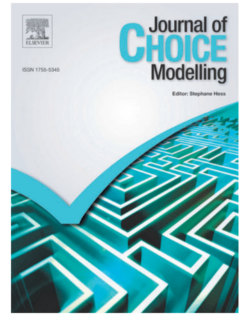
Journal Pre-proof



Multiple discrete choice and quantity with order statistic marginal utilities

Scott Webster

PII: S1755-5345(22)00052-5
DOI: https://doi.org/10.1016/j.jocm.2022.100395
Reference: JOCM 100395

To appear in: *Journal of Choice Modelling*

Received Date: 30 July 2020
Revised Date: 1 October 2022
Accepted Date: 15 November 2022

Multiple Discrete Choice and Quantity with Order Statistic Marginal Utilities

This paper presents a random utility maximization model for individuals selecting discrete quantities from a set of *n* alternatives. Multiple alternatives with positive quantities may be selected. Diminishing marginal utility to quantity of each alternative is modeled via order statistics of independent Gumbel random variables. The model is parsimonious and tractable, admitting closed-form expressions for choice probabilities. As such, the model is amenable to maximum likelihood estimation of structural parameters from observed choices.

Probability functions recover binary logit probabilities under binary choice and a maximum quantity of one unit, and probability is monotonic in the quantity of each alternative. The monotonic property likely restricts the application of the model to a narrow range of settings. The property is a manifestation of a recursive relationship among Gumbel order statistic probabilities. This relationship and related properties may lead to new models for capturing important complexities in a tractable manner.

Keywords: Order Statistics; Random Utility Maximization; Choice Models

1. Introduction

The literature on multiple discrete choice (MDC) with continuous quantity decisions is extensive. The literature on MDC with discrete quantity decisions is not. This can be at least partially explained by tractability challenges that arise with the requirement of discrete quantities.

For many real-world applications that exhibit discrete quantity choices, a continuous decision variable can be a reasonable approximation. This is more likely to be case in applications where observed choice quantities tend to be relatively large. For other settings, such as shopping where consumers tend to purchase a few units of several products, the continuous approximation may be problematic.

Gallego and Wang (2020) propose a tractable formulation for analyzing MDC models with discrete quantities called the threshold utility model (TUM). TUM specifies how alternatives are selected at a choice event. TUM underlies MDC models with continuous quantity decisions and includes all generalized extreme value models (McFadden 1978) as a special case.

In this paper, I present a novel modeling strategy (Section 4) that defines random marginal utilities as differences between Gumbel order statistics. Order statistics capture diminishing marginal utility to consumption. I present properties related to Gumbel order statistics and use these properties to derive choice probability expressions for this model within a TUM framework (Section 6). Included is a theorem showing that a property of the difference of Gumbel random variables extends to a form of conditional Gumbel order statistics. This result is one key to obtaining closed-form expressions for choice probabilities. The characterizations of Gumbel order statistics in this paper may have implications for other application areas. I briefly discuss parameter estimation and normative analysis, then conclude with a summary and reflection. Proofs are in the appendix.

2. Quantity Decisions as Nonnegative Continuous Variables

2.1. General Model

Individual j from a population selects quantities among n alternatives. Let $u_{ij}(x_i)$ denote individual j utility from $x_i \geq 0$ units of alternative i. Function $u_{ij}(x_i)$ is concave and increasing for all i and j. Individual j's total utility as a function of continuous quantity vector $\mathbf{x} = (x_1, \ldots, x_n)$ is $\sum_{i=1}^{n} u_{ij}(x_i)$, i.e., utility is additively separable in the alternatives.

Individual j chooses $\mathbf{x}$ to maximize total utility subject to an upper limit on the total quantity selected, denoted B_j:

$$\max_{\mathbf{x}\geq\mathbf{0}} \left\{ \sum_{i=1}^{n} u_{ij}(x_i) : \sum_{i=1}^{n} x_i \leq B_j \right\}. \tag{1}$$

The corresponding Lagrangian is

$$L_j(\mathbf{x}, \lambda_j) = \sum_{i=1}^{n} u_{ij}(x_i) + \lambda_j \left(B_j - \sum_{i=1}^{n} x_{ij} \right).$$

If u_{ij} are differentiable and the constraint is binding, then marginal utility at quantity x_i is $\Delta u_{ij}(x_i) = u_{ij}'(x_i)$ and at optimal solution $(\mathbf{x}^*, \lambda_j) = \arg\max_{\mathbf{x},\lambda_j} L(\mathbf{x}, \lambda_j)$,

$$\Delta u_{ij}(x_i^*) = \lambda_j \text{ for all } x_i^* > 0,\ \Delta u_{ij}(0) \leq \lambda_j \text{ for all } x_i^* = 0,\ \sum_{i=1}^{n} x_i^* = B_j. \tag{2}$$

The value of λ_j represents the individual's threshold marginal utility in the utility maximization problem; for any i satisfying $\Delta u_{ij}(0) > \lambda_j$, the value of x_i is increased until marginal utility matches the threshold λ_j.

From a microeconomic decision-making perspective, the decision model given in (1) can be viewed as a problem in the second stage of a consumer's two-stage budgeting process proposed by Strotz (1957). In the first stage, an individual allocates total expenditures for an upcoming period across groups of products (e.g., food, clothing, recreation, etc.). In the second stage, the individual selects the quantities of products within each group. For example, Hausman et al. (1995) apply two-stage budgeting in an empirical estimation of recreation choice; individuals decide the total number of recreational trips (for an upcoming period) in the first stage, then choose the specific activities in the second stage.

The two-stage budgeting process can be formalized as an iterative process wherein feedback from the previous cycle informs budget allocation in the next period. Gorman (1959) identifies necessary and sufficient conditions for which the first-stage allocation decision only requires price indices for each product group (i.e., conditions under which detailed product prices are not needed for the first-stage decision). A sufficient condition, for example, is additive separability of total utility across product groups (e.g., if there a N product groups, total utility = $U_1(\cdot) + \ldots + U_N(\cdot)$ where U_i is the utility function of product group i). Gorman (1959) shows that errors in an iterative budgeting cycle are small if price changes from one period to the next are small.

2.2. Multiple Discrete-Continuous Extreme Value Model

Bhat (2005) proposes the multiple discrete-continuous extreme value (MDCEV) model (for recent discussions of MDCEV extensions and applications see, e.g., Bhat et al. 2020, Palma and Hess 2020, Saxena et al. 2020). Because **x** is continuous and u_{ij} are continuous, increasing, differentiable functions, the budget constraint is binding, which implies

$$x_i > 0 \text{ for some } i \in \{1, \ldots, n\} \tag{3}$$

for each individual. This is not restrictive, e.g., if $\mathbf{x} = 0$ among alternatives 1 through n is possible, then an alternative 0 corresponding to not selecting 1 through n can be introduced to the choice set.

MDCEV is appealing because (1) it admits closed-form probabilities for choice decisions among a population of individuals and (2) it reduces to the widely accepted multinomial logit (MNL) choice probability if only one alternative is selected. I summarize this model below.

The utility of quantity x_i of alternative i of a randomly selected individual from the population is

$$\tilde{u}_i(x_i) = e^{a_i + \tilde{\varepsilon}_i}(b_i + x_i)^{\alpha_i}$$

where $\tilde{\varepsilon}_i$ are iid Gumbel random variables that capture idiosyncratic (unknown to the researcher) preferences of individuals. The remaining parameters, a_i, b_i, α_i reflect observed characteristics of alternative i utility, and $\alpha_i \in (0, 1]$ to reflect positive and diminishing marginal utility to quantity.

Note that for realization ε_i of $\tilde{\varepsilon}_i$, $\Delta u_i(x_i) = e^{a_i + \varepsilon_i + \ln\left(\alpha_i (b_i + x_i)^{\alpha_i - 1}\right)}$. Assuming the same Lagrange multiplier λ among individuals in the population, optimal $\mathbf{x}^*$ satisfies

$$a_i + \varepsilon_i + \ln(\alpha_i) + (\alpha_i - 1)\ln(b_i + x_i^*) = \ln \lambda \text{ for all } x_i^* > 0$$

$$a_i + \varepsilon_i + \ln(\alpha_i) + (\alpha_i - 1)\ln(b_i) < \ln \lambda \text{ for all } x_i^* = 0.$$

Define

$$v_i(x_i) = a_i + \ln(\alpha_i) + (\alpha_i - 1)\ln(b_i + x_i)$$

$$v_0 = \ln \lambda$$

$\tilde{\mathbf{x}} = (\tilde{x}_1, \ldots, \tilde{x}_n)$ = quantity choice vector for a randomly selected individual from the population.

Then the fraction of the population that selects quantity vector $\mathbf{x} = (x_1, \ldots, x_k, 0, \ldots, 0)$ where $x_i > 0$ for $i = 1, \ldots, k$ and $k \in \{1, \ldots, n\}$ is

$$P\left[\tilde{\mathbf{x}}=\mathbf{x}\right]=P\left[v_1\left(x_1\right)+\tilde{\varepsilon}_1=v_0,\ldots,v_k\left(x_k\right)+\tilde{\varepsilon}_k=v_0,v_{k+1}\left(0\right)+\tilde{\varepsilon}_{k+1}<v_0,\ldots,v_n\left(0\right)+\tilde{\varepsilon}_n<v_0\right].$$

(Due to (3), $P\left[\tilde{\mathbf{x}}=\mathbf{0}\right]=0$.). Since $\tilde{\varepsilon}_i$ are iid, $P\left[\tilde{\mathbf{x}}=\mathbf{x}\right]$ can be expressed as a product of probabilities, i.e.,

$$P\left[\tilde{\mathbf{x}}=\mathbf{x}\right]=\left(\prod_{i=1}^{k}P\left[\tilde{\varepsilon}_i=v_0-v_i\left(x_i\right)\right]\right)\left(\prod_{i=k+1}^{n}P\left[\tilde{\varepsilon}_i<v_0-v_i\left(0\right)\right]\right).$$

Substituting Gumbel probabilities into the above, normalizing, and simplifying yields

$$P\left[\tilde{\mathbf{x}}=\mathbf{x}\right]=\left(\prod_{i=1}^{k}\frac{1-\alpha_i}{b_i+x_i}\right)\left(\sum_{i=1}^{k}\frac{b_i+x_i}{1-\alpha_i}\right)\left(\frac{\prod_{i=1}^{k}e^{v_i\left(x_i\right)}}{\left(\sum_{j=1}^{n}e^{v_j\left(x_j\right)}\right)^{k}}\right)\left(k-1\right)! \tag{4}$$

(Bhat 2005). If $k = 1$, then (4) reduces to the form of MNL choice probability, i.e.,

$$P\left[\tilde{\mathbf{x}}=\left(x_1,0,\ldots,0\right)\right]=e^{v_1\left(x_1\right)}\left(\sum_{j=1}^{n}e^{v_j\left(x_j\right)}\right)^{-1}. \tag{5}$$

The closed-form expression for choice probabilities allows for efficient estimation of utility function parameters, e.g., via maximum likelihood estimation (MLE) methods.

3. Discrete Quantity Choices

3.1. Adaptation of Classic Discrete Choice Models

One alternative for modeling multiple discrete choice with discrete quantities is the MNL model with a choice set that includes all combinations of quantity decisions (Train 2009). For example, if $m = n = 2$, then each individual selects one of $(1 + m)^n = 9$ alternatives corresponding to $\mathbf{X} = \{(0, 0), (0, 1), (0, 2), (1, 0), (1, 1), (1, 2), (2, 0), (2, 1), (2, 2)\}$. The MNL-based model has the advantage of simple probability expressions. However, the model has the disadvantage of overfitting when estimating the parameters from the data when m is not small, or when observed data cover only a fraction of the quantity-choice set. The utilities associated with alternative $i = 1, \ldots, n$ have $m + 1$ parameters, one for each possible discrete quantity (and $(m + 1)^n - 1$ parameters in total, after normalizing a parameter to value 1). For comparison, MDCEV has up to three parameters (i.e., a_i, b_i, α_i) for each u_i function, reflecting a parsimonious relationship governing marginal utility. This observation motivates exploration of alternatives to the above MNL-based model that explicitly account for discrete quantity decisions.

The next section considers the special case where the quantity available in any alternative is limited to one. The properties and intuition associated with this special case provide a foundation for generalized models in subsequent sections.

3.2. Binary Quantity Decisions

In this section and the remainder of the paper, I present model elements in a manner that allows for two interpretations of the set of choice alternatives: (1) including an *outside* option against which the utilities of *inside* options are compared, (2) not including an outside option for comparison. I clarify the distinction between these interpretations after introducing necessary background and notation.

There are n alternatives. Each alternative i for $i = 1, \ldots, n$ is either selected or not selected, i.e., quantity decisions are binary. Let

$$\tilde{U}_i = \tilde{u}_i - \tilde{u}_0 \,,\; i = 1, \ldots, n$$

denote the (net) utility of alternative i where $\tilde{u}_i$ for $i = 0, 1, \ldots, n$ are random variables (model primitives).

