would drop-off rapidly above a latent threshold price, making the negatively skewed distribution a better alternative to the symmetric normal or positively skewed Gumbel distribution[1]. Also, the EC model has applicability in choice contexts where the perceived attractiveness of different options is bounded[15]. An attractive feature of the EC model is that it has a closed-form probability expression but without the limiting IIA property of the MNL model[1]. Also, the heteroscedastic extension of the EC model has a closed-form choice probability unlike the HEV model which entails evaluation of a one-dimensional integral[2, 4].

Along similar lines, researchers developed choice models using distributions with restricted range for $\varepsilon_{d,i}$ terms akin to the exponential distribution used in the EC model. However, closed-form probability in these choice models is obtained only when ideal utilities $V_{d,i}$ are equal across all choice alternatives, i.e. $V_{d,i} = V_d \forall i$. For example, Beilner and Jacobs developed a choice model with Weibull-distributed utility functions for analyzing route choice preferences where ideal utility $V_{d,i} = 0$ for all route alternatives. In the route choice context, it is reasonable to assume that route choice utilities are bounded from above by 0 given that the main factors affecting route choice are travel times and costs[15]. This model is equivalent to the model developed by Castillo under the random disutility minimization (RDM) decision rule with independent Weibull distributed route costs and the traveler is assumed to choose the least cost route[11]. This is because the RDM decision rule with disutility $\tilde{U}_{d,i} = \varepsilon_{d,i}$ is equivalent to the RUM decision rule with additive inverse utility, i.e. $U_{d,i} = -\varepsilon_{d,i}$.

Another model that bears close resemblance to the Weibit model is the choice model with multiplicative errors[20]. In the multiplicative errors model, the utility associated with alternative i is specified as: $U_{d,i} = V_{d,i}\varepsilon_{d,i}$, where $V_{d,i} < 0$ and $\varepsilon_{d,i} > 0$ is assumed to be an exponential random variable. The negative sign assumption of $V_{d,i}$ is reasonable in applications where observed utility can be interpreted as generalized cost. This model with multiplicative errors can be viewed as MNL model with non-linear utility function given by $-\ln\left(-V_{d,i}\right)$ because natural logarithm of exponential random variable has reverse Gumbel (or additive inverse of Gumbel) distribution. Among continuous random variables with positive domain, the exponential distribution is the

Table 1: Past literature on alternate kernel error distributions.

Studies	Utility structure	Kernel error distribution	Behavioral reasons	Key features
Daganzo[15] and Alptekinoglu and Semple[1]	Additive	Negative exponential	1. Utility functions are bounded from above in pricing choice contexts	1. Does not have the independence from irrelevant alternatives (IIA) property
			2. Willingness-to-pay (WTP) is negatively skewed across travelers	2. Closed-form probability expression in heteroscedastic generalization
Beilner and Jacobs[3] and Castillo et al.[11]	Additive	Negative Weibul	Route choice utilities are bounded from above at 0 given that any non-zero travel cost is bound to result in some disutility to the traveler	Relaxes the monotonic profile of negative exponential distribution
Fosgerau and Bierlaire[20]	Multiplicative	Generalized exponential	In certain choice contexts, the additive utility assumption can be restrictive, in which case the logarithmic utility function consistent with multiplicative errors is better suited	1. Equivalent to additive MNL model with logarithmic specification for the observed part of the utility
				2. Exponential distribution is the maximum entropy distribution implying that is assumes minimal information in addition to the observed part of the utility
Chikaraishi and Nakayama[13]	q-product	q-Generalized reverse Gumbel	The parameter q indicates the decision maker's risk attitude. $q=0$: risk neutral; $q>0$: risk loving; and $q<0$: risk averse	Encompasses Weibul and Gumbel distribution as special cases for different values of q
del Castillo[12]	Additive	Additive combination of Gumbel and exponential/logistic errors	May be better suited in cases with heteroscedastic errors	Does not have the independence from irrelevant alternatives (IIA) property
Dubey et al.[16]	Additive	Multivariate student-T	Better accounts for decision-uncertainty relative to the variation in the indirect utility	Better suited for un-balanced choice datasets that are common in transportation
Paleti[34]	Additive	Multivariate truncated normal	Utility functions are bounded	Can accommodate any combination of skewness and bounds on utilities across choice alternatives
Li[28]	Additive	Semi-parametric	Encompasses linear, logarithmic, and any arbitrary non-linear effects of choice attributes	Allows heteroscedastic variances across choice alternatives while retaining closed-form probability
Wang et al.[41]	Additive	Non-parametric	Each peak in the multi-modal distribution can be interpreted as representative of latent population groups with different unobserved attitudes and preferences toward the choice alternative	Does not have the Independence from Irrelevant Alternatives (IIA) property

