A Multivariate Distribution with Pareto Tails and Pareto Maxima

Costas Arkolakis
Yale University

Andrés Rodríguez-Clare
UC Berkeley and NBER

Jiun-Hua Su
Institute of Economics, Academia Sinica *

This Version: April, 2026

Abstract

We present a new multivariate distribution with Pareto distributed tails and maxima. Compared to uncorrelated univariate Pareto distributions, it features one additional parameter that governs the covariance of its realizations. We argue that it has a number of aggregation properties that make it useful for applied work alongside multivariate Gumbel and Fréchet distributions. In particular, it is well suited for applications with a participation threshold. Finally, we show that this distribution is indeed valid by proving a general result about n-increasing functions.

Keywords: Multivariate Pareto, Pareto tails, Pareto maxima, participation threshold. **JEL code**: C46.

*We are grateful to the Co-Editor Theodore Papageorgiou and an anonymous referee for their constructive comments and suggestions that substantially improved the paper. We also thank Barry Arnold, Xiaohong Chen, Roger Nelsen, and Tianhao Wu for useful suggestions. All errors are, of course, our own.
Corresponding author: Jiun-Hua Su, jhsu@econ.sinica.edu.tw

1. Introduction

The Pareto size distribution is one of the most ubiquitous empirical relationships in the natural and social sciences. It has been used to describe the distributions of, among other things, incomes, wealth, productivity, firm sizes, stock returns, and city populations (see e.g. Birkner *et al.* (2023), Hsu *et al.* (2023), Gabaix (2009) and the references therein). Because of its empirical prevalence, but also its mathematical simplicity, the Pareto distribution has become an extremely important statistical tool for scientists across disciplines (see e.g. Gabaix (2009) and Gabaix (2016)). Typically, the modeling of these statistical processes implies independence of the different Pareto realizations. However, for a large number of empirical and theoretical applications, such as natural disasters, stock returns, and firm sales across multiple markets, realizations could be closely correlated while Pareto size distributions still prevail.[1]

In this paper we describe a multivariate distribution that explicitly allows for correlation across different draws and exhibits Pareto marginals. In addition, the maximum is distributed Pareto, and conditional distributions have a convenient, tractable form. In particular, we show that the function

$$H(\mathbf{z}) = 1 - \left(\sum_{i=1}^{n} \left(T_i z_i^{-\theta} \right)^{1/(1-\rho)} \right)^{1-\rho} \tag{1}$$

with support

$$z_i \geq \widetilde{T}^{1/\theta} \text{ for all } i, \text{ where}$$

$$\widetilde{T} \equiv \left(\sum_{i=1}^{n} T_i^{1/(1-\rho)} \right)^{1-\rho}, T_i > 0 \text{ for all } i,$$

$\theta > 0$ and $\rho \in [0, 1)$, is a joint cumulative distribution function for the vector of random variables $(Z_1, \ldots, Z_n)$, where $H(\mathbf{z}) = H(z_1, \ldots, z_n) = \Pr(Z_1 \leq z_1, \ldots, Z_n \leq z_n)$.[2] We also show that this distribution has marginals that have Pareto tails, and that $\max\{Z_1, ..., Z_n\}$ is distributed Pareto. The different parameters of the distribution have natural interpretations: the parameter θ determines the heterogeneity across realizations of different

[1] For example, Jupp and Mardia (1982, p. 1023) characterize the joint distribution of incomes across successive years, allowing for correlation and Pareto tails. Chiragiev and Landsman (2007) model correlated losses with Pareto tails and study their implications for tail risk and capital allocation.

[2] The distribution (1) has finite mean if $\theta > 1$ and finite variance if $\theta > 2$.

vectors, while ρ determines the heterogeneity of the realizations of a single vector.

Since the introduction of multivariate Pareto distributions of the first and second kind in Mardia (1962), a wide variety of multivariate Pareto distributions have been proposed. Many of these examples and generalizations are surveyed in Arnold (1990), Arnold *et al.* (1992), and Kotz *et al.* (2000). Our distribution generalizes the bivariate construction in Nelsen (2006) that results from combining a non-strict Archimedean copula with Pareto type-I marginal distributions.

To illustrate the relevance of our multivariate Pareto distribution (MVP), we study discrete choice problems and compare the implications of using the MVP with those obtained under commonly used alternatives, including the multivariate Gumbel and multivariate Fréchet distributions.[3] A key distinction is that, under the MVP, both the distribution of the maximum utility and the joint tails of the choice-specific shocks exhibit Pareto behavior, whereas the Gumbel distribution has Gumbel distributed maximum and exponentially decreasing tails, and the Fréchet distribution has Fréchet distributed maximum and tails that exhibit polynomial decay.

We furthermore discuss two key statistics that are useful for applied work and aggregation, and illustrate the properties of the MVP vis-a-vis multivariate versions of the Gumbel and Fréchet: the joint distribution of the maximizer and the maximum being above a certain threshold, and the distribution of the maximizer conditional on the maximum being above a threshold. We argue that while the conditional probability has the same tractable algebraic structure in both MVP and Fréchet cases, their joint probabilities are different.

We finally analyze the resulting implications for choice probabilities, aggregate outcomes, and the suitability of each specification across different economic applications. We conclude that the MVP is best suited for economic problems with a participation threshold, such as firm entry in monopolistic competition models following Melitz (2003). By contrast, it is less convenient in settings without such thresholds. Thus, for example, the MVP is not well-suited to Roy-type labor market models unless one modifies the standard framework by assuming that workers choose to participate in the labor mar-

[3]The multivariate Gumbel distribution is a member of the GEV class introduced into discrete choice by McFadden (1978). The multivariate Fréchet distribution corresponds to a max-stable extreme-value specification with Fréchet marginals and was popularized in quantitative trade by Eaton and Kortum (2002).

ket only if the resulting income is above some threshold. Such an extension, however, is not unrealistic. For example, a participation constraint can be quite relevant to explain the decline in the labor force participation of low-skilled workers, the topic of an emerging literature in labor economics (Binder and Bound (2019), Aguiar *et al.* (2021)).

2. A Bivariate Example

To understand the properties of our MVP and the difference with known distributions we first study a known bivariate example. Consider the joint cumulative distribution

$$G(z_1, z_2) = 1 - \left[(T_1 z_1^{-\theta})^{1/(1-\rho)} + (T_2 z_2^{-\theta})^{1/(1-\rho)}\right]^{1-\rho},$$

with support implicitly defined by (z_1, z_2) such that $(T_1 z_1^{-\theta})^{1/(1-\rho)} + (T_2 z_2^{-\theta})^{1/(1-\rho)} \leq 1$. This distribution can be seen as resulting from the Archimedean copula

$$C(u_1, u_2) \equiv \max\left\{0, 1 - \left[(1-u_1)^{1/(1-\rho)} + (1-u_2)^{1/(1-\rho)}\right]^{1-\rho}\right\},$$

where $G(z_1, z_2) = C(F_1(z_1), F_2(z_2))$ and where Z_i is distributed Pareto with $F_i(z_i) = \Pr(Z_i \leq z_i) = 1 - T_i z_i^{-\theta}$ for $i = 1, 2$.[4]

In order to better understand the role of ρ in this bivariate example, we can study the upper tail dependence (λ_U) of the copula. Roughly speaking, λ_U gives the probability that Z_1 is large given that Z_2 is large, and vice versa. Formally, this is found by taking the limit of the conditional probability $\Pr(F_1(Z_1) > u \mid F_2(Z_2) > u)$,

$$\lambda_U \equiv \lim_{u \uparrow 1} \Pr(F_1(Z_1) > u \mid F_2(Z_2) > u) = \lim_{u \uparrow 1} \frac{1 - 2u + C(u, u)}{1 - u} = 2 - 2^{1-\rho}.$$

As $\rho \to 0$, $\lambda_U \to 0$, so that if Z_2 is large then Z_1 is large with probability 0, while as $\rho \to 1$, $\lambda_U \to 1$, so that if Z_2 is large then Z_1 will also be large with probability 1.

The distribution $G(z_1, z_2)$ has the property that $Z \equiv \max(Z_1, Z_2)$ is distributed Pareto with the cumulative distribution $F(z) = 1 - \widetilde{T} z^{-\theta}$ with support $z \geq \widetilde{T}^{1/\theta}$. Moreover, the

[4]This is copula 4.2.2 in Nelsen (2006).

joint probability that $j = \arg\max_i Z_i$ and that $Z_j \geq z$ is simply

$$\Pr\left(\arg\max_i Z_i = j \,\cap \max_i Z_i \geq z\right) = \frac{T_j^{1/(1-\rho)}}{\widetilde{T}^{1/(1-\rho)}} \widetilde{T} z^{-\theta}. \tag{2}$$

This is a convenient property for many economic applications. For example, as explored in Arkolakis *et al.* (2018), multinational firms can produce a good in many locations and will choose the location with higher productivity (controlling for other determinants of cost). If productivity in location i is z_i and if $G(z_1, z_2, ..., z_n)$ is the distribution of productivity across locations, then we may need $Z \equiv \max(Z_1, ..., Z_n)$ to be distributed Pareto so that sales across firms in a market is approximately Pareto, which is what we tend to observe in the data (at least in the right tail). In addition, having property (2) proves very convenient for analytical tractability.