At each choice event, an individual selects the alternatives with the highest utilities subject to a constraint that, on average, the number of alternatives selected at each event is no more than B. The problem can be expressed as

$$\max_{\lambda \geq 0} \left\{ \sum_{i:\hat{x}\left(\lambda, \tilde{U}_i\right)=1} E\left[\tilde{U}_i\right] : \sum_{i=1}^{n} E\left[\hat{x}\left(\lambda, \tilde{U}_i\right)\right] \leq B \right\} \tag{6}$$

where $\hat{x}\left(\lambda, U_i\right) = 1$ if $U_i \geq \lambda$; otherwise $\hat{x}\left(\lambda, U_i\right) = 0$. Note that $\hat{x}\left(\lambda, U_i\right)$ is consistent with the decision rule given in (2). The problem is referred to as the threshold utility model in Gallego and Wang (2020).[1] If $\sum_{i=1}^{n} E\left[\tilde{U}_i\right] \leq B$, then the constraint is nonbinding and the threshold is $\lambda = 0$; otherwise the threshold λ is the solution to

$$\sum_{i=1}^{n} E\left[\hat{x}\left(\lambda, \tilde{U}_i\right)\right] = B. \tag{7}$$

The probability that alternative i is selected is

$$P\left[\tilde{x}_i = 1\right] = P\left[\tilde{U}_i \geq \lambda\right],\; i = 1, \ldots, n. \tag{8}$$

If $\tilde{u}_i = a_i + \tilde{\varepsilon}_i$, $i = 0, 1, \ldots, n$ where $\tilde{\varepsilon}_i$ are iid Gumbel random variables (normalized with location parameter 0 and scale parameter 1), then the choice probabilities have the following binary logit form:

$$P\left[\tilde{x}_i = 1\right] = P\left[\tilde{\varepsilon}_0 - \tilde{\varepsilon}_i \leq a_i - a_0 - \lambda\right] = \frac{e^{a_i - a_0 - \lambda}}{1 + e^{a_i - a_0 - \lambda}} \,,\; i = 1, \ldots, n. \tag{9}$$

Note that $\tilde{U}_i$ can be interpreted as the utility of alternative i. In this case, an individual selects among n alternatives; characteristics of any outside option are not relevant. Alternatively, one may interpret $\tilde{u}_i$ as

[1] A more general formulation allows the threshold to be different for each alternative, i.e., maximize over $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)$ instead of scalar λ. Gallego and Wang (2020, Theorem 1) show that the optimal solution is a scalar, i.e., $\lambda_i = \lambda$ for all i.

the utility of inside option i that is compared against the best outside option with utility $\tilde{u}_0$ when an individual makes choice decisions; under this interpretation, alternative i is not selected if the utility does not exceed the outside option as shown in (8). Consideration of a no-purchase option within a choice model is relatively commonplace in the literature and this interpretation may be useful for some applications. For example, a firm interested in estimating parameters of a choice model will likely have data on choice decisions of its own products and may be able to develop reasonable estimates of market size. However, the firm is unlikely to have access to choice data on competitor products. In this case, it may be convenient to recognize the best outside option in the model. It is important to emphasize that this approach to estimation requires an implicit assumption on consumers' microeconomic decision-making model: consumers view the firm's offerings within the product category as a group that warrants its own constraint on choice decisions (see (6)), e.g., competitor products, if considered for selection, fall into a separate group.

One might infer from (8) or (9) that the model is not appropriate for normative analysis, e.g., not capturing how changes in one alternative may affect choice probabilities of other alternatives. However, as with MDCEV, interactions are captured through the relationship between the threshold parameter and B. Furthermore, normative analysis of the model under Gumbel error terms is relatively tractable. For example, let $\overline{\mathbf{a}}_i = \left(\overline{a}_{i1},...,\overline{a}_{iM}\right)$ denote the vector of predictors for alternative i. Let $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_M)$ denote the utility coefficient vector estimated from the data, e.g., $\boldsymbol{\alpha}^T\overline{\mathbf{a}}_i$ is an estimate of $a_i - a_0 - \lambda$ in (9) for $i = 1, \ldots, n$. Consider a change in attribute l of alternative i from $\overline{a}_{il}$ to $\overline{a}_{il} + \Delta_{il}$. If one accepts an assumption that the value of B remains stable as changes in alternatives are introduced (or can be predicted), then the updated choice probabilities are

$$P\left[\tilde{x}_j = 1\right] = \frac{e^{\boldsymbol{\alpha}^T\overline{\mathbf{a}}_j + \alpha_l \Delta_{il} I(j=i)}}{e^{\Delta_\lambda} + e^{\boldsymbol{\alpha}^T\overline{\mathbf{a}}_j + \alpha_l \Delta_{il} I(j=i)}}, j = 1, \ldots, n$$

where $I(\cdot)$ in an indicator function and Δ_λ is the solution to

$$\sum_{j=1}^{n} \frac{e^{\boldsymbol{\alpha}^T\overline{\mathbf{a}}_j + \alpha_l \Delta_{il} I(j=i)}}{e^{\Delta_\lambda} + e^{\boldsymbol{\alpha}^T\overline{\mathbf{a}}_j + \alpha_l \Delta_{il} I(j=i)}} = B. \tag{10}$$

While Δ_λ cannot be expressed in closed-form, the left-hand side of (10) is decreasing in Δ_λ, and thus can be efficiently obtained via bisection search.

3.3. Quantity Decisions as Nonnegative Integers

Up to m units of each alternative may be selected where $m > 1$. For quantity $x \in \{1, \ldots, m\}$, let $\Delta\tilde{u}_i(x) = \tilde{u}_i(x) - \tilde{u}_i(x-1)$ where $\tilde{u}_i(0) := 0$ for $i = 1, \ldots, n$. The random marginal (net) utility of the x^{th} unit of alternative i is $\Delta\tilde{u}_i(x)-\tilde{u}_0$. The x^{th} unit of alternative i will not be selected if marginal utility is negative, i.e., if $\Delta\tilde{u}_i(x)-\tilde{u}_0 < 0$. Note that $\sum_{k=1}^{x}\left(\Delta\tilde{u}_i(k)-\tilde{u}_0\right)$ is the total gain in utility from each additional unit of alternative i up to x units. The threshold optimization problem can be expressed as

$$\max_{\lambda\geq 0}\left\{\sum_{i=1}^{n}\sum_{x=1}^{m}E\left[\Delta\tilde{u}_i(x)-\tilde{u}_0 \mid \Delta\tilde{u}_i(x)-\tilde{u}_0\geq\lambda\right]P\left[\Delta\tilde{u}_i(x)-\tilde{u}_0\geq\lambda\right]:\sum_{i=1}^{n}\sum_{x=1}^{m}P\left[\Delta\tilde{u}_i(x)-\tilde{u}_0\geq\lambda\right]\leq B\right\}$$

The probability that x units of alternative i are selected by a random individual in the population is

$$P\left[\tilde{x}_i=x\right]=P\left[\tilde{x}_i\geq x\right]-P\left[\tilde{x}_i\geq x+1\right]=P\left[\Delta\tilde{u}_i(x)-\tilde{u}_0\geq\lambda\right]-P\left[\Delta\tilde{u}_i(x+1)-\tilde{u}_0\geq\lambda\right]. \tag{11}$$

Empirical estimation of utility parameters requires specification of a model governing random marginal utilities. One alternative appears in Gallego and Wang (2020): For $i = 1, \ldots, n$, $x = 1, \ldots, m$,

$$\Delta\tilde{u}_i(x)-\tilde{u}_0=\Delta g_i(x)\left(\tilde{u}_i-\tilde{u}_0\right)-\Delta h_i(x)$$

$$\Delta g_i(x)=g_i(x)-g_i(x-1)$$

$$\Delta h_i(x)=h_i(x)-h_i(x-1).$$

Function $g_i(x)$ is concave increasing, function $h_i(x)$ is convex, and $g_i(0)= h_i(0) = 0$. These properties imply that marginal utility $\Delta u_i(x)-u_0$ is decreasing in quantity. Diminishing marginal utility assures that the x^{th} unit of an alternative will not be selected by an individual unless the $(x\text{-}1)^{\text{th}}$ unit is selected, e.g., $P\left[\tilde{x}_i\geq x\right]$ is decreasing in x. Threshold λ is the solution to

$$\sum_{i=1}^{n}\sum_{x=1}^{m}\mathrm{P}\left[\Delta\tilde{u}_i(x)-\tilde{u}_0\geq\lambda\right]=B;$$

$\lambda = 0$ if the constraint is nonbinding.[2] If $\tilde{\varepsilon}_i$ are iid Gumbel normalized random variables, then choice probabilities can be expressed in closed form:

$$P\left[\tilde{x}_i\geq x\right]=P\left[\tilde{\varepsilon}_0-\tilde{\varepsilon}_i\leq a_i-a_0-\frac{\lambda+\Delta h_i(x)}{\Delta g_i(x)}\right]=e^{a_i-a_0-\frac{\lambda+\Delta h_i(x)}{\Delta g_i(x)}}\left(1+e^{a_i-a_0-\frac{\lambda+\Delta h_i(x)}{\Delta g_i(x)}}\right)^{-1} \tag{12}$$

$$P\left[\tilde{\mathbf{x}}=\mathbf{x}\right]=\prod_{i=1}^{n}\left(P\left[\tilde{x}_i\geq x_i\right]-P\left[\tilde{x}_i\geq x_i+1\right]\right). \tag{13}$$

The simplicity of (12) motivates an alternative to MLE based on (13): each observed event $\tilde{x}_i = x$ for $x \geq 1$ in the data generates events $\tilde{x}_i\geq 1$, …, $\tilde{x}_i\geq x$ in an augmented dataset. While the form and parameters of

---

[2] Gallego and Wang (2020, Theorem 6) consider the generalized threshold optimization problem of maximizing over $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)$ and show the optimal solution is a scalar for their model.

functions $\Delta g_i(x)$ and $\Delta h_i(x)$ must be specified, the choice probably model can be efficiently estimated from the augmented dataset due to the simple binary logit form in (12).

In summary, the formulation yielding choice probabilities given by (11) defines a class of models for multiple discrete choice and discrete quantity in a manner consistent with random utility maximization. Specific models within this class depend upon the model primitives that govern that random marginal utility of each alternative *i* and quantity x_i, i.e., $\Delta\tilde{u}_i(x)-\tilde{u}_0$. One subclass of (11) specifies functions that map a random variable for net utility of each alternative to marginal utility at different values of x_i, e.g., as in (13). An alternative subclass specifies probability distribution functions for random variables, $\Delta\tilde{u}_i(x)-\tilde{u}_0$, $i = 1, \ldots, n$ and $x = 1, \ldots, m$. In the following sections, I introduce such a model and derive choice probability expressions. The model relies on order statistics, which capture the property of diminishing marginal return to quantity for each alternative.

4. Diminishing Marginal Utility as Order Statistics

The *law of diminishing marginal utility*, as originally proposed by Marshall (1920), states that the gain in utility from each unit of consumption decreases in quantity. This feature can be captured through order statistics, as described and illustrated in (14) below.

For $x_i \in \{1, \ldots, m\}$ units of alternative $i \in \{1, \ldots, n\}$, let

$$\tilde{u}_i(x_i) = a_i x_i + \sum_{j=1}^{x_i} \tilde{\varepsilon}_i^{(j)}$$

$$\Delta\tilde{u}_i(x_i) = a_i + \tilde{\varepsilon}_i^{(x_i)}$$

$$\tilde{u}_0 = a_0 + \tilde{\varepsilon}_0$$

where $\tilde{\varepsilon}_i^{(1)} = \max\{\tilde{\varepsilon}_{i1},\ldots,\tilde{\varepsilon}_{im}\}$, …, $\tilde{\varepsilon}_i^{(m-1)} = \max\left\{\{\tilde{\varepsilon}_{i1},\ldots,\tilde{\varepsilon}_{im}\} \setminus \left\{\tilde{\varepsilon}_i^{(1)},\ldots,\tilde{\varepsilon}_i^{(m-2)}\right\}\right\}$, $\tilde{\varepsilon}_i^{(m)} = \min\{\tilde{\varepsilon}_{i1},\ldots,\tilde{\varepsilon}_{im}\}$

and $\tilde{\varepsilon}_0$, $\tilde{\varepsilon}_{ij}$ for $i = 1, \ldots, n$, $j = 1, \ldots, m$ are iid random variables.[3] During a choice event, an individual observes the realizations of random variables $\tilde{u}_0$ and $\Delta\tilde{u}_i(x_i)$ for $i = 1, \ldots, n$ and $x_i = 1, \ldots, m$, then selects quantities of alternatives for which marginal utility exceeds threshold λ. While consecutive marginal utilities of an alternative exhibit randomness over choice events, realizations exhibit diminishing marginal utility, i.e.,

$$\Delta\tilde{u}_i(1) \geq \ldots \geq \Delta\tilde{u}_i(m). \tag{14}$$

[3] This indexing of largest-to-smallest is the reverse of the standard order-statistic convention of smallest-to-largest, but is more convenient for our setting.

The specific character of random diminishing marginal utility depends on the probability distribution of error terms $\tilde{\varepsilon}_0$, $\tilde{\varepsilon}_{ij}$.

This paper presents results for the case where $\tilde{\varepsilon}_0$ and $\tilde{\varepsilon}_{ij}$ are Gumbel random variables, which I refer to as OS-G model (order statistics – Gumbel). The location parameter is normalized to 0 without loss of generality. The scale parameter is β. Figure 1 illustrates expected marginal net utility for the OS-G model at several values of β as quantity x_i ranges between 1 and 25.

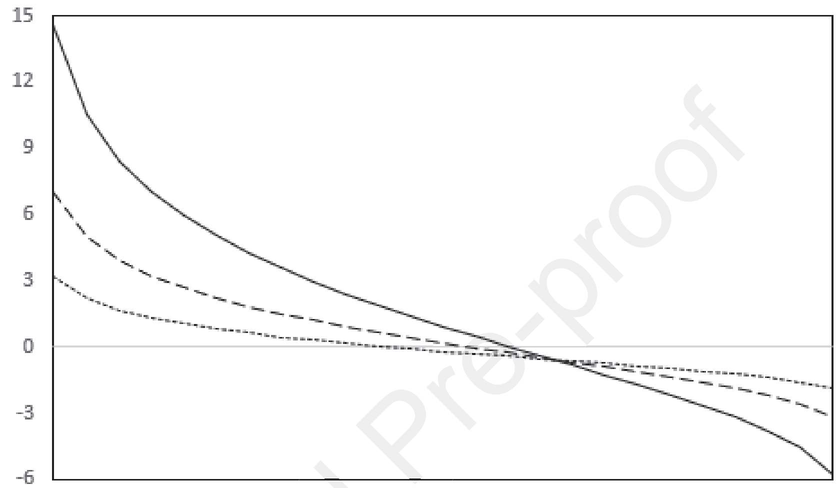


Figure 1. Plot of $E\left[\Delta\tilde{u}_i\left(x_i\right)-\tilde{u}_0\right]$ with $m = 25$, $a_0 = a_i = 0$, and $\beta = 1$ (dotted curve), $\beta = 2$ (dashed curve), $\beta = 4$ (solid curve). The x-axis ranges from $x_i = 1$ to $x_i = 25$. Expressions for $E\left[\Delta\tilde{u}_i\left(x_i\right)-\tilde{u}_0\right]$ appear in Lemma 6.