maximum entropy distribution implying that it assumes minimum information about unknown factors that influence the utility function[20]. The choice model with multiplicative errors was found to have a better statistical fit and different willingness-to-pay (WTP) measures compared to standard MNL model in Danish and Swiss mode choice contexts.

More recently, a generalized choice model with q-product disutility functions was developed that encompasses choice models with additive, multiplicative, or in-between relationship between systematic and random utility components depending on the value of parameter q. The disutility associated with alternative i is specified as: $U_{d,i} = V_{d,i} \otimes_q \varepsilon_{d,i}$, where the q-product operator $\otimes_q$ is defined as follows: $V_{d,i} \otimes_q \varepsilon_{d,i} = \left[V_{d,i}^{1-q} + \varepsilon_{d,i}^{1-q} - 1\right]^{\frac{1}{1-q}}$ and $V_{d,i}^{1-q} + \varepsilon_{d,i}^{1-q} - 1 > 0$. The parameter q determines the degree of dependence between $V_{d,i}$ and variance of $U_{d,i}$. Specifically, when $q = 1$, $U_{d,i} = V_{d,i}\varepsilon_{d,i}$ and variance increases at the rate of $V_{d,i}^2$ and the distribution is equivalent to the Weibull distribution Alternatively, when $q = 0$, $U_{d,i} = V_{d,i} + \varepsilon_{d,i} - 1$ and the distribution is equivalent to the reverse Gumbel distribution. This implies that both logit and Weibit choice models can be re-framed as special cases of choice model with q-product disutility functions in the random disutility minimization (RDM) framework[13]. The q-product choice models were found to outperform the standard MNL model in both route and travel mode choice travel contexts.

More recently, a new class of RUM models with kernel errors that are sum of two independent random variables, one of which has Gumbel distribution were developed. Four different choice models that belong to this new class of RUM models were proposed. These include choice models where (a) all kernel errors are difference of independent Gumbel and exponential random variables; (b) a subset of kernel errors are Gumbel random variables and remaining kernel errors are different of Gumbel and exponential random variables; (c) all kernel errors are logistic random variables; and (d) a subset of kernel errors are logistic and remaining are Gumbel random variables. These choice models do not have the restrictive 'Independence from Irrelevant Alternatives' (IIA) property, accommodate heteroscedasticity between choice alternatives with Gumbel and non-Gumbel error terms, and have closed-form probability expressions[12]. Also, models with non-Gumbel errors either have symmetric (in the case of logistic errors) or positively skewed utility functions (in the case when errors are difference of Gumbel and exponential random variables). These models were found to be good competitors to the standard MNL model in the mode choice context[12]. These new choice models were found to marginally better or at least as good model fit as the standard MNL model in mode choice context in Japan.

More recently, choice models with heavy tailed t-distributed kernel errors were developed and found to perform better in choice contexts with class-imbalanced datasets[16]. For instance, the model with t-distributed errors was found to have better data fit in new vehicle purchase choice data with class-imbalance where the share of EVs was (32%) compared to gasoline-fueled vehicles (64%) amd opt-out alternative (4%). In this model, the vector of kernel errors $\varepsilon_d = \left(\varepsilon_{d,1}, \varepsilon_{d,2}, \ldots \varepsilon_{d,J}\right)'$ is assumed to be a multivariate t-distributed random vector with δ degrees of freedom (DOF). The marginal distribution of kernel error $\varepsilon_{d,i}$ is univariate t-distribution with δ DOF. However, the difference of two t-distributed random variables is not t-distributed even in the special case when the t-distributions are independent. So, it is unclear how researchers resolved this inconsistency in their model formulation because calculation of choice probabilities involves working with the cumulative distribution function of utility differences.

