A New Structural Decomposition of the Multivariate Extreme Value (MEV) Distribution in Nested Logit Models

Chandra R. Bhat
The University of Texas at Austin
Department of Civil, Architectural and Environmental Engineering
301 E. Dean Keeton St. Stop C1761, Austin TX 78712, USA
Tel: +1-512-471-4535; Email: bhat@mail.utexas.edu

ABSTRACT
This paper develops a new structural decomposition of the multivariate extreme value (MEV) distribution underlying the nested logit model. Starting directly from the MEV cumulative distribution function of the single-level nested logit, we derive an explicit polynomial recursion for the density and obtain a decomposition in which each disturbance is expressed as the sum of a common log-aggregation term and a log-Dirichlet allocation component. This representation yields closed-form expressions for the moment generating function and cumulant generating function of the common component, enabling a complete higher-order characterization of the MEV distribution. A key result is that while the variance of the common aggregation term decreases with the number of alternatives in a nest, harmonic terms in the common aggregation term and the log-Dirichlet components cancel exactly in pairwise covariances, leaving within-nest correlation in the MEV invariant at $\left(1-\rho^2\right)$. More generally, all multivariate cumulants admit a compact factorial–zeta form depending only on the dissimilarity parameter. These results provide new analytical and computational tools for studying dependence in the MEV distributions underlying nested logit systems. To our knowledge, such a complete higher-order characterization of the MEV distribution has not previously been derived in closed form. The new decomposition developed in the paper further implies a finite Gamma-mixture representation that enables exact and computationally efficient simulation for MEV errors. The new simulation approach is validated at both the univariate and multivariate levels.

Keywords: Multivariate extreme-value distribution (MEV); MEV simulation; nested logit; Dirichlet distribution; Gamma-mixture.

1. INTRODUCTION

The multinomial logit (MNL) model has been, for decades now, the workhorse of random utility-maximization (RUM)-based discrete choice analysis in transportation and many other fields. This model is based on assuming that the stochastic (or error) terms in the utility functions of the alternatives are identically and independently distributed (IID). This IID assumption leaves the MNL model saddled with the "independence of irrelevant alternatives" (IIA) property at the individual level (Luce and Suppes, 1965; see also Ben-Akiva and Lerman, 1985 for a detailed discussion of this property). The identically distributed assumption can be relaxed using non-identical random components, such as the negative exponential model of Daganzo (1979), the oddball alternative model of Recker (1995), or the heteroscedastic extreme-value (HEV) model of Bhat (1995). The focus of this paper, however, is on relaxing the independently distributed assumption of the stochastic terms, with the nested logit (NL) model being an archetypical representative in this family of structures. In particular, the NL model introduces covariance among the random components of subsets (or nests) of alternatives (each alternative being exclusively assigned to one and only one nest). The result is that alternatives within a nest share a common dissimilarity parameter governing their substitution patterns relative to alternatives outside the nest, and their collective attractiveness can be represented by the logsum (inclusive value), which serves as a measure of accessibility (Williams, 1977; McFadden, 1978; Daly and Zachary, 1978; Ben-Akiva and Lerman, 1979; see Hensher et al., 2015 for a detailed exposition). The intensity of correlation (due to unobserved factors) across any pair of alternatives within each nest of alternatives is determined by a dissimilarity parameter that needs to be between 0 and 1 if the NL model is to remain globally consistent with the random utility maximizing principle. This NL model is typically motivated from the construction of a generalized extreme value (GEV) generating function (that must satisfy some strict conditions to be consistent with RUM), which, in turn, induces a multivariate extreme-value joint cumulative distribution function (CDF). Choice probabilities can be obtained directly from the generating function or equivalently by direct integration from the induced CDF. In this NL formulation, the model's dependence structure is governed by the generating function rather than being parameterized in terms of disturbance covariances (see, for example, McFadden, 1978).

In the special case of the nested logit model when there is only one nest with two alternatives, the situation becomes simpler because the nested logit error structure becomes identical to what Johnson and Kotz (1972, page 256) have identified as the Type B bivariate extreme value (EV) distribution. In particular, in this bivariate case, the cumulative distribution function (CDF) of the nested logit error structure (assuming a single level nesting, which is the focus of the current paper) takes the following form (Johnson and Kotz parameterize ρ as $1/m$):

$$F_{\varepsilon_1,\varepsilon_2}(w_1,w_2)=\exp\left[-\left(e^{-w_1/\rho}+e^{-w_2/\rho}\right)^{\rho}\right] \text{ for } 0<\rho\leq 1. \tag{1}$$

Each of the univariate errors ε_1 and ε_2 have a standard extreme value distribution as $F_{\varepsilon_1}(w_1)=\exp-\left(e^{-w_1/\rho}\right)^{\rho}=\exp-\left(e^{-w_1}\right)$ and $F_{\varepsilon_2}(w_2)=\exp-\left(e^{-w_2/\rho}\right)^{\rho}=\exp-\left(e^{-w_2}\right)$. Johnson and Kotz proceed to attribute the closed-form Pearson correlation of $\left(1-\rho^2\right)$ between ε_1 and ε_2 to Oliveira (1961), though Oliveira (1961) only develops a general representation of bivariate extreme-value (BEV) distributions and does not derive moment-based dependence measures. Later, Oliveira (1962–63) showed that for a particular parametric subclass of the bivariate extreme value distribution, the difference of the two variables follows a logistic distribution, implying a closed-form expression for the Pearson correlation as a function of the dependence parameter. Gumbel and Mustafi (1967) later rederived this result and made the correlation–parameter relationship explicit as $\text{Correlation} = \left(1-\alpha^2\right)$, where α in their notation corresponds to ρ in Equation (1).[1] Clearly, this result must hold even for the single-level nested logit form of the multivariate extreme value (MEV) error terms with a common dissimilarity parameter ρ in which the CDF (with more than two alternatives in a nest) is given by:

$$F_{\varepsilon_1,\varepsilon_2,...,\varepsilon_K}(w_1, w_2, ..., w_K) = \exp\left[-\left(\sum_{k=1}^{K} e^{-w_k/\rho}\right)^{\rho}\right] \text{ for } 0 < \rho \leq 1. \quad (2)$$

This is because the above MEV CDF collapses in the bivariate marginal case to the BEV CDF of Equation (1) (in the rest of this paper, any reference to the MEV distribution will refer strictly to the distribution of the form in Equation (2), which characterizes a single-level nesting structure). Thus, without any ambiguity, the pairwise correlation between any two alternatives in a single nest in a single-level nested logit model is $\left(1-\rho^2\right)$. For good measure, Bhat (1998) validates this closed-form correlation between each pair of alternatives in a nested logit model in the process of extending the nested logit model to a more advanced nested logit-ordered model structure. But, while the pairwise dependence structure of the MEV distribution is well understood, a general higher-order multivariate characterization has remained analytically intractable.

In related literature, Williams (1977) recognized that the correlation among the utilities of alternatives within a nest may be engendered through a decomposition of the overall error terms of alternatives into a common shared component within a nest and a separate independent alternative-specific component. Specifically, using the notation from earlier, for all alternatives *j* within a nest (say nest *b*), the error terms may be written as follows:

$$\varepsilon_{j,nest\,b} = v_{nest\,b} + v_j^*, \quad (3)$$

where the error term $v_{nest\,b}$ is the common error term and v_j^* is a separate independent Gumbel distributed error term with a location parameter of zero and a scale parameter ρ less than 1. With the overall error term $\varepsilon_{j,nest\,b}$ assumed to be standard Gumbel (with a location parameter of zero and a scale parameter of 1), and the assumption that the v_j^* error terms are independent across alternatives *j*, this formulation becomes consistent with the means, variances and the pairwise correlation implied by the multivariate nested logit error structure. However, there are two issues with this error decomposition perspective for the nested logit. The first, and the most important one from the standpoint of this paper, is that there is no theoretical link between the overall distribution of the error terms in the MEV structure of Equation (2) and the overall distribution of the error terms implied by the decomposition approach of Equation (3). That is, while the scale of

[1] If the difference of ε_1 and ε_2 follows a logistic distribution with scale ρ, this implies that:
$Var(\varepsilon_1-\varepsilon_2) = \frac{\pi^2}{3}\rho^2 = Var(\varepsilon_1) + Var(\varepsilon_2) - 2Cov(\varepsilon_1,\varepsilon_2) = \frac{\pi^2}{6} + \frac{\pi^2}{6} - 2Cov(\varepsilon_1,\varepsilon_2)$, which then implies that $Cov(\varepsilon_1,\varepsilon_2) = \frac{\pi^2}{3}\left(1-\rho^2\right)$. And $Cor(\varepsilon_1,\varepsilon_2) = \frac{Cov(\varepsilon_1,\varepsilon_2)}{\sqrt{Var(\varepsilon_1)}\sqrt{Var(\varepsilon_2)}} = \frac{\frac{\pi^2}{3}\left(1-\rho^2\right)}{\frac{\pi^2}{3}} = \left(1-\rho^2\right)$.

each error term in the decomposition approach is set to 1 and the implied pairwise error correlation is $1-\rho^2$, nothing guarantees that the partitioning approach will produce the same exact profile of the full multivariate distribution of the MEV errors (including the higher order moments). This is because matching on the first two moments of two multivariate distributions is not adequate to show equality of the two distributions. The second issue is whether there is even any continuous distribution for $v_{nest\,b}$ that satisfies the requirements embedded in Equation (3). In this regard, Carrasco and Ortuzar (2002) note that "such a distribution may not exist". However, on this second assumption, Cardell (1997) observed that such a distribution for $v_{nest\,b}$ does indeed exist. Specifically, Cardell identifies a class of distributions that he refers to as the $C(\lambda)$ family with $0 \le \lambda \le 1$ such that when a $C(\lambda)$-distributed variate is added to a Type 1 extreme-value (Gumbel) variate, the result is again a Gumbel variate. Further, the Gumbel distribution is itself a member of the $C(\lambda)$ class, with $C(0)$ corresponding to the standard Gumbel distribution. The probability density function of a $C(\lambda)$ distributed variable takes the form of an infinite sum, with no established and obvious closed form expressions for both the probability density function (PDF) and the CDF (except, of course, for $C(0)$, which corresponds to the standard univariate Gumbel distribution).

In summary, there is a lack of clarity on the exact equivalence of the decomposition approach of Equation (3) with the MEV distribution of Equation (2) (the first issue identified above), even using the Cardell decomposition. That is, the decomposition method of Equation (3) relies on starting from a variance components approach and then working toward any equivalence to the MEV distribution. Further, the Cardell decomposition results in a distribution that has no known closed-form PDF or CDF. From a practical standpoint, this implies that the Cardell decomposition only allows for approximations (in an analytic sense) to generate realizations from the MEV distribution (as may be needed for predicting population distributions of counterfactual policies or even overall aggregate market change computations). For example, Ye (2011) use a combination of a bivariate extreme value distribution and the Gaussian copula to generate realizations from Cardell's $C(\lambda)$ distribution family for the common error term $v_{nest\,b}$, while Bunch and Rocke (2016) use the characteristic function of the family $C(\lambda)$ (which also is not in closed-form by itself) to identify a "flexible" approximation to the $C(\lambda)$ distribution by matching on the first four moments. More recently, Townsend (2025) uses a similar approach, except to approximate the $C(\lambda)$ distribution using a numeric approach to match the non-closed form $C(\lambda)$ PDF.

Different from the Cardell approach that develops a distribution for the extreme value that has formed the basis for a variance-components perspective for the nested logit (and that earlier studies have employed to (approximately) generate variates from the MEV distribution), we introduce a new decomposition approach for the MEV that involves a convolution of two different types of error terms, each with closed-form PDFs and CDFs, as well as closed form moment generating functions with closed form higher order moments. In particular, the decomposition we propose involves the logarithm of the multivariate Dirichlet combined with the logarithm of a single new scalar Z_k distribution . We refer to the induced distribution $S_k = \log(Z_k)$ as the S_k distribution. The form of the S_k distribution (for any k) is obtained using an exact recursion formula proposed in the paper. Another salient feature of our approach is that rather than start from a variance decomposition approach and then attempt to show how this may conform to the generator function derivation for the MEV, we start directly from the CDF of the MEV distribution derived from the generator function (which is available in closed form) and then show how this MEV distribution can be exactly characterized using our new decomposition approach. To our knowledge, this is the first such derivation for the extreme value distribution in both its univariate and multivariate forms. We also discuss how the resulting closed form for our S_k distribution can be used to generate realizations from the MEV distribution using an exact analytic approach. Overall, we propose a novel decomposition of the MEV distribution, offering a new structural perspective on its stochastic formulation that provides alternative pathways for simulation, analysis, and interpretation. Our results should be of considerable interest not only to applied microeconomic researchers in the transportation field and beyond, but also to economic and statistical theoreticians due to the more transparent interpretation of the MEV distribution.