5. Existing Theory

5.1. Gumbel Distribution

The pdf and cdf of a Gumbel random variable $\tilde{z}$ with location parameter ν and scale parameter β are

$$f(z) = \frac{e^{-(z-\nu)/\beta}}{\beta} e^{-e^{-(z-\nu)/\beta}}, \; z \in (-\infty, \infty) \tag{15}$$

$$F(z) = e^{-e^{-(z-\nu)/\beta}}, \; z \in (-\infty, \infty) \tag{16}$$

with mean and variance

$$E\left[\tilde{z}\right] = \nu + \gamma\beta \text{ where } \gamma = -\int_0^\infty e^{-t} \ln t dt \approx 0.577 \text{ is the Euler-Mascheroni constant} \tag{17}$$

$$V\left[\tilde{z}\right] = \pi^2\beta^2 / 6.$$

Lemma 1 (Gumbel 1954). *The Gumbel distribution is closed under maximization. That is, for independent Gumbel random variables* $\tilde{z}_1, \ldots, \tilde{z}_m$ *with scale parameter* β *and location parameters* $\nu_1, \ldots,$

v_m, $\tilde{z}^{(1)} = \max\{\tilde{z}_1, ..., \tilde{z}_m\}$ *is a Gumbel random variable with scale parameter* β *and location parameter* $v = \beta \ln \sum_{i=1}^{m} e^{v_i/\beta}$.

Corollary 1. $\tilde{\varepsilon}_i^{(1)} = \max\{\tilde{\varepsilon}_{i1}, ..., \tilde{\varepsilon}_{im}\}$ *is a Gumbel random variable with scale parameter* β *and location parameter* $\beta \ln m$.

Lemma 2 (Train 2009, p. 35). *The difference between two independent Gumbel random variables with the same scale parameter is a logistic random variable. That is, for Gumbel* $\tilde{z}_1$ *and* $\tilde{z}_2$ *with scale parameter* β *and location parameters* v_1 *and* v_2, $\tilde{z} = \tilde{z}_2 - \tilde{z}_1$ *is logistic with mean* $v_2 - v_1$, *variance* $\pi^2 \beta^2 / 3$, *and cdf*

$$F_{\tilde{z}}(z) = \frac{1}{1 + e^{-(z-(v_2 - v_1))/\beta}}.$$

5.2. Order Statistic Distribution Functions

Let $\tilde{z}_1, ..., \tilde{z}_m$ denote iid continuous random variables with pdf f and cdf F. Let $\tilde{z}^{(1)} = \max\{\tilde{z}_1, ..., \tilde{z}_m\}$, $\tilde{z}^{(2)} = \max\left\{\{\tilde{z}_1, ..., \tilde{z}_m\} \setminus \{\tilde{z}^{(1)}\}\right\}$, $\tilde{z}^{(3)} = \max\left\{\{\tilde{z}_1, ..., \tilde{z}_m\} \setminus \{\tilde{z}^{(1)}, \tilde{z}^{(2)}\}\right\}$, …, $\tilde{z}^{(m)} = \min\{\tilde{z}_1, ..., \tilde{z}_m\}$. Then

$$f_{\tilde{z}^{(x)}}(z) = \frac{m!}{(x-1)!(m-x)!} F(z)^{m-x} (1 - F(z))^{x-1} f(z) \tag{18}$$

$$F_{\tilde{z}^{(x)}}(z) = \sum_{j=0}^{x-1} \binom{m}{j} F(z)^{m-j} (1 - F(z))^{j} \tag{19}$$

$$f_{\tilde{z}^{(1)}, ..., \tilde{z}^{(x)}}(\mathbf{z}) = \frac{m!}{(m-x)!} F(z_x)^{m-x} \prod_{j=1}^{x} f(z_j), \quad z_1 \geq z_2 \geq \ldots \geq z_x,\ x \leq n \tag{20}$$

(see Chapter 2 in David 1981). Expressions (18) and (19) show how the probability density function and the cumulative distribution function of an order statistic $\tilde{z}^{(x)}$ relate to the corresponding functions of random variable $\tilde{z}_i$. Expression (20) shows the joint density for the x^{th} largest order statistics. These results are used in the next section.

6. Theory Development

To simplify presentation, parameters a_0 and λ are normalized to zero, e.g., by redefining $a_i = a_i - a_0 - \lambda$ for all i. Thus

$$P[\tilde{x}_i \geq x] = P[\Delta\tilde{u}_i(x) - \tilde{u}_0 \geq \lambda] = P\left[\tilde{\varepsilon}_0 - \tilde{\varepsilon}_i^{(x)} \leq a_i\right], \quad i = 1, \ldots, n,\ x = 1, \ldots, m. \tag{21}$$

Define

$$A_i = e^{a_i/\beta} \text{ for } i = 0, 1, \ldots, n$$

(e.g., $A_0 = 1$). Corollary 2 follows from (15) – (20).

Corollary 2. *For* $x \in \{1, \ldots, m\}$,

$$F_{\Delta\tilde{u}_0}(t) = e^{-e^{-t/\beta}}$$

$$f_{\Delta\tilde{u}_0}(t) = \frac{e^{-t/\beta}}{\beta} e^{-e^{-t/\beta}}$$

$$F_{\Delta\tilde{u}_i(x)}(t) = \sum_{j=0}^{x-1} \binom{m}{j} e^{-(m-j)A_i e^{-t/\beta}} \left(1 - e^{-A_i e^{-t/\beta}}\right)^j$$

$$f_{\Delta\tilde{u}_i(x)}(t) = x\binom{m}{x} \frac{A_i e^{-t/\beta}}{\beta} e^{-(m-x+1)A_i e^{-t/\beta}} \left(1 - e^{-A_i e^{-t/\beta}}\right)^{x-1}$$

$$f_{\Delta\tilde{u}_i(1),\ldots,\Delta\tilde{u}_i(x)}(\mathbf{t}) = \frac{m!}{(m-x)!}\left(\frac{A_i}{\beta}\right)^x e^{-\sum_{j=1}^{x} t_j/\beta} e^{-A_i\left(\sum_{j=1}^{x} e^{-t_j/\beta} + (m-x)e^{-t_x/\beta}\right)}, \; t_1 \geq \ldots \geq t_x.$$

Lemmas 3 – 5 provide the foundation for theorems 1 through 3 that pertain to the probability distribution of random choice vector $\tilde{\mathbf{x}} = (\tilde{x}_1, \ldots, \tilde{x}_n)$. Lemma 6 describes the expected value of marginal utility.

Lemma 3. *For any* $i \in \{1, \ldots, n\}$,

$$P[\tilde{x}_i \geq 1] = \frac{mA_i}{1 + mA_i} \tag{22}$$

$$P[\tilde{x}_i = 0] = \frac{1}{1 + mA_i}. \tag{23}$$

Lemma 4. *Let* $\Delta\tilde{u}_0(1) = \tilde{u}_0$ *in* (25) *and* (27) *below. For any* $i \in \{0, 1, \ldots, n\}$ *and* t,

$$P\left[\Delta\tilde{u}_i(x) > t \geq \Delta\tilde{u}_i(x+1)\right] = \binom{m}{x} e^{-(m-x)A_i e^{-t/\beta}} \left(1 - e^{-A_i e^{-t/\beta}}\right)^x, \; x \in \{1, \ldots, m\} \tag{24}$$

$$P\left[\Delta\tilde{u}_i(1) \leq t\right] = e^{-mA_i e^{-t/\beta}} \tag{25}$$

$$P\left[\Delta\tilde{u}_i(m) > t\right] = P\left[\Delta\tilde{u}_i(m) > t \geq \Delta\tilde{u}_i(m+1)\right] = \left(1 - e^{-A_i e^{-t/\beta}}\right)^m \tag{26}$$

$$P\left[\Delta\tilde{u}_i(1) > t\right] = 1 - e^{-mA_i e^{-t/\beta}}. \tag{27}$$

Lemma 5. *For a, b* > 0 *and nonnegative integers* $m \geq x$,

$$(m-x)\binom{m}{x}\sum_{j=0}^{x}\binom{x}{j}(-1)^j\left(\frac{b}{a+(m-x+j)b}\right) = \prod_{j=0}^{x}\frac{(m-j)b}{a+(m-j)b}. \tag{28}$$

Lemma 6. *For* $x = 1, \ldots, m$,

$$E\left[\Delta\tilde{u}_i(x)\right] = a_i + \beta x \binom{m}{x} \sum_{j=0}^{x-1} \binom{x-1}{j} (-1)^j \left(\frac{\gamma + \ln(m-x+1+j)}{m-x+1+j} \right) \tag{29}$$

Theorem 1. *Marginal probability functions are*

$$P\left[\tilde{x}_i \geq x\right] = \prod_{j=0}^{x-1} \frac{(m-j)A_i}{1+(m-j)A_i}, \; x \in \{1, \ldots, m\} \tag{30}$$

$$P\left[\tilde{x}_i \geq x+1 \mid \tilde{x}_i \geq x\right] = \frac{(m-x)A_i}{1+(m-x)A_i}, \; x \in \{1, \ldots, m-1\} \tag{31}$$

$$P\left[\tilde{x}_i = x \mid \tilde{x}_i \geq x\right] = \frac{1}{1+(m-x)A_i}, \; x \in \{1, \ldots, m\} \tag{32}$$

$$P\left[\tilde{x}_i = x\right] = \left(\frac{1}{1+(m-x)A_i} \right) \prod_{j=0}^{x-1} \frac{(m-j)A_i}{1+(m-j)A_i}, \; x \in \{0, \ldots, m\} \tag{33}$$

$$= \left(\frac{A_i + (m-x)A_i}{1+(m-x)A_i} \right) P\left[\tilde{x}_i = x-1\right], \; x \in \{1, \ldots, m\}. \tag{34}$$

Notice in (34) that the probability that x units is selected is proportional to the probability that $x-1$ units is selected. Whether the probability is increasing in x or decreasing in x depends on the value of A_i, as formalized in the following corollary.

Corollary 3. *If $A_i = 1$, then $P\left[\tilde{x}_i = x\right] = \frac{1}{1+m}$ for all $x \in \{0, \ldots, m\}$ and $E\left[\tilde{x}_i\right] = \frac{m}{2}$. If $A_i < 1$, then $P\left[\tilde{x}_i = 0\right] > \ldots > P\left[\tilde{x}_i = m\right]$ and $E\left[\tilde{x}_i\right] < \frac{m}{2}$. If $A_i > 1$, then $P\left[\tilde{x}_i = 0\right] < \ldots < P\left[\tilde{x}_i = m\right]$ and $E\left[\tilde{x}_i\right] > \frac{m}{2}$.*

Notice that (23) and (33) recover the binary logit choice probabilities when $m = 1$:

$$P\left[\tilde{x}_i = 0\right] = \frac{1}{1+e^{a_i/\beta}}$$

$$P\left[\tilde{x}_i = 1\right] = \frac{e^{a_i/\beta}}{1+e^{a_i/\beta}}.$$

(e.g., see (9)). Furthermore, if $n = 1$ and $m > 1$, so that an individual selects a quantity (possibly zero) up to m of a single alternative, then the probability distribution of random choice $\tilde{x}$ is fully specified by (23) and (33):

$$P\left[\tilde{x} = 0\right] = \frac{1}{1+mA_1}$$

$$P\left[\tilde{x} = 1\right] = \left(\frac{mA_1}{1+(m-1)A_1} \right) P\left[\tilde{x} = 0\right]$$

$$P\left[\tilde{x}=2\right]=\left(\frac{(m-1)A_1}{1+(m-2)A_1}\right)P\left[\tilde{x}=1\right]$$

$$\vdots$$

$$P\left[\tilde{x}=m-1\right]=\left(\frac{2A_1}{1+A_1}\right)P\left[\tilde{x}=m-2\right]$$

$$P\left[\tilde{x}=m\right]=A_1P\left[\tilde{x}=m-1\right]$$

Finally, observe that (31) in Theorem 1 hints at a rather remarkable result on the character of the probability function of conditional Gumbel order statistics, which I state below as a theorem. The theorem illuminates a simple structure underlying the probability distribution of the difference between Gumbel order statistics.

Theorem 2. *Let $\tilde{z}_0, \tilde{z}_1, ..., \tilde{z}_m$ be independent Gumbel random variables with the same scale parameter β and location parameter ν_i with $\nu_i = \nu$ for $i \geq 1$. Let $\tilde{z}^{(x:m)}$ denote the x^{th} largest value in m-dimensional random vector $\left(\tilde{z}_1, ..., \tilde{z}_m\right)$. Define $\tilde{y}^{(x:m)} = \tilde{z}_0 - \tilde{z}^{(x:m)}$ for $x \in \{1, \ldots, m\}$ and let $\tilde{y}^{(x+1:m)|(x:m)}$ denote a conditional random variable with probability distribution $P\left[\tilde{y}^{(x+1:m)|(x:m)} \leq t\right] = P\left[\tilde{y}^{(x+1:m)} \leq t \mid \tilde{y}^{(x:m)} \leq t\right]$ for $x \in \{1, \ldots, m-1\}$. Then $\tilde{y}^{(x+1:m)|(x:m)}$ is a logistic random variable with mean $\nu_0 - \nu^{(m\text{-x})}$ where $\nu^{(m\text{-x})} = [\nu + \beta\ln(m-x)]$ is the location parameter of $\tilde{z}^{(1:m-x)}$, variance $\pi^2\beta^2/3$, and cdf*

$$P\left[\tilde{y}^{(x+1:m)|(x:m)} \leq t\right] = P\left[\tilde{y}^{(1:m-x)} \leq t\right] = P\left[\tilde{z}_0 - \tilde{z}^{(1:m-x)} \leq t\right] = \frac{1}{1+e^{-\left(t-\left(\nu_0-\nu^{(m-x)}\right)\right)/\beta}}. \tag{35}$$

Theorem 2 generalizes Lemma 2 to non-extreme, but conditional, order statistics. For example, at x = 0, then $\tilde{y}^{(0:m)}$ doesn't exist, and (35) becomes

$$P\left[\tilde{z}_0 - \tilde{z}^{(1:m)} \leq t\right] = \frac{1}{1+e^{-\left(t-\left(\nu_0-\nu^{(m)}\right)\right)/\beta}},$$

which is the result in Lemma 2. While $\tilde{z}^{(x:m)}$ is not a Gumbel random variable when $x > 1$, it retains the Gumbel character when appropriately conditioned on the next largest order statistic. The effect of conditioning is akin to the creation of an unconditional Gumbel obtained from the maximum of a smaller set of iid Gumbels. This special structure is key to the relatively simple and recursive expressions that appear in Theorem 1. A manifestation of this special structure inherent to Gumbel random variables appears in Beggs et al. (1981). They show that the probability of an ordering of the largest k of $n > k$ independent Gumbel random variables is independent of the $n-k$ smallest random variables. The next result builds on theorems 1 and 2 and presents the choice probability functions.