Unfortunately, extending the distribution $G(z_1, z_2)$ to three or more variables cannot be done since the copula $C(u_1, u_2)$ is not strict (see Nelsen, 2006). Thus, the function

$$H(\mathbf{z}) = 1 - \left(\sum_{i=1}^{n} \left(T_i z_i^{-\theta}\right)^{1/(1-\rho)}\right)^{1-\rho}$$

with domain defined implicitly by $\sum_{i=1}^{n} \left(T_i z_i^{-\theta}\right)^{1/(1-\rho)} \leq 1$ is not a distribution for $n \geq 3$. Moreover, to the best of our knowledge, there are no other multivariate distributions for $n > 2$ that have Pareto tails and Pareto distributed maximum.[5] In this paper we show that we can modify the support of the distribution to make it an N-box defined by $z_i \geq \widetilde{T}^{1/\theta}$ for all i, where $\widetilde{T} \equiv \left(\sum_{i=1}^{n} T_i^{1/(1-\rho)}\right)^{1-\rho}$, as in the distribution introduced in (1). We also discuss a number of properties of this distribution that may be useful for applied work.

The distribution we present in this paper has a number of convenient features that have already been exploited in recent work. Arkolakis *et al.* (2018) uses it to model the decision by individual firms about where to locate plants to serve different destinations in a way that yields tractable expressions for aggregate analysis. They also describe a procedure to estimate the two key parameters, θ and ρ using international and

[5] We have evaluated the multivariate Pareto distributions in Mardia (1962), Arnold (1983), Hanagal (1996), Li (2006), Asimit *et al.* (2010), Su and Furman (2017), as well as the copulas in Table 4.1 on page 116 in Nelsen (2006). In none of these cases the maximum is distributed Pareto.

United States data on trade and multinational production. This has been followed up by Wang (2017) and Shen (2024), who extend the model in Arkolakis *et al.* (2018) to allow for economies of scale and corporate taxation, respectively. In a separate application, Eaton and Kortum (2018) use it to model trade in goods and services.

In the next section we discuss the properties of this distribution that make it useful for applied work, while in Section 4 we prove that $H(\mathbf{z})$ is a proper distribution. Section 5 discusses potential applications of the MVP in economic problems of firm trade and multinational production and labor supply to multiple occupations or sectors.

3. Distributional Properties

3.1. Properties of the Proposed Distribution

Below we state some important properties of the distribution in (1). All the proofs can be found in the Appendix.

i. **The maximum is distributed Pareto.** $Z \equiv \max(Z_1, ..., Z_n)$ is distributed Pareto of type I (Arnold, 2015) with shape parameter θ and scale parameter $\widetilde{T}^{1/\theta}$ – that is, $\Pr(Z \leq z) = 1 - \widetilde{T} z^{-\theta}$.

ii. **Conditional probabilities.** The joint probability that $\arg\max_i Z_i = j$ and $Z_j \geq z$ for $z > \widetilde{T}^{1/\theta}$ has the following convenient form:

$$\Pr\left(\arg\max_i Z_i = j \cap \max_i Z_i \geq z\right) = \frac{T_j^{1/(1-\rho)}}{\widetilde{T}^{1/(1-\rho)}} \widetilde{T} z^{-\theta}.$$

Combined with property (i), this implies that

$$\Pr\left(\arg\max_i Z_i = j | \max_i Z_i \geq z\right) = \frac{T_j^{1/(1-\rho)}}{\widetilde{T}^{1/(1-\rho)}}.$$

iii. **Marginal distributions.** To study marginal distributions we need some additional notation. Let $\mathbb{N}_n \equiv \{1, 2, \ldots, n\}$, let ξ be some nonempty proper subset of $\mathbb{N}_n$, and let $m(\xi)$ be the cardinality of set ξ. For $m(\xi) < n$ the lower-dimension marginals are

$$H(\mathbf{z};\xi) = 1 - \left(\sum_{i\in\xi} \left(T_i z_i^{-\theta} \right)^{1/(1-\rho)} \right)^{1-\rho}$$

with support $z_i \geq \widetilde{T}^{1/\theta}$ for $i \in \xi$. For the special case in which $m(\xi) = 1$, so $\xi = \{i\}$ for some $i \in \mathbb{N}_n$, then $H(\mathbf{z};\xi) = F_i(z_i) = 1 - T_i z_i^{-\theta}$.

iv. **Discontinuities at the boundary.** The lower-dimension marginals $H(\mathbf{z};\xi)$ (with $m(\xi) < n$) are discontinuous at any point $\mathbf{z}$ with $z_i = \widetilde{T}^{1/\theta}$ for $i \in \xi$. Focusing again on the special case with $\xi = \{i\}$ for some $i \in \mathbb{N}_n$, then the discontinuity at $\mathbf{z}$ with $z_i = \widetilde{T}^{1/\theta}$ can be seen by noting that $F_i(\widetilde{T}^{1/\theta}) = 1 - T_i/\widetilde{T} > 0$ while $F_i(\widetilde{T}^{1/\theta} - \varepsilon) = 0$ for any positive ε. The distribution $H(\cdot)$ is also discontinuous at any point $\mathbf{z}$ in which $z_i = \widetilde{T}^{1/\theta}$ for some i except if $z_i = \widetilde{T}^{1/\theta}$ for all i – that is, $H(\cdot)$ is continuous in $(\widetilde{T}^{1/\theta}, \infty]^n \cup \{z^*\}$ where $z^* \equiv (\widetilde{T}^{1/\theta}, \cdots, \widetilde{T}^{1/\theta})$.

v. **Pareto tails.** The conditional marginal distribution for $z_i \geq a > \widetilde{T}^{1/\theta}$ is Pareto of type I (Arnold, 2015),

$$\Pr(Z_i \geq z_i \mid Z_i \geq a) = \left(\frac{z_i}{a} \right)^{-\theta}.$$

We interpret this result to mean that the distribution $H(\cdot)$ has Pareto tails. Of course, the same property holds in the Archimedian copula of Section 1.

vi. **The role of ρ.** The upper tail dependence parameter between Z_i and Z_j is

$$\lambda_U \equiv \lim_{u\uparrow 1} \Pr(F_i(Z_i) > u \mid F_j(Z_j) > u) = 2 - 2^{(1-\rho)},$$

which is the same as that of the Archimedian copula in Section 1. Note also that, as in the bivariate distribution in Section 1, as $\rho \to 1$, we have

$$H(\mathbf{z}) \to \min\{F_1(z_1), \ldots, F_n(z_n)\},$$

where $F_i(\cdot)$ is the one dimensional marginal for each i. This implies that as $\rho \to 1$ our distribution $H(\mathbf{z})$ converges to the Fréchet-Hoeffding upper bound copula and in this case Z_i's are not only comonotonic (Dhaene *et al.*, 2002) but also pairwise perfectly correlated. On the other hand, if $\rho = 0$, then the proposed distribution $H(\cdot)$ is singular; concretely, the density is zero (i.e., $h(\mathbf{z}) = 0$) almost everywhere on $\overline{\mathbb{R}}^n$ and the distribution is concentrated on the set $\bigcup_{i=1}^n \{(z_1, \ldots, z_n) : z_i \geq z_j = \widetilde{T}^{1/\theta}$ for all $j \neq i\}$, which is of Lebesgue measure zero. Note that with $\rho = 0$ we can write

$$H(\mathbf{z}) = \sum_{i=1}^{n} (T_i/\tilde{T})(1 - \tilde{T} z_i^{-\theta}),$$

with $\tilde{T} = \sum_{j=1}^n T_j$. This is equivalent to choosing i with probability $T_i/\tilde{T}$ and then having $Z_j = \tilde{T}^{1/\theta}$ for all $j \neq i$ and Z_i distributed according to $\Pr(Z_i \leq z_i) = 1 - \tilde{T} z_i^{-\theta}$.

vii. **Stochastic dominance with respect to T.** If $T_i' \geq T_i$ for all i then $H_{T'}(\mathbf{z})$ stochastically dominates $H_T(\mathbf{z})$: $H_{T'}(\mathbf{z}) \leq H_T(\mathbf{z})$.

viii. **Stochastic dominance with respect to θ.** Let $\theta' \geq \theta$ and $\widetilde{T}^{1/\theta} \geq 1$; then $H_\theta(\mathbf{z})$ stochastically dominates $H_{\theta'}(\mathbf{z})$: $H_{\theta'}(\mathbf{z}) \geq H_\theta(\mathbf{z})$.

3.2. A Brief Comparison with Multivariate Gumbel and Fréchet Distributions

Our proposed MVP features a Pareto-distributed maximum (Property i above) and Pareto tails (Property v above). These two properties are particularly useful for applied work but are absent from other multivariate distributions commonly used in applications.

To facilitate comparisons with two important multivariate distributions widely used in economics, we collect the relevant results in the next two Propositions. In contrast to our proposed distribution, we show that the multivariate Gumbel has a Gumbel-distributed maximum and exponentially decreasing tails, whereas the multivariate Fréchet has a Fréchet-distributed maximum and tails that exhibit polynomial decay.[6]

Let $(Z_1, \ldots, Z_n)$ follow the distribution with its cumulative distribution function F. We have the following two results.

Proposition 1. *Consider the generalized extreme value (GEV) distribution in McFadden (1978, p. 80)*

$$F(z_1, \ldots, z_n) \equiv \exp\left[-G(\exp\{-z_1\}, \ldots, \exp\{-z_n\})\right] \quad \textit{for} \quad (z_1, \ldots, z_n) \in \mathbb{R}^n,$$

[6] Although these results may not be new in the literature, they are collected in Propositions 1 and 2 for ease of comparison, with proofs in the Appendix.

where $G : \mathbb{R}_+^n \to \mathbb{R}_+$ is a nonnegative function satisfying the following assumptions:

a. G is homogeneous of degree one; b. $G(e_1, \ldots, e_n) \to \infty$ if any $e_i \to \infty$; c. the cross-partial derivatives of G satisfy,

$$\frac{\partial G}{\partial e_i} \geq 0 \ \textit{ for all } \ i, \ \ \frac{\partial^2 G}{\partial e_j \partial e_i} \leq 0 \ \textit{ for all } \ j \neq i, \ \ \frac{\partial^3 G}{\partial e_k \partial e_j \partial e_i} \geq 0 \ \textit{ for all } \ k \neq j \neq i,$$

and so on for higher-order cross-partial derivatives.