2. A RECURSION FORM FOR THE MEV DENSITY FUNCTION

We first develop a recursion formula for the MEV density function, which is the central support for many important later derivations. Specifically, define a random variable $Z=\sum_{k=1}^{K} e^{-\varepsilon_k/\rho}$ with $z=\sum_{k=1}^{K} e^{-w_k/\rho}$. Then, the MEV CDF of Equation (2) may then be written as follows:

$$F_{\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_K}(w_1,w_2,\ldots,w_K)=\exp\left[-(z)^{\rho}\right] \text{ for } 0<\rho\leq 1. \tag{4}$$

Through continued differentiation, the multivariate density function for the MEV with *k* variates may be written as follows:

$$f_{\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_N}(w_1,w_2,\ldots,w_K)=\left\{\exp\left[-(z)^{\rho}\right]z^{\rho-k}\exp\left[-\left(\frac{1}{\rho}\sum_{k=1}^{K}w_k\right)\right]\right\}\times P_k(z) \text{ for } 0<\rho\leq 1, \tag{5}$$

where the polynomial $P_k(z)$ is a polynomial in *z* with its highest degree being *k*–1. This polynomial may be written in recursive form as:

$$P_k(z)=\left(z^{\rho}-\left[\frac{\rho-(k-1)}{\rho}\right]\right)P_{k-1}(z)-\left(\frac{z}{\rho}\right)\frac{dP_{n-1}(z)}{dz},\ k\geq 2 \text{ and } P_1(z)=1. \tag{6}$$

The polynomial above can be further simplified by writing $v=z^{\rho}$, and developing an equivalent recursive relationship $Q_k(v)$ for $P_k(z)$. That is, we need $Q_k(v)=P_k(z)$ and $Q_{k-1}(v)=P_{k-1}(z)$. Consider the differentiation in the third term of the recursion above, and write:

$$\left(\frac{z}{\rho}\right)\frac{dP_{k-1}(z)}{dz}=\left(\frac{z}{\rho}\right)\frac{dQ_{k-1}(v)}{dv}\left(\frac{dv}{dz}\right)=\left(\frac{z}{\rho}\right)\frac{dQ_{k-1}(v)}{dv}\left(\rho z^{\rho-1}\right)=\frac{dQ_{k-1}(v)}{dv}z^{\rho}=\frac{dQ_{k-1}(v)}{dv}v \tag{7}$$

We can then write:

$$Q_k(v)=\left(v-\left[\frac{\rho-(k-1)}{\rho}\right]\right)Q_{k-1}(v)-v\frac{dQ_{k-1}(v)}{dv},\ k\geq 2 \text{ and } Q_1(v)=1;\ v=z^{\rho}, v\geq 0 \tag{8}$$

Plugging back into the MEV PDF, we get:

$$f_{\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_N}(w_1,w_2,\ldots,w_K)=\left\{\exp\left[-v\right]z^{\rho-k}\exp\left[-\frac{1}{\rho}\left(\sum_{k=1}^{K}w_k\right)\right]\right\}\times Q_k(v)\ \text{ for } 0<\rho\le 1,\ v=z^{\rho} \tag{9}$$

The recursion in Equation (8) provides a computationally tractable mechanism for constructing the MEV density for arbitrary *k*, as illustrated in Equation (9), and serves as the structural backbone for the decomposition developed in the next section. However, for practical use in estimation or simulation from the MEV distribution, the recursion must be expressed in closed form by eliminating the differentiation operator on the right-hand side of Equation (8). This can be accomplished by explicitly constructing the polynomial coefficients embedded in $Q_k(v)$ and deriving a recursion directly in those coefficients, as we now proceed to do.

Proposition 1: **Recursivity of Polynomial Coefficients of** v **in** $Q_k(v)$

For each number of variates $k\,(k\ge 2)$ in the MEV, let $Q_k(v)=\sum_{j=0}^{k-1}a_{k,j}\rho^{-(k-1-j)}v^j$, where, for each *k* and *j*, $a_{k,j}$ is the coefficient on v^j for a given ρ (note that $a_{k,j}$ is itself a polynomial in ρ of degree at most $k-1-j$). The factor $\rho^{-(k-1-j)}$ is introduced explicitly so that $Q_k(v)$ admits a finite polynomial representation in v. Then, the coefficients $a_{k,j}$ satisfy, for $j=0,1,\ldots,k-1$, the following recursion (with the base convention that $a_{1,0}=1$):

$$a_{k,j}=\left[a_{k-1,j-1}+\left(k-1-(j+1)\rho\right)a_{k-1,j}\right],\ a_{k-1,-1}=0,\ a_{k-1,k-1}=0 \tag{10}$$

Proof: See Appendix A.

The recursion in Equation (10), given ρ>0, also implies that $a_{k,j}\ge 0\ \forall k$ and *j*. For numerical evaluation at a given fixed value of ρ, the recursion above for the polynomial $a_{k,j}$ may be implemented directly for fast computation of $Q_k(v)$. However, if one wants to retain the full symbolic dependence on the correlation parameter ρ, or show closed-form expressions, or maintain a single codebase for any ρ value $(0<\rho\le 1)$ for a given number of variates (valuable for estimation purposes with different ρ values across iterations), one can compute, for given *k*, the specific coefficient $a_{k,j,h}$ corresponding to each polynomial degree of ρ within the polynomial $a_{k,j}$ such that $a_{k,j}=\sum_{h=0}^{k-1-j}a_{k,j,h}\rho^{k-1-j-h}$ ($j=0,1,\ldots,k-1,\ h=0,1,\ldots,k-1-j$) (with the following notational constructions: $a_{k-1,-1,h}=0$ (*j* cannot take the value –1), $a_{k-1,j,-1}=0$ (*h* cannot take the value –1), $a_{k-1,k-1,h}=0$ (as discussed earlier, the entire polynomial $a_{k-1,k-1}$ does not even appear in $Q_{k-1}(v)$), and the base $a_{1,0,0}=1$. The recursivity relationship for the $a_{k,j,h}$ is provided in the proposition below.

Proposition 2: Recursivity of polynomial coefficients on ρ **within the polynomial** $a_{k,j}$

The scalar $a_{k,j,h}$ coefficients (that are entirely free of ρ) follow the recursion:

$$a_{k,j,h}=a_{k-1,j-1,h}+(k-1)a_{k-1,j,h-1}-(j+1)a_{k-1,j,h} \tag{11}$$

Proof: See Appendix B. With the $a_{k,j,h}$ coefficients, the following holds:

$$Q_k(v)=\sum_{j=0}^{k-1}a_{k,j}\rho^{-(k-1-j)}v^j=\sum_{j=0}^{k-1}\left(\sum_{h=0}^{k-1-j}a_{k,j,h}\rho^{k-1-j-h}\right)\rho^{-(k-1-j)}v^j=\sum_{j=0}^{k-1}\left(\sum_{h=0}^{k-1-j}a_{k,j,h}\rho^{-h}\right)v^j \tag{12}$$

3. A NEW CHARACTERIZATION OF THE BIVARIATE EXTREME VALUE (BEV) DISTRIBUTION

In advance of presenting the new characterization for the MEV distribution with any $k\geq 2,$ we first present the conceptual basis with the simpler setting of $k=2.$ With the notations already introduced, the probability density function of the bivariate extreme value (BEV) distribution is:

$$f_{\varepsilon_1,\varepsilon_2}(w_1,w_2)=\left\{\exp[-v]z^{\rho-2}\exp\left[-\frac{1}{\rho}\left(\sum_{k=1}^{2}w_k\right)\right]\right\}\times Q_2(v)\ \text{ for } 0<\rho\leq 1, \tag{13}$$

$$\text{where } Q_2(v)=v-\left(\frac{\rho-1}{\rho}\right)\text{ with } z=\exp\left[-\left(\frac{w_1}{\rho}\right)\right]+\exp\left[-\left(\frac{w_2}{\rho}\right)\right],\ v=z^{\rho}. \tag{14}$$

Let $U_k=\dfrac{e^{-\varepsilon_k/\rho}}{Z_2}$ and, correspondingly, $u_k=\dfrac{e^{-w_k/\rho}}{z}$; $k=1,2$. Note also that $U_1+U_2=1$.

This also implies that $\varepsilon_1=-\rho\left(\log U_1+\log Z_2\right)$ and $\varepsilon_2=-\rho\left(\log(U_2)+\log Z_2\right)=-\rho\left(\log(1-U_1)+\log Z_2\right)$. Throughout the paper, log denotes the natural logarithm.

Using Equation (13) for the BEV distribution, we may write the corresponding bivariate density function for (ε_1,Z_2). The desired determinant of the Jacobian for the transformation is:

$$|J|_{(\varepsilon_1,\varepsilon_2)\to(U_1,Z_2)}=\begin{vmatrix}\dfrac{-\rho}{U_1} & \dfrac{-\rho}{(1-U_1)}\\ \dfrac{-\rho}{Z_2} & \dfrac{-\rho}{Z_2}\end{vmatrix}=\frac{\rho^2}{U_1Z_2}+\frac{\rho^2}{(1-U_1)Z_2}=\rho^2\left(\frac{Z_2}{U_1Z_2(1-U_1)Z_2}\right),\quad\text{so that}$$

$$|J|_{(\varepsilon_1,\varepsilon_2)\to(U_1,Z)}\Big|_{U_1=u_1,Z=z}=\rho^2\left(\frac{z}{u_1z(1-u_1)z}\right)=\rho^2\left(\frac{z}{\exp\left[-\left(\frac{1}{\rho}\sum_{k=1}^{2}w_k\right)\right]}\right) \tag{15}$$

Then, we may write:

$$f_{U_1,Z_2}(u_1,z)=\left\{\exp[-v]z^{\rho-2}\exp\left[-\left(\frac{1}{\rho}\sum_{k=1}^{2}w_k\right)\right]\right\}\times Q_2(v)\times\rho^2\left(\frac{z}{\exp\left[-\frac{1}{\rho}\left(\sum_{k=1}^{2}w_k\right)\right]}\right)\ \text{for } 0<\rho\leq 1, \tag{16}$$

$$=\rho^2\exp[-v]z^{\rho-1}\times Q_2(v)$$

Importantly, the above bivariate distribution of (U_1, Z_2) is entirely independent of U_1, implying that U_1 and Z_2 are independent of each other as well as that U_1 is standard uniform distributed (which then also implies that U_2 is standard uniform distributed such that $U_1 + U_2 = 1$). Further, the only way that $f_{U_1,Z_2}(u_1, z)$ can be entirely independent of U_1 is if the following holds:

$f_{U_1,Z_2}(u_1,z) = f_{U_1}(u_1) \times \dfrac{f_{U_1,Z_2}(u_1,z)}{f_{U_1}(u_1)} \Rightarrow f_{Z_2}(z) = \dfrac{f_{U_1,Z_2}(u_1,z)}{f_{U_1}(u_1)}$. That is,

$$f_{Z_2}(z) = \frac{\rho^2 \exp\left[-v\right] z^{\rho-1} \times Q_2(v)}{f_{U_1}(u_1)} = \frac{\rho^2 \exp\left[-v\right]\left(\dfrac{v}{z}\right) \times Q_2(v)}{f_{U_1}(u_1)}, \; Z_2 \in (0,\infty); f_{U_1}(u_1) = 1, \; U_1 \in (0,1); \tag{17}$$

For generalization purposes, we can also equivalently write the distribution of Z_2 as:

$$f_{Z_2}(z) = \frac{\rho^2}{f_{Dirichlet\,(U_1,U_2)}(1,1)} \exp\left[-v\right]\left(\frac{v}{z}\right) \times Q_2(v) = \frac{\rho^2}{(2-1)!} \exp\left[-v\right]\left(\frac{v}{z}\right) \times Q_2(v), \tag{18}$$

because the bivariate Dirichlet distribution Dirichlet(1,1) is exactly the same as that of two standard uniform variables whose sum is equal to 1 (and $f_{Dirichlet\,(U_1,U_2)}(1,1)$ =(2-1)!=1; for later use, note also that, in multivariate space, $f_{Dirichlet\,(U_1,U_2,\ldots U_k)}(1,1,\ldots,1) = (k-1)!$ on $\sum_{i=1}^{k} U_i = 1,\, U_i \geq 0$). The above analysis shows that the BEV distribution is exactly equivalent to the decomposition of each of the overall error terms ε_1 and ε_2 as $\varepsilon_1 = -\rho\left(\log U_1 + \log Z_2\right)$ and $\varepsilon_2 = -\rho\left(\log U_2 + \log Z_2\right)$. From well-known properties of the Dirichlet(1,1) distribution (or equivalently the uniform distribution for U_1, with $U_2 = 1 - U_1$), $E(\log U_1) = -1$, $E(\log U_2) = -1$, $Cov(\log U_1, \log U_2) = 1 - \dfrac{\pi^2}{6}$. This leaves the probability density function of $S_2 = \log(Z_2)$ to be derived in the decomposition of ε_1 and ε_2. Using the usual probability density transform result, this density function is as follows (note that $Z_2 = e^{S_2}$):

$$\begin{aligned} f_{S_2}(s) &= f_Z(z)\big|_{z=e^s} \times \frac{dz}{ds} = f_{Z_2}(z)\big|_{z=e^s} \times z = \frac{\rho^2}{(2-1)!} \exp\left[-(e^s)^{\rho}\right]\left(\frac{\left(e^s\right)^{\rho}}{z}\right) \times Q_2(v) \times z \\ &= \frac{\rho^2}{(2-1)!} \exp\left[-v\right]\left(v\right) \times Q_2(v), \text{with } v = z^{\rho} = \left(e^s\right)^{\rho} = e^{s\rho} \\ &= \rho^2 \exp\left[-v\right]\left(v\right) \times \left(v - \left(\frac{\rho-1}{\rho}\right)\right) \end{aligned} \tag{19}$$