Theorem 3. *The probability mass function of* $\tilde{\mathbf{x}}$ *is*

$$P\left[\tilde{\mathbf{x}}=\mathbf{x}\right]=\left(\prod_{i=1}^{n}\binom{m}{x_i}\right)\sum_{j_1=0}^{x_1}\cdots\sum_{j_n=0}^{x_n}\left(\frac{\binom{x_1}{j_1}\cdots\binom{x_n}{j_n}(-1)^{\left(\sum_{i=1}^{n}j_i\right)}}{1+\sum_{i=1}^{n}\left(m-x_i+j_i\right)A_i}\right), \mathbf{x}\in\{0,\ldots,m\}^n. \tag{36}$$

or in vector notation,

$$P\left[\tilde{\mathbf{x}}=\mathbf{x}\right]=\left(\prod_{i=1}^{m}\binom{m}{x_i}\right)\vec{C}_2\left(\mathbf{x}\right)^T\vec{\alpha}_2\left(\mathbf{x}\right) \tag{37}$$

where $I(\bullet)$ *is an indicator function returning* 1 *if condition* $\bullet$ *holds and*

$$\vec{C}\left(\mathbf{x}\right)=\left(\left(\prod_{i=1}^{n}\binom{x_i}{j_i}I\left(j_i\le x_i\right)\right)(-1)^{\sum_{i=1}^{n}j_i}\right)_{\mathbf{j}\in\{0,\ldots,m\}^n}$$

$$\vec{\alpha}\left(\mathbf{x}\right)=\left(\frac{1}{1+\sum_{i=1}^{n}\left(m-x_i+j_i\right)A_i}\right)_{\mathbf{j}\in\{0,\ldots,m\}^n}.$$

The single-choice and null-choice probabilities exhibit a particularly simple structure. For $\mathbf{x} = (x,0,\ldots,0)$ with $x\in\{1,\ldots,m\}$, I apply the identity in Lemma 5, and (36) simplifies to

$$P\left[\tilde{\mathbf{x}}=\left(x,0,\ldots,0\right)\right]=\left(\frac{1}{1+m\sum_{i=2}^{n}A_i+\left(m-x\right)A_1}\right)\prod_{j=0}^{x-1}\frac{\left(m-j\right)A_1}{1+m\sum_{i=2}^{n}A_i+\left(m-j\right)A_1}, \tag{38}$$

which follows the marginal probability structure in (33). If $\mathbf{x} = \mathbf{0}$, then (36) simplifies to

$$P\left[\tilde{\mathbf{x}}=\mathbf{0}\right]=\frac{1}{1+m\sum_{i=1}^{n}A_i}, \tag{39}$$

which recovers the MNL null-choice probability.

7. Parameter Estimation

Without loss of generality, the scale parameter is normalized to $\beta = 1$, e.g., $A_i = e^{\hat{a}_i/\beta} = e^{a_i}$.[4] Estimation of parameters of the OS-G model will rely on an estimate of m, which represents an upper limit on the number of units selected of an alternative. The selection of this value may be informed by some

[4] Parameter a_i can be interpreted as the nominal marginal utility of alternative i, which may be modeled as a linear combination of alternative i attributes/characteristics $\overline{\mathbf{a}}_i=\left(\overline{a}_{i1},\ldots,\overline{a}_{iM}\right)$ and a vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_M)$ that reflects characteristics of the population e.g., $a_i=\boldsymbol{\alpha}^T\overline{\mathbf{a}}_i$.

combination of judgement, observed choice quantities in the data, and computational considerations, e.g., the complexity of probability expressions increases in m.

For a given value of m, one may estimate vector $\mathbf{A} = (A_1, \ldots, A_n)$ by maximizing the log-likelihood function (i.e., MLE) of observations $\mathbf{x}^1, \ldots, \mathbf{x}^N$:

$$\hat{\mathbf{A}} = \arg\max_{\mathbf{Y}} \ln\left(\prod_{k=1}^{N} P\left[\tilde{\mathbf{x}} = \mathbf{x}^k \mid \mathbf{A} = \mathbf{Y}\right]\right) = \arg\max_{\mathbf{Y}} \left(\sum_{k=1}^{N} \ln\left(\left(\prod_{i=1}^{m} \binom{m}{x_i^k}\right) \vec{C}\left(\mathbf{x}^k\right)^T \vec{\alpha}\left(\mathbf{x}^k \mid \mathbf{Y}\right)\right)\right)$$

where

$$\vec{\alpha}\left(\mathbf{x}^k \mid \mathbf{Y}\right) = \left(\frac{1}{1 + \sum_{i=1}^{n}\left(m - x_i^k + j_i\right) Y_i}\right)_{\mathbf{j} \in \{0,\ldots,m\}^n} .$$

An alternative estimation method that provides less theoretical precision (because it ignores some information) but is computationally more efficient is MLE using the marginal probability functions instead of the joint probability functions:

$$\hat{\mathbf{A}} = \arg\max_{\mathbf{Y}} \ln\left(\prod_{i=1}^{n}\prod_{k=1}^{N} P\left[\tilde{x}_i = x_i^k \mid A_i = Y_i\right]\right) = \arg\max_{\mathbf{Y}} \sum_{i=1}^{n}\sum_{k=1}^{N} \ln\left(P\left[\tilde{x}_i = x_i^k \mid A_i = Y_i\right]\right)$$

$$= \arg\max_{\mathbf{Y}} \sum_{i=1}^{n}\sum_{k=1}^{N} \ln\left(\left(\frac{1}{1 + \left(m - x_i^k\right) Y_i}\right) \prod_{j=0}^{x_i^k - 1} \frac{(m-j) Y_i}{1 + (m-j) Y_i}\right)$$

$$= \arg\max_{\mathbf{Y}} \sum_{i=1}^{n}\sum_{k=1}^{N} \left(\ln\left(\frac{1}{1 + \left(m - x_i^k\right) Y_i}\right) + \sum_{j=0}^{x_i^k - 1} \ln\left(\frac{(m-j) Y_i}{1 + (m-j) Y_i}\right)\right).$$

With this approach, the parameter for each alternative can be estimated independently, i.e., for $i = 1, \ldots, n$,

$$\hat{A}_i = \arg\max_{Y_i} \sum_{k=1}^{N} \left(\ln\left(\frac{1}{1 + \left(m - x_i^k\right) Y_i}\right) + \sum_{j=0}^{x_i^k - 1} \ln\left(\frac{(m-j) Y_i}{1 + (m-j) Y_i}\right)\right).$$

I run a small numerical experiment to provide a sense of how the two methods compare in terms of accuracy. I randomly generate individual choices according to the OS-G model with $m = 3$, $n = 3$, and true parameters $A_1 = 7 \approx e^2$ $A_2 = 1 = e^0$, $A_3 = 0.13 \approx e^{-2}$, and $a_0 = \lambda = 0$. I set the sample size to $N = 1{,}000$ and run 30 replications. Table 1 compares the mean absolute percentage error (MAPE) of MLE parameter estimates $\hat{A}_i$ for the two methods. The results illustrate the decrease in precision from ignoring correlation in quantity choices of individuals in a population. In this example, the reduction in accuracy of parameter estimates is relatively small.

		Sample Size = 1000 Choices			
		Method 1		Method 2	
	True Value	Mean MLE Estimate	MAPE (Std Err)	Mean MLE Estimate	MAPE (Std Err)
A_1	7.00	6.942	5.21% (0.61%)	6.874	6.11% (0.69%)
A_2	1.00	1.004	4.21% (0.61%)	1.004	4.23% (0.61%)
A_3	0.13	0.132	5.25% (0.70%)	0.132	5.21% (0.70%)

Table 1. Mean parameter estimate and mean absolute percentage error over 30 replications with sample size $N = 1{,}000$ using the probability mass function (Method 1) and the marginal probability function (Method 2). Standard errors are in parentheses.

8. Normative Analysis

Recall that B is the average number of units selected during a choice event by individuals in the population, which may be estimated from observed choices. The value of B plays a role when measuring how choice probabilities change as features of alternatives change. If the value of B remains stable (or can be predicted) as changes in alternatives are introduced, then the effects of changes in alternatives on choice probabilities can be measured (e.g., in a manner similar to Section 3.2). I illustrate this relationship for the OS-G model below, after presenting Theorem 4 on optimal thresholds for the general OS model. Following Gallego and Wang (2020), it is straightforward to show that the solution to the threshold optimization problem with marginal utilities as order statistics of iid error terms is a scalar.

Theorem 4. *OS model:* $\tilde{u}_0 = a_0 + \tilde{\varepsilon}_0$ *and* $\Delta\tilde{u}_i(x_i) = a_i + \tilde{\varepsilon}_i^{(x_i)}$ *where* $\tilde{\varepsilon}_0, \tilde{\varepsilon}_{ij}$ *are iid continuous random variables. For the OS model, the solution to the threshold optimization problem*

$$\max_{\boldsymbol{\lambda}\geq\mathbf{0}}\left\{\sum_{i=1}^{n}\sum_{x=1}^{m}E\left[\Delta\tilde{u}_i(x)-\tilde{u}_0 \mid \Delta\tilde{u}_i(x)-\tilde{u}_0\geq\lambda_i\right]P\left[\Delta\tilde{u}_i(x)-\tilde{u}_0\geq\lambda_i\right]:\sum_{i=1}^{n}\sum_{x=1}^{m}P\left[\Delta\tilde{u}_i(x)-\tilde{u}_0\geq\lambda_i\right]\leq B\right\}$$

is the unique solution to

$$\sum_{i=1}^{n}\sum_{x=1}^{m}P\left[\Delta\tilde{u}_i(x)-\tilde{u}_0\geq\lambda\right]=B \tag{40}$$

if a solution exists; *otherwise* $\lambda = 0$.

Assume that $\tilde{\varepsilon}_0, \tilde{\varepsilon}_{ij}$ are iid normalized Gumbel random variables (OS-G model). As in Section 3.2, let $\bar{\mathbf{a}}_i = (\bar{a}_{i1},...,\bar{a}_{iM})$ denote the vector of predictors for alternative i, let $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_M)$ denote the utility coefficient vector estimated from the data (e.g., $\boldsymbol{\alpha}^T\bar{\mathbf{a}}_i$ is an estimate of a_i), and let Δ_{il} denote a change in attribute l of alternative i. Prior to the change in the attribute,

$$P\left[\tilde{x}_j\geq x\right]=\prod_{k=0}^{x-1}\frac{(m-k)e^{\boldsymbol{\alpha}^T\bar{\mathbf{a}}_j}}{1+(m-k)e^{\boldsymbol{\alpha}^T\bar{\mathbf{a}}_j}},\ j = 1, \ldots, n,\ x = 1, \ldots, m$$

(see (21) and (30)), and after the change in attribute l of alternative i to $\bar{a}_{il} + \Delta_{il}$,

$$P\left[\tilde{x}_j \geq x\right] = \prod_{k=0}^{x-1} \frac{(m-k)e^{\boldsymbol{\alpha}^T\bar{\mathbf{a}}_j + \alpha_l \Delta_{il} I(j=i)}}{e^{\Delta_\lambda} + (m-k)e^{\boldsymbol{\alpha}^T\bar{\mathbf{a}}_j + \alpha_l \Delta_{il} I(j=i)}}$$

where Δ_λ is the solution to

$$\sum_{j=1}^{n}\sum_{x=1}^{m}\prod_{k=0}^{x-1} \frac{(m-k)e^{\boldsymbol{\alpha}^T\bar{\mathbf{a}}_j + \alpha_l \Delta_{il} I(j=i)}}{e^{\Delta_\lambda} + (m-k)e^{\boldsymbol{\alpha}^T\bar{\mathbf{a}}_j + \alpha_l \Delta_{il} I(j=i)}} = B. \tag{41}$$

The value of Δ_λ can be efficiently obtained via bisection search because the left-hand side of (41) is decreasing in Δ_λ.

9. Summary and Reflection

This paper presents a model of discrete quantity decisions over a set of alternatives that is consistent with random utility maximization. The model employs order statistics of Gumbel random variables to capture individuals' idiosyncratic and diminishing marginal utility to consumption. The underlying mathematical structure admits closed-form marginal quantity-choice probability functions and closed-form joint probability functions.

An important question is whether, or under what settings, the model may capture the essence of real-world behavior . . . whether the model can provide useful predictions or expose new insights for guiding decisions in public- or private-sector settings. I suspect that the range of meaningful application is narrow, and one reason stems from a property of the probability distribution of Gumbel order statistics identified in Theorem 1. One manifestation of this property is the probability mass function of quantity choice of each alternative is monotonic (see Corollary 3); as A_i increases from 0, the distribution of mass shifts from an extreme right skew (all mass at 0) to uniform at $A_i = 1$, to extreme left skew with all mass at m as A_i approaches infinity. The model restricts the mode of the quantity-choice distribution to be the extreme left, the extreme right, or the entire sample space (in the case of $A_i = 1$), a property that may not fit with reality in some settings. This limitation is illustrated in Figure 1 that shows how the expected value of marginal utility changes in quantity. The form of the curves that are initially convex decreasing and later concave decreasing may align with behavior in some settings but not in others.

Interestingly, the monotonic feature of the probability distribution is characteristic of a power-law relation where probability is scale-invariant and proportional to x^{-k}. For example, the continuous analog of the probability distribution in Theorem 1 has density proportional to $(1 + m - x)^k$ where $k >$ (<) 0 if $A_i <$ (>) 1 (see the appendix). A wide variety of phenomena (physical, biological, man-made) exhibit power-law relations.

An additional potential weakness of the model is the form of the constraint in an individual's choice optimization problem – an individual maximizes expected utility for a given average purchase quantity at each choice event. A key advantage is that, under this constraint, an individual's optimization problem at each choice event is simple: choose alternatives with the highest utility that exceed a given threshold. An alternative formulation is to maximize expected utility for a given average spend amount at each choice event. This formulation translates to a more complex (i.e., combinatorial) decision problem for an individual, e.g., a simple rule of choosing alternatives with the highest utility per dollar that exceed a given threshold is not necessarily optimal. Consideration of this, or other, alternative formulations is one avenue for future research.