The following properties hold:

(i) The recentered maximum, $\max\{Z_1, \ldots, Z_n\} - \log\{G(1, \ldots, 1)\}$, follows the univariate standard Gumbel distribution; in particular,

$$\Pr\left(\max\{Z_1, \ldots, Z_n\} \leq z\right) = \exp\{-\exp\{-(z - \log\{G(1, \ldots, 1)\})\}\}$$

for every real number z.

(v) The tails exhibit exponential decay: For every z, a with $z \geq a \geq 0$,

$$(1 + g_i)^{-1} \exp\{-(z - a)\} \leq \Pr(Z_i \geq z | Z_i \geq a) \leq (1 + g_i) \exp\{-(z - a)\},$$

where $g_i \equiv G(\iota_i)$ with ι_i being the n-dimensional vector with its ith element equal to one and the remaining elements equal to zero.

Next, we present corresponding results for the multivariate Fréchet distribution (e.g. Eaton and Kortum 2002; Lind and Ramondo 2023).

Proposition 2. *Consider the Fréchet distribution*

$$F(z_1, \ldots, z_n) \equiv \exp\left[-H_\theta(z_1, \ldots, z_n)\right] \ \ \textit{for} \ \ (z_1, \ldots, z_n) \in \mathbb{R}_+^n,$$

where $H_\theta : \mathbb{R}_+^n \to \mathbb{R}_+$ is a nonnegative function with $\theta \geq 1$ and satisfies the following assumptions:

a. H_θ is homogeneous of degree $-\theta$; b. $H_\theta(z_1, \ldots, z_n) \to \infty$ if any $z_i \to 0$, and $H_\theta(z_1, \ldots, z_n) = T_i z_i^{-\theta}$ for some constant $T_i > 0$ when $z_j = \infty$ for all $j \neq i$; c. the *cross-partial derivatives of H_θ satisfy,*

$$\frac{\partial H_\theta}{\partial z_i} \le 0 \;\; \textit{for all} \;\; i, \quad \frac{\partial^2 H_\theta}{\partial z_j \partial z_i} \le 0 \;\; \textit{for all} \;\; j \neq i, \quad \frac{\partial^3 H_\theta}{\partial z_k \partial z_j \partial z_i} \le 0 \;\; \textit{for all} \;\; k \neq j \neq i,$$

and so on for higher-order cross-partial derivatives.

The following properties hold:

(i) The rescaled maximum, $\max\{Z_1, \ldots, Z_n\}\,[H_\theta(1, \ldots, 1)]^{(-1/\theta)}$*, follows the univariate Fréchet distribution with shape parameter* θ*; in particular,*

$$\Pr(\max\{Z_1, \ldots, Z_n\} \le z) = \left(\exp\left\{-z^{-\theta}\right\}\right)^{H_\theta(1,\ldots,1)}$$

for every positive real number z*.*

(v) The tails of this multivariate Fréchet distribution exhibit polynomial decay. For every real numbers z *and* a *with* $z \ge a > 0$*,*

$$\frac{a^\theta}{T_i + z^\theta} \le \Pr(Z_i \ge z | Z_i \ge a) \le \frac{T_i + a^\theta}{z^\theta}.$$

To summarize, these popular cases in the literature have analytical formulas for properties (i), (v). But the rescaled maximum and the tails of these distributions inherit the shapes of their corresponding distributions while the MVP exhibits Pareto maximum and tails.

Furthermore, the joint and conditional probabilities of the multivariate Gumbel and Fréchet distributions, which bear some resemblance to Property ii in Section 3.1, might be of interest in their own right. A convenient form for those probabilities, however, necessitates more mathematical structure. In what follows, we present the relevant results for the joint and conditional probabilities associated with multivariate Gumbel and Fréchet distributions that are commonly used in literature.

A leading example of McFadden's (1978) GEV distribution is the multinomial logit model with $G(e_1, \ldots, e_n) = \sum_{i=1}^n e_i$, in which the joint and conditional probabilities with respect to a minimum value z are, respectively,

$$\Pr\left(\arg\max_i Z_i = j \cap \max_i Z_i \ge z\right) = \frac{1}{n}\left[1 - \exp\{-n \exp\{-z\}\}\right]$$

and

$$\Pr\left(\arg\max_i Z_i = j | \max_i Z_i \geq z\right) = \frac{1}{n}$$

for every $z \in \mathbb{R}$ and every $j = 1, \ldots, n$. Another example, given in Bhat and Guo (2004, pp. 151-152), is the spatially correlated logit model with

$$G(e_1, \ldots, e_n) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left[(\alpha_{i,ij} e_i)^{1/\tilde{\rho}} + (\alpha_{j,ij} e_j)^{1/\tilde{\rho}}\right]^{\tilde{\rho}},$$

where $0 < \alpha_{i,ij} < 1$ for all i and j, $0 < \tilde{\rho} \leq 1$, $\sum_{j=1}^{n} \alpha_{i,ij} = 1$ for all i, and $e_i > 0$ for all i. Letting $c_\alpha \equiv G(1, \ldots, 1) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left[\alpha_{i,ij}^{1/\tilde{\rho}} + \alpha_{j,ij}^{1/\tilde{\rho}}\right]^{\tilde{\rho}}$ and

$$c_{\alpha,j} \equiv \begin{cases} \sum_{k=2}^{n} \left[\alpha_{1,1k}^{1/\tilde{\rho}} + \alpha_{k,1k}^{1/\tilde{\rho}}\right]^{(\tilde{\rho}-1)} \alpha_{1,1k}^{1/\tilde{\rho}}, & \text{if } j = 1, \\ \sum_{k=j+1}^{n} \left[\alpha_{j,jk}^{1/\tilde{\rho}} + \alpha_{k,jk}^{1/\tilde{\rho}}\right]^{(\tilde{\rho}-1)} \alpha_{j,jk}^{1/\tilde{\rho}} + \sum_{i=1}^{j-1} \left[\alpha_{i,ij}^{1/\tilde{\rho}} + \alpha_{j,ij}^{1/\tilde{\rho}}\right]^{(\tilde{\rho}-1)} \alpha_{j,ij}^{1/\tilde{\rho}}, & \text{if } 2 \leq j \leq n-1, \\ \sum_{i=1}^{n-1} \left[\alpha_{i,in}^{1/\tilde{\rho}} + \alpha_{n,in}^{1/\tilde{\rho}}\right]^{(\tilde{\rho}-1)} \alpha_{n,in}^{1/\tilde{\rho}}, & \text{if } j = n, \end{cases}$$

we have

$$\Pr\left(\arg\max_i Z_i = j \cap \max_i Z_i \geq z\right) = \frac{c_{\alpha,j}}{c_\alpha} \left[1 - \exp\{-c_\alpha \exp\{-z\}\}\right]$$

and

$$\Pr\left(\arg\max_i Z_i = j | \max_i Z_i \geq z\right) = \frac{c_{\alpha,j}}{c_\alpha}$$

for every $z \in \mathbb{R}$ and every $j = 1, \ldots, n$.

Taking correlation among $Z_1, \ldots, Z_n$ into account, Eaton and Kortum (2002) consider the multivariate Fréchet distribution,

$$H_\theta(z_1, \ldots, z_n; \varrho) = \left[\sum_{i=1}^{n} \left(T_i z_i^{-\theta}\right)^{1/\varrho}\right]^{\varrho}$$

for some $\theta > 1$ and $\varrho \in (0, 1]$.[7] The corresponding joint and conditional probabilities

[7] A special case of this distribution is the case of uncorrelated draws, $\varrho = 1$ considered by Hsieh *et al.* (2019, p. 1444).

under this distribution are

$$\Pr\left(\arg\max_i Z_i = j \cap \max_i Z_i \geq z\right) = \frac{T_j^{1/\varrho}}{\sum_{i=1}^n T_i^{1/\varrho}}\left[1 - \exp\left\{-\left[\sum_{i=1}^n T_i^{1/\varrho}\right]^{\varrho} z^{-\theta}\right\}\right]$$

and

$$\Pr\left(\arg\max_i Z_i = j | \max_i Z_i \geq z\right) = \frac{T_j^{1/\varrho}}{\sum_{i=1}^n T_i^{1/\varrho}}.$$

Focusing on property ii, the joint probability under the multivariate Fréchet distribution differs from that under the MVP, while the conditional probability has the same algebraic structure in both cases.

4. A Multivariate Pareto Distribution

To simplify the exposition, we assume that $\widetilde{T} = 1$. This is without loss of generality, since we can apply the simple transformation $Z_i \to \widetilde{T}^{(-1/\theta)} Z_i$ for each i to get this form. Then, defining the distribution on $\overline{\mathbb{R}}^n$, we have

$$H(\mathbf{z}) = \begin{cases} 1 - \left[\sum_{i=1}^n (T_i z_i^{-\theta})^{(1/(1-\rho))}\right]^{1-\rho} & \text{if } \mathbf{z} \in [1, \infty]^n; \\ 0 & \text{if } z_i < 1 \text{ for some } i. \end{cases}$$

In the rest of this note we show that $H(\cdot)$ is indeed a distribution function. To do so, we need to show that $H(\cdot)$ satisfies the following two conditions (see Nelsen (2006) page 46).