Note that the above distributions for S_k is not a log-Gamma distribution because of the presence of the polynomial $Q_2(v)$, which is dimension driven (in the case of the BEV, the dimension is 2). This distinction is important because a log-Gamma representation would imply a simpler exponential-family structure that does not capture the nested dependence induced by the MEV construction. But, as shown later, the S_k distribution for general k exhibits a collection of remarkably structured properties—most notably a closed-form moment generating function and cumulant expressions—that align in an unexpected elegant and highly coherent manner with the geometry of the Dirichlet distribution. These properties provide a novel and parsimonious characterization of the MEV distribution that is not apparent from standard formulations. But, following through first on the BEV distribution, and using basic integration principles, it is straightforward (even if tedious) to show the following:

$$\begin{aligned} E(S_2) &= \int_{s=-\infty}^{s=\infty} s \times \rho^2 \ \exp[-v] v \times \left(v - \left(\frac{\rho - 1}{\rho} \right) \right) ds, \ v = e^{s\rho} \\ &= \left[-(\rho s v + s + 1) e^{-v} - \frac{\Gamma(0, v)}{\rho} \right] \Bigg|_{v=-\infty}^{v=\infty}, \end{aligned} \tag{20}$$

where $\Gamma(0,h) = \int_{t=h}^{t=\infty} \frac{e^{-t}}{t} dt$ is the upper incomplete Gamma function. Evaluating the many terms at the limits, we get the following result:

$$E(S) = 1 - \frac{\gamma}{\rho}, \quad \text{where } \gamma \text{ is the Euler's constant.} \tag{21}$$

Similarly, the following result may be obtained:

$$Var(S_2) = E(S_2^2) - \left[E(S_2) \right]^2 = \frac{1}{\rho} \left[\left(\frac{\pi^2}{6} + \gamma^2 \right) \left(\frac{1}{\rho} \right) - 2\gamma \right] - \left(1 - \frac{\gamma}{\rho} \right)^2 = \frac{\pi^2}{6\rho^2} - 1 \tag{22}$$

As required then,

$$\begin{aligned} E(\varepsilon_1) &= E\left[-\rho \left(\log U_1 + \log Z_2 \right) \right] = -\rho \left[E(\log U_1) + E\left(\log Z_2 \right) \right] = -\rho \left[-1 + 1 - \frac{\gamma}{\rho} \right] = \gamma, \\ \text{Var}(\varepsilon_1) &= \rho^2 \left[Var(\log U_1) + Var(\log Z_2) \right] = \rho^2 \left[1 + \frac{\pi^2}{6\rho^2} - 1 \right] = \frac{\pi^2}{6}, \\ Cov(\varepsilon_1, \varepsilon_2) &= \rho^2 \left[Cov\left[\left(\log U_1 + \log Z_2 \right), \left(\log U_2 + \log Z_2 \right) \right] \right] \\ &= \rho^2 \left[Cov\left(\log U_1, \log U_2 \right) + Var(\log Z_2) \right] = \rho^2 \left[1 - \frac{\pi^2}{6} + \frac{\pi^2}{6\rho^2} - 1 \right] = \frac{\pi^2}{6} \left[1 - \rho^2 \right], \text{ and} \\ \text{Cor}(\varepsilon_1, \varepsilon_2) &= 1 - \rho^2 \end{aligned} \tag{23}$$

The CDF of S_2 is also available in a closed form as follows (this can be obtained by straightforward integration of the density function; the derivation of a general expression for any number of variates in the MEV distribution is provided in the next section, also in closed form):

$$\begin{aligned} F_{S_2}(s) &= \int_{t=-\infty}^{t=s} \frac{\rho^2}{(1-1)!} \exp\left[-(e)^{t\rho} \right] e^{t\rho} \times \left(e^{t\rho} - \left(\frac{\rho - 1}{\rho} \right) \right) dt \\ &= 1 - \exp\left[-(e)^{s\rho} \right] \left[\rho e^{s\rho} + 1 \right] = 1 - \exp(-v) \left[\rho v + 1 \right], \ v = e^{s\rho} \end{aligned} \tag{24}$$

4. A NEW CHARACTERIZATION OF THE MEV DISTRIBUTION

We next generalize the bivariate transformation and decomposition to the full multivariate MEV distribution, showing that the same structural logic carries through for arbitrary *k*. To do so, we first characterize the PDF and CDF of the variable S_k corresponding to the MEV distribution with *k*-variates. Starting from Equation (4) for the MEV distribution, and following exactly the same procedure as for the BEV distribution, we may write:

$$\varepsilon_i = -\rho\left(\log(U_i) + \log Z_k\right) = -\rho\left(\log(U_i) + S_k\right),\ i = 1,2,...,k\ . \tag{25}$$

The PDF for $U_1, U_2, ..., U_k, Z_k$ may be obtained, again following the same procedure as for the BEV distribution in the previous section, as:

$$f_{U_1,U_2,...,U_k,Z_k}(1,1,...1,z) = f_{Dirichlet\,(U_1,U_2,...,U_k)}(1,1,...,1) \times f_{Z_k}(z),\ \text{where}$$

$$\begin{aligned} f_{Z_k}(z) &= \frac{\rho^k}{f_{Dirichlet\,(U_1,U_2,...,U_k)}(1,1,...,1)} \exp[-v]\left(\frac{v}{z}\right) \times Q_k(v) \\ &= \frac{\rho^k}{(k-1)!} \exp[-v]\left(\frac{v}{z}\right) \times Q_k(v),\ v = z^{\rho} \end{aligned} \tag{26}$$

Importantly, the PDF for Z_k changes based on the number of variates *k* in the MEV distribution because of the presence of the polynomial $Q_k(v)$ that is dependent on *k*. Then, the probability density function of $S_k = \log(Z_k)$ obtained from a variable transform is:

$$\begin{aligned} f_{S_k}(s) &= f_{Z_k}(e^s) \times \frac{dz}{ds} = f_{Z_k}(e^s) \times z = \frac{\rho^k}{(k-1)!} \exp[-v]\left(\frac{v}{z}\right) \times Q_k(v) \times z \\ &= \frac{\rho^k}{(k-1)!} \exp[-v](v) \times Q_k(v)\ ,\ \text{with } v = z^{\rho} = \left(e^s\right)^{\rho} = e^{s\rho} \end{aligned} \tag{27}$$

This PDF is easily generated because the form of $Q_k(v)$ is the same as in Equation (8) and the coefficients within $Q_k(v)$ can be generated in a straightforward manner (see Equations (10)-(12)).

Theorem 1: (Closed Form CDF for S_k)

The CDF for S_k admits a closed form $F_{S_k}(s) = \left(1 - e^{-v}\left[\left(\frac{\rho^{k-1}}{(k-1)!}\right) R_k(v)\right]\right)$, where $R_k(v)$ is a polynomial of degree *k*–1 that may be written as $R_k(v) = \sum_{j=0}^{k-1} b_{k,j}\left(\frac{1}{\rho^{(k-j-1)}}\right) v^j$ with $R_1(v) = 1$. With the coefficients $a_{k,j}$ from the recursion of $Q_k(v)$, the polynomial (in coefficients $b_{k,j}$ in the CDF satisfy the backward recursion $b_{k,j} - (j+1)b_{k,j+1} = a_{k,j}$ for $j = 0,1,\ldots,k-2$, with boundary condition $b_{k,k} = 0$. Note that $b_{k,k} = 0$ because $R_k(v)$ is a polynomial in *v* of degree *k*–1, and has no v^k term (so $b_{k,k-1} = a_{k,k-1}$). Also, $b_{k,j}$ is itself a polynomial in ρ of degree at most $k-1-j$.

Proof: See Appendix C.

From the recursion for the $b_{k,j}$ coefficients, one can compute the specific scalar coefficient $b_{k,j,h}$ corresponding to each polynomial degree of ρ within the polynomial $b_{k,j}$.

Proposition 3: Recursivity of polynomial coefficients on ρ within the polynomial $b_{k,j}$

Write $b_{k,j} = \sum_{h=0}^{k-1-j} b_{k,j,h}\rho^{k-1-j-h}$ ($j = 0,1,\ldots,k-1,\ h = 0,1,\ldots,k-1-j$). The $b_{k,j,h}$ coefficients are now scalar values entirely free of ρ, and follow the recursion below:

$$b_{k,j,h} - (j+1)b_{k,j+1,h-1} = a_{k,j,h} \text{ with } b_{k,j,0} = a_{k,j,0},\ b_{k,k,h} = 0, \text{ and } b_{k,j,-1} = 0. \tag{28}$$

Proof: See Appendix D.

The polynomial $R_1(v)$ in $F_{S_k}(s)$ may be computed directly from the $b_{k,j,h}$ scalar coefficients as:

$$R_k(v) = \sum_{j=0}^{k-1} b_{k,j}\left(\frac{1}{\rho^{(k-j-1)}}\right)v^j = \left[\sum_{j=0}^{k-1}\sum_{h=0}^{k-1-j} b_{k,j,h}\rho^{k-1-j-h}\left(\frac{1}{\rho^{(k-j-1)}}\right)\right]v^j = \sum_{j=0}^{k-1}\left(\sum_{h=0}^{k-1-j} b_{k,j,h}\rho^{-h}v^j\right). \tag{29}$$

Using the recursions for the coefficients in the polynomials $Q_k(v)$ for the PDF of S_k and $R_k(v)$ as just developed above, we present the table of coefficients for *k*=2 through *k*=6 in Table 1.

Next, to provide a closed-form expression for the moment generating function (MGF) of S_k (needed to obtain the closed-form expression for the cumulant generating function (CGF) of the MEV distribution), the following Lemma is needed.

Lemma 1: Define the following integral for $x > -1$:

$$I_k(x) = \int_0^\infty e^{-v}v^x Q_k(v)dv. \tag{30}$$

Then $I_k(x)$ satisfies the following regression:

$$I_k(x) = \left(x + \frac{k-1}{\rho}\right)I_{k-1}(x),\ k \geq 2. \tag{31}$$

Proof: See Appendix E.

Theorem 2: The Moment generating function of S_k is:

$$\frac{1}{(k-1)!}\Gamma\left(1+\frac{t}{\rho}\right)\prod_{j=1}^{k-1}(t+j),\ t > -\rho, \text{ and where } \Gamma(h) = \int_{l=0}^{l=\infty} l^{h-1}e^{-l}dl,\ h > 0 \tag{32}$$

Proof: See Appendix F.

Theorem 3: Let S_k have moment generating function as provided and proved in Proposition 1. Define generalized Harmonic numbers as follows:

$$H_{k-1}^{(m)} = \sum_{j=1}^{k-1} \frac{1}{j^m}. \tag{33}$$

Then, the cumulant generating function of S_k is:

$$\kappa_n(S_k) = \frac{1}{\rho^n}\psi^{(n-1)}(1) + (-1)^{n-1}(n-1)! H_{k-1}^{(n)},\ n \geq 1. \tag{34}$$

where ψ is the digamma function $\psi(x) = \frac{d}{dx}\log\Gamma(x)$, with its derivatives being the polygama functions given by $\psi^m(x) = \frac{d^m}{dx^m}\psi(x) = \frac{d^{m+1}}{dx^{m+1}}\log\Gamma(x)$.

Proof: See Appendix G.

Corollary 1: Using the above expression for the cumulants, the first four cumulants of S_k for any number of variates *k* may be obtained as follows, with γ being the Euler's constant and $\zeta(.)$ the Riemann zeta function[2].

$$\mathbb{E}[S_k] = \kappa_1(S_k) = H_{k-1} - \frac{\gamma}{\rho},\ \text{because } \psi^0(1) = \psi(1) = -\gamma$$

$$\text{Var}(S_k) = \kappa_2(S_k) = \frac{\pi^2}{6\rho^2} - H_{k-1}^{(2)},\ \text{because } \psi^1(1) = \frac{\pi^2}{6} \tag{35}$$

$$\text{Skew}(S_k) = \frac{\kappa_3(S_k)}{\left(\kappa_2(S_k)\right)^{3/2}}\ ;\ \kappa_3(S_k) = 2H_{k-1}^{(3)} - \frac{2\zeta(3)}{\rho^3},\ \text{because } \psi^2(1) = -2\zeta(3)$$

$$\text{Kurt}(S_k) = \frac{\kappa_4(S_k) + 3\left[\kappa_2(S_k)\right]^2}{\left[\kappa_2(S_k)\right]^2};\ \kappa_4(S_k) = -6H_{k-1}^{(4)} + \frac{\pi^4}{15\rho^4},\text{because } \psi^3(1) = \frac{\pi^4}{15}.$$

Figure 1 plots the density distribution of S_k for $k = 1,2,...,5$. As may be observed, S_k is left-skewed, so that $-\rho S_k$ (as it appears in the decomposition of the extreme value error in Equation (25)) is right-skewed and adds up with the $-\rho\log U_k$ term to provide the univariate standard extreme value error). Also, as the number of variates *k* in the MEV distribution increases, the variance of S_k decreases, because $\text{Var}(S_k) = \frac{\pi^2}{6\rho^2} - H_{k-1}^{(2)}$, and $H_{k-1}^{(2)} = \sum_{j=1}^{k-1}\frac{1}{j^2}$ increases with *k*. Essentially, S_k constitutes an increasing aggregation of extreme-value components, which leads to greater stability (less volatility) in the common nest-level factor as *k* increases. At the same time, the variance of each log-Dirichlet component, which is given by $Var(\log U_i) = H_{k-1}^{(2)}\ \forall i = 1,2,...k$, increases, as does the pairwise covariance between the log-Dirichlet components

[2] The Reimann zeta function takes the following form:
$\zeta(d) = \sum_{n=1}^{\infty}\frac{1}{n^d},\ d > 1;\ \zeta(2) = \frac{\pi^2}{6},\ \zeta(3) \approx 1.2020569,\ \zeta(4) = \frac{\pi^4}{90}.$

$\left(Cov(\log U_i, \log U_j) = H_{k-1}^{(2)} - \frac{\pi^2}{6} = Var(\log U_i) - \frac{\pi^2}{6} \right)$. Thus, enlarging the nest increases both the marginal dispersion and the cross-covariance of the log-Dirichlet components, implying richer and more diffuse substitution patterns across alternatives.[3] The intriguing structural feature of the decomposition is that the harmonic terms appearing in the variance of S_k are exactly offset by the harmonic terms in the pairwise covariance of the log-Dirichlet components. That is, extending Equation (23) to the more general case beyond the BEV, we get the following for the MEV:

$$\begin{aligned} Cov(\varepsilon_i, \varepsilon_j) &= \rho^2 \left[Cov\left[\left(\log U_i + S_k\right), \left(\log U_j + S_k\right) \right] \right] \\ &= \rho^2 \left[Cov\left(\log U_i, \log U_j\right) + Var(S_k) \right] = \rho^2 \left[H_{k-1}^{(2)} - \frac{\pi^2}{6} + \frac{\pi^2}{6\rho^2} - H_{k-1}^{(2)} \right] = \frac{\pi^2}{6}\left[1 - \rho^2\right], \end{aligned} \tag{36}$$

Because of this exact cancellation in $Cov(\varepsilon_1, \varepsilon_2)$, the pairwise correlation of the resulting MEV errors remains invariant at $1-\rho^2$, independent of the number of alternatives in the nest. Importantly, this cancellation is not confined to second moments. The harmonic contributions also offset in the higher-order marginal and multivariate cumulants of the log-Dirichlet and S_k components. This property enables a complete higher-order characterization of the MEV distribution through simple closed-form expressions, as established in the next two theorems.