While there have been several attempts to explore alternate distributions for kernel errors, none of these models work with the normal distribution. Normal distribution has two key properties that makes it perfectly suited for complex dependency patterns in choice outcomes—(1) affine property which implies that linear transformations of multivariate normal random variables also have normal distribution, i.e. if $\boldsymbol{X} \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ then $\boldsymbol{Y} = \mathbf{A}\boldsymbol{X} + \boldsymbol{b} \sim MVN\left(\mathbf{A}\boldsymbol{\mu} + \boldsymbol{b}, \mathbf{A}\Sigma\mathbf{A}^T\right)$; and (2) linear combinations of two multivariate normal random variables also has multivariate normal distribution. These two properties of normal random variables is highly desirable in choice scenarios including complex spatio-temporal dependency patterns where choices of consumers are inter-dependent and choices of the same consumer across multiple choice occasions are correlated, and integrated choice and latent variable (ICLV) models that incorporate latent psychological constructs in choice models[8, 17, 25, 39]. Recently, the MNP model with multivariate truncated normal distribution for kernel errors, referred to as the multinomial truncated probit (MNTP) model was developed[34]. The resulting model can accommodate utility functions with any arbitrary bounds and skewness patterns across choice alternatives and can be estimated using the traditional GHK simulator. The ability of the maximum simulated likelihood (MSL) inference method to retrieve the model parameters was shown using simulation analysis. Also, the practical applicability of the proposed model was demonstrated in airline itinerary choice context where the kernel errors of the inside itinerary options were normal random variables truncated from above at zero and the kernel error of the outside option of not choosing to fly was regular normal random variable[34].

Alternatively, researchers developed a semi-parametric method that does not require any prior assumptions regarding the distribution of kernel error terms. These models have the logit probability structure where the linear deterministic utility component $V_{q,i}$ is replaced by a sensitivity function $S_{q,i}$ whose functional form is determined by the data[28]. These methods were developed in the context of random disutility minimization where alternative costs are assumed to be random variables independent across choice alternatives. The kernel errors in this model are assumed to belong to a large class of distributions with the following functional form for the cumulative distribution function, $F\left(\varepsilon_{d,i} < t\right) = 1 - [1 - F_b(t)]^{\alpha_{d,i}}$, where $F_b(t)$ is a base distribution function and $\alpha_{d,i}$ is a parameter specific to decision-maker d and alternative i. This method is based on the idea that any random variable with a cumulative distribution function of the form described above can be transformed into a Gumbel random variables as follows: $h(t) = \frac{1}{\theta} \log\left(-\log\left[1 - F_b(t)\right]\right)$ where θ is the scale parameter.

So, random costs $\varepsilon_{d,i}$ in the original formulation are transformed into Gumbel random variables $\eta_{d,i} = h\left(\varepsilon_{d,i}\right)$ with cumulative density function given by: $G\left(\eta_{d,i} < t\right) = 1 - e^{-\alpha_{d,i} e^{\theta t}}$. The sensitivity function $S_{q,i}$ in the probability expression is defined as $ln\left(H_{q,i}\right)$ where $H_{q,i}$ is the relationship between the means of kernel error before and after the transformation $h\left(\varepsilon_{d,i}\right)$. Assuming random costs $\varepsilon_{d,i}$ to have exponential distribution is equivalent to the MNL model with logarithmic transformation of random costs as was the case in the Weibit model and choice model with multiplicative errors[11, 20]. In addition to exponential distributed random costs, Li provided the non-linear transformation function $h(t)$ and corresponding sensitivity function $S_q(t)$ for a wide array of parametric distributions including exponential, Pareto, type II generalized logistic, Gompertz, Rayleigh, and Weibull distributions. Although this method subsumes several other choices models as special cases, it is not necessary to assume a specific distribution for random costs. In fact, Li used p-splines to approximate the unknown sensitivity function $S_{q,i}$ as a linear combination of several B-spline basis functions and the resulting model was estimated using Bayesian methods. The performance of the proposed semi-parametric method was compared against the standard MNL and choice model with multiplicative errors for analyzing mode choice preferences in two different contexts in Denmark. The model where the functional form of $S_{q,i}$ was obtained from data clearly outperformed the parametric models in one of the datasets whereas there were not any significant differences across models in the other dataset.