In sum, the main value of this paper is perhaps less in the OS-G model of multiple discrete choice—for which application may be narrow—and more in lemmas 3 – 6, theorems 1 – 3, and corollaries 2 – 3. These lemmas, theorems, and corollaries present probability distribution functions and properties related to the difference of a pair of Gumbel order statistics. Order statistics arise in a variety of real-world phenomena that extend beyond choice decisions. Thus, the results presented in Section 6 have potential to spur new applications and research on order-statistic-based models.

Acknowledgement

I thank and recognize the contributions of the editor and the entire review team. Their insightful comments and suggestions have significantly improved the paper.

10. References

Beggs S, Cardell S, Hausman J (1981) Assessing the potential demand for electric cars. *Journal of Econometrics* 17(1):1-19.

Bhat CR (2005) A multiple discrete–continuous extreme value model: formulation and application to discretionary time-use decisions. *Transportation Research Part B* 39:679-707.

Bhat CR, Mondal A, Asmussen K, Bhat AC (2020) A multiple discrete extreme value choice model with grouped consumption data and unobserved budgets. *Transportation Research Part B* to appear.

David HA (1981) *Order Statistics*, 2nd edition, John Wiley & Sons, New York, NY.

Gallego G, Wang R (2020) Threshold utility model with applications to retailing and discrete choice models. Working paper: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3420155

Gorman WM (1959) Separable utility and aggregation. *Econometrica* 27(3):469-481.

Gumbel EJ (1954) *Statistical Theory of Extreme Values and Some Practical Applications*, U.S. Department of Commerce, National Bureau of Standards.

Hausman JA, Leonard GK, McFadden D (1995) A utility-consistent, combined discrete choice and count data model. Assessing recreational use losses due to natural resource damage. *Journal of Public Economics* 56(1):1-30.

Marshall A (1920) *Principles of Economics* (revised edition). Macmillan, London (reprinted by Prometheus Books).

McFadden D (1978) Modeling the choice of residential location. Cowles Foundation Discussion Paper No. 477.

Palma D, Hess S (2020) Some adaptations of Multiple Discrete-Continuous Extreme Value (MDCEV) models for a computationally tractable treatment of complementarity and substitution effects, and reduced influence of budget assumptions. Working paper: http://www.stephanehess.me.uk/papers/working_papers/Palma_Hess_2020.pdf

Saxena S, Pinjari AR, Paleti R (2020) An ordered multiple discrete-continuous extreme value (OMDCEV) modelling framework for episode-level activity participation and time-use analysis. Working paper: https://abdulpinjari.weebly.com/uploads/9/6/7/8/9678119/saxena_pinjari_paleti_15april2020.pdf

Strotz RH (1957) The empirical implications of a utility tree. *Econometrica* 25(2): 269-280.

Train KE (2009) *Discrete Choice Methods with Simulation*. Cambridge University Press, New York, NY.

11. Appendix

11.1. Proofs

Proof of Corollary 2. To simplify notation, $f(t)=f_{\Delta\tilde{u}_0}(t)$, $F(t)=F_{\Delta\tilde{u}_0}(t)$, $f_i(t)=f_{a_i+\tilde{\varepsilon}_{ij}}(t)$, and $F_i(t)=F_{a_i+\tilde{\varepsilon}_{ij}}(t)$. From (15) and (16),

$$F(t)=e^{-e^{-t/\beta}} \qquad f(t)=\frac{e^{-t/\beta}}{\beta}e^{-e^{-t/\beta}}$$

$$F_i(t)=e^{-A_i e^{-t/\beta}} \qquad f_i(t)=\frac{A_i e^{-t/\beta}}{\beta}e^{-A_i e^{-t/\beta}}$$

and from (15) – (20),

$$f_{\Delta\tilde{u}_i(x)}(t)=\frac{m!}{(x-1)!(m-x)!}\frac{A_i e^{-t/\beta}}{\beta}e^{-(m-x+1)A_i e^{-t/\beta}}\left(1-e^{-A_i e^{-t/\beta}}\right)^{x-1}$$

$$F_{\Delta\tilde{u}_i(x)}(t)=\sum_{j=0}^{x-1}\binom{m}{j}e^{-(m-j)A_i e^{-t/\beta}}\left(1-e^{-A_i e^{-t/\beta}}\right)^{j}$$

$$\begin{aligned}
f_{\Delta\tilde{u}_i(1),\ldots,\Delta\tilde{u}_i(x)}(\mathbf{t}) &= \frac{m!}{(m-x)!}F(t_x)^{m-x}\prod_{j=1}^{x}f(t_j) \\
&= \frac{m!}{(m-x)!}\left(e^{-A_i e^{-t_k/\beta}}\right)^{m-x}\left(\frac{A_i e^{-t_1/\beta}}{\beta}e^{-A_i e^{-t_1/\beta}}\right)\times\cdots\times\left(\frac{A_i e^{-t_x/\beta}}{\beta}e^{-A_i e^{-t_x/\beta}}\right) \\
&= \frac{m!}{(m-x)!}\left(e^{-A_i\left(\sum_{j=1}^{x}e^{-t_j/\beta}+(m-x)e^{-t_x/\beta}\right)}\right)\left(\frac{A_i e^{-t_1/\beta}}{\beta}\right)\times\cdots\times\left(\frac{A_i e^{-t_x/\beta}}{\beta}\right)
\end{aligned}$$

$$= x\binom{m}{x}\left(\frac{A_i}{\beta}\right)^x e^{-\sum_{j=1}^{x} t_j/\beta}\, e^{-A_i\left(\sum_{j=1}^{x} e^{-t_j/\beta}+(m-x)e^{-t_x/\beta}\right)}. \square$$

Proof of Lemma 3. Recall that $\tilde{u}_0$ is Gumbel with scale β and location 0. It follows from Lemma 1 that $\Delta\tilde{u}_i(1)$ is an independent Gumbel with scale β and location $a_i + \beta\ln m$. Therefore, from Lemma 2, it follows that $\tilde{u}_0 - \Delta\tilde{u}_i^{(1)}$ is a logistic random variable with mean $-(a_i + \beta\ln m)$, and thus,

$$P\left[\tilde{x}_i \geq 1\right] = P\left[\Delta\tilde{u}_i(1) \geq \tilde{u}_0\right] = P\left[\tilde{u}_0 - \Delta\tilde{u}_i(1) \leq 0\right] = \frac{1}{1+e^{-(0-(0-a_i-\beta\ln m))/\beta}} = \frac{1}{1+e^{-(a_i+\beta\ln m)/\beta}}$$

$$= \frac{me^{a_i/\beta}}{1+me^{a_i/\beta}} = \frac{mA_i}{1+mA_i}$$

$$P\left[\tilde{x}_i = 0\right] = 1 - P\left[\tilde{x}_i \geq 1\right] = \frac{1}{1+mA_i}. \square$$

Proof of Lemma 4. To simplify notation, let $\tilde{\mathbf{u}}_i(x) = \left(\Delta\tilde{u}_i(1),\ldots,\tilde{u}_i(x)\right)$. Then

$$A = P\left[\Delta\tilde{u}_i(x) > t \geq \Delta\tilde{u}_i(x+1)\right] = \int_t^{\infty}\int_t^{t_1}\cdots\int_t^{t_{x-1}}\int_{-\infty}^{t} f_{\tilde{\mathbf{u}}_i(x+1)}(t_1,\ldots,t_{x+1})\,dt_{x+1}\ldots dt_1$$

$$= \int_t^{\infty}\int_t^{t_1}\cdots\int_t^{t_{x-1}}\int_{-\infty}^{t} \frac{m!}{(m-x-1)!}\left(\frac{A_i}{\beta}\right)^{x+1} e^{-\sum_{j=1}^{x+1} t_j/\beta}\, e^{-A_i\left(\sum_{j=1}^{x+1} e^{-t_j/\beta}+(m-x)e^{-t_{x+1}/\beta}\right)} dt_{x+1}\ldots dt_1$$

$$= \frac{m!}{(m-x-1)!}\int_t^{\infty}\int_t^{t_1}\cdots\int_t^{t_{x-1}}\left(\frac{A_i}{\beta}\right)^{x} e^{-\sum_{j=1}^{x} t_j/\beta}\, e^{-A_i\sum_{j=1}^{x} e^{-t_j/\beta}}\left(\int_{-\infty}^{t}\frac{A_i e^{-t_{x+1}/\beta}}{\beta}e^{-A_i(m-x)e^{-t_{x+1}/\beta}}\,dt_{x+1}\right)dt_x\ldots dt_1$$

Let $\delta = -A_i e^{-t_{x+1}/\beta}$. Then $d\delta = \frac{A_i e^{-t_{x+1}/\beta}}{\beta}dt$, $t_{x+1} = -\infty \Rightarrow \delta = -\infty$, $t_{x+1} = t \Rightarrow \delta = -A_i e^{-t/\beta}$, and

$$A = \frac{m!}{(m-x-1)!}\int_t^{\infty}\int_t^{t_1}\cdots\int_t^{t_{x-1}}\left(\frac{A_i}{\beta}\right)^{x} e^{-\sum_{j=1}^{x} t_j/\beta}\, e^{-A_i\sum_{j=1}^{x} e^{-t_j/\beta}}\left(\int_{-\infty}^{-A_i e^{-t/\beta}} e^{(m-x)\delta}\,d\delta\right)dt_x\ldots dt_1$$

$$= \frac{m!}{(m-x-1)!}\int_t^{\infty}\int_t^{t_1}\cdots\int_t^{t_{x-1}}\left(\frac{A_i}{\beta}\right)^{x} e^{-\sum_{j=1}^{x} t_j/\beta}\, e^{-A_i\sum_{j=1}^{x} e^{-t_j/\beta}}\left(\frac{e^{-(m-x)A_i e^{-t/\beta}}}{m-x}\right)dt_x\ldots dt_1$$

$$= \frac{m!e^{-(m-x)A_i e^{-t/\beta}}}{(m-x)!}\int_t^{\infty}\int_t^{t_1}\cdots\int_t^{t_{x-1}}\left(\frac{A_i}{\beta}\right)^{x} e^{-\sum_{j=1}^{x} t_j/\beta}\, e^{-A_i\sum_{j=1}^{x} e^{-t_j/\beta}}\,dt_x\ldots dt_1$$

$$= \frac{m!e^{-(m-x)A_i e^{-t/\beta}}}{(m-x)!}\int_t^{\infty}\int_t^{t_1}\cdots\int_t^{t_{x-2}}\left(\frac{A_i}{\beta}\right)^{x-1} e^{-\sum_{j=1}^{x-1} t_j/\beta}\, e^{-A_i\sum_{j=1}^{x-1} e^{-t_j/\beta}}\left(\int_t^{t_{x-1}}\frac{A_i}{\beta}e^{-t_x/\beta}e^{-A_i e^{-t_x/\beta}}\,dt_x\right)dt_{x-1}\ldots dt_1$$

$$=\frac{m!e^{-(m-x)A_i e^{-t/\beta}}}{(m-x)!}\int_t^{\infty}\int_t^{t_1}\cdots\int_t^{t_{x-2}}\left(\frac{A_i}{\beta}\right)^{x-1} e^{-\sum_{j=1}^{x-1}t_j/\beta} e^{-A_i\sum_{j=1}^{x-1}e^{-t_j/\beta}}\left(\int_{-A_i e^{-t/\beta}}^{-A_i e^{-t_{x-1}/\beta}} e^{\delta}d\delta\right)dt_{x-1}\ldots dt_1$$

$$=\frac{m!e^{-(m-x)A_i e^{-t/\beta}}}{(m-x)!}\int_t^{\infty}\int_t^{t_1}\cdots\int_t^{t_{x-2}}\left(\frac{A_i}{\beta}\right)^{x-1} e^{-\sum_{j=1}^{x-1}t_j/\beta} e^{-A_i\sum_{j=1}^{x-1}e^{-t_j/\beta}}\left(e^{-A_i e^{-t_{x-1}/\beta}}-e^{-A_i e^{-t/\beta}}\right)dt_{x-1}\ldots dt_1$$

$$=\frac{m!e^{-(m-x)A_i e^{-t/\beta}}}{(m-x)!}\int_t^{\infty}\int_t^{t_1}\cdots\int_t^{t_{x-3}}\left(\frac{A_i}{\beta}\right)^{x-2} e^{-\sum_{j=1}^{x-2}t_j/\beta} e^{-A_i\sum_{j=1}^{x-2}e^{-t_j/\beta}}\left(\begin{array}{l}\displaystyle\int_t^{t_{x-2}}\frac{A_i e^{-t_{x-1}/\beta}}{\beta}e^{-2A_i e^{-t_{x-1}/\beta}}dt_{x-1}\\ \displaystyle -e^{-A_i e^{-t/\beta}}\int_t^{t_{x-2}}\frac{A_i e^{-t_{x-1}/\beta}}{\beta}e^{-A_i e^{-t_{x-1}/\beta}}dt_{x-1}\end{array}\right)dt_{x-2}\ldots dt_1$$

$$=\frac{m!e^{-(m-x)A_i e^{-t/\beta}}}{(m-x)!}\int_t^{\infty}\int_t^{t_1}\cdots\int_t^{t_{x-3}}\left(\frac{A_i}{\beta}\right)^{x-2} e^{-\sum_{j=1}^{x-2}t_j/\beta} e^{-A_i\sum_{j=1}^{x-2}e^{-t_j/\beta}}\left(\int_{-A_i e^{-t/\beta}}^{-A_i e^{-t_{x-2}/\beta}} e^{2\delta}d\delta-e^{-A_i e^{-t/\beta}}\int_{-A_i e^{-t/\beta}}^{-A_i e^{-t_{x-2}/\beta}} e^{\delta}d\delta\right)dt_{x-2}\ldots dt_1$$