Theorem 4: Let $(\varepsilon_1, \varepsilon_2, ... \varepsilon_k)$ follow the multivariate extreme value (MEV) distribution with dependence parameter $\rho \in (0,1]$. Then, for any *k*, the univariate marginal cumulant is given by: $(-1)^n \psi^{(n-1)}(1)$.

Proof: See Appendix H.

Corollary 2: The first four univariate cumulants of the MEV errors satisfy the following, as one would expect given the univariate distributions correspond to the univariate Type 1 extreme-value (Gumbel) distribution:

$$E(\varepsilon_i) = \kappa_1(\varepsilon_i) = (-1)^1 \psi^0(1) = (-1)\psi(1) = -1(-\gamma) = \gamma$$

$$Var(\varepsilon_i) = \kappa_2(\varepsilon_i) = (-1)^2 \psi^1(1) = \frac{\pi^2}{6}$$

$$\text{Skew}(\varepsilon_i) = \frac{\kappa_3(\varepsilon_i)}{\left(\kappa_2(\varepsilon_i)\right)^{3/2}} \; ; \; \kappa_3(\varepsilon_i) = (-1)^3 \psi^2(1) = 2\zeta(3) \tag{37}$$

$$\text{Kurt}(\varepsilon_i) = \frac{\kappa_4(\varepsilon_i) + 3\left[\kappa_2(\varepsilon_i)\right]^2}{\left[\kappa_2(\varepsilon_i)\right]^2} ; \; \kappa_4(\varepsilon_i) = (-1)^4 \psi^3(1) = \frac{\pi^4}{15}.$$

[3] Intuitively speaking, the common component S_k captures the overall intensity of dependence across the MEV distributed terms (controlling common extreme-value aggregation), while the Dirichlet components (controlling allocation geometry across alternatives in the nest) reflect the slicing of that overall intensity across the many dependent components.

The MEV errors can also be characterized in a surprisingly simple closed-form in higher dimensions also as follows:

Theorem 5: Let $\left(\varepsilon_1,\varepsilon_2,...,\varepsilon_k\right)$ follow the multivariate extreme value (MEV) distribution with dependence parameter $\rho \in (0,1]$. Then, for any k and for distinct indices $i_1, i_2, ..., i_m$, $m \geq 2$, the joint multivariate cumulant of order *m* is given by:

$$(m-1)!\zeta(m)\left[1-\rho^m\right]. \tag{38}$$

Proof: See Appendix I.

This result reveals that the MEV dependence structure admits a remarkably compact higher-order representation, independent of *k*, reinforcing the structural parsimony of the decomposition.

Corollary 3: The second-order multivariate cumulant (that is, the bivariate cumulant or the pairwise covariance) for any pairing ε_i and $\varepsilon_j \; (i \neq j)$ in an MEV distribution is:

$$\kappa_2\left(\varepsilon_i,\varepsilon_j\right) = Cov(\varepsilon_i,\varepsilon_j) = (2-1)!\zeta(2)\left[1-\rho^2\right] = \frac{\pi^2}{6}\left[1-\rho^2\right] \tag{39}$$

The correlation between any two errors ε_i and $\varepsilon_j \; (i \neq j)$ in an MEV distribution is unambiguously given by:

$$Cor(\varepsilon_i,\varepsilon_j) = \frac{Cov(\varepsilon_i,\varepsilon_j)}{\sqrt{Var(\varepsilon_i)}\sqrt{Var(\varepsilon_j)}} = \frac{\frac{\pi^2}{6}\left[1-\rho^2\right]}{\sqrt{\frac{\pi^2}{6}}\sqrt{\frac{\pi^2}{6}}} = 1-\rho^2 \tag{40}$$

The third-order multivariate cumulant (and the skew) for any distinct combination of three errors is:

$$\begin{aligned} &\kappa\left(\varepsilon_i,\varepsilon_j,\varepsilon_l\right) = (3-1)!\zeta(3)\left[1-\rho^3\right] = 2\zeta(3)\left[1-\rho^3\right], \\ &\text{with skew} = \frac{\kappa_3\left(\varepsilon_i,\varepsilon_j,\varepsilon_l\right)}{\sqrt{\kappa_2\left(\varepsilon_i\right)\kappa_2\left(\varepsilon_j\right)\kappa_2\left(\varepsilon_l\right)}} = \frac{2\zeta(3)\left[1-\rho^3\right]}{\left(\pi^2/6\right)^{3/2}} = \frac{12\sqrt{6}}{\pi^3}\zeta(3)\left[1-\rho^3\right] \end{aligned} \tag{41}$$

The fourth-order multivariate cumulant (and the Kurtosis) for any distinct combination of four errors is:

$$\begin{aligned} &\kappa_4\left(\varepsilon_i,\varepsilon_j,\varepsilon_l,\varepsilon_h\right) = (4-1)!\zeta(4)\left[1-\rho^4\right] = 6\zeta(4)\left[1-\rho^4\right] = \frac{\pi^4}{15}\left[1-\rho^4\right], \\ &\text{with kurtosis} = \frac{\kappa_4\left(\varepsilon_i,\varepsilon_j,\varepsilon_l,\varepsilon_h\right)}{\sqrt{\kappa_2\left(\varepsilon_i\right)\kappa_2\left(\varepsilon_j\right)\kappa_2\left(\varepsilon_l\right)\kappa_2\left(\varepsilon_h\right)}} = \frac{36}{15}\left[1-\rho^4\right] = \frac{12}{5}\left[1-\rho^4\right] \end{aligned} \tag{42}$$

Importantly, while the MEV structure is typically characterized by its covariance structure, its higher-order dependence reveals a remarkably simple closed factorial zeta form entirely independent of *k*. With this, we have now fully characterized the MEV error structure.

5. GENERATING REALIZATIONS FROM THE MEV DISTRIBUTION

An exact draw from the MEV distribution with *k* variates can be obtained by first sampling from a Dirichlet $(1,1,...,1)$ with *k* variates, then sampling the scalar S_k from its derived distribution, and finally applying the transformation $\varepsilon_i = -\rho(\log U_i + S_k)$. The nonstandard component of this procedure is the S_k-distribution, which has closed form CDF and moments. One approach is to employ the inverse transform method, while an alternative is to exploit a finite Gamma-mixture representation of the S_k-distribution. The first employs inverse transform sampling based on the closed-form cumulative distribution function. This approach requires solving a one-dimensional nonlinear equation for each draw, typically via safeguarded Newton or bracketing methods. While analytically exact (up to floating-point precision), its computational cost grows with the number of required realizations due to repeated root-finding operations. The second exploits the finite Gamma-mixture representation of the S_k-distribution. In this formulation, each draw reduces to (i) sampling a discrete mixture component and (ii) generating a Gamma variate with integer shape parameter. This method avoids root-finding entirely and relies only on standard uniform and Gamma random number generators. Consequently, it is computationally more efficient and numerically stable for large-scale simulation, particularly when the number of required draws is substantial.

Both approaches are analytically exact, with discrepancies arising solely from finite-precision arithmetic, and both can be implemented using standard random number generation routines, as discussed next.

5.1. The Inverse Transform Simulation

This is a rather straightforward, albeit mechanical, approach, starting from the closed form CDF expression given in Theorem 1. The inverse transform method involves solving the following function for s:

$$F_{S_k}(s) = \left(1 - e^{-v}\left[\left(\frac{\rho^{k-1}}{(k-1)!}\right)R_k(v)\right]\right) = \tilde{U}, \text{ or } \left(e^{-v}\left[\left(\frac{\rho^{k-1}}{(k-1)!}\right)R_k(v)\right]\right) = 1 - \tilde{U}, \text{ where } \tilde{U} \in (0,1) \quad (43)$$

The steps are as follows:

(1) Draw a *k*-dimensional Dirichlet random vector: $(U_1,\ldots,U_k) \sim \text{Dirichlet}(1,\ldots,1)$

(2) Draw a standard uniform random variate $\tilde{U} \sim \text{Uniform}(0,1)$

(3) Solve for $v > 0$, the function $G(v) = \left(e^{-v}\left[\left(\frac{\rho^{k-1}}{(k-1)!}\right)R_k(v)\right]\right) - \left(1 - \tilde{U}\right) = 0$, where the polynomial $R_k(v)$ is developed as presented in Theorem 1. Note that because $dG(v)/dv < 0 \ \forall \ v > 0$ (given $\rho > 0$), the function $G(v)$ is strictly monotone, and the equation admits a unique solution $v^* > 0$. Any standard root-finding method (such as bisection or Newton-Raphson) may be used. The equation may be solved for v $(v > 0)$ by taking logarithms to yield the following:

$$\log G(v) = \log(C) + \log R_k(v) - v - \log(1 - \tilde{U}) = 0, \text{ where } C = \frac{\rho^{k-1}}{(k-1)!} \quad (44)$$

(4) Transform to obtain a realization of $S_k = \frac{1}{\rho}\log v$

(5) Construct MEV errors $\varepsilon_i = -\rho\left(\log U_i + S_k\right),\ i = 1,\ldots,k$

The inverse transform constitutes an exact simulation approach that is conceptually transparent. As discussed earlier, it does require numerical root finding for each draw in the third step, though it may be used for small-scale, small-dimensional MEV models. Given its well-established approach, it can also serve as a validation check for other simulation mechanisms.

5.2. Finite Gamma-Mixture Approach

The availability of a finite Gamma-mixture representation is not incidental but follows directly from the log Dirichlet/ S_k decomposition developed in this paper. Specifically, the polynomial structure of $Q_k(v)$, which emerges from the recursive characterization of the S_k density, implies that the distribution of $v = e^{S_k \rho}$ can be expressed as a finite mixture of Gamma distributions with integer shape parameters. In turn, it yields a simulation mechanism that is both exact and computationally efficient, avoiding iterative inversion procedures. Specifically, from Equation (27), the density function of S_k is:

$f_{S_k}(s) = \frac{\rho^k}{(k-1)!}\exp\left[-v\right]\left(v\right)\times Q_k(v)$, with $v = z^\rho = \left(e^s\right)^\rho = e^{s\rho}$; that is $s = \frac{1}{\rho}\log v$. Then,

$$\begin{aligned} f_{V_k}(v) &= f_{S_k}\left(\frac{1}{\rho}\log v\right)\frac{ds}{dv} = f_{S_k}\left(\frac{1}{\rho}\log v\right)\frac{1}{\rho v} \\ &= \frac{\rho^{k-1}}{(k-1)!}\exp(-v)Q_k(v),\ v > 0. \end{aligned} \tag{45}$$

Substituting $Q_k(v) = \sum_{j=0}^{k-1} a_{k,j}\rho^{-(k-1-j)}v^j$,

$$f_{V_k}(v) = \sum_{j=0}^{k-1}\frac{\rho^{k-1}}{(k-1)!}\exp(-v)a_{k,j}\rho^{-(k-1-j)}v^j = \sum_{j=0}^{k-1}\frac{\rho^j}{(k-1)!}a_{k,j}\exp(-v)v^j \tag{46}$$

Consider the gamma distribution with shape j+1 (j being a positive integer) and rate 1:

$g_{j+1}(v) = \frac{\exp(-v)v^j}{j!}$, $v > 0$, so that

$$\begin{aligned} f_{V_k}(v) &= \sum_{j=0}^{k-1}\left(\frac{\rho^j}{(k-1)!}a_{k,j}j!\right)g_{j+1}(v) \\ &= \sum_{j=0}^{k-1}\pi_{k,j}g_{j+1}(v),\ \text{with mixture weights } \pi_{k,j} = \left(\frac{\rho^j}{(k-1)!}a_{k,j}j!\right) \end{aligned} \tag{47}$$

Importantly, note that $\sum_{j=0}^{k-1} \pi_{k,j} = 1$ automatically, because

$$\int_{v=0}^{\infty} f_{V_k}(v) = 1 = \int_{v=0}^{\infty} \sum_{j=0}^{k-1} \pi_{k,j} g_{j+1}(v) = \sum_{j=0}^{k-1} \pi_{k,j} \int_{v=0}^{\infty} g_{j+1}(v) = \sum_{j=0}^{k-1} \pi_{k,j} .\, 1 = \sum_{j=0}^{k-1} \pi_{k,j} \tag{48}$$

Also, each $\pi_{k,j} \geq 0$ from the fact that $a_{k,j} \geq 0 \; \forall \; k, j$ and $\rho > 0$. Thus, the mixture weights are finite, positive, and, as required, sum to one. With the above, the steps for generating MEV errors are as follows:

(1) Draw a k-dimensional Dirichlet random vector: $(U_1, \ldots, U_k) \sim \text{Dirichlet}(1, \ldots, 1)$. This may be accomplished via exponential draws. So, first draw independent uniform random variates $W_1, W_2, \ldots W_i, \ldots, W_k$, next construct exponential draws as $E_i = -\log(W_i)$, and finally obtain realizations $U_i \; (i = 1, 2, \ldots, k)$ as $E_i / \sum_{i=1}^{k} E_i$.