Also, semi-non-parametric choice model based on orthonormal Legendre polynomial was developed to construct a semi-parametric distribution function that encompasses the Gumbel distribution as a special case[41]. The resulting model can accommodate multi-modal density functions for kernel errors and was used to analyze commute mode choice preferences among four alternatives (auto, transit, bicycle, and walk) in Switzerland. Unlike unimodal density functions that dominated the literature, this study found evidence for bi-modal density function of the kernel error term in utility function of transit alternative. The multiple peaks in multi-modal distribution of the kernel error of an alternative can be interpreted as the presence of latent population groups with different unobserved preferences and attitudes toward that choice alternative. For instance, a dominant group of commuters with positive attitudes toward transit are associated to major mode of kernel distribution whereas a smaller group of commuters with negative attitudes toward transit are associated with the minor mode of the bi-modal kernel distribution. However, as the authors note, one of the limitations of this study is that the computation complexity of the model increases substantially with the number of polynomial functions used for constructing the semi-parametric density function and the number of choice alternatives.

4 Conclusion

Choices models with unbounded and positively skewed (e.g., multinomial logit (MNL)) and symmetric (e.g., multinomial probit (MNP)) utility functions dominated the empirical literature. The closed-form probability expression with Gumbel errors and the affine property of normal random variables combined with efficient estimation methods are the primary reasons for the dominance of MNL/MNP models and their variants. However, there is renewed interest in developing choice models with more flexible kernel errors for both behavioral and accuracy reasons. While the research in this arena is relatively sparse, there have been quite varied and unique approaches to accommodate alternate kernel error structures in choice models. This paper provided an overview of these methods that span both parametric and non-parametric distributions for kernel errors. While some of these models are easily estimable given their closed-form probability structure, others require simulation techniques such as the GHK simulator or Bayesian methods in case of semi-parametric methods. In summary, these choice models with flexible kernel errors were shown to perform better or at least as good as existing models with flexible heteroscedasticity and substitution patterns across choice alternatives. It is interesting that, in spite of their wider applicability beyond transportation, most of these studies were undertaken in the travel mode and route choice contexts.

However, there are a couple of avenues for further improvement in this area. Firstly, the distribution of kernel error of an alternative is closely tied to the specification of the observed part of the utility function for that alternative. Non-linear systematic utility functions and utility functions that violate the regularity assumption (which implies that systematic utility of an alternative does not depend on other alternatives) can accommodate a wide range of behavioral and substitution patterns without necessarily changing the standard assumptions underlying kernel error terms[26]. For instance, in the random regret minimization (RRM) framework, the systematic utility of an alternative takes a non-linear form where the utility of an alternative depends on better alternatives in the choice set that would be forgone if that alternative was chosen[14]. The RRM model with reverse Gumbel distribution has closed-form probabilities and does not have the restrictive IIA property unlike the MNL model in RUM framework. Also, any choice model with choice probabilities $P_{d,i}$ can be written as a special case of "Mother Logit" model with *iid* Gumbel errors and systematic utility function given by $\ln\left(P_{d,i}\right)$ [26]. The resulting mother logit model has a highly non-linear utility function where the effective systematic utility of an alternative $\ln\left(P_{d,i}\right)$ depends on all alternatives in the choice set. While this may be a pure mathematical artifact, it does allude to the possibility that non-linear utility functions and kernel error structures are interdependent. So, further research is needed to understand the relationship between specification of the systematic utility function and the necessity to explore flexible kernel errors. Secondly, random taste heterogeneity and kernel error terms can be confounded. For instance, the mixed logit model with Gumbel kernel error terms can closely approximate any discrete choice model derived from random utility maximization (RUM)[32]. It is possible that models with alternate kernel error distributions are more parsimonious than mixed logit models that attempt to approximate these models using random parameters. So, further research is needed to provide guidance on the choice between mixed logit and choice models with alternate kernel error structures.

Publisher's Note

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

Received: 12 August 2019 Accepted: 23 September 2019
Published online: 28 November 2019

Observed or systematic utility: The part of the utility function of an alternative that is composed of factors known to the analyst; typically parameterized as a linear function of attributes describing the choice alternative and decision-maker.