$$=\frac{m!e^{-(m-x)A_i e^{-t/\beta}}}{(m-x)!}\int_t^{\infty}\int_t^{t_1}\cdots\int_t^{t_{x-3}}\left(\frac{A_i}{\beta}\right)^{x-2} e^{-\sum_{j=1}^{x-2}t_j/\beta} e^{-A_i\sum_{j=1}^{x-2}e^{-t_j/\beta}}\left(\begin{array}{l}\displaystyle\frac{e^{-2A_i e^{-t_{x-2}/\beta}}-e^{-2A_i e^{-t/\beta}}}{2}\\ -e^{-A_i e^{-t/\beta}}\left(e^{-A_i e^{-t_{x-2}/\beta}}-e^{-A_i e^{-t/\beta}}\right)\end{array}\right)dt_{x-2}\ldots dt_1$$

$$=\frac{m!e^{-(m-x)A_i e^{-t/\beta}}}{(m-x)!}\int_t^{\infty}\int_t^{t_1}\cdots\int_t^{t_{x-3}}\left(\frac{A_i}{\beta}\right)^{x-2} e^{-\sum_{j=1}^{x-2}t_j/\beta} e^{-A_i\sum_{j=1}^{x-2}e^{-t_j/\beta}}\left(\begin{array}{l}\displaystyle\frac{e^{-2A_i e^{-t_{x-2}/\beta}}}{2}\\ \displaystyle -e^{-A_i e^{-t/\beta}}e^{-A_i e^{-t_{x-2}/\beta}}+\frac{e^{-2A_i e^{-t/\beta}}}{2}\end{array}\right)dt_{x-2}\ldots dt_1$$

$$=\frac{m!e^{-(m-x)A_i e^{-t/\beta}}}{(m-x)!}\int_t^{\infty}\int_t^{t_1}\cdots\int_t^{t_{x-3}}\left(\frac{A_i}{\beta}\right)^{x-2} e^{-\sum_{j=1}^{x-2}t_j/\beta} e^{-A_i\sum_{j=1}^{x-2}e^{-t_j/\beta}}\frac{1}{2}\left(e^{-A_i e^{-t_{x-2}/\beta}}-e^{-A_i e^{-t/\beta}}\right)^2 dt_{x-2}\ldots dt_1$$

$$=\frac{m!e^{-(m-x)A_i e^{-t/\beta}}}{2(m-x)!}\int_t^{\infty}\int_t^{t_1}\cdots\int_t^{t_{x-4}}\left(\frac{A_i}{\beta}\right)^{x-3} e^{-\sum_{j=1}^{x-3}t_j/\beta} e^{-A_i\sum_{j=1}^{x-3}e^{-t_j/\beta}}\left(\int_t^{t_{x-3}}\left(\begin{array}{l}\displaystyle\frac{A_i e^{-t_{x-2}/\beta}}{\beta}\\ \times e^{-A_i e^{-t_{x-2}/\beta}}\left(\begin{array}{l}e^{-A_i e^{-t_{x-2}/\beta}}\\ -e^{-A_i e^{-t/\beta}}\end{array}\right)^2\end{array}\right)dt_{x-2}\right)dt_{x-3}\ldots dt_1$$

$$=\frac{m!e^{-(m-x)A_i e^{-t/\beta}}}{2(m-x)!}\int_t^{\infty}\int_t^{t_1}\cdots\int_t^{t_{x-4}}\left(\frac{A_i}{\beta}\right)^{x-3} e^{-\sum_{j=1}^{x-3}t_j/\beta} e^{-A_i\sum_{j=1}^{x-3}e^{-t_j/\beta}}\left(\int_{-A_i e^{-t/\beta}}^{-A_i e^{-t_{x-3}/\beta}} e^{\delta}\left(e^{\delta}-e^{-A_i e^{-t/\beta}}\right)^2 d\delta\right)dt_{x-3}\ldots dt_1$$

$$=\frac{m!e^{-(m-x)A_i e^{-t/\beta}}}{2(m-x)!}\int_t^{\infty}\int_t^{t_1}\cdots\int_t^{t_{x-4}}\left(\frac{A_i}{\beta}\right)^{x-3} e^{-\sum_{j=1}^{x-3}t_j/\beta} e^{-A_i\sum_{j=1}^{x-3}e^{-t_j/\beta}}\left(\left.\frac{\left(e^{\delta}-e^{-A_i e^{-t/\beta}}\right)^3}{3}\right|_{-A_i e^{-t/\beta}}^{-A_i e^{-t_{x-3}/\beta}}\right)dt_{x-3}\ldots dt_1$$

$$= \frac{m!\, e^{-(m-x)A_i e^{-t/\beta}}}{2(m-x)!} \int_t^{\infty} \int_t^{t_1} \cdots \int_t^{t_{x-4}} \left(\frac{A_i}{\beta}\right)^{x-3} e^{-\sum_{j=1}^{x-3} t_j/\beta}\, e^{-A_i \sum_{j=1}^{x-3} e^{-t_j/\beta}}\, \frac{\left(e^{-A_i e^{-t_{x-3}/\beta}} - e^{-A_i e^{-t/\beta}}\right)^3}{3}\, dt_{x-3} \ldots dt_1$$

$$= \frac{m!\, e^{-(m-x)A_i e^{-t/\beta}}}{3!(m-x)!} \int_t^{\infty} \int_t^{t_1} \cdots \int_t^{t_{x-5}} \left(\frac{A_i}{\beta}\right)^{x-4} e^{-\sum_{j=1}^{x-4} t_j/\beta}\, e^{-A_i \sum_{j=1}^{x-4} e^{-t_j/\beta}} \left(\int_{-A_i e^{-t/\beta}}^{-A_i e^{-t_{x-4}/\beta}} e^{\delta} \left(e^{\delta} - e^{-A_i e^{-t/\beta}}\right)^3 d\delta \right) dt_{x-4} \ldots dt_1$$

$$= \frac{m!\, e^{-(m-x)A_i e^{-t/\beta}}}{3!(m-x)!} \int_t^{\infty} \int_t^{t_1} \cdots \int_t^{t_{x-5}} \left(\frac{A_i}{\beta}\right)^{x-4} e^{-\sum_{j=1}^{x-4} t_j/\beta}\, e^{-A_i \sum_{j=1}^{x-4} e^{-t_j/\beta}} \left(\left. \frac{\left(e^{\delta} - e^{-A_i e^{-t/\beta}}\right)^4}{4} \right|_{-A_i e^{-t/\beta}}^{-A_i e^{-t_{x-4}/\beta}} \right) dt_{x-4} \ldots dt_1$$

$$= \frac{m!\, e^{-(m-x)A_i e^{-t/\beta}}}{4!(m-x)!} \int_t^{\infty} \int_t^{t_1} \cdots \int_t^{t_{x-5}} \left(\frac{A_i}{\beta}\right)^{x-4} e^{-\sum_{j=1}^{x-4} t_j/\beta}\, e^{-A_i \sum_{j=1}^{x-4} e^{-t_j/\beta}} \left(e^{-A_i e^{-t_{x-4}/\beta}} - e^{-A_i e^{-t/\beta}}\right)^4 dt_{x-4} \ldots dt_1$$

$\vdots$ (repeating the pattern up to t_1, e.g., $x-5=1$, $x-4=2$, $4=x-2$)

$$= \frac{m!\, e^{-(m-x)A_i e^{-t/\beta}}}{(x-2)!(m-x)!} \int_t^{\infty} \left(\int_t^{t_1} \left(\frac{A_i}{\beta}\right)^{2} e^{-\sum_{j=1}^{2} t_j/\beta}\, e^{-A_i \sum_{j=1}^{2} e^{-t_j/\beta}} \left(e^{-A_i e^{-t_2/\beta}} - e^{-A_i e^{-t/\beta}}\right)^{x-2} dt_2 \right) dt_1$$

$$= \frac{m!\, e^{-(m-x)A_i e^{-t/\beta}}}{(x-2)!(m-x)!} \int_t^{\infty} \frac{A_i e^{-t_1/\beta}}{\beta} e^{-A_i e^{-t_1/\beta}} \left(\int_t^{t_1} \frac{A_i e^{-t_2/\beta}}{\beta} e^{-A_i e^{-t_2/\beta}} \left(e^{-A_i e^{-t_2/\beta}} - e^{-A_i e^{-t/\beta}}\right)^{x-2} dt_2 \right) dt_1$$

$$= \frac{m!\, e^{-(m-x)A_i e^{-t/\beta}}}{(x-2)!(m-x)!} \int_t^{\infty} \frac{A_i e^{-t_1/\beta}}{\beta} e^{-A_i e^{-t_1/\beta}} \left(\int_{-A_i e^{-t/\beta}}^{-A_i e^{-t_1/\beta}} e^{\delta} \left(e^{\delta} - e^{-A_i e^{-t/\beta}}\right)^{x-2} d\delta \right) dt_1$$

$$= \frac{m!\, e^{-(m-x)A_i e^{-t/\beta}}}{(x-2)!(m-x)!} \int_t^{\infty} \frac{A_i e^{-t_1/\beta}}{\beta} e^{-A_i e^{-t_1/\beta}} \left(\left. \frac{\left(e^{\delta} - e^{-A_i e^{-t/\beta}}\right)^{x-1}}{x-1} \right|_{-A_i e^{-t/\beta}}^{-A_i e^{-t_1/\beta}} \right) dt_1$$

$$= \frac{m!\, e^{-(m-x)A_i e^{-t/\beta}}}{(x-1)!(m-x)!} \int_t^{\infty} \frac{A_i e^{-t_1/\beta}}{\beta} e^{-A_i e^{-t_1/\beta}} \left(e^{-A_i e^{-t_1/\beta}} - e^{-A_i e^{-t/\beta}}\right)^{x-1} dt_1$$

$$= \frac{m!\, e^{-(m-x)A_i e^{-t/\beta}}}{(x-1)!(m-x)!} \int_{-A_i e^{-t/\beta}}^{0} e^{\delta} \left(e^{\delta} - e^{-A_i e^{-t/\beta}}\right)^{x-1} d\delta$$

$$= \frac{m!\, e^{-(m-x)A_i e^{-t/\beta}}}{(x-1)!(m-x)!} \left(\left. \frac{\left(e^{\delta} - e^{-A_i e^{-t/\beta}}\right)^{x}}{x} \right|_{-A_i e^{-t/\beta}}^{0} \right)$$

$$= \frac{m!\, e^{-(m-x)A_i e^{-t/\beta}}}{x!(m-x)!} \left[\left(1 - e^{-A_i e^{-t/\beta}}\right)^x - \left(e^{-A_i e^{-t/\beta}} - e^{-A_i e^{-t/\beta}}\right)^x \right]$$

$$=\binom{m}{x} e^{-(m-x)A_i e^{-t/\beta}} \left(1-e^{-A_i e^{-t/\beta}}\right)^x,$$

which is (24). Note that

$$P\left[\Delta\tilde{u}_i(1) \le t\right] = \int_{-\infty}^{t} f_{\Delta\tilde{u}_i(1)}(t_1)\,dt_1 = \int_{-\infty}^{t} \frac{mA_i e^{-t_1/\beta}}{\beta} e^{-mA_i e^{-t_1/\beta}}\,dt_1$$

$$=\left(e^{-mA_i e^{-t_1/\beta}}\Big|_{-\infty}^{t}\right) = e^{-mA_i e^{-t/\beta}}$$

$$P\left[\Delta\tilde{u}_i(m) > t\right] = \int_{t}^{\infty} f_{\Delta\tilde{u}_i(m)}(t_1)\,dt_1 = \int_{t}^{\infty} \frac{mA_i e^{-t_1/\beta}}{\beta} e^{-A_i e^{-t_1/\beta}} \left(1-e^{-A_i e^{-t_1/\beta}}\right)^{m-1} dt_1$$

$$=\left(-\left(1-e^{-A_i e^{-t_1/\beta}}\right)^m\right)\Bigg|_{t}^{\infty} = \left(1-e^{-A_i e^{-t/\beta}}\right)^m$$

and it is apparent that (24) includes (25) and (26) as special cases. Finally,

$$P\left[\Delta\tilde{u}_i(1) > t\right] = 1 - P\left[\Delta\tilde{u}_i(1) \le t\right] = 1 - e^{-mA_i e^{-t/\beta}}. \square$$

Proof of Lemma 5. It is clear that (28) holds at $x = m$. I evaluate over increasing values of x to establish the pattern that generalizes to (28). Let $g(x)$ be the LHS of (28) and $h(x)$ be the RHS of (28). Then for increasing values of $x < m$,

$$g(0) = \frac{mb}{a+mb} = h(0)$$

$$g(1) = m(m-1)b\left(\frac{1}{a+(m-1)b} - \frac{1}{a+mb}\right) = \left(\frac{(m-1)b}{a+(m-1)b}\right)\left(\frac{mb}{a+mb}\right) = h(1)$$

$$g(2) = \frac{m!b}{2(m-3)!}\left(\frac{1}{a+(m-2)b} - \frac{2}{a+(m-1)b} + \frac{1}{a+mb}\right)$$

$$= \frac{m!b}{2(m-3)!}\left(\frac{1}{a+(m-2)b} - \frac{1}{a+(m-1)b} - \left(\frac{1}{a+(m-1)b} - \frac{1}{a+mb}\right)\right)$$

$$= \frac{m!b}{2(m-3)!}\left(\frac{b}{(a+(m-2)b)(a+(m-1)b)} - \frac{b}{(a+(m-1)b)(a+mb)}\right)$$

$$= \frac{m!b^2}{2(m-3)!}\left(\frac{2b}{(a+(m-2)b)(a+(m-1)b)(a+mb)}\right)$$

$$= \left(\frac{(m-2)b}{(a+(m-2)b)}\right)\left(\frac{(m-1)b}{a+(m-1)b}\right)\left(\frac{mb}{a+mb}\right) = h(2)$$

$$g(3)=\frac{m!b}{3!(m-4)!}\left(\frac{1}{a+(m-3)b}-\frac{3}{a+(m-2)b}+\frac{3}{a+(m-1)b}-\frac{1}{a+mb}\right)$$

$$=\frac{m!b}{3!(m-4)!}\left(\begin{array}{l}\frac{1}{a+(m-3)b}-\frac{1}{a+(m-2)b}\\ -2\left(\frac{1}{a+(m-2)b}-\frac{1}{a+(m-1)b}\right)+\frac{1}{a+(m-1)b}-\frac{1}{a+mb}\end{array}\right)$$

$$=\frac{m!b^2}{3!(m-4)!}\left(\begin{array}{l}\frac{1}{(a+(m-3)b)(a+(m-2)b)}-\frac{1}{(a+(m-2)b)(a+(m-1)b)}\\ -\left(\frac{1}{(a+(m-2)b)(a+(m-1)b)}-\frac{1}{(a+(m-1)b)(a+mb)}\right)\end{array}\right)$$

$$=\frac{m!b^2}{3!(m-4)!}\left(\begin{array}{l}\frac{2b}{(a+(m-3)b)(a+(m-2)b)(a+(m-1)b)}\\ -\frac{2b}{(a+(m-2)b)(a+(m-1)b)(a+mb)}\end{array}\right)$$

$$=\frac{m!b^3}{3(m-4)!}\left(\frac{3b}{(a+(m-3)b)(a+(m-2)b)(a+(m-1)b)(a+mb)}\right)$$

$$=\left(\frac{(m-3)b}{a+(m-3)b}\right)\left(\frac{(m-2)b}{a+(m-2)b}\right)\left(\frac{(m-1)b}{a+(m-1)b}\right)\left(\frac{mb}{a+mb}\right)=h(3)$$