(2) Develop the mixture weights $\pi_{k,j} \; (j = 0, 1, 2, \ldots k-1)$, and construct $c_{k,j} = \sum_{h=0}^{j} \pi_{k,j}$, $c_{k,k-1} = 1$.

(3) Draw a uniform number $\tilde{U} \sim \text{Uniform}(0,1)$, and select a specific component mixture J as $\min\{j : \tilde{U} \leq c_{k,j}\}$

(4) Draw a realization v from $\text{Gamma}(J+1, 1)$.

(5) Transform to obtain a realization of $S_k = \frac{1}{\rho} \log v$

(6) Construct MEV errors $\varepsilon_i = -\rho\left(\log U_i + S_k\right),\; i = 1, \ldots, k$

As such, in our testing of the inverse transform and the Gamma mixture methods, the Gamma mixture method is easily the generation mechanism of choice for realizations from the MEV distribution.

6. VALIDATION EXERCISES

The exact generation mechanism described above may be validated in a number of ways. For illustration, Figure 2 begins by showing the density function for $A_i = -\rho \log(U_i)$, which is an exponential variable with rate $1/\rho$, so that $f_A(x) = \left(1/\rho\right) e^{-(x/\rho)}$. Also, the density function of $B_k = -\rho S_k$ is $f_{B_k}(x) = \left(1/\rho\right) f_{S_k}(-x/\rho)$. The sum of these provides the density function of the standard extreme value variate (note that the B_k distribution lies to the left of the extreme-value distribution, but then gets pulled by the positive exponential distribution of A_i to provide the extreme-value distribution). To validate the generation mechanism, we simulate MEV variates (with $k = 5$, $\rho = 0.7$, and 1000,000 realizations) using the gamma-mixture approach, and take the first variate and overlay the resulting distribution of the realizations as a step function with a bin size of 0.025. As can be observed from Figure 2, the step function distribution tracks the standard extreme value function accurately (except for the minor coarseness due to the binning of the generated realizations).

Next, Table 2 computes the mean, variance, skew, and kurtosis at the univariate level (again, and without any loss of generalizability, taking the realizations for the first variate) after generating from a *k*-variate MEV distribution (all generations are done using the Gamma-mixture approach, because the inverse transform is much slower, especially at $k \geq 5$). These univariate statistics are compared relative to their known true values, as given in the expressions in Equation (36) for the univariate standard extreme value distribution. Table 2 also computes the pairwise correlation (between the realizations for the first two variates), three-way skew (among the realizations for the first three variates), and four-way kurtosis (among the realizations for the first four variates) after generating from a *k*-variate MEV distribution, and then compares these to known theoretical values as provided in Equations (39)-(41). These evaluations are done for three different values of the dissimilarity parameter ρ (for $\rho = 0.2, 0.5$, and 0.8), which are arranged as three row panels in Table 2. For each ρ value, we generate realizations for five different values of *k* ($k = 5, 10, 15, 20$ and 100), as shown in the second column. The third main column provides the true univariate statistics and the computed (from 1000,000 generated realizations of the *k*-variate MEV distribution) statistics, while the fourth main column provides the corresponding multivariate statistics up to order four. As may be observed, both the univariate and multivariate statistics line up very well with their true values, regardless of the number of variates *k*, demonstrating the robustness and exactness of our generating procedure.[4]

The last column of Table 2 provides the computational speed of the exact Gamma-mixture generator. All runs were undertaken on a notebook computer with a 12th Gen Intel(R) Core(TM) i7-12800H 2.40 GHz processor with an installed RAM of 64 GB. The dissimilarity parameter ρ has literally no effect on computational speed. The number of variates, however, does, with an additional 0.21-0.23 seconds for each additional five variates from $k = 5$ to $k = 20$. But the increase in computational time is less than linear as we go to even higher *k* values.

7. CONCLUSIONS

This paper develops a structural decomposition of the multivariate extreme value (MEV) distribution underlying the nested logit model. While some earlier studies have already recognized that the nested structure may be viewed from a decomposition-of-errors standpoint, there has been no clear theoretical link between the decomposition and the MEV CDF structure of Equation (2). In contrast, by working directly from the MEV cumulative distribution function, we derive an explicit polynomial recursion for the density and obtain a representation in which each disturbance decomposes into a common log-aggregation component and a log-Dirichlet allocation term. This decomposition yields closed-form expressions for the moment generating and cumulant generating

[4] We should note here that the univariate statistics lined up very well with the true univariate values of mean and variance for even *k*=100 with as few as 10,000 draws. As a sample illustration, from one run with 10,000 draws for *k*=100 and *ρ*=0.2, the univariate mean for the first variate was 0.58264 (compared to the true value of 0.57722) and the variance was 1.6448 (relative to the true value of 1.6449)). However, even at this univariate level, the computed values were quite off for skew and kurtosis (skew computed value of 1.2091 relative to the true value of 1.1395, and the kurtosis computed value of 6.0102 relative the true value of 5.4000). At the multivariate level, the correlation between the first two variates was close to the true value (computed value of 0.9603 relative to the true value of 0.9600), but the three-way skew and four-way kurtosis were again quite off (computed skew value of 1.1834 relative to the true value of 1.1304 and computed kurtosis value of 2.8633 relative to the true value of 2,3962). These results caution against making any substantial conclusions from validation exercises for MEV generators that only compare the mean and variance at the univariate level, and the pairwise correlations at the bivariate level. It is important to undertake validation for MEV generators at higher orders at both the univariate and multivariate levels, as undertaken in the current paper.

functions, enabling a complete higher-order characterization of the MEV distribution. An easy-to-implement and analytically exact simulation procedure is proposed based on the decomposition, using a Gamma-mixture approach. Codes for computing the PDF and generating MEV errors are available upon request from the author.

A central insight is that while the variance of the common aggregation term decreases with the number of alternatives in a nest, harmonic terms cancel exactly in pairwise covariances, leaving within-nest correlation invariant at $\left(1-\rho^2\right)$(assuming a simple nesting structure). More broadly, all multivariate cumulants admit a compact factorial–zeta form depending only on the dissimilarity parameter. These results clarify the structural role of the generating function and reveal a higher-order dependence structure that is not visible from second-moment analysis alone.

The analysis in this paper has focused on the standard one-level nested logit model. However, the structural decomposition derived here suggests a natural pathway for extending the approach to more flexible MEV systems, including multi-level nesting and other general nested logit structures such as the Ordered GEV (OGEV) model (Small, 1987), the Paired Combinatorial Logit (PCL) model (Chu, 1989; Koppelman and Wen, 2000), the Cross-Nested Logit (CNL) model (Vovsha, 1997), the Product Differentiation Logit (PDL) model (Breshanan et al., 1997), the Multinomial Logit-Ordered GEV (MNL-OGEV) model (Bhat, 1998), the spatially correlated logit (Bhat and Guo, 2004), the ordered GEV-nested logit (OGEV-NL) model (Whelan et al., 2002), the network Generalized Extreme Value (NGEV) model (Daly and Bierlaire, 2006), and the Generalized Nested Logit model (Wen and Koppelman, 2001). In multi-level nesting structures, the generating function is constructed recursively from lower-level aggregates. The present results indicate that at each level of aggregation, an analogous decomposition into a common log-sum component and a Dirichlet-type allocation term may be obtainable. Such a hierarchical decomposition would imply a layered common-factor representation, with harmonic cancellations occurring at each nesting level. In the general nested logit structures, where alternatives exhibit richer dependency patterns, the generating function no longer decomposes into disjoint power sums. Nevertheless, the polynomial recursion derived here suggests that a generalized allocation component—possibly involving weighted Dirichlet structures—could be constructed to represent the joint distribution. Establishing such a decomposition would allow closed or semi-closed form for higher-order cumulants and exact simulation procedures (from an analytic standpoint) for these general nested logit structures. Exploring these extensions would deepen understanding of dependence structures and may yield new analytical tools for complex substitution systems.

ACKNOWLEDGMENTS: This research was partially supported by the U.S. Department of Transportation through the Center for Understanding Future Travel Behavior and Demand (TBD) (Grant No. 69A3552344815 and No. 69A3552348320). The author would like to thank Anna Beliveau for help with plotting Figures 1 and 2, and Lisa Macias for help in formatting this document. The author would also like to thank Prof. Patricia Mokhtarian for recent discussions related to dependence patterns in nested logit models.

REFERENCES

Ben-Akiva, M. and Lerman, S.R. (1979). Disaggregate travel and mobility choice models and measures of accessibility. In Hensher, D.A. and Stopher, P.R. (Eds.) *Behavioral Travel Modelling*. 1st ed. Routledge, London. https://doi.org/10.4324/9781003156055.

Ben-Akiva, M. and Lerman, S.R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*. The MIT Press, Cambridge.

Bhat, C.R. (1995). A heteroscedastic extreme-value model of intercity mode choice. *Transportation Research B*, 29(6), 471–483. https://doi.org/10.1016/0191-2615(95)00015-6.

Bhat, C. R. (1998). Accommodating variations in responsiveness to level-of-service measures in travel mode choice modeling. *Transportation Research Part A*, 32(7), 495–507. https://doi.org/10.1016/S0965-8564(98)00011-1.

Bhat, C.R. and Guo, J. (2004) A mixed spatially correlated logit model: Formulation and application to residential choice modeling. *Transportation Research Part B*, 38(2), 147–68. https://doi.org/10.1016/S0191-2615(03)00005-5.

Bresnahan, T.F., Stern, S., and Trajtenberg, M. (1997). Market segmentation and the sources of rents from innovation: Personal computers in the late 1980s. *RAND Journal of Economics*, 28(0), S17–S44. https://www.jstor.org/stable/3087454.

Bunch, D.S. and Rocke, D.M. (2016) Variance-component-based nested logit specifications: Improved formulation, and practical microsimulation of random disturbance terms. *Journal of Choice Modelling*, 21, 30–35. https://doi.org/10.1016/j.jocm.2016.04.001.

Carrasco, J.A. and Ortúzar, J. de. D. (2002). Review and assessment of the nested logit model. *Transport Reviews*, 22(2), 197–218. https://doi.org/10.1080/01441640110091224.

Cardell, N.S. (1997). Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity. *Economic Theory*, 13(2), 185–213. http://www.jstor.org/stable/3532724.

Chu, C. (1989) A paired combinational logit model for travel demand analysis, in *Transport Policy, Management and Technology Towards 2001: Selected Proceedings of the Fifth World Conference on Transport Research*, 4, 295–309, Western Periodicals, Ventura, CA.

Daganzo, C. (1979). *Multinomial Probit: The Theory and its Application to Demand Forecasting*. Academic Press, New York. https://doi.org/10.1016/C2009-0-22050-2.

Daly, A. and Zachary, S. (1978). Improved multiple choice models. In Hensher, D.A. and Dalvi, M.Q. (Eds.), *Determinants of Travel Choice*, Saxon House, Westmead.

Daly, A. and Bierlaire, M. (2006). A general and operational representation of Generalised Extreme Value models. *Transportation Research Part B*, 40(4), 285–305. https://doi.org/10.1016/j.trb.2005.03.003.

Gumbel, E.J. and Mustafi, C.K. (1967). Some analytical properties of bivariate extremal distributions. *Journal of the American Statistical Association*, 62(318), 569–588. https://doi.org/10.1080/01621459.1967.10482930.

Hensher, D.A., Rose, J.M., and Greene, W.H. (2015). *Applied Choice Analysis*. 2nd ed. Cambridge University Press. https://doi.org/10.1017/CBO9781316136232.

Johnson, N.L. and Kotz, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions*. John Wiley and Sons.

Koppelman, F.S. and Wen, C-H. (2000). The paired combinatorial logit model: Properties, estimation and application. *Transportation Research Part B*, 34(2), 75–89. https://doi.org/10.1016/S0191-2615(99)00012-0.

Luce, R.D., and Suppes, S. (1965). Preference, utility and subjective probability. In Luce, R.D., Bush, R.R., and Galanter, E.H. (Eds.), *Handbook of Mathematical Psychology*, Wiley, New York, 249–410.

McFadden, D. (1978). Modeling the choice of residential location. In Karlquist, A. et al. (Eds.) *Spatial Interaction Theory and Planning Models*, North Holland, Amsterdam, 75–96.

Oliveira, J. Tiago de (1961). La representation des distributions extremales bivariees. *Bulletin of the International Statistical Institute*, 33, 477–480.

Oliveira, J. Tiago de (1962-1963). Structure Theory of Bivariate Extremes; Extensions. *Estudos de Matematica, Estatistica e Econometria*, 7, Lisbon, pp. 165–195.