Condition 1 (Lower and upper limits)

$\lim_{\mathbf{z}\to+\infty} H(\mathbf{z}) = 1$ and $H(\mathbf{z}) = 0$ for all $\mathbf{z} \in \overline{\mathbb{R}}^n$ such that $z_i = -\infty$ for at least one i.

Condition 2 (n-increasing)

$H(\cdot)$ is *n-increasing*. Formally, for any two vectors $\mathbf{a}, \mathbf{b} \in \overline{\mathbb{R}}^n$ with $\mathbf{a} \leq \mathbf{b}$ and $B \equiv [\mathbf{a}, \mathbf{b}] = [a_1, b_1] \times ... \times [a_n, b_n]$ the *n-box* formed by the vectors a and b, H is *n-increasing* if and only if

$$V_H(B) \equiv \sum_{c \in \Theta(B)} sgn(\mathbf{c})\, H(\mathbf{c}) \geq 0,$$

where $\Theta(B)$ is the set of vertices of B and $sgn\,(\mathbf{c})$ is given by

$$sgn\,(\mathbf{c}) = \begin{cases} 1, & \text{if } c_k = a_k \text{ for an even number of } k\text{'s;} \\ -1, & \text{if } c_k = a_k \text{ for an odd number of } k\text{'s.} \end{cases}$$

Condition 1 holds because $\lim_{\mathbf{z}\to\infty} H(\mathbf{z}) = 1$ follows directly from $\lim_{z_i\to\infty} z_i^{-\theta/(1-\rho)} = 0$ for all i and $H(\mathbf{z}) = 0$ for $\mathbf{z} \notin [1,\infty]^n$. Indeed, we have $H(\mathbf{z}) \in [0,1]$ for all $\mathbf{z} \in \overline{\mathbb{R}}^n$. Since $\left[\sum_{i=1}^n (T_i z_i^{-\theta})^{1/(1-\rho)}\right]^{1-\rho} \geq 0$ then $H(\mathbf{z}) \leq 1$, and since $H(\mathbf{z})$ is increasing in each z_i (for $\mathbf{z} \in [1,\infty]^n$) then to show that $H(\mathbf{z})$ is non-negative it is sufficient to note that $H(1,1,...,1) = 0$.

The challenge in proving that $H(\cdot)$ is a proper distribution lies in showing that Condition 2 holds. It is of course easy to establish that Condition 2 holds if $\frac{\partial^n H(\mathbf{z})}{\partial z_1 \partial z_2 \ldots \partial z_n}$ exists at the entire support of the distribution. But as we explained in the previous section, the function is discontinuous at the boundary of the support, and thus this derivative does not exist there. This observation shows that the *n*-increasing property is not obvious because of discontinuity at the boundary of the support.

Since this discontinuity is irrespective of Paretian properties, we consider a generic function $G(\cdot)$ with discontinuity at the boundary of the support $[1,\infty]^n$ and first establish a general theorem, which shows that if the function $G(\cdot)$ has a discrete component along a square, and is smooth everywhere else, then it is n-increasing. We then check conditions of this theorem to show $H(\cdot)$ is *n*-increasing.

4.1. Main Result

We introduce some additional definitions. Let $\mathbb{N}_n \equiv \{1, 2, \ldots, n\}$ and consider an n-Box B. Let $A(\mathbf{c}; B) \equiv \{k | c_k = a_k\}$ and $m(S)$ the cardinality of the set S, then we can write

$$V_G\,(B) = \sum_{\mathbf{c}\in\Theta(B)} (-1)^{m(\mathrm{A}(\mathbf{c};B))} G\,(\mathbf{c})\,.$$

For future reference, we refer to $V_G(B)$ as the G-volume of the box B. For any proper subset $\xi \subset \mathbb{N}_n$, let $l(\xi) = n - m(\xi)$. Let $\mathbf{x} \in \overline{\mathbb{R}}^n$, let $\mathbf{x}(\xi)$ be the vector in $\overline{\mathbb{R}}^{l(\xi)}$ defined by taking out from x all the elements $i \in \xi$, and let $\bar{\mathbf{x}}(\xi)$ be the vector in $\overline{\mathbb{R}}^n$ with *i*-th element equal to one if $i \in \xi$ and x_i otherwise. Let G_ξ be a mapping from $\overline{\mathbb{R}}^{l(\xi)}$ to $[0,1]$ defined as $G_\xi : \mathbf{x}(\xi) \mapsto G(\bar{\mathbf{x}}(\xi))$. Finally, define ϕ to be the empty set.[8] Loosely speaking, we can think of $V_{G_\xi}([\mathbf{a}(\xi), \mathbf{b}(\xi)])$ as the G-volume of $[\mathbf{a}, \mathbf{b}]$ in $l(\xi)$ dimensions.[9]

To prove that $G(\cdot)$ is *n*-increasing, we split any n-box B into sub-boxes, each of which is either all outside or all inside the box $[1,\infty]^n$, but never crossing the 1-axis. We show that the G-volume of the box B is the sum of the G_ξ-volume of these sub-boxes and that each G_ξ-volume is non-negative. Before we prove the main result below, let us first illustrate the split of a 2-box.

Figure 1 shows two possible 2-boxes. In Panel A, we consider a 2-box $B \equiv [a_1, b_1] \times [a_2, b_2]$ with $a_1 < 1 < b_1$ and $1 < a_2 < b_2$. Setting $\mathbf{a}^* = (1, a_2)$, we can decompose the G-volume of the box B as follows:

$$\begin{aligned} V_G(B) &= G(b_1, b_2) - G(b_1, a_2) - G(a_1, b_2) + G(a_1, a_2) \\ &= [G(b_1, b_2) - G(b_1, a_2) - G(1, b_2) + G(1, a_2)] + [G(1, b_2) - G(1, a_2)] \\ &= V_G([\mathbf{a}^*, \mathbf{b}]) + V_{G_{\{1\}}}([\mathbf{a}^*(\{1\}), \mathbf{b}(\{1\})]). \end{aligned}$$

Similarly, in Panel B, we consider a 2-box $B \equiv [a_1, b_1] \times [a_2, b_2]$ with $a_1 < 1 < b_1$ and $a_2 < 1 < b_2$ and let $\mathbf{a}^* = (1,1)$. The G-volume of this box can similarly be decomposed as follows:

$$V_G(B) = V_G([\mathbf{a}^*, \mathbf{b}]) + V_{G_{\{1\}}}([\mathbf{a}^*(\{1\}), \mathbf{b}(\{1\})]) + V_{G_{\{2\}}}([\mathbf{a}^*(\{2\}), \mathbf{b}(\{2\})]).$$

Note that we need not add the box corresponding to $\xi = \{1,2\} = \mathbb{N}_2$ since this box obviously has zero volume. As proved in Theorem 1 below, the decomposition of a general n-box is valid and each G_ξ volume is non-negative as long as its density g_ξ is non-negative on $(1,\infty]^{l(\xi)}$. It is thus obvious that the function $G(\cdot)$ is *n*-increasing. A general theorem is stated as follows.

Theorem 1. *Suppose a function* $G : \overline{\mathbb{R}}^n \to \mathbb{R}$ *satisfies (i)* $G(\mathbf{z}) = 0$ *for* $\mathbf{z} \notin [1,\infty]^n$*, (ii)* $g(\mathbf{z}) = \frac{\partial^n G(\mathbf{z})}{\partial z_1 \partial z_2 \ldots \partial z_n} \geq 0$ *on* $(1,\infty]^n$*, and (iii)* $g_\xi(\cdot) = \frac{\partial G_\xi(\cdot)}{\partial x(\xi)} \geq 0$ *on* $(1,\infty]^{l(\xi)}$ *for each nonempty* $\xi \neq \mathbb{N}_n$*. Then* $G(\cdot)$ *is n-increasing.*

[8]For example, for $n = 2$ and $\xi = \{1\}$, we have $\mathbf{x}(\xi) = \mathrm{x}_2$, $\bar{\mathbf{x}}(\xi) = (1, \mathrm{x}_2)$ and $G_\xi(x_2) = G(1, x_2)$. If $\xi = \phi$ the empty set, we have $\mathbf{x}(\xi) = \bar{\mathbf{x}}(\xi) = \mathbf{x}$ and $G_\xi = G_\phi = G$.

[9]In the example above with $n = 2$ and $\xi = \{1\}$, $V_{G_\xi}([\mathbf{a}(\xi), \mathbf{b}(\xi)]) = G(1, b_2) - G(1, a_2)$.