References

1. Alptekinoglu A, Semple JH (2016) The exponomial choice model: a new alternative for assortment and price optimization. Opera Res 64(1):79–93. https://doi.org/10.2139/ssrn.2210478
2. Alptekinoglu A, Semple JH (2018) Heteroscedastic exponomial choice. SSRN Electron J. https://doi.org/10.2139/ssrn.3232788
3. Beilner H, Jacobs F (1974) Probabilistic aspects of traffic assignment. In: Proceeedings of 5th International Symposium on the Theory of Traffic Flow and Transportation, Berkely, pp 183–194
4. Bhat CR (1995) A heteroscedastic extreme value model of intercity travel mode choice. Transp Res Part B. https://doi.org/10.1016/0191-2615(95)00015-6
5. Bhat CR (2011) The maximum approximate composite marginal likelihood (MACML) estimation of multinomial probit-based unordered response choice models. Transp Res Part B Methodol. https://doi.org/10.1016/j.trb.2011.04.005
6. Bhat CR (2018) New matrix-based methods for the analytic evaluation of the multivariate cumulative normal distribution function. Transp Res Part B Methodol. https://doi.org/10.1016/j.trb.2018.01.011
7. Bhat CR, Astroza S, Hamdi AS (2017) A spatial generalized ordered-response model with skew normal kernel

error terms with an application to bicycling frequency. Transp Res Part B Methodol. https://doi.org/10.1016/j.trb.2016.10.014

8. Bhat CR, Dubey SK (2014) A new estimation approach to integrate latent psychological constructs in choice modeling. Transp Res Part B Methodol. https://doi.org/10.1016/j.trb.2014.04.011
9. Bhat CR, Dubey SK, Nagel K (2015) Introducing non-normality of latent psychological constructs in choice modeling with an application to bicyclist route choice. Transp Res Part B Methodol. https://doi.org/10.1016/j.trb.2015.04.005
10. Bhat CR, Sidharthan R (2012) A new approach to specify and estimate non-normally mixed multinomial probit models. Transp Res Part B Methodol. https://doi.org/10.1016/j.trb.2012.02.007
11. Castillo E et al (2008) Closed form expressions for choice probabilities in the Weibull case. Transp Res Part B Methodol. https://doi.org/10.1016/j.trb.2007.08.002
12. del Castillo J (2016) A class of RUM choice models that includes the model in which the utility has logistic distributed errors. Transp Res Part B Methodol 91:1–20. https://doi.org/10.1016/j.trb.2016.04.022
13. Chikaraishi M, Nakayama S (2016) Discrete choice models with q-product random utilities. Transp Res Part B Methodol. https://doi.org/10.1016/j.trb.2016.08.013
14. Chorus CG, Arentze TA, Timmermans HJP (2008) A random regret-minimization model of travel choice. Transp Res Part B Methodol. https://doi.org/10.1016/j.trb.2007.05.004
15. Daganzo C (1979) Multinomial probit: the theory and its application to demand forecasting. Academic Press, Elsevier, Amsterdam. https://doi.org/10.2307/2287751
16. Dubey S. et al. (2019) A generalized continuous-multinomial response model with a t-distributed error kernel, pp. 1–39. Available at: http://arxiv.org/abs/1904.08332. Accessed 15 June 2019
17. Eidsvik J et al (2014) Estimation and prediction in spatial models with block composite likelihoods. J Comput Graph Stat. https://doi.org/10.1080/10618600.2012.760460
18. Eluru N, Bhat CR, Hensher DA (2008) A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. Accid Anal Prev. https://doi.org/10.1016/j.aap.2007.11.010
19. Fosgerau M (2006) Investigating the distribution of the value of travel time savings. Transp Res Part B Methodol. https://doi.org/10.1016/j.trb.2005.09.007
20. Fosgerau M, Bierlaire M (2009) Discrete choice models with multiplicative error terms. Transp Res Part B Methodol. https://doi.org/10.1016/j.trb.2008.10.004
21. Geweke J, Keane M, Runkle D (2006) Alternative computational approaches to inference in the multinomial probit model. Rev Econ Stat. https://doi.org/10.2307/2109766
22. Greene W et al (2014) Heterogeneity in ordered choice models: a review with applications to self-assessed health. J Econ Surv. https://doi.org/10.1111/joes.12002
23. Greene WH, Hensher DA (2010) Modeling ordered choices: a primer. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511845062
24. Hausman JA, Wise DA (1978) A conditional probit model for qualitative choice: discrete decisions recognizing interdependence and heterogeneous preferences. Econometrica 46(2):403–426. https://doi.org/10.2307/1913909
25. Heagerty PJ, Lele SR (1998) A composite likelihood approach to binary spatial data. J Am Stat Assoc. https://doi.org/10.1080/01621459.1998.10473771
26. Hess S, Daly A, Batley R (2018) Revisiting consistency with random utility maximisation: theory and implications for practical work. Theory Decis. https://doi.org/10.1007/s11238-017-9651-7
27. Koppelman FS, Wen CH (2000) The paired combinatorial logit model: properties, estimation and application. Transp Res Part B Methodol. https://doi.org/10.1016/s0191-2615(99)00012-0
28. Li B (2011) The multinomial logit model revisited: a semi-parametric approach in discrete choice analysis. Transp Res Part B Methodol. https://doi.org/10.1016/j.trb.2010.09.007
29. McCulloch R, Polson NG, Rossi PE (2000) A Bayesian analysis of the multinomial probit model with fully identified parameters. J Econ 99(1):173–193. https://doi.org/10.1016/S0304-4076(00)00034-8
30. McCulloch R, Rossi PE (1994) An exact likelihood analysis of the multinomial probit model. J Econ 64(1–2):207–240. https://doi.org/10.1016/0304-4076(94)90064-7
31. McFadden D (1973) Conditional logit analysis of qualitative choice behavior. In: Zaremb P (ed) Frontiers in econometrics. Academic Press, New York. https://doi.org/10.1108/eb028592
32. McFadden D, Train K (2002) Mixed MNL models for discrete response. J Appl Econo 15(5):447–470. https://doi.org/10.1002/1099-1255(200009/10)15:5%3c447:aid-jae570%3e3.3.co;2-t
33. Paleti R (2018) Generalized multinomial probit model: accommodating constrained random parameters. Transp Res Part B Methodol. https://doi.org/10.1016/j.trb.2018.10.019
34. Paleti R (2019) Multinomial probit model with truncated normal kernel errors: analysis of airline itinerary choices. Technical Paper, The Pennsylvania State University, University Park
35. Paleti R, Pinjari A (2019) A new class of ordered response models with stochastic thresholds. Technical Paper, The Pennsylvania State University, University Park
36. Patil PN et al (2017) Simulation evaluation of emerging estimation techniques for multinomial probit models. J Choice Model. https://doi.org/10.1016/j.jocm.2017.01.007