Recker, W.W. (1995). Discrete choice with an oddball alternative. *Transportation Research B*, 29(3), 201–211. https://doi.org/10.1016/0191-2615(95)00002-U.

Small, K.A. (1987). A discrete choice model for ordered alternatives. *Econometrica*, 55(2), 409–424. https://doi.org/10.2307/1913243.

Townsend, W. (2025). A novel algorithm for drawing nested extreme value random variables. *Journal of Choice Modelling*, 57, 100575. https://doi.org/10.1016/j.jocm.2025.100575.

Vovsha, P. (1997). Application of cross-nested logit model to mode choice in Tel Aviv, Israel, metropolitan area. *Transportation Research Record: Journal of the Transportation Research Board*, 1607(1), 6–15. https://doi.org/10.3141/1607-02.

Wen, C-H. and Koppelman, F.S. (2001). The generalized nested logit model, *Transportation Research Part B*, 35(7), 627–641. https://doi.org/10.1016/S0191-2615(00)00045-X.

Whelan, G., Batley, R., Fowkes, T., and Daly, A. (2002). Flexible models for analyzing route and departure time choice. European Transport Conference Proceedings, Association for European Transport, Cambridge.

Williams, H. C. W. L. (1977). On the formation of travel demand models and economic evaluation measures of user benefit. *Environment and Planning A*, 9(3), 285–344. https://doi.org/10.1068/a090285.

Ye, X. (2011). Investigation of underlying distributional assumptions in nested logit model using copula-based simulation and numerical approximation. *Transportation Research Record: Journal of the Transportation Research Board*, 2254, 36–43. https://doi.org/10.3141/2254-05.

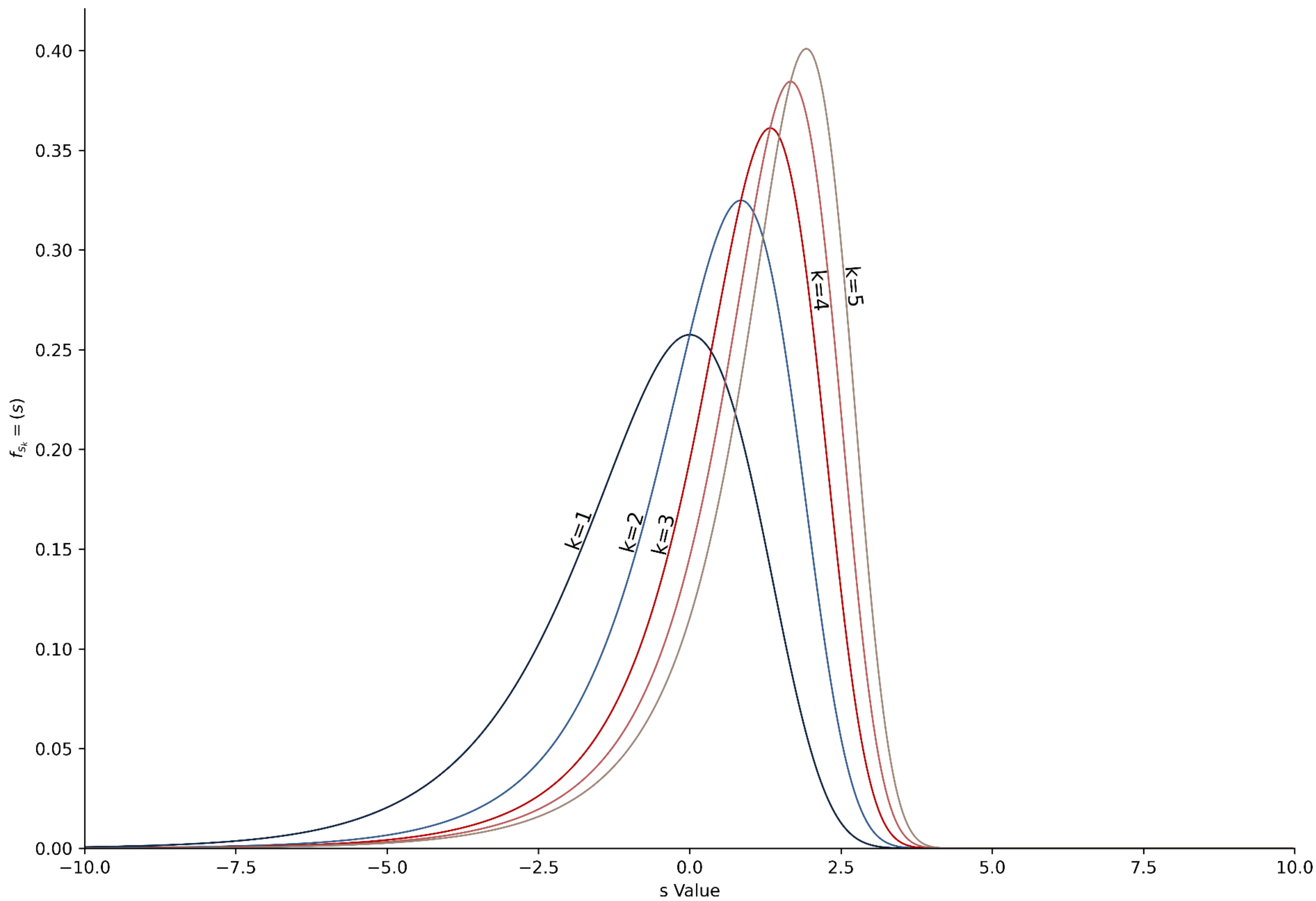

Figure 1: Probability Density Function of s_k for k=1,2,3,4,5

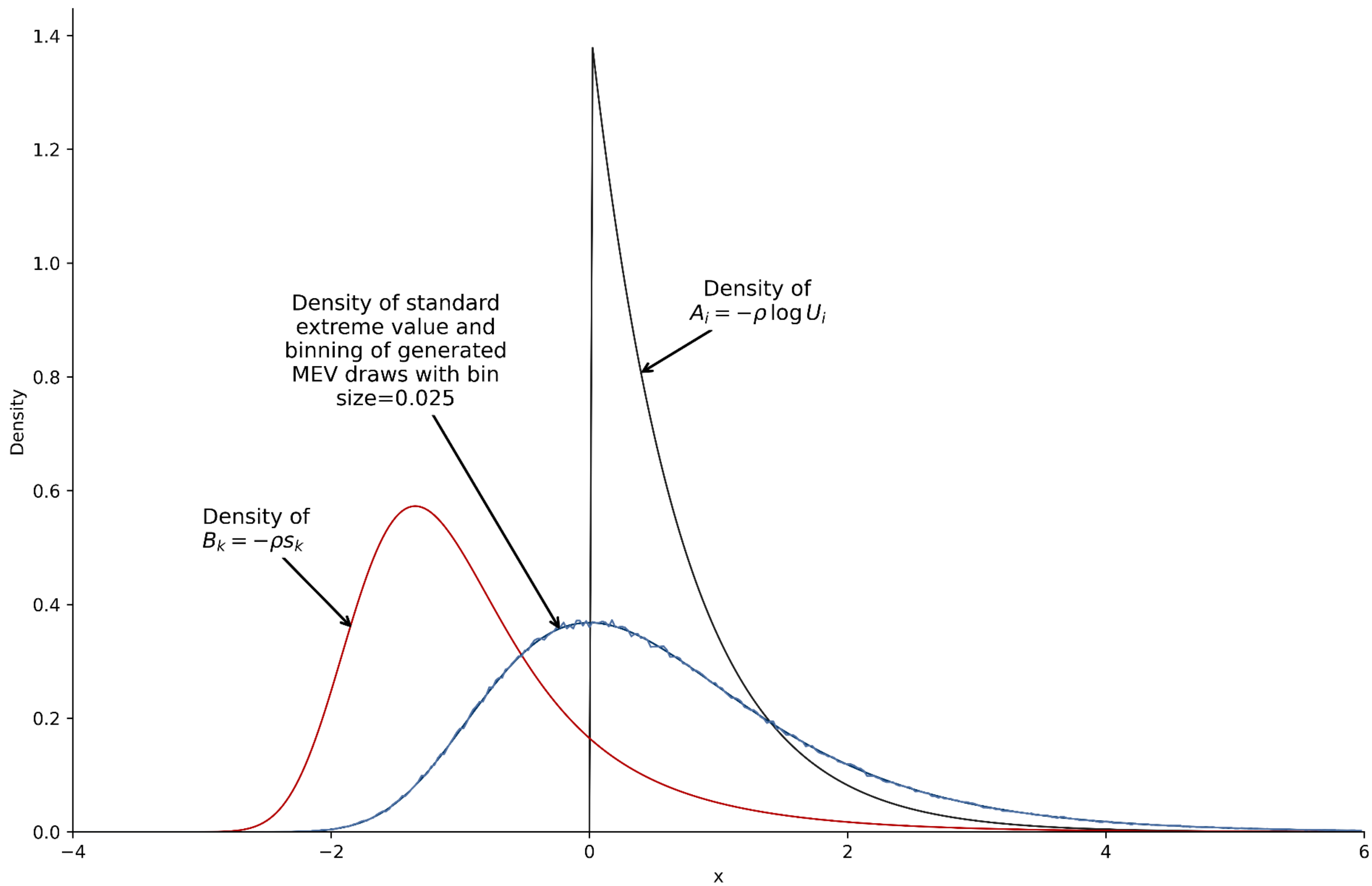

Figure 2: Decomposition-Based Validation
$\varepsilon_i\,(k=5,\ \rho=0.7) = -\rho \log U_i - \rho\, s_k$

Table 1: $Q_k(v)$ Polynomial Expressions for Each Number of Variates *k* in the MEV Distribution (for *k*=1, 2, 3, 4, 5, and 6)

K	$Q_k(v)$ **Polynomial Expression**
2	$v+\frac{1}{\rho}-1$
3	$v^2+\left(-3+\frac{3}{\rho}\right)v+\left(1-\frac{3}{\rho}+\frac{2}{\rho^2}\right)$
4	$v^3+\left(-6+\frac{6}{\rho}\right)v^2+\left(7-\frac{18}{\rho}+\frac{11}{\rho^2}\right)v+\left(-1+\frac{6}{\rho}-\frac{11}{\rho^2}+\frac{6}{\rho^3}\right)$
5	$v^4+\left(-10+\frac{10}{\rho}\right)v^3+\left(25-\frac{60}{\rho}+\frac{35}{\rho^2}\right)v^2+\left(-15+\frac{70}{\rho}-\frac{105}{\rho^2}+\frac{50}{\rho^3}\right)v+\left(1-\frac{10}{\rho}+\frac{35}{\rho^2}-\frac{50}{\rho^3}+\frac{24}{\rho^4}\right)$
6	$v^5+\left(-15+\frac{15}{\rho}\right)v^4+\left(65-\frac{150}{\rho}+\frac{85}{\rho^2}\right)v^3+\left(-90+\frac{375}{\rho}-\frac{510}{\rho^2}+\frac{225}{\rho^3}\right)v^2+\left(31-\frac{225}{\rho}+\frac{595}{\rho^2}-\frac{675}{\rho^3}+\frac{274}{\rho^4}\right)v+\left(-1+\frac{15}{\rho}-\frac{85}{\rho^2}+\frac{225}{\rho^3}-\frac{274}{\rho^4}+\frac{120}{\rho^5}\right)$

Table 2: Comparison of Statistics from Generated MEV Realizations with True MEV Values

	Number of Variates (k)	Univariate Statistics (first variate)				Multivariate Statistics			Generation Time (secs)
		Mean	Variance	Skew	Kurtosis	Pairwise Correlation (first two variates)	Three-Way Skew (first three variates)	Four-Way Kurtosis (first four variates)	
ρ **value = 0.2**									
True Value		0.5772	1.6449	1.1395	5.4000	0.9600	1.1304	2.3962	--
Simulated Value	5	0.5774	1.6462	1.1465	5.4428	0.9600	1.1367	2.4301	1.00
	10	0.5779	1.6443	1.1452	5.4354	0.9600	1.1356	2.4296	1.23
	15	0.5772	1.6464	1.1432	5.4120	0.9600	1.1340	2.4139	1.45
	20	0.5773	1.6465	1.1495	5.4635	0.9601	1.1393	2.4554	1.66
	100	0.5754	1.6431	1.1384	5.3852	0.9599	1.1298	2.3823	4.59
ρ **value = 0.5**									
True Value		0.5772	1.6449	1.1395	5.4000	0.7500	0.9971	2.2500	--
Simulated Value	5	0.5769	1.6463	1.1354	5.3861	0.7499	0.9959	2.2615	1.00
	10	0.5772	1.6475	1.1450	5.4527	0.7508	1.0040	2.3044	1.24
	15	0.5765	1.6427	1.1371	5.3758	0.7495	0.9929	2.2153	1.45
	20	0.5780	1.6488	1.1377	5.3790	0.7504	1.0004	2.2617	1.66
	100	0.5764	1.6423	1.1325	5.3331	0.7498	0.9893	2.1762	4.74
ρ **value = 0.8**									
True Value		0.5772	1.6449	1.1395	5.4000	0.3600	0.5561	1.4170	--
Simulated Value	5	0.5770	1.6450	1.1438	5.4357	0.3607	0.5571	1.4087	1.00
	10	0.5755	1.6448	1.1363	5.3747	0.3585	0.5470	1.3758	1.23
	15	0.5772	1.6428	1.1391	5.3989	0.3606	0.5588	1.4174	1.45
	20	0.5754	1.6385	1.1368	5.3642	0.3603	0.5561	1.4035	1.64
	100	0.5782	1.6457	1.1368	5.3713	0.3586	0.5566	1.4120	4.73