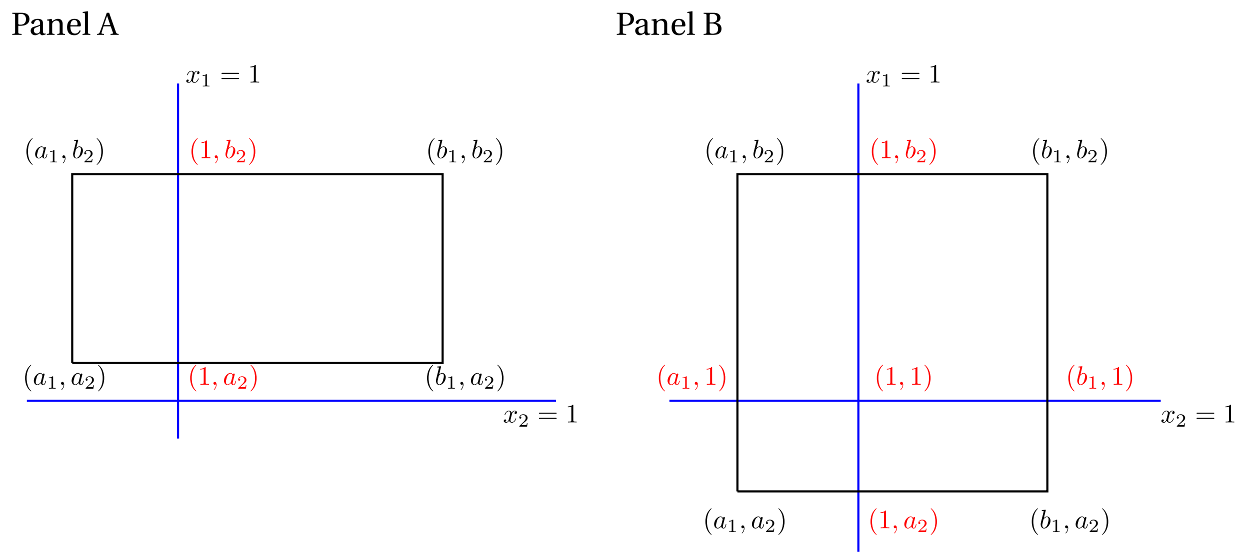


Figure 1: Decomposition of the G-volume of a 2-box

Proof. The proof is deferred to the Appendix. □

Remark 1. *Condition (iii) in Theorem 1 ensures that functions G_ξ are smooth on $(1, \infty]^{l(\xi)}$ so that the G_ξ-volumes are non-negative by the fundamental theorem of calculus. The functions g_ξ may be regarded as densities on $(1, \infty]^{l(\xi)}$ and can be calculated from the function G so that Condition (iii) is easy to check from the function G. For example, consider a function*

$$G(z_1, z_2; \rho) = \begin{cases} \int_{-\infty}^{z_2} \int_{-\infty}^{z_1} \phi(x_1, x_2; \rho) dx_1 dx_2, & \textit{if } z_1 \geq 1 \textit{ and } z_2 \geq 1; \\ 0, & \textit{otherwise}, \end{cases}$$

where $\phi(\cdot, \cdot; \rho)$ is the density function of the centered bivariate normal with variances one and covariance $\rho \in [0, 1)$. Conditions (i) and (ii) hold clearly. It is also easy to show that

$$g_{\{1\}}(u; \rho) = \int_{-\infty}^{1} \phi(x_1, u; \rho) dx_1 \geq 0 \quad \textit{and} \quad g_{\{2\}}(u; \rho) = \int_{-\infty}^{1} \phi(u, x_2; \rho) dx_2 \geq 0.$$

Applying Theorem 1, we conclude that $G(\cdot, \cdot; \rho)$ is 2-increasing.

Equipped with Theorem 1, we now turn to show that the function $H(\cdot)$ defined in (1) is indeed a distribution function. As stated in the beginning of Section 3, $H(\cdot)$ satisfies Condition 1. To establish that $H(\cdot)$ is n-increasing, we need to check conditions (*ii*) and (*iii*) of Theorem 1. Note that the density of H on $(1,\infty]^n$ is

$$h(\mathbf{z}) = (1-\rho)\left(\frac{\theta}{1-\rho}\right)^n \left[\prod_{i=1}^{n-1}(i-(1-\rho))\right]\left[\prod_{i=1}^{n} z_i^{-1}\left(T_i z_i^{-\theta}\right)^{\left(\frac{1}{1-\rho}\right)}\right]\left[\sum_{i=1}^{n}\left(T_i z_i^{-\theta}\right)^{\left(\frac{1}{1-\rho}\right)}\right]^{(1-\rho-n)},$$

which is non-negative. Moreover, simple calculation shows that for each nonempty $\xi \neq \mathbb{N}_n$,

$$\begin{aligned} h_\xi(\mathbf{x}(\xi)) &= \frac{\partial H_\xi(\mathbf{x}(\xi))}{\partial \mathbf{x}(\xi)} \\ &= \begin{cases} (1-\rho)\left(\frac{\theta}{1-\rho}\right)^{l(\xi)}\left[\prod_{i=1}^{l(\xi)-1}(i-(1-\rho))\right]Q(\xi), & \text{if } l(\xi)\geq 2, \\ \theta Q(\xi), & \text{otherwise,} \end{cases} \end{aligned}$$

where

$$Q(\xi) = \left[\prod_{i\in\mathbb{N}_n\backslash\xi} \mathbf{x}_i^{-1}\left(T_i\mathbf{x}_i^{-\theta}\right)^{\left(\frac{1}{1-\rho}\right)}\right]\left[\sum_{i\in\xi} T_i^{\left(\frac{1}{1-\rho}\right)} + \sum_{i\in\mathbb{N}_n\backslash\xi}\left(T_i\mathbf{x}_i^{-\theta}\right)^{\left(\frac{1}{1-\rho}\right)}\right]^{(1-\rho-l(\xi))}.$$

These results are summarized in the following corollary.

Corollary 1. *The function $H(\cdot)$ defined in (1) is indeed a distribution function.*

5. Applications

The multivariate Pareto distribution analyzed here was introduced in Arkolakis *et al.* (2018), where its key properties were exploited in a model of multinational production. Subsequent work has built on this structure to model firm heterogeneity and location sorting in environments with correlated shocks, leveraging features such as Pareto marginals, a Pareto-distributed maximum, and a tractable dependence structure. In particular, the Pareto marginals have been used in the multinational production literature to capture the heavy-tailed distribution of firm productivity across regions, shaping endogenous firm location choices (Lan, 2019; Li, 2025; Liu and Ma, 2023). Xiang (2023) embeds this structure into a semi-endogenous growth framework, where plants draw ideas with different productivities across locations governed by the MVP. The Pareto-distributed maximum delivers closed-form solutions for a firm's choice of its most profitable market, as illustrated in studies of online retailers (Li, 2025) and multinational firms (Fan and Luo, 2026).

Property (ii) of the MVP follows from structural features of extreme-value distributions that it shares with the multivariate Fréchet distribution, as shown above. As a result, both the MVP and the Fréchet yield convenient CES-like expressions for conditional shares, but with one key distinction. The MVP delivers such expressions only in models with truncation, and only when the truncation threshold is sufficiently high that we do not have to integrate over the boundary. By contrast, the Fréchet yields CES-like expressions only in the absence of truncation. We illustrate the distinct applicability of these distributions with two examples.

Trade and multinational production. Consider the trade and multinational production model in Arkolakis *et al.* (2018), which assumes that productivity is distributed MVP. The model features a Melitz-type fixed marketing cost to serve any given market n. Focusing on sales by firms from a given Home country i to a destination n, Property (ii) of the MVP implies that the share of those sales produced in location l takes a CES-like form as long as the lower boundary of the support of the MVP entails productivities that are too low to be profitable when serving market n. Specifically, ignoring trade and MP frictions here for simplicity, and letting w_l denote the wage in country l and c_n^* denote the unit-cost cutoff to serve market n, the key assumption for tractability in Arkolakis *et al.* (2018) is $w_l > \left(\sum_l T_{il}^{1/(1-\rho)}\right)^{1-\rho}$ for all i, l, n. As is typical in Melitz (2003)-style models, this assumption implies that for any i, l, n there are always productivity draws for firms from i producing in l that are too low to profitably serve market n. In this case, aggregate sales in market n by affiliates in j of firms from i, denoted by X_{ijn}, satisfy

$$\frac{X_{ijn}}{\sum_l X_{iln}} = \frac{\left(T_{ij} w_j^{-\theta}\right)^{\frac{1}{1-\rho}}}{\sum_l \left(T_{il} w_l^{-\theta}\right)^{\frac{1}{1-\rho}}}.$$

The same expression holds under the Fréchet specification, except that one must assume away fixed marketing costs. This is the approach followed by Ramondo and Rodríguez-Clare (2013), who assume perfect competition with constant returns to scale and no fixed costs.

Roy model. Next, consider the Roy model as in Lagakos and Waugh (2013), Hsieh *et al.* (2019), and Galle *et al.* (2010). This literature typically assumes that sectoral productivity $(Z_1, \ldots, Z_K)$ follows a multivariate Fréchet distribution. Given wages per efficiency unit $(w_1, \ldots, w_K)$, a key implication is that the share of workers choosing sector k is given by

$$\pi_k = \frac{\left(T_k w_k^\theta\right)^{\frac{1}{1-\rho}}}{\sum_l \left(T_l w_l^\theta\right)^{\frac{1}{1-\rho}}}.$$

Moreover, the Fréchet assumption implies equalization of average income across sectors, $\frac{w_k \mathcal{E}_k}{\pi_k L} = \xi \Psi(w)$, where $\Psi(w) \equiv \left(\sum_j (T_j w_j^\theta)^{\frac{1}{1-\rho}}\right)^{\frac{1-\rho}{\theta}}$, $\mathcal{E}_k$ is the supply of efficiency units to sector k, L is the total number of workers, and ξ is a constant.

Assuming that productivity is distributed MVP instead of Fréchet yields the same allocation once a cutoff is introduced. For example, suppose workers drop out of the labor force if $\max_i w_i z_i \leq m$. If m is sufficiently large—so that no agent would be willing to work with a productivity at the boundary of the distribution—the CES-like expression for π_k remains valid under MVP.[10] The income-equalization result $\frac{w_k \mathcal{E}_k}{\pi_k L} = \tilde{\xi} \tilde{\Psi}(w)$ continues to hold as well, except that $\tilde{\xi}$ now depends on the cutoff m and $\tilde{\Psi}(w) = [\Psi(w)]^\theta$.