$\vdots$

continuing the pattern …

$\vdots$

$$g(m-1)=\left(\frac{b}{a+b}\right)\left(\frac{2b}{a+2b}\right)\cdots\left(\frac{(m-1)b}{a+(m-1)b}\right)\left(\frac{mb}{a+mb}\right)=h(m-1).\ \square$$

Proof of Lemma 6. Recall that $\Delta\tilde{u}_i(x)=a_i+\tilde{\varepsilon}_i^{(x)}$, and thus

$$E\left[\Delta\tilde{u}_i(x)\right]=a_i+E\left[\tilde{\varepsilon}_i^{(x)}\right]$$

where $\tilde{\varepsilon}_i^{(x)}$ is the x^{th} largest value among m independent Gumbel random variables with location parameter $\nu = 0$ and scale parameter β. From Lemma 1, it follows that $E\left[\tilde{\varepsilon}_i^{(1)}\right]=\beta(\ln m+\gamma)$, and thus,

$$E\left[\Delta\tilde{u}_i(1)\right]=a_i+\beta(\ln m+\gamma). \tag{A1}$$

Suppose $x > 1$. For $b > 0$, let $\delta = be^{-t/\beta}$, and note that $d\delta = \frac{-be^{-t/\beta}}{\beta}dt$, $t = \beta\ln\left(\frac{b}{\delta}\right)$, $t = -\infty \Rightarrow \delta = \infty$, $t = \infty \Rightarrow \delta = 0$. Therefore,

$$\int_{-\infty}^{\infty}\frac{tbe^{-t/\beta}}{\beta}e^{-be^{-t/\beta}}dt = -\int_{\infty}^{0}\beta\ln\left(\frac{b}{\delta}\right)e^{-\delta}d\delta = \int_{0}^{\infty}\beta\ln\left(\frac{b}{\delta}\right)e^{-\delta}d\delta = \beta\int_{0}^{\infty}\left[\ln b - \ln\delta\right]e^{-\delta}d\delta$$

$$= \beta\left(-e^{-\delta}\ln b\Big|_0^{\infty} - \int_0^{\infty}e^{-\delta}\ln\delta d\delta\right) = \beta\left(\ln b + \gamma\right). \qquad \text{(see (17))}$$

Note that $f_{\tilde{\varepsilon}_i^{(x)}}(t) = x\binom{m}{x}\frac{e^{-t/\beta}}{\beta}e^{-(m-x+1)e^{-t/\beta}}\left(1-e^{-e^{-t/\beta}}\right)^{x-1}$ (see Corollary 2). Therefore, for $x = 2, \ldots, m$,

$$E\left[\tilde{\varepsilon}_i^{(x)}\right] = \int_{-\infty}^{\infty}tf_{\tilde{\varepsilon}_i^{(x)}}(t)dt = x\binom{m}{x}\int_{-\infty}^{\infty}\frac{te^{-t/\beta}}{\beta}e^{-(m-x+1)e^{-t/\beta}}\left(1-e^{-e^{-t/\beta}}\right)^{x-1}dt.$$

Applying the binomial expansion, $\left(1-e^{y}\right)^{x} = \sum_{j=0}^{x}\binom{x}{j}(-1)^{j}e^{jy}$, $x = 0, 1, 2, 3, \ldots$,

$$E\left[\tilde{\varepsilon}_i^{(x)}\right] = x\binom{m}{x}\int_{-\infty}^{\infty}\frac{te^{-t/\beta}}{\beta}e^{-(m-x+1)e^{-t/\beta}}\sum_{j=0}^{x-1}\binom{x-1}{j}(-1)^{j}e^{-je^{-t/\beta}}dt$$

$$= x\binom{m}{x}\sum_{j=0}^{x-1}\binom{x-1}{j}\int_{-\infty}^{\infty}\frac{te^{-t/\beta}}{\beta}e^{-(m-x+1+j)e^{-t/\beta}}(-1)^{j}dt$$

$$= \binom{m}{x}\sum_{j=0}^{x-1}\frac{x}{m-x+1+j}\binom{x-1}{j}\int_{-\infty}^{\infty}\frac{t(m-x+1+j)e^{-t/\beta}}{\beta}e^{-(m-x+1+j)e^{-t/\beta}}(-1)^{j}dt$$

$$= \beta x\binom{m}{x}\sum_{j=0}^{x-1}\binom{x-1}{j}(-1)^{j}\left(\frac{\gamma+\ln(m-x+1+j)}{m-x+1+j}\right).$$

Thus

$$E\left[\Delta\tilde{u}_i(x)\right] = a_i + \beta x\binom{m}{x}\sum_{j=0}^{x-1}\binom{x-1}{j}(-1)^{j}\left(\frac{\gamma+\ln(m-x+1+j)}{m-x+1+j}\right),$$

which, as a consistency check, reduces to (A1) when $x = 1$. □

Proof of Theorem 1. From Corollary 2,

$$A = P\left[\tilde{x}_i \geq x\right] = P\left[\Delta\tilde{u}_i(x) - \tilde{u}_0 > 0\right] = \int_{-\infty}^{\infty}P\left[\tilde{u}_0 \leq t\right]f_{\Delta\tilde{u}_i(x)}(t)dt$$

$$= \frac{m!}{(x-1)!(m-x)!}\int_{-\infty}^{\infty}e^{-e^{-t/\beta}}\frac{A_ie^{-t/\beta}}{\beta}e^{-(m-x+1)A_ie^{-t/\beta}}\left(1-e^{-A_ie^{-t/\beta}}\right)^{x-1}dt.$$

Applying the binomial expansion, $\left(1-e^{y}\right)^{x} = \sum_{j=0}^{x}\binom{x}{j}(-1)^{j}e^{jy}$, $x = 0, 1, 2, 3, \ldots$,

$$A = \frac{m!}{(x-1)!(m-x)!}\int_{-\infty}^{\infty}\frac{A_i e^{-t/\beta}}{\beta}e^{-\left[1+(m-x+1)A_i\right]e^{-t/\beta}}\sum_{j=0}^{x-1}\binom{x-1}{j}(-1)^j e^{-jA_i e^{-t/\beta}}dt$$

$$= \frac{m!}{(x-1)!(m-x)!}\sum_{j=0}^{x-1}\int_{-\infty}^{\infty}\frac{A_i e^{-t/\beta}}{\beta}\binom{x-1}{j}(-1)^j e^{-\left[1+(m-x+1+j)A_i\right]e^{-t/\beta}}dt .$$

Let $\delta = -e^{-t/\beta}$. Then $d\delta = \frac{e^{-t/\beta}}{\beta}dt$, $t = -\infty \Rightarrow \delta = -\infty$, $t = \infty \Rightarrow \delta = 0$, and

$$A = \frac{m!}{(x-1)!(m-x)!}\sum_{j=0}^{x-1}\int_{-\infty}^{0}\binom{x-1}{j}(-1)^j A_i e^{\left[1+(m-x+1+j)A_i\right]\delta}d\delta$$

$$= \left(m-(x-1)\right)\binom{m}{x-1}\sum_{j=0}^{x-1}\binom{x-1}{j}(-1)^j\left(\frac{A_i}{1+\left(m-(x-1)+j\right)A_i}\right)$$

and it follows from Lemma 5 that

$$P\left[\tilde{x}_i \geq x\right] = \left(m-(x-1)\right)\binom{m}{x-1}\sum_{j=0}^{x-1}\binom{x-1}{j}(-1)^j\left(\frac{A_i}{1+\left(m-(x-1)+j\right)A_i}\right)$$

$$= \prod_{j=0}^{x-1}\frac{(m-j)A_i}{1+(m-j)A_i},$$

which is (30).

Note that $P\left[\tilde{x}_i \geq 0\right] = 1$. From Lemma 3,

$$P\left[\tilde{x}_i \geq 1\right] = P\left[\tilde{x}_i \geq 1 \mid \tilde{x}_i \geq 0\right] = \frac{mA_i}{1+mA_i}.$$

Thus, for $x \in \{1, \ldots, m-1\}$, I rearrange

$$P\left[\tilde{x}_i \geq x+1\right] = P\left[\tilde{x}_i \geq x+1 \mid \tilde{x}_i \geq x\right]P\left[\tilde{x}_i \geq x\right]$$

and apply (30) to obtain

$$P\left[\tilde{x}_i \geq x+1 \mid \tilde{x}_i \geq x\right] = \frac{P\left[\tilde{x}_i \geq x+1\right]}{P\left[\tilde{x}_i \geq x\right]} = \frac{(m-x)A_i}{1+(m-x)A_i},$$

which is (31). Applying the above,

$$P\left[\tilde{x}_i = x \mid \tilde{x}_i \geq x\right] = 1 - P\left[\tilde{x}_i \geq x+1 \mid \tilde{x}_i \geq x\right] = \frac{1}{1+(m-x)A_i},$$

which is (32). Finally, applying (30) and (32),

$$P\left[\tilde{x}_i = x\right] = P\left[\tilde{x}_i = x \mid \tilde{x}_i \geq x\right]P\left[\tilde{x}_i \geq x\right] = \left(\frac{1}{1+(m-x)A_i}\right)\prod_{j=0}^{x-1}\frac{(m-j)A_i}{1+(m-j)A_i}$$

yields (33), or alternatively (as a consistency check),

$$P[\tilde{x}_i = x] = P[\tilde{x}_i \geq x] - P[\tilde{x}_i \geq x+1] = \prod_{j=0}^{x-1} \frac{(m-j)A_i}{1+(m-j)A_i} - \prod_{j=0}^{x} \frac{(m-j)A_i}{1+(m-j)A_i}$$

$$= \left(\frac{1}{1+(m-x)A_i}\right) \prod_{j=0}^{x-1} \frac{(m-j)A_i}{1+(m-j)A_i} \qquad \text{(which is (33))}$$

$$= \left(\frac{1}{1+(m-x)A_i}\right)\left(\frac{(m-x+1)A_i}{1+(m-x+1)A_i}\right) \prod_{j=0}^{x-2} \frac{(m-j)A_i}{1+(m-j)A_i}$$

$$= \left(\frac{(m-x+1)A_i}{1+(m-x)A_i}\right)\left(\frac{1}{1+(m-x+1)A_i}\right) \prod_{j=0}^{x-2} \frac{(m-j)A_i}{1+(m-j)A_i}$$

$$= \left(\frac{(m-x+1)A_i}{1+(m-x)A_i}\right) P[\tilde{x}_i = x-1],$$

which is (34). □

Proof of Corollary 3. The relationship among probabilities follows directly from (34) and observation that

$$\frac{A_i + (m-x)A_i}{1+(m-x)A_i} \begin{cases} <1, & A_i > 1 \\ =1, & A_i = 1 \\ >1, & A_i < 1 \end{cases}.$$

If $A_i = 1$, then

$$E[\tilde{x}_i] = \frac{1}{1+m}\sum_{j=0}^{m} j = \frac{m}{2},$$

and $E[\tilde{x}_i] < \frac{m}{2}$ for $A_i < 1$ and $E[\tilde{x}_i] > \frac{m}{2}$ for $A_i > 1$ follow from the monotonicity of the probability mass function. □

Proof of Theorem 2. The proof relies on (31) in Theorem 1. To simplify notation, I suppress the subscript i, and (31) becomes

$$P[\tilde{x} \geq x+1 \mid \tilde{x} \geq x] = P\left[\tilde{\varepsilon}_0 - \tilde{\varepsilon}^{(x+1:m)} \leq a \mid \tilde{\varepsilon}_0 - \tilde{\varepsilon}^{(x:m)} \leq a\right] = \frac{(m-x)e^{a/\beta}}{1+(m-x)e^{a/\beta}},\ x \in \{1, \ldots, m-1\}.$$

From Lemma 1, $\tilde{\varepsilon}^{(1:k)}$ is a Gumbel random variable with scale parameter β and location parameter $\nu = \beta \ln k$. From Lemma 2, $\tilde{z} = \tilde{\varepsilon}_0 - \tilde{\varepsilon}^{(1:k)}$ is a logistic random variable with cdf

$$F_{\tilde{z}}(z) = \frac{1}{1+e^{-(z-(0-\beta\ln k))/\beta}} = \frac{1}{1+e^{-z/\beta+\ln k}} = \frac{ke^{z/\beta}}{1+ke^{z/\beta}}.$$

Therefore

$$P\left[\tilde{\varepsilon}_0-\tilde{\varepsilon}^{(x+1:m)}\le a \mid \tilde{\varepsilon}_0-\tilde{\varepsilon}^{(x:m)}\le a\right]=P\left[\tilde{\varepsilon}_0-\tilde{\varepsilon}^{(1:m-x)}\le a\right]. \tag{1}$$

We interpret (1) terms of $\tilde{z}_0,\tilde{z}_1,...,\tilde{z}_m$ given in the theorem, which can be rewritten as $\tilde{z}_0=\nu_0+\tilde{\varepsilon}_0$ and $\tilde{z}_i=\nu+\tilde{\varepsilon}_i$ where $\tilde{\varepsilon}_i$ are iid Gumbel random variables with location 0 and scale β. Furthermore, $\tilde{z}^{(1:m-x)}=\nu+\tilde{\varepsilon}^{(1:m-x)}$ with location parameter $\nu^{(m-x)}=\nu+\beta\ln(m-x)$ (see Lemma 1). Let $a=\nu^{(m-x)}+t-\nu_0$.