APPENDIX A: Recursivity of Polynomial Coefficients of v in $Q_k(v)$

From Equation (8), we have:

$$Q_k(v) = \left(v - \frac{\rho-(k-1)}{\rho} \right) Q_{k-1}(v) - v\frac{d}{dv}Q_{k-1}(v) \text{ for } k \geq 2$$

Consider the polynomial $Q_{k-1}(v) = \sum_{j=0}^{k-2} a_{k-1,j}\, \rho^{-(k-2-j)} v^j$. This is a polynomial of degree k–2. That is, the coefficient on v^{k-1} in $Q_{k-1}(v)$ is necessarily zero -- $a_{k-1,k-1} = 0$. Also, the coefficient $a_{k-1,j}$ applies to index j. Then, it follows that:

$$\begin{aligned}
\left(v - \frac{\rho-(k-1)}{\rho} \right) Q_{k-1}(v) &= v\sum_{j=0}^{k-2} a_{k-1,j}\rho^{-(k-2-j)} v^j - \sum_{j=0}^{k-2} a_{k-1,j}\rho^{-(k-2-j)} v^j \quad + \rho^{-1}\sum_{j=0}^{k-2} (k-1) a_{k-1,j}\rho^{-(k-2-j)} v^j \\
&= \sum_{j=1}^{k-1} a_{k-1,j-1}\rho^{-(k-1-j)} v^j - \sum_{j=0}^{k-2} a_{k-1,j}\rho\rho^{-(k-1-j)} v^j + \sum_{j=0}^{k-2} (k-1) a_{k-1,j}\rho^{-(k-1-j)} v^j \\
&= \sum_{j=0}^{k-1} a_{k-1,j-1}\rho^{-(k-1-j)} v^j - \sum_{j=0}^{k-1} a_{k-1,j}\rho\rho^{-(k-1-j)} v^j + \sum_{j=0}^{k-1} (k-1) a_{k-1,j}\rho^{-(k-1-j)} v^j \\
&\qquad \text{because } a_{k-1,-1} = a_{k-1,k-1} = 0 \\
&= \sum_{j=0}^{k-1} \left[a_{k-1,j-1} + \left(k-1-\rho\right) a_{k-1,j} \right] \rho^{-(k-1-j)} v^j \text{, and}
\end{aligned}$$

$$\begin{aligned}
v\frac{d}{dv}Q_{k-1}(v) &= v\sum_{j=0}^{k-2} j a_{k-1,j}\rho^{-(k-2-j)} v^{j-1} = \sum_{j=0}^{k-2} j a_{k-1,j}\rho\rho^{-(k-1-j)} v^j \\
&= \sum_{j=0}^{k-1} j a_{k-1,j}\rho\rho^{k-1-j} v^j \text{ because } a_{k-1,k-1} = 0.
\end{aligned}$$

Collecting all the three contributions to the coefficient of $\rho^{-(k-1-j)} v^j$, we get

$$\begin{aligned}
Q_k(v) &= \sum_{j=0}^{k-1} a_{k,j}\rho^{-(k-1-j)} v^j \\
&= \sum_{j=1}^{k-1} \left[a_{k-1,j-1} + \left(k-1-\rho\right) a_{k-1,j} - j a_{k-1,j}\rho \right] \rho^{-(k-1-j)}\, v^j \\
&= \sum_{j=1}^{k-1} \left[a_{k-1,j-1} + \left(k-1-(j+1)\rho\right) a_{k-1,j} \right] \rho^{-(k-1-j)}\, v^j \text{, which implies}
\end{aligned}$$

$a_{k,j} = a_{k-1,j-1} + \left(k-1-(j+1)\rho\right) a_{k-1,j}$, $a_{1,0} = 1$, $a_{k-1,-1}$=0 and $a_{k-1,k-1}$=0

APPENDIX B: Recursivity of Polynomial Coefficients on ρ within the Polynomial $a_{k,j}$

From Equation (10), the following holds:

$$a_{k,j} = a_{k-1,j-1} + \left(k-1-(j+1)\rho\right) a_{k-1,j}, a_{k-1,-1} = 0 \text{ and } a_{k-1,k-1} = 0$$

To obtain a recursion for the scalar coefficients $a_{k,j,h}$, we may write $a_{k-1,j-1} = \sum_{h=0}^{k-1-j} a_{k-1,j-1,h}\rho^{k-1-j-h}$ and $a_{k-1,j} = \sum_{h=0}^{k-2-j} a_{k-1,j,h}\rho^{k-2-j-h}$.

Then, using the recursion at the polynomial level as in Equation (10), we get:

$$\sum_{h=0}^{k-1-j} a_{k,j,h}\rho^{k-1-j-h} = \sum_{h=0}^{k-1-j} a_{k-1,j-1,h}\rho^{k-1-j-h} + \sum_{h=0}^{k-2-j} (k-1)a_{k-1,j,h}\rho^{k-2-j-h} - \sum_{h=0}^{k-2-j} (j+1)\rho a_{k-1,j,h}\rho^{k-2-j-h}$$

Consider the second term and change the index from h to $h = h+1$. Then,

$$\sum_{h=0}^{k-1-j} a_{k,j,h}\rho^{k-1-j-h} = \sum_{h=0}^{k-1-j} a_{k-1,j-1,h}\rho^{k-1-j-h} + \sum_{h=1}^{k-1-j} (k-1)a_{k-1,j,h-1}\rho^{k-1-j-h} - \sum_{h=0}^{k-2-j} (j+1)a_{k-1,j,h}\rho^{k-1-j-h}$$

By notational convention, because $a_{k-1,-1} = 0$, $a_{k-1,j,-1}=0\ \forall\ j$. Thus,

$$\sum_{h=0}^{k-1-j} a_{k,j,h}\rho^{k-1-j-h} = \sum_{h=0}^{k-1-j} a_{k-1,j-1,h}\rho^{k-1-j-h} + \sum_{h=0}^{k-1-j} (k-1)a_{k-1,j,h-1}\rho^{k-1-j-h} - \sum_{h=0}^{k-2-j} (j+1)a_{k-1,j,h}\rho^{k-1-j-h}$$

Collecting coefficients of equal powers of ρ

$$a_{k,j,h}=a_{k-1,j-1,h} + (k-1)a_{k-1,j,h-1} - (j+1)a_{k-1,j,h}$$

APPENDIX C: Closed Form CDF for S_k

The CDF of S_k is:

$$F_{S_k}(s) = \int_{t=-\infty}^{s} f_{S_k}(t)dt = \int_{t=-\infty}^{s} \frac{\rho^k}{(k-1)!}\exp\left[-e^{t\rho}\right]\left(e^{t\rho}\right)\times Q_k(e^{t\rho})dt,$$

Let $m = e^{t\rho}$, so that $\frac{dm}{dt} = e^{t\rho}\rho$ and $dt = \frac{dm}{m\rho}$. Then, the CDF may be rewritten with the change of variables as:

$$F_{S_k}(s) = \int_{m=0}^{m=e^{s\rho}=v} \frac{\rho^k}{(k-1)!}\exp[-m](m)\times Q_k(m)\frac{dm}{\rho m}$$

$$= \int_{m=0}^{m=v} \frac{\rho^{k-1}}{(k-1)!}\exp[-m]\times Q_k(m)dm = \frac{\rho^{k-1}}{(k-1)!}\int_{m=0}^{m=v} e^{-m}\times Q_k(m)dm$$

Next, since $Q_k(m)$ is a polynomial of degree k–1, repeated integration by parts implies that the integral can be written in the form $\int_{m=0}^{m=v} e^{-m}Q_k(m)dm = C - e^{-v}R_k(v)$ for some polynomial $R_k(v)$ of degree k–1 and constant C so that $F_{S_k}(s) = \frac{\rho^{k-1}}{(k-1)!}\left(C - e^{-v}R_k(v)\right)$. Now as $\underset{s\to\infty}{Lim}\left[e^{-v}R_k(v)\right]\to 0$, and any CDF should satisfy $\underset{s\to\infty}{Lim}\left[F_{S_k}(s)\right]=1$, it must be the case that $\frac{\rho^{k-1}}{(k-1)!}C = 1$ or $C = \frac{(k-1)!}{\rho^{k-1}}$

Thus, $F_{S_k}(s) = \left(1 - e^{-v}\left[\frac{\rho^{k-1}}{(k-1)!}R_k(v)\right]\right)$. Differentiating this expression with respect to s, we get the following:

$$\frac{d}{ds}F_{S_k}(s) = \frac{d}{dv}\left(1 - e^{-v}\left(\frac{\rho^{k-1}}{(k-1)!}\right)R_k(v)\right)\frac{dv}{ds}$$

$$= \frac{\rho^k}{(k-1)!}\left[e^{-v}R_k(v) - e^{-v}\frac{dR_k(v)}{dv}\right]v \quad \text{(because } v = e^{s\rho}\text{)}$$

$$= \frac{\rho^k}{(k-1)!}e^{-v}v\left[R_k(v) - e^{-v}\frac{dR_k(v)}{dv}\right]$$

The above should be the same as the density function $f_{S_k}(s) = \frac{\rho^k}{(k-1)!}e^{-v}v\times Q_k(v)$. That is,

$$\frac{\rho^k}{(k-1)!}e^{-v}v\left[R_k(v)-\frac{dR_k(v)}{dv}\right]=\frac{\rho^k}{(k-1)!}e^{-v}v\times Q_k(v)\text{, or}$$

$$R_k(v)-\frac{dR_k(v)}{dv}=Q_k(v)\text{.}$$[5]

Next replace the left and right sides of the above expression by the corresponding coefficients:

$$R_k(v)=\sum_{j=0}^{k-1}b_{k,j}\frac{1}{\rho^{(k-j-1)}}v^j,\ \frac{dR_k(v)}{dv}=\sum_{j=1}^{k-1}jb_{k,j}\frac{1}{\rho^{(k-j-1)}}v^{j-1},\ Q_k(v)=\sum_{j=0}^{k-1}a_{k,j}\frac{1}{\rho^{(k-j-1)}}v^j$$

Consider the derivative expression, and let index $n=j-1$. Then

$$\frac{dR_k(v)}{dv}=\sum_{n=0}^{k-2}(n+1)\,b_{k,n+1}\frac{1}{\rho^{(k-n-2)}}v^n\text{.}$$

Since j and n are mere accounting indices, we can rename n as j to obtain

$$\frac{dR_k(v)}{dv}=\sum_{j=0}^{k-2}(j+1)\,b_{k,j+1}\frac{1}{\rho^{(k-j-2)}}v^j=\sum_{j=0}^{k-1}(j+1)\,b_{k,j+1}\frac{1}{\rho^{(k-j-1)}}v^j \text{ (because } b_{k,k}=0\text{).}$$

The net result is $\sum_{j=0}^{k-1}b_{k,j}\frac{1}{\rho^{(k-j-1)}}v^j$ - $\sum_{j=0}^{k-1}(j+1)\,b_{k,j+1}\frac{1}{\rho^{(k-j-1)}}v^j=\sum_{j=0}^{k-1}a_{k,j}\frac{1}{\rho^{(k-j-1)}}v^j$, which implies that

$$b_{k,j}-(j+1)b_{k,j+1}=a_{k,j}$$

[5] The particularly simple polynomial–exponential structure obtained here is a direct consequence of the geometry of the Dirichlet(1,…,1) distribution. Because Dirichlet(1,…,1) is uniform on the (k−1)-simplex, its density is constant and equal to (k−1)!, the reciprocal of the simplex volume. This constancy is what gives rise, after the log–exponential change of variables $v=e^{s\rho}$, to a density of the form $e^{-v}v$ multiplied by a finite-degree polynomial $Q_k(v)$. As a result, the associated cumulative distribution function necessarily admits the closed form $F_{S_k}(s)=\left(1-e^{-v}\left[\left(\frac{\rho^{k-1}}{(k-1)!}\right)R_k(v)\right]\right)$ where $R_k(v)$ is another polynomial of the same degree. In this setting, differentiation of the CDF yields the exact operator identity $R_k(v)-\frac{dR_k(v)}{dv}=Q_k(v)$, leading to a linear, closed recursion for the polynomial coefficients. This polynomial closure and the resulting simplicity of the moment-generating and cumulative distribution functions rely critically on the uniform simplex geometry.

Appendix D: Recursivity of Polynomial Coefficients on ρ within the Polynomial

From the recursion $b_{k,j} - (j+1)b_{k,j+1} = a_{k,j}$, and substituting $b_{k,j} = \sum_{h=0}^{k-1-j} b_{k,j,h}\rho^{k-1-j-h}$ and $a_{k,j} = \sum_{h=0}^{k-1-j} a_{k,j,h}\rho^{k-1-j-h}$, we obtain:

$$\sum_{h=0}^{k-1-j} b_{k,j,h}\rho^{k-1-j-h} - (j+1)\sum_{h=0}^{k-2-j} b_{k,j+1,h}\rho^{k-2-j-h} = \sum_{h=0}^{k-1-j} a_{k,j,h}\rho^{k-1-j-h} \tag{D.1}$$

Setting $n = h+1$ in the second term, we get $(j+1)\sum_{n=1}^{k-1-j} b_{k,j+1,n-1}\rho^{k-1-j-n}$. Resetting h to be n (both are merely indices), and noting that $b_{k,j+1,-1} = 0$, we can write Equation (D.1) as:

$$\sum_{h=0}^{k-1-j} b_{k,j,h}\rho^{k-1-j-h} - (j+1)\sum_{h=0}^{k-1-j} b_{k,j+1,h-1}\rho^{k-1-j-h} = \sum_{h=0}^{k-1-j} a_{k,j,h}\rho^{k-1-j-h}$$

Collecting coefficients on the same power of ρ, we obtain the recursion:

$b_{k,j,h} - (j+1)b_{k,j+1,h-1} = a_{k,j,h}$, with $b_{k,j,0} = a_{k,j,0}$,

$j = 0,1,\ldots,k-1,\ h = 0,1,\ldots,k-1-j;\ b_{k,k,h} = 0;\ b_{k,j,-1} = 0.$

APPENDIX E: Proof for Lemma 1.