6. Conclusion

Since its introduction in Arkolakis *et al.* (2018), the multivariate Pareto distribution has become a practical tool across a small but growing body of applied and theoretical research in international trade, spatial economics, and innovation. Across these diverse applications, the distribution has proven to be both flexible and tractable, offering researchers a robust framework to model rich patterns of heterogeneity and spatial sorting while preserving analytical tractability. We demonstrate the key properties of this distribution and how its properties differ from the well-known multivariate Gumbel and Fréchet distributions. Together, these distributions constitute a complete toolbox that offers tractable and intuitive solutions for the economics of discrete choice.

[10] Specifically, the condition for this result is $m > \left(\sum_j T_j^{1/(1-\rho)}\right)^{\frac{1-\rho}{\theta}} \max_k w_k$.

Appendix

1. Properties of the Proposed Distribution

The maximum

Let $Z \equiv \max(Z_1, ..., Z_n)$. The distribution function of Z is

$$\Pr(Z \leq z) = H(z, \ldots, z) = \begin{cases} 1 - \widetilde{T} z^{-\theta} & \text{if } z \geq \widetilde{T}^{1/\theta}; \\ 0 & \text{otherwise} . \end{cases}$$

This shows Z is distributed Pareto with shape parameter θ and scale parameter $\widetilde{T}^{1/\theta}$.

Conditional probabilities

Let $h_j(z)$ be the marginal density of Z_j. Then

$$\Pr\left(\arg\max_i Z_i = j \cap \max_i Z_i \geq z\right) = \int_z^\infty \Pr(Z_i \leq u, \forall i \neq j | Z_j = u) h_j(u) du.$$

Noting that, for $u > \widetilde{T}^{1/\theta}$,

$$\begin{aligned} \Pr(Z_i \leq u, \forall i \neq j | Z_j = u) h_j(u) &= \left.\frac{\partial \Pr\left(Z_1 \leq u, ..., Z_j \leq z_j, ..., Z_n \leq u\right)}{\partial z_j}\right|_{z_j = u} \\ &= \frac{T_j^{1/(1-\rho)}}{\widetilde{T}^{1/(1-\rho)}} \widetilde{T} \theta u^{-\theta-1}, \end{aligned}$$

it follows that, for $z > \widetilde{T}^{1/\theta}$,

$$\Pr\left(\arg\max_i Z_i = j \cap \max_i Z_i \geq z\right) = \frac{T_j^{1/(1-\rho)}}{\widetilde{T}^{1/(1-\rho)}} \widetilde{T} z^{-\theta}.$$

Marginal distributions

For any nonempty proper subset ξ of $\mathbb{N}_n$ and $\mathbf{z} \in \overline{\mathbb{R}}^{m(\xi)}$, let $\mathbf{z}^*(\xi)$ be the vector in $\overline{\mathbb{R}}^n$ with i-th element equal to z_i if $i \in \xi$ and ∞ otherwise. The lower-dimension marginal corresponding to ξ is

$$H(\mathbf{z};\xi) = H(\mathbf{z}^*(\xi)) = \begin{cases} 1 - \left(\sum_{i\in\xi}\left(T_i z_i^{-\theta}\right)^{1/(1-\rho)}\right)^{1-\rho} & \text{if } z_i \geq \widetilde{T}^{1/\theta} \text{ for all } i \in \xi; \\ 0 & \text{otherwise .} \end{cases}$$

Pareto tails

The distribution of $Z_i | Z_i \geq a$ is Pareto because

$$\Pr(Z_i \geq z \mid Z_i \geq a) = \frac{\Pr(Z_i \geq z)}{\Pr(Z_i \geq a)} = \left(\frac{z}{a}\right)^{-\theta}.$$

for $z \geq a > \widetilde{T}^{1/\theta}$.

The role of ρ

Fix a point $\mathbf{z} = (z_1, \ldots, z_n) \in \overline{\mathbb{R}}^n$. If $z_i < \widetilde{T}^{1/\theta}$ for some i, then $H_\rho(\mathbf{z}) = 0 = \min_j F_j(z_j)$ because $F_i(z_i) = 0$. Suppose $z_i \geq \widetilde{T}^{1/\theta}$ for all i. Without loss of generality, we assume $T_1 z_1^{-\theta} = \max_j T_j z_j^{-\theta}$. It follows that as $\rho \to 1$,

$$\begin{aligned} H_\rho(\mathbf{z}) &= 1 - T_1 z_1^{-\theta}\left(\sum_{i=1}^n \left(\frac{T_i z_i^{-\theta}}{T_1 z_1^{-\theta}}\right)^{1/(1-\rho)}\right)^{1-\rho} \\ &\to 1 - T_1 z_1^{-\theta} \\ &= \min_j\{1 - T_j z_j^{-\theta}\} \\ &= \min_j F_j(z_j). \end{aligned}$$

This shows that for any $\mathbf{z}$, we have $H_\rho(\mathbf{z}) \to \min\{F_1(z_1), \ldots, F_n(z_n)\}$ as $\rho \to 1$.

It also follows that for every pair (i,j), $H(\mathbf{z}; \{i,j\}) \to 1 - \max\{T_i z_i^{-\theta}, T_j z_j^{-\theta}\}$ as $\rho \to 1$. This implies that all mass points (Z_i, Z_j) are on the line

$$\left\{(z_i, z_j) : z_i = \left(\frac{T_j}{T_i}\right)^{-\frac{1}{\theta}} z_j\right\},$$

and hence that Z_i and Z_j are perfectly correlated as $\rho \to 1$.

Stochastic dominance with respect to T

Consider $T_i' \geq T_i$ for all i. It is clear that $H_{T'}(\mathbf{z}) = 0 \leq H_T(\mathbf{z})$ if $z_i < \left(\widetilde{T}'\right)^{1/\theta}$ for some i. Suppose $z_i \geq \left(\widetilde{T}'\right)^{1/\theta}$ for all i. We have

$$\left(\sum_{i=1}^{n}\left(T_i z_i^{-\theta}\right)^{1/(1-\rho)}\right)^{1-\rho} \leq \left(\sum_{i=1}^{n}\left(T_i' z_i^{-\theta}\right)^{1/(1-\rho)}\right)^{1-\rho}$$

and thus

$$H_{T'}(\mathbf{z}) \leq H_T(\mathbf{z})$$

by definition of $H(\cdot)$.

Stochastic dominance with respect to θ

Consider $\theta' \geq \theta$. It is clear that $H_\theta(\mathbf{z}) = 0 \leq H_{\theta'}(\mathbf{z})$ if $z_i < \widetilde{T}^{1/\theta}$ for some i. Suppose $z_i \geq \widetilde{T}^{1/\theta}$ for all i. We have

$$\left(\sum_{i=1}^{n}\left(T_i z_i^{-\theta'}\right)^{1/(1-\rho)}\right)^{1-\rho} \leq \left(\sum_{i=1}^{n}\left(T_i z_i^{-\theta}\right)^{1/(1-\rho)}\right)^{1-\rho}$$

and thus

$$H_\theta(\mathbf{z}) \leq H_{\theta'}(\mathbf{z})$$

by definition of $H(\cdot)$.

2. Proof of Proposition 1

Suppose that $(Z_1, \ldots, Z_n)$ follows the GEV distribution.

The maximum

For every real number z,

$$\begin{aligned}\Pr\left(\max\left\{Z_1, \ldots, Z_n\right\} \leq z\right) &= \Pr\left(Z_1 \leq z, \ldots, Z_n \leq z\right) \\ &= \exp\left[-G(\exp\{-z\}, \ldots, \exp\{-z\})\right]\end{aligned}$$

$$
\begin{aligned}
&= \exp\left[-\exp\{-z\}G(1,\ldots,1)\right] \\
&= \exp\{-\exp\left\{-\left(z-\log\{G(1,\ldots,1)\}\right)\right\}\}
\end{aligned}
$$

because G is homogeneous of degree one.

Tail behavior

We use the basic inequality that $\exp\{x\} \geq 1+x$ for every real number x. We first note that for every real numbers z and a with $z \geq a$,

$$
\Pr(Z_i \geq z | Z_i \geq a) = \frac{\Pr(Z_i \geq z)}{\Pr(Z_i \geq a)} = \frac{1-\exp\{-g_i\exp\{-z\}\}}{1-\exp\{-g_i\exp\{-a\}\}}.
$$

In view of the basic inequality with $x = -g_i\exp\{-u\}$,

$$
1-\exp\left\{-g_i\exp\{-u\}\right\} \leq g_i\exp\{-u\}
$$

for every real number u. Applying the basic inequality with $x = g_i\exp\{-u\}$ yields

$$
\begin{aligned}
\exp\{-g_i\exp\{-u\}\} - 1 &= \frac{1}{\exp\{g_i\exp\{-u\}\}} - 1 \\
&\leq \frac{1}{1+g_i\exp\{-u\}} - 1 \\
&= -\frac{g_i\exp\{-u\}}{1+g_i\exp\{-u\}}
\end{aligned}
$$

for every real number u. It follows that whenever $u \geq 0$,

$$
\begin{aligned}
1-\exp\{-g_i\exp\{-u\}\} &\geq \frac{g_i\exp\{-u\}}{1+g_i\exp\{-u\}} \\
&\geq \frac{g_i}{1+g_i}\exp\{-u\}.
\end{aligned}
$$

Finally, combining these results, we obtain

$$
\begin{aligned}
\frac{1}{1+g_i}\exp\{-(z-a)\} &= \frac{\frac{g_i}{1+g_i}\exp\{-z\}}{g_i\exp\{-a\}} \\
&\leq \frac{1-\exp\{-g_i\exp\{-z\}\}}{1-\exp\{-g_i\exp\{-a\}\}}
\end{aligned}
$$

$$\leq \frac{g_i \exp\{-z\}}{\frac{g_i}{1+g_i}\exp\{-a\}}$$
$$= (1+g_i)\exp\{-(z-a)\}$$

for every real numbers z and a with $z \geq a \geq 0$.