37. Small KA (1987) A discrete choice model for ordered alternatives. Econometrica 55(2):409. https://doi.org/10.2307/1913243
38. Train K (2003) Discrete choice methods with simulation, discrete choice methods with simulation. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511753930
39. Vij A, Walker JL (2016) How, when and why integrated choice and latent variable models are latently useful. Transp Res Part B Methodol. https://doi.org/10.1016/j.trb.2016.04.021
40. Vovsha P (2007) Application of cross-nested logit model to mode choice in Tel Aviv, Israel, Metropolitan area. Transp Res Rec J Transp Res Board. https://doi.org/10.3141/1607-02
41. Wang K et al (2017) On the development of a semi-nonparametric generalized multinomial logit model for travel-related choices. PLoS One 12(10):e0186689
42. Wen CH, Koppelman FS (2001) The generalized nested logit model. Transp Res Part B Methodol. https://doi.org/10.1016/s0191-2615(00)00045-x
43. Ye X et al (2017) A practical method to test the validity of the standard Gumbel distribution in logit-based multinomial choice models of travel behavior. Transp Res Part B Methodol. https://doi.org/10.1016/j.trb.2017.10.009

Rajesh Paleti graduated from the University of Texas at Austin with a doctoral degree in Civil Engineering in 2012. Prior to joining Penn State, Dr. Paleti worked as an assistant professor at Old Dominion University from 2014 to 2018, and as a transportation consultant at Parsons Brinckerhoff from 2013 to 2014. Dr. Paleti's research is primarily focused on generating novel econometric and statistical models with potential applications in three key areas: (1) travel behavior analysis, (2) traffic safety, and (3) transportation systems modeling and simulation. He worked as a consultant for 2 years during which he was involved in the development of advanced activity-based travel demand models for southern California, Jerusalem, Ohio, and Dallas. Dr. Paleti received several honors, including the 2016 Chan Wui & Yunyin Rising Star Award for Early Career Professionals by Transportation Research Board (TRB), the 2013 Pikarsky Award for outstanding Ph.D. Dissertation in Transportation Science & Technology, and the 2013 Pyke Johnson award for outstanding paper in transportation systems planning.