Substitute the recursive definition of $Q_k(v)$ into $I_k(x)$:

$$I_k(x) = \int_0^\infty e^{-v} v^x \left[\left(v - \frac{\rho - (k-1)}{\rho} \right) Q_{k-1}(v) - v \frac{d}{dv} Q_{k-1}(v) \right] dv = A - B, \text{ where} \tag{E.1}$$

$$A = \int_0^\infty e^{-v} v^{x+1} Q_{k-1}(v) dv - \frac{\rho - (k-1)}{\rho} I_{k-1}(x) \text{ and } B = \int_0^\infty e^{-v} v^{x+1} \frac{d}{dv} Q_{k-1}(v) dv.$$

Consider the second integral B and using integration by parts with $db = \frac{d}{dv} Q_{k-1}(v) dv$ and $a = e^{-v} v^{x+1}$.

$$B = \left[e^{-v} v^{x+1} Q_{k-1}(v) \right]_0^\infty - \int_0^\infty \frac{d}{dv} \left(e^{-v} v^{x+1} \right) Q_{k-1}(v) dv$$

Considering the first term above, $\lim_{v \to \infty} e^{-v} v^{x+1} Q_{k-1}(v) = 0$ because e^{-v} dominates any polynomial growth in $Q_{k-1}(v)$ as $v \to \infty$. For the limit at zero, $Q_{k-1}(v)$ is finite, while $v^{x+1} \to 0$ for $x > -1$. Thus, $\lim_{v \to 0^+} e^{-v} v^{x+1} Q_{k-1}(v) = 0$ for $x > -1$, and

$$\begin{aligned} B &= -\int_0^\infty \frac{d}{dv} \left(e^{-v} v^{x+1} \right) Q_{k-1}(v) dv = -\int_0^\infty e^{-v} \left[(x+1) v^x - v^{x+1} \right] Q_{k-1}(v) dv \\ &= -\int_0^\infty e^{-v} \left[(x+1) v^x \right] Q_{k-1}(v) + \int_0^\infty e^{-v} v^{x+1} Q_{k-1}(v) dv \\ &= -(x+1) I_{k-1}(x) + \int_0^\infty e^{-v} v^{x+1} Q_{k-1}(v) dv. \end{aligned}$$

Substituting back in Equation (E.1), we get the desired result:

$$\begin{aligned} I_k(x) &= \int_0^\infty e^{-v} v^{x+1} Q_{k-1}(v) dv - \frac{\rho - (k-1)}{\rho} I_{k-1}(x) - \left[-(x+1) I_{k-1}(x) + \int_0^\infty e^{-v} v^{x+1} Q_{k-1}(v) dv \right] \\ &= \left(x + \frac{k-1}{\rho} \right) I_{k-1}(x), \; k \ge 2. \end{aligned}$$

APPENDIX F: Moment Generating Function (MGF) for S_k

$$M_k(t) = \int_{-\infty}^{\infty} e^{ts} f_k(s)ds = \int_{-\infty}^{\infty} \frac{\rho^k}{(k-1)!} e^{-e^{\rho s}} e^{\rho s} e^{ts} Q_{k-1}(e^{\rho s})ds.$$

Letting $v = e^{\rho s}$, $ds = \frac{1}{\rho}\frac{dv}{v}$ and $e^{ts} = v^{t/\rho}$, and

$$M_k(t) = \frac{\rho^{k-1}}{(k-1)!}\int_0^{\infty} e^{-v} v^{t/\rho} Q_{k-1}(v)dv = \frac{\rho^{k-1}}{(k-1)!} I_{k-1}\left(\frac{t}{\rho}\right)$$
$$= \frac{\rho^{k-1}}{(k-1)!} I_{k-1}(x) \text{ with } x = \frac{t}{\rho}.$$

Using Lemma 1 for $x > -1$,

$$I_{k-1}(x) = \left(x + \frac{k-2}{\rho}\right) I_{k-2}(x) = \left(x + \frac{k-2}{\rho}\right)\left(x + \frac{k-3}{\rho}\right) I_{k-3}(x)$$
$$= \left(x + \frac{k-2}{\rho}\right)\left(x + \frac{k-3}{\rho}\right) \ldots \left(x + \frac{1}{\rho}\right) I_1(x)$$
$$= \prod_{j=1}^{k-1}\left(x + \frac{j}{\rho}\right) I_1(x) = \prod_{j=1}^{k-1}\left(\frac{t}{\rho} + \frac{j}{\rho}\right) I_1(x) = \frac{1}{\rho^{k-1}}\prod_{j=1}^{k-1}(t+j) I_1(x)$$

Next, noting that $I_1(x) = \int_0^{\infty} e^{-v} v^x\, dv = \Gamma(1+x)$ because $Q_1(v) = 1$,

$$I_{k-1}(x) = \frac{1}{\rho^{k-1}}\prod_{j=1}^{k-1}(t+j) I_1(x) = \frac{1}{\rho^{k-1}}\prod_{j=1}^{k-1}(t+j)\Gamma(1+x) = \frac{1}{\rho^{k-1}}\Gamma(1+\frac{t}{\rho})\prod_{j=1}^{k-1}(t+j).$$ Finally,

$$M_k(t) = \frac{\rho^{k-1}}{(k-1)!} I_{k-1}(x)$$
$$= \frac{\rho^{k-1}}{(k-1)!}\frac{1}{\rho^{k-1}}\Gamma\left(1 + \frac{t}{\rho}\right)\prod_{j=1}^{k-1}(t+j)$$
$$= \frac{1}{(k-1)!}\Gamma\left(1 + \frac{t}{\rho}\right)\prod_{j=1}^{k-1}(t+j),\ t > -\rho$$

APPENDIX G: Cumulant Generating Function (CGF) for S_k

The CGF for S_k may be obtained from the MGF for S_k as follows:

$$K_k(t) = \log M_k(t) = -\log((k-1)!) + \log\Gamma\left(1+\frac{t}{\rho}\right) + \sum_{j=1}^{k-1}\log(t+j).$$

The n^{th} cumulant may be obtained as $\kappa_n(S_k) = \left.\frac{d^n}{dt^n}K_k(t)\right|_{t=0}$. Consider the derivative $\frac{d^n}{dt^n}K_k(t)$.

$$\frac{d^n}{dt^n}K_k(t) = \frac{d^n}{dt^n}\log\Gamma\left(1+\frac{t}{\rho}\right) + \left(\sum_{j=1}^{k-1}\frac{d^n}{dt^n}\log(t+j)\right) = \frac{1}{\rho^n}\psi^{(n-1)}\left(1+\frac{t}{\rho}\right) + \left(\sum_{j=1}^{k-1}(-1)^{n-1}(n-1)!(t+j)^{-n}\right)$$

$$\begin{aligned}\kappa_n(S_k) = \left.\frac{d^n}{dt^n}K_k(t)\right|_{t=0} &= \frac{1}{\rho^n}\psi^{(n-1)}(1) + \left((-1)^{n-1}(n-1)!\sum_{j=1}^{k-1}(j)^{-n}\right), \\ &= \frac{1}{\rho^n}\psi^{(n-1)}(1) + (-1)^{n-1}(n-1)!H_{k-1}^{(n)},\ n \geq 1.\end{aligned}$$

APPENDIX H: Univariate Marginal Cumulants of MEV Errors

For the log-Dirichlet(1,…,1) distribution of $\log U_1, \log U_2, ... \log U_k$, a standard identity for the *n*th cumulant is as follows:

$$\kappa_n(\log U_i) = \psi^{(n-1)}(1) - \psi^{(n-1)}(k), n \geq 1.$$

For integers $n \geq 2$, $\psi^{(n-1)}(1) = (-1)^n (n-1)! \zeta(n)$ and $\psi^{(n-1)}(k) = (-1)^n (n-1)! \left[\zeta(n) - H_{k-1}^{(n)}\right]$, with the special cases for $n = 1$ being $\psi^{(0)}(1) = \psi(1) = -\gamma$ and $\psi^{(0)}(k) = \psi(k) = H_{k-1} - \gamma$. This provides the general identity:

$$\psi^{(n-1)}(1) - \psi^{(n-1)}(k) = (-1)^n (n-1)! \left(-H_{k-1}^{(n)}\right) = (-1)^{n+1} (n-1)! H_{k-1}^{(n)}.$$

Using the above identify,

$$\kappa_n(\log U_i) = (-1)^n (n-1)! H_{k-1}^{(n)},\ n \geq 1$$

Then, the marginal moments of each ε_i may be recovered for any number of variates *k*, just as in the bivariate case of Section 3. Specifically, given the established independence of the U_i variates $(i = 1, 2, \ldots, k)$ and S_k, the n^{th} cumulant of ε_i may be written as:

$$\begin{aligned}
\kappa_n(\varepsilon_i) &= (-\rho)^n \left[\kappa_n(\log(U_i) + \kappa_n(S_k)\right],\ n \geq 1 \\
&= (-\rho)^n \left[\frac{1}{\rho^n}\psi^{(n-1)}(1) + (-1)^n (n-1)! H_{k-1}^{(n)} + (-1)^{n-1}(n-1)! H_{k-1}^{(n)}\right],\ n \geq 1 \\
&= (-\rho)^n \left[\frac{1}{\rho^n}\psi^{(n-1)}(1)\right], n \geq 1 \\
&= (-1)^n \psi^{(n-1)}(1), n \geq 1
\end{aligned}$$

APPENDIX I: Joint Multivariate Cumulants of MEV Errors

Start with the decomposition $\varepsilon_i = -\rho\big(\log U_i + S_k\big),\ i = 1,2,...,k,$ $\mathbf{U} = (U_1,\ldots,U_k) \sim \text{Dirichlet}(1,\ldots,1),\ \mathbf{U} \perp S_k$. The general formula for the joint cumulant generating function for the multivariate log-Dirichlet distribution $(\log U_1,\ldots,\log U_k)$ is as follows:

$$K_{\ln U}(t_1,\ldots,t_k) = \log\Gamma(k) - \log\Gamma\left(k + \sum_{i=1}^{k} t_i\right) + \sum_{i=1}^{k}\log\Gamma(1+t_i).$$

Taking the mixed derivatives, and observing that only the second term on the right side of the equation above contributes for distinct indices $i_1, i_2, ..., i_m,\ m \geq 2$, the mixed cumulants for any m is:

$$\kappa_m\big(\log U_{i_1},\ldots,\log U_{i_m}\big) = \left.\frac{\partial^m K_{\ln U}}{\partial t_{i_1}\cdots\partial t_{i_m}}\right|_{t=0} = -\psi^{(m-1)}(k),\ i_1,\ldots,i_m \text{ distinct},\ m \geq 2.$$

Substituting $\psi^{(m-1)}(k) = (-1)^m (m-1)!\left[\zeta(m) - H_{k-1}^{(m)}\right]$ for integer k,

$$\kappa_m\big(\log U_{i_1},\ldots,\log U_{i_m}\big) = (-1)(-1)^m (m-1)!\left[\zeta(m) - H_{k-1}^{(m)}\right] = (-1)^{m+1}(m-1)!\left[\zeta(m) - H_{k-1}^{(m)}\right],\ H_{k-1}^{(m)} = \sum_{j=1}^{k-1}\frac{1}{j^m}$$

Next to obtain the higher-order cross-cumulants of the MEV errors, by using multilinearity and the fact that $\mathbf{U}\big(= U_1,\ldots,U_k\big) \perp S_k$,

$$\begin{aligned}
\kappa_m\big(\varepsilon_{i_1},\ldots,\varepsilon_{i_m}\big) &= (-\rho)^m\Big[\kappa_m\big(\log U_{i_1},\ldots,\log U_{i_m}\big) + \kappa_m(S_k)\Big],\ m \geq 2 \\
&= (-\rho)^m\left((-1)^{m+1}(m-1)!\left[\zeta(m) - H_{k-1}^{(m)}\right] + (-1)^{m-1}(m-1)!H_{k-1}^{(m)} + \frac{1}{\rho^m}\psi^{(m-1)}(1)\right) \\
&= (-\rho)^m\left[(-1)^{m+1}(m-1)!\zeta(m) \cancel{-(-1)^{m+1}(m-1)!H_{k-1}^{(m)}} + \cancel{(-1)^{m-1}(m-1)!H_{k-1}^{(m)}} + \frac{1}{\rho^m}\psi^{(m-1)}(1)\right] \\
&= (-\rho)^m(-1)^{m+1}(m-1)!\zeta(m) + (-1)^m\psi^{(m-1)}(1) \\
&= (-1)^{2m+1}(m-1)!\rho^m\zeta(m) + (-1)^m(-1)^m(m-1)!\zeta(m) \text{ because } \psi^{(m-1)}(1) = (-1)^m(m-1)!\zeta(m) \\
&= (-1)(m-1)!\rho^m\zeta(m) + (m-1)!\zeta(m) \\
&= (m-1)!\zeta(m)\left[1 - \rho^m\right].
\end{aligned}$$

The remarkable simplification happens because of the cancellation of the terms involving $H_{k-1}^{(m)}$.