3. Proof of Proposition 2

Suppose that $(Z_1, \dots, Z_n)$ follows the multivariate Fréchet distribution.

The maximum

For every positive real number z,

$$\begin{aligned}
\Pr(\max\{Z_1,\dots,Z_n\} \leq z) &= \Pr(Z_1 \leq z, \dots, Z_n \leq z) \\
&= \exp[-H_\theta(z,\dots,z)] \\
&= \exp\left[-\frac{H_\theta(1,\dots,1)}{z^\theta}\right] \\
&= \left(\exp\left\{-z^{-\theta}\right\}\right)^{H_\theta(1,\dots,1)}
\end{aligned}$$

because H_θ is homogeneous of degree $-\theta$.

Tail behavior

We use the basic inequality that $\exp\{x\} \geq 1+x$ for every real number x. We first note that for every real numbers z and a with $z \geq a$,

$$\Pr(Z_i \geq z | Z_i \geq a) = \frac{\Pr(Z_i \geq z)}{\Pr(Z_i \geq a)} = \frac{1-\exp\{-T_i z^{-\theta}\}}{1-\exp\{-T_i a^{-\theta}\}}.$$

In view of the basic inequality with $x = -T_i u^{-\theta}$,

$$1 - \exp\left\{-T_i u^{-\theta}\right\} \leq T_i u^{-\theta}$$

for every positive real number u. Applying the basic inequality with $x = T_i u^{-\theta}$ yields $1 + T_i u^{-\theta} \leq \exp\{T_i u^{-\theta}\}$, leading to $\exp\{-T_i u^{-\theta}\} \leq (1 + T_i u^{-\theta})^{-1}$ and

$$1 - \exp\left\{-T_i u^{-\theta}\right\} \geq 1 - \frac{1}{1 + T_i u^{-\theta}} = \frac{T_i}{T_i + u^{\theta}}$$

for every positive real number u. Finally, combining these results, we obtain

$$\begin{aligned}
\frac{a^{\theta}}{T_i + z^{\theta}} &= \frac{\frac{T_i}{T_i + z^{\theta}}}{T_i a^{-\theta}} \\
&\leq \frac{1 - \exp\{-T_i z^{-\theta}\}}{1 - \exp\{-T_i a^{-\theta}\}} \\
&\leq \frac{T_i z^{-\theta}}{\frac{T_i}{T_i + a^{\theta}}} \\
&\leq \frac{T_i + a^{\theta}}{z^{\theta}}.
\end{aligned}$$

for every real numbers z and a with $z \geq a > 0$.

4. Derivation of the Conditional Probabilities in Section 3.2.

- Suppose that $(Z_1, \ldots, Z_n)$ follows the GEV distribution in McFadden (1978) with $G(e_1, \ldots, e_n) = \sum_{i=1}^{n} e_i$. Let $h_j(z)$ be the marginal density of Z_j. Then

$$\begin{aligned}
\Pr\left(\arg\max_i Z_i = j \cap \max_i Z_i \geq z\right) &= \int_z^{\infty} \Pr(Z_i \leq u, \forall i \neq j | Z_j = u) h_j(u) du \\
&= \int_z^{\infty} \exp\{-n \exp\{-u\}\} \exp\{-u\} du \\
&= \int_0^{\exp\{-z\}} \exp\{-nv\} dv \\
&= -\frac{1}{n} \exp\{-nv\}\Big|_{v=0}^{\exp\{-z\}} \\
&= \frac{1}{n} \left[1 - \exp\{-n \exp\{-z\}\}\right].
\end{aligned}$$

Since $\Pr\left(\max_i Z_i \geq z\right) = 1 - \exp\{-n \exp\{-z\}\}$, we obtain

$$\Pr\left(\arg\max_i Z_i = j | \max_i Z_i \geq z\right) = \frac{\frac{1}{n}\left[1 - \exp\{-n \exp\{-z\}\}\right]}{1 - \exp\{-n \exp\{-z\}\}} = \frac{1}{n}.$$

- Suppose that $(Z_1, \ldots, Z_n)$ follows the distribution in Bhat and Guo (2004) with

$$F(z_1, \ldots, z_n) = \exp\left\{-G(\exp\{-z_1\}, \ldots, \exp\{-z_n\})\right\} \quad \text{for} \quad (z_1, \ldots, z_n) \in \mathbb{R}^n,$$

where

$$G(e_1, \ldots, e_n) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left[(\alpha_{i,ij} e_i)^{1/\tilde{\rho}} + (\alpha_{j,ij} e_j)^{1/\tilde{\rho}}\right]^{\tilde{\rho}}$$

with $0 < \alpha_{i,ij} < 1$ for all i and j, $0 < \tilde{\rho} \leq 1$, $\sum_{j=1}^{n} \alpha_{i,ij} = 1$ for all i, and $e_i > 0$ for all i. For every $j = 1, 2, \ldots, n$,

$$\frac{\partial F(z_1, \ldots, z_n)}{\partial z_j} = F(z_1, \ldots, z_n) \exp\{-z_j\} \frac{\partial G(\exp\{-z_1\}, \ldots, \exp\{-z_n\})}{\partial e_j}$$

and

$$\begin{aligned}
&\frac{\partial G(e_1, \ldots, e_n)}{\partial e_j} \\
&= \begin{cases}
\frac{\partial}{\partial e_1} \sum_{k=2}^{n} \left[(\alpha_{1,1k} e_1)^{1/\tilde{\rho}} + (\alpha_{k,1k} e_k)^{1/\tilde{\rho}}\right]^{\tilde{\rho}}, & \\
& \text{if } j = 1; \\
\frac{\partial}{\partial e_j} \left\{\sum_{k=j+1}^{n} \left[(\alpha_{j,jk} e_j)^{1/\tilde{\rho}} + (\alpha_{k,jk} e_k)^{1/\tilde{\rho}}\right]^{\tilde{\rho}} + \sum_{i=1}^{j-1} \left[(\alpha_{i,ij} e_i)^{1/\tilde{\rho}} + (\alpha_{j,ij} e_j)^{1/\tilde{\rho}}\right]^{\tilde{\rho}}\right\}, & \\
& \text{if } 2 \leq j \leq n-1; \\
\frac{\partial}{\partial e_n} \sum_{i=1}^{n-1} \left[(\alpha_{i,in} e_i)^{1/\tilde{\rho}} + (\alpha_{n,in} e_n)^{1/\tilde{\rho}}\right]^{\tilde{\rho}}, & \\
& \text{if } j = n;
\end{cases} \\
&= \begin{cases}
\sum_{k=2}^{n} \left[(\alpha_{1,1k} e_1)^{1/\tilde{\rho}} + (\alpha_{k,1k} e_k)^{1/\tilde{\rho}}\right]^{(\tilde{\rho}-1)} \alpha_{1,1k}^{1/\tilde{\rho}} e_1^{(1-\tilde{\rho})/\tilde{\rho}}, & \text{if } j = 1; \\
\sum_{k=j+1}^{n} \left[(\alpha_{j,jk} e_j)^{1/\tilde{\rho}} + (\alpha_{k,jk} e_k)^{1/\tilde{\rho}}\right]^{(\tilde{\rho}-1)} \alpha_{j,jk}^{1/\tilde{\rho}} e_j^{(1-\tilde{\rho})/\tilde{\rho}} & \\
\quad + \sum_{i=1}^{j-1} \left[(\alpha_{i,ij} e_i)^{1/\tilde{\rho}} + (\alpha_{j,ij} e_j)^{1/\tilde{\rho}}\right]^{(\tilde{\rho}-1)} \alpha_{j,ij}^{1/\tilde{\rho}} e_j^{(1-\tilde{\rho})/\tilde{\rho}}, & \text{if } 2 \leq j \leq n-1; \\
\sum_{i=1}^{n-1} \left[(\alpha_{i,in} e_i)^{1/\tilde{\rho}} + (\alpha_{n,in} e_n)^{1/\tilde{\rho}}\right]^{(\tilde{\rho}-1)} \alpha_{n,in}^{1/\tilde{\rho}} e_n^{(1-\tilde{\rho})/\tilde{\rho}}, & \text{if } j = n.
\end{cases}
\end{aligned}$$