Substituting into (1),

$$P\left[\tilde{\varepsilon}_0-\tilde{\varepsilon}^{(x+1:m)}\le \nu^{(m-x)}+t-\nu_0 \mid \tilde{\varepsilon}_0-\tilde{\varepsilon}^{(x:m)}\le \nu^{(m-x)}+t-\nu_0\right]=P\left[\tilde{z}_0-\tilde{z}^{(x+1:m)}\le t \mid \tilde{z}_0-\tilde{z}^{(x:m)}\le t\right]$$

$$P\left[\tilde{\varepsilon}_0-\tilde{\varepsilon}^{(1:m-x)}\le \nu+t-\nu_0\right]=P\left[\tilde{z}_0-\tilde{z}^{(1:m-x)}\le t\right]=\frac{(m-x)e^{(\nu+t-\nu_0)/\beta}}{1+(m-x)e^{(\nu+t-\nu_0)/\beta}}=\frac{e^{(\nu+\beta\ln(m-x)+t-\nu_0)/\beta}}{1+e^{(\nu+\beta\ln(m-x)+t-\nu_0)/\beta}}$$

$$=\frac{1}{1+e^{-\left(t-\left(\nu_0-\nu^{(m-x)}\right)\right)/\beta}}$$

Therefore, $\tilde{y}^{(x+1:m)|(x:m)}$ is a logistic random variable with mean $\nu_0-\nu^{(m-x)}$ and variance $\pi^2\beta^2/3$. □

Proof of Theorem 3. From Corollary 2 and Lemma 4,

$$A=P[\tilde{\mathbf{x}}=\mathbf{x}]=\int_{-\infty}^{\infty} f_{\Delta\tilde{u}_0}(t)\left(\prod_{i=1}^{n}P\left[\Delta\tilde{u}_i(x_i)\ge t\ge \Delta\tilde{u}_i(x_i+1)\right]\right)dt$$

$$=\int_{-\infty}^{\infty}\frac{e^{-t/\beta}}{\beta}e^{-e^{-t/\beta}}\left(\prod_{i=1}^{n}\binom{m}{x_i}e^{-(m-x_i)A_ie^{-t/\beta}}\left(1-e^{-A_ie^{-t/\beta}}\right)^{x_i}\right)dt$$

$$=\left(\prod_{i=1}^{n}\binom{m}{x_i}\right)\int_{-\infty}^{\infty}\frac{e^{-t/\beta}}{\beta}\left(e^{-\left(1+\sum_{i=1}^{n}(m-x_i)A_i\right)e^{-t/\beta}}\right)\left(\prod_{i=1}^{n}\left(1-e^{-A_ie^{-t/\beta}}\right)^{x_i}\right)dt\,.$$

Applying the binomial expansion,

$$\left(1-e^{y}\right)^{x}=\sum_{j=0}^{x}\binom{x}{j}(-1)^{j}e^{jy}\,,\ x=0,\,1,\,2,\,3,\,\ldots,$$

$$A=\left(\prod_{i=1}^{n}\binom{m}{x_i}\right)\int_{-\infty}^{\infty}\frac{e^{-t/\beta}}{\beta}\left(e^{-\left(1+\sum_{i=1}^{n}(m-x_i)A_i\right)e^{-t/\beta}}\right)\left(\sum_{j_1=0}^{x_1}\cdots\sum_{j_n=0}^{x_n}\binom{x_1}{j_1}\cdots\binom{x_n}{j_{kn}}(-1)^{\left(\sum_{i=1}^{n}j_i\right)}e^{-\left(\sum_{i=1}^{n}j_iA_i\right)e^{-t/\beta}}\right)dt$$

$$=\left(\prod_{i=1}^{n}\binom{m}{x_i}\right)\left(\sum_{j_1=0}^{x_1}\cdots\sum_{j_n=0}^{x_n}\int_{-\infty}^{\infty}\binom{x_1}{j_1}\cdots\binom{x_n}{j_n}(-1)^{\left(\sum_{i=1}^{n}j_i\right)}\frac{e^{-t/\beta}}{\beta}\left(e^{-\left(1+\sum_{i=1}^{n}(m-x_i)A_i\right)e^{-t/\beta}}\right)e^{-\left(\sum_{i=1}^{n}j_iA_i\right)e^{-t/\beta}}dt\right)$$

$$=\left(\prod_{i=1}^{n}\binom{m}{x_i}\right)\left(\sum_{j_1=0}^{x_1}\cdots\sum_{j_n=0}^{x_n}\int_{-\infty}^{\infty}\binom{x_1}{j_1}\cdots\binom{x_n}{j_n}(-1)^{\left(\sum_{i=1}^{n}j_i\right)}\frac{e^{-t/\beta}}{\beta}\left(e^{-\left(1+\sum_{i=1}^{n}(m-x_i+j_i)A_i\right)e^{-t/\beta}}\right)dt\right).$$

Let $\delta = -e^{-t/\beta}$. Then $d\delta = \dfrac{e^{-t/\beta}}{\beta}dt$, $t = -\infty \Rightarrow \delta = -\infty$, $t = \infty \Rightarrow \delta = 0$, and

$$A = P\left[\tilde{\mathbf{x}} = \mathbf{x}\right] = \left(\prod_{i=1}^{n}\binom{m}{x_i}\right)\left(\sum_{j_1=0}^{x_1}\cdots\sum_{j_n=0}^{x_k}\int_{-\infty}^{0}\binom{x_1}{j_1}\cdots\binom{x_n}{j_n}(-1)^{\left(\sum_{i=1}^{n} j_i\right)}\left(e^{\left(1+\sum_{i=1}^{n}(m-x_i+j_i)A_i\right)\delta}\right)d\delta\right)$$

$$= \left(\prod_{i=1}^{n}\binom{m}{x_i}\right)\sum_{j_1=0}^{x_1}\cdots\sum_{j_n=0}^{x_n}\left(\frac{\binom{x_1}{j_1}\cdots\binom{x_n}{j_n}(-1)^{\left(\sum_{i=1}^{n} j_i\right)}}{1+\sum_{i=1}^{n}(m-x_i+j_i)A_i}\right).$$

As a consistency check, recall that $\tilde{u}_0, \Delta\tilde{u}_1(1), \ldots, \Delta\tilde{u}_n(1)$ are independent Gumbel random variables with scale β and location parameters 0 and $a_i + \beta \ln m$ for $i = 1, \ldots, n$. Thus, it follows from lemmas 1 and 2 that

$$P\left[\tilde{\mathbf{x}} = \mathbf{0}\right] = P\left[\tilde{u}_0 \geq \max_i\left\{\Delta\tilde{u}_i(1)\right\}\right] = \frac{1}{1+e^{-\left(0-\left(\sum_{i=1}^{n}(a_i+\beta\ln m)-0\right)\right)/\beta}} = \frac{1}{1+m\sum_{i=1}^{n}A_i},$$

which aligns with the general expression above. □

Proof of Theorem 4. Let $\boldsymbol{\tau}$ denote a threshold vector. The Lagrangian of the optimization problem is

$$L(\boldsymbol{\tau},\lambda) = \lambda B + \sum_{i=1}^{n}\sum_{x=1}^{m}\left(E\left[\Delta\tilde{u}_i(x)-\tilde{u}_0 \mid \Delta\tilde{u}_i(x)-\tilde{u}_0 \geq \tau_i\right]-\lambda\right)P\left[\Delta\tilde{u}_i(x)-\tilde{u}_0 \geq \tau_i\right].$$

Let $f_i(t\,|x)$ denote the probability density function of $\Delta\tilde{u}_i(x)-\tilde{u}_0$. Then

$$\frac{\partial L(\boldsymbol{\tau},\lambda)}{\partial \tau_i} = \frac{\partial}{\partial \tau_i}\sum_{x=1}^{m}\int_{\tau_i}^{\infty}(t-\lambda)f_i(t\,|\,x)dt = -\sum_{x=1}^{m}(\tau_i-\lambda)f_i(\tau_i\,|\,x)$$

$$\frac{\partial^2 L(\boldsymbol{\tau},\lambda)}{\partial \tau_i^2} = -\sum_{x=1}^{m}\left[(\tau_i-\lambda)\frac{\partial f_i(\tau_i\,|\,x)}{\partial \tau_i}+f_i(\tau_i\,|\,x)\right]$$

$$\frac{\partial^2 L(\boldsymbol{\tau},\lambda)}{\partial \tau_i \partial \tau_j} = 0,\ i \neq j$$

The first-order condition (necessary condition for interior optimal) is

$\tau_i = \lambda$ for all i

and at this point,

$$\frac{\partial^2 L(\boldsymbol{\lambda},\lambda)}{\partial \tau_i^2} < 0,$$

which implies that $\tau_i = \tau = \lambda$ is a unique global maximum. If $\sum_{i=1}^{n}\sum_{x=1}^{m} P\left[\Delta\tilde{u}_i(x) - \tilde{u}_0 \geq 0\right] < B$, then the constraint is nonbinding and $\tau = \lambda = 0$. Otherwise τ is the unique solution to (40). □

11.2. Hard Constraint versus Soft Constraint

Estimation of MDCEV (via (4)) and the models in sections 3.2 and 3.3 include an assumption on how individuals operationalize choice decisions. I describe the assumption in this section and compare it to an alternative assumption.

At each choice event, an individual makes decisions to maximize utility. There are at least two ways in which this may be operationalized. The following two formulations show optimal expected utility from a future choice event.

$$\text{P1: } U_1^* = E\left[\max_{\mathbf{x}}\left\{\sum_{i=1}^{n}\tilde{u}_i(x_i) : \sum_{i=1}^{n} x_i \leq B\right\}\right] \qquad \text{(hard constraint)}$$

$$\text{P2: } U_2^* = \max_{\lambda\geq 0}\left\{\sum_{i=1}^{n} E\left[\tilde{u}_i\left(\hat{x}\left(\lambda,\tilde{u}_i(\cdot)\right)\right)\right] : \sum_{i=1}^{n} E\left[\hat{x}\left(\lambda,\tilde{u}_i(\cdot)\right)\right] \leq B\right\} \qquad \text{(soft constraint)}$$

where for realization $u_i(\cdot)$ of random utility function $\tilde{u}_i(\cdot)$

$$\hat{x}\left(\lambda, u_i(\cdot)\right) = \max_{x\in\mathbb{R}_+}\left\{x : \Delta u_i(x) \geq \lambda\right\}.$$

Note that $\hat{x}\left(\lambda, u_i(\cdot)\right)$ is the decision rule given in (2). For a given realization of utility functions, the decision the rule yields the quantity decisions $\mathbf{x}$ that maximize total utility subject to an upper limit on consumption that is determined by the dual value λ. The optimal value of λ is set to maximize expected utility subject to an upper limit on expected consumption (B).

Under P1, the decision maker chooses $\mathbf{x}$ that maximizes utility subject to the constraint for each realization of utility functions. P1 enforces the constraint at each choice event, i.e., the inequality is treated as a hard constraint.

Under P2, the decision maker chooses $\mathbf{x}$ according to the threshold decision rule; the decision maker identifies alternatives with marginal utility above the threshold λ at the origin, and then increases quantity until marginal utility matches the threshold. The threshold is set to maximize expected utility while satisfying the constraint in expectation, i.e., the inequality is treated as a soft constraint. That is, the decision maker identifies the threshold that maximizes expected utility while assuring that the constraint is satisfied "on average" (e.g., tuned heuristically from experience over time). As shown in the Section 2.2, MDCEV is based on P2, as are the models in sections 3.2 and 3.3.

Intuition may suggest that P2 results in higher expected utility than P1, i.e., $U_2^* \geq U_1^*$. This intuition is correct. The reason is that, at the optimal threshold under P2, the decision maker selects more than B

units when realized choice utilities are high, which is offset (in expectation) by selecting fewer than B units when realized choice utilities are low.[5] It may be that P2 is more aligned with human nature as well, e.g., tend to choose more than B when highly desirable to do so.

In summary, in the context of discrete quantity choice, P2 has appealing characteristics relative to P1 across several dimensions: individual decision-making, empirical analysis, and normative analysis. From the perspective of an individual's choice decision, P2 results in higher average utility for an individual while being practical to operationalize. Choice probabilities have a simple form that is amenable to efficient estimation under alternative assumptions for probability distributions of idiosyncratic error terms (e.g., probit, logit). And while perhaps not apparent from the choice probability expression, the model captures the interaction of alternatives through the threshold parameter. Normative analysis based on the model is relatively tractable. Lastly, it is important to note that the entire class of generalized extreme value models (McFadden 1978) is a special case of the model given in P2 with $B = 1$, i.e., P2 is consistent with GEV models. This result is proved in Gallego and Wang (2020, Theorem 3).

11.3. Continuous Analog of the Discrete Distribution in Theorem 1.[6]

Define $g(x) = P\left[\tilde{x} = x\right]$, $x = 0, \ldots, m$. From Theorem 1 (with the subscript for alternative i suppressed),

$$g(0) = \frac{1}{1+mA}$$

$$g(x) = \left(\frac{1+m-x}{1/A+m-x}\right) g\left(x-1\right),\ x = 1, \ldots, m.$$

Thus,

$$\frac{\Delta g\left(x\right)}{g\left(x\right)} = \frac{g\left(x\right)-g\left(x-1\right)}{g\left(x\right)} = \frac{1+m-x-\left(1/A+m-x\right)}{1+m-x} = \frac{1-1/A}{1+m-x}.$$

The continuous analog, $h(x)$, satisfies

$$\frac{h'\left(x\right)}{h\left(x\right)} = \frac{1-1/A}{1+m-x}.$$

For $A \neq 1$, the solution to the differential equation is

$$h(x) = c\left(1+m-x\right)^{1/A-1}$$

---

[5] To briefly formalize this point, suppose that the conclusion is not true. Consider a pair of realizations, one for which the optimal threshold under P1 is low (with threshold λ_L) and the other for which the optimal threshold is high (with threshold $\lambda_H > \lambda_L$). The constraint is satisfied at equality for both realizations under P1. Then there exists a threshold λ between λ_L and λ_H for which the constraint is satisfied in expectation with respect to these two realizations and with higher expected utility, which leads to a contradiction. See Gallego and Wang (2020; Theorem 2) for a detailed proof for the case of discrete quantity decisions.

[6] This continuous analog was identified by Harish Guda, W. P. Carey School of Business, Arizona State University.