Note that for ease of notation, we hereafter adopt the convention that

$$\sum_{k=n+1}^{n} \left[(\alpha_{j,jk} e_j)^{1/\tilde{\rho}} + (\alpha_{k,jk} e_k)^{1/\tilde{\rho}}\right]^{(\tilde{\rho}-1)} \alpha_{j,jk}^{1/\tilde{\rho}} e_j^{(1-\tilde{\rho})/\tilde{\rho}} = 0$$

and

$$\sum_{i=1}^{0}\left[(\alpha_{i,ij}e_i)^{1/\tilde{\rho}}+(\alpha_{j,ij}e_j)^{1/\tilde{\rho}}\right]^{(\tilde{\rho}-1)}\alpha_{j,ij}^{1/\tilde{\rho}}e_j^{(1-\tilde{\rho})/\tilde{\rho}}=0,$$

and thus simply write

$$\begin{aligned}\frac{\partial G(e_1,\ldots,e_n)}{\partial e_j}=&\sum_{k=j+1}^{n}\left[(\alpha_{j,jk}e_j)^{1/\tilde{\rho}}+(\alpha_{k,jk}e_k)^{1/\tilde{\rho}}\right]^{(\tilde{\rho}-1)}\alpha_{j,jk}^{1/\tilde{\rho}}e_j^{(1-\tilde{\rho})/\tilde{\rho}}\\&+\sum_{i=1}^{j-1}\left[(\alpha_{i,ij}e_i)^{1/\tilde{\rho}}+(\alpha_{j,ij}e_j)^{1/\tilde{\rho}}\right]^{(\tilde{\rho}-1)}\alpha_{j,ij}^{1/\tilde{\rho}}e_j^{(1-\tilde{\rho})/\tilde{\rho}}\end{aligned}$$

for every $j=1,\ldots,n$. It follows that

$$\begin{aligned}&\frac{\partial F(u,\ldots,u)}{\partial z_j}\\&=F(u,\ldots,u)\exp\{-u\}\frac{\partial G(\exp\{-u\},\ldots,\exp\{-u\})}{\partial e_j}\\&=F(u,\ldots,u)\exp\{-u\}\left\{\sum_{k=j+1}^{n}\left[\alpha_{j,jk}^{1/\tilde{\rho}}+\alpha_{k,jk}^{1/\tilde{\rho}}\right]^{(\tilde{\rho}-1)}\alpha_{j,jk}^{1/\tilde{\rho}}+\sum_{i=1}^{j-1}\left[\alpha_{i,ij}^{1/\tilde{\rho}}+\alpha_{j,ij}^{1/\tilde{\rho}}\right]^{(\tilde{\rho}-1)}\alpha_{j,ij}^{1/\tilde{\rho}}\right\}.\end{aligned}$$

For ease of notation, let

$$c_\alpha\equiv\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\left[(\alpha_{i,ij})^{1/\tilde{\rho}}+(\alpha_{j,ij})^{1/\tilde{\rho}}\right]^{\tilde{\rho}}$$

and

$$c_{\alpha,j}\equiv\sum_{k=j+1}^{n}\left[\alpha_{j,jk}^{1/\tilde{\rho}}+\alpha_{k,jk}^{1/\tilde{\rho}}\right]^{(\tilde{\rho}-1)}\alpha_{j,jk}^{1/\tilde{\rho}}+\sum_{i=1}^{j-1}\left[\alpha_{i,ij}^{1/\tilde{\rho}}+\alpha_{j,ij}^{1/\tilde{\rho}}\right]^{(\tilde{\rho}-1)}\alpha_{j,ij}^{1/\tilde{\rho}}$$

for every $j=1,2,\ldots,n$. Since

$$\begin{aligned}F(u,\ldots,u)=\Pr\left(\max_i Z_i\leq u\right)&=\exp\{-G(\exp\{-u\},\ldots,\exp\{-u\})\}\\&=\exp\{-c_\alpha\exp\{-u\}\},\end{aligned}$$

we have

$$\frac{\partial F(u,\ldots,u)}{\partial z_j} = c_{\alpha,j}\exp\{-c_\alpha \exp\{-u\}\}\exp\{-u\}.$$

Therefore, denoting the marginal density of Z_j by $h_j(z)$, we obtain

$$\begin{aligned}
\Pr\left(\arg\max_i Z_i = j \cap \max_i Z_i \geq z\right) &= \int_z^\infty \Pr(Z_i \leq u, \forall i \neq j | Z_j = u) h_j(u) du \\
&= c_{\alpha,j}\int_z^\infty \exp\{-c_\alpha \exp\{-u\}\}\exp\{-u\} du \\
&= c_{\alpha,j}\int_0^{\exp\{-z\}} \exp\{-c_\alpha v\} dv \\
&= \frac{c_{\alpha,j}}{c_\alpha}\left[1 - \exp\{-c_\alpha \exp\{-z\}\}\right]
\end{aligned}$$

and

$$\begin{aligned}
\Pr\left(\arg\max_i Z_i = j | \max_i Z_i \geq z\right) &= \frac{\Pr\left(\arg\max_i Z_i = j \cap \max_i Z_i \geq z\right)}{1 - F(z,\ldots,z)} \\
&= \frac{c_{\alpha,j}}{c_\alpha}.
\end{aligned}$$

Now, we verify that $\sum_{j=1}^n c_{\alpha,j} = c_\alpha$:

$$\begin{aligned}
\sum_{j=1}^n c_{\alpha,j} &= \sum_{j=1}^n \sum_{k=j+1}^n \left[\alpha_{j,jk}^{1/\tilde{\rho}} + \alpha_{k,jk}^{1/\tilde{\rho}}\right]^{(\tilde{\rho}-1)} \alpha_{j,jk}^{1/\tilde{\rho}} + \sum_{j=1}^n \sum_{i=1}^{j-1} \left[\alpha_{i,ij}^{1/\tilde{\rho}} + \alpha_{j,ij}^{1/\tilde{\rho}}\right]^{(\tilde{\rho}-1)} \alpha_{j,ij}^{1/\tilde{\rho}} \\
&= \sum_{i=1}^{n-1} \sum_{k=i+1}^n \left[\alpha_{i,ik}^{1/\tilde{\rho}} + \alpha_{k,ik}^{1/\tilde{\rho}}\right]^{(\tilde{\rho}-1)} \alpha_{i,ik}^{1/\tilde{\rho}} + \sum_{k=2}^n \sum_{i=1}^{k-1} \left[\alpha_{i,ik}^{1/\tilde{\rho}} + \alpha_{k,ik}^{1/\tilde{\rho}}\right]^{(\tilde{\rho}-1)} \alpha_{k,ik}^{1/\tilde{\rho}} \\
&= \sum_{i=1}^{n-1} \sum_{k=i+1}^n \left[\alpha_{i,ik}^{1/\tilde{\rho}} + \alpha_{k,ik}^{1/\tilde{\rho}}\right]^{(\tilde{\rho}-1)} \alpha_{i,ik}^{1/\tilde{\rho}} + \sum_{i=1}^{n-1} \sum_{k=i+1}^n \left[\alpha_{i,ik}^{1/\tilde{\rho}} + \alpha_{k,ik}^{1/\tilde{\rho}}\right]^{(\tilde{\rho}-1)} \alpha_{k,ik}^{1/\tilde{\rho}} \\
&= \sum_{i=1}^{n-1} \sum_{k=i+1}^n \left[\alpha_{i,ik}^{1/\tilde{\rho}} + \alpha_{k,ik}^{1/\tilde{\rho}}\right]^{(\tilde{\rho}-1)} \left(\alpha_{i,ik}^{1/\tilde{\rho}} + \alpha_{k,ik}^{1/\tilde{\rho}}\right) \\
&= \sum_{i=1}^{n-1} \sum_{k=i+1}^n \left[\alpha_{i,ik}^{1/\tilde{\rho}} + \alpha_{k,ik}^{1/\tilde{\rho}}\right]^{\tilde{\rho}} \\
&= c_\alpha.
\end{aligned}$$

Thus, we obtain

$$\sum_{j=1}^{n} \frac{c_{\alpha,j}}{c_\alpha} = 1 \text{ and } 0 \leq \frac{c_{\alpha,j}}{c_\alpha} \leq 1 \text{ for every } j = 1, 2, \ldots, n.$$

- Suppose that $(Z_1, \ldots, Z_n)$ follows the multivariate Fréchet distribution in Eaton and Kortum (2002):

$$F_\theta(z_1, \ldots, z_n; \varrho) = \exp\left[-H_\theta(z_1, \ldots, z_n; \varrho)\right] \text{ for } (z_1, \ldots, z_n) \in \mathbb{R}_+^n,$$

where $H_\theta(z_1, \ldots, z_n; \varrho) = \left[\sum_{i=1}^{n} \left(T_i z_i^{-\theta}\right)^{1/\varrho}\right]^\varrho$ for some $\theta > 1$ and $\varrho \in (0, 1]$. Simple algebra shows that

$$\begin{aligned}\frac{\partial F_\theta(z_1, \ldots, z_n; \varrho)}{\partial z_j} &= F_\theta(z_1, \ldots, z_n; \varrho) \cdot (-1) \frac{\partial H_\theta(z_1, \ldots, z_n; \varrho)}{\partial z_j} \\ &= F_\theta(z_1, \ldots, z_n; \varrho) \left[\sum_{i=1}^{n} (T_i z_i^{-\theta})^{1/\varrho}\right]^{(\varrho-1)} \theta T_j^{1/\varrho} z_j^{(-1-\theta/\varrho)}.\end{aligned}$$

It follows that for every $u > 0$,

$$\begin{aligned}&\frac{\partial F_\theta(u, \ldots, u; \varrho)}{\partial z_j} \\ &= F_\theta(u, \ldots, u; \varrho) \left[\sum_{i=1}^{n} (T_i u^{-\theta})^{1/\varrho}\right]^{(\varrho-1)} \theta T_j^{1/\varrho} u^{(-1-\theta/\varrho)} \\ &= \exp\left\{-\left[\sum_{i=1}^{n} T_i^{1/\varrho}\right]^\varrho u^{-\theta}\right\} \left[\sum_{i=1}^{n} T_i^{1/\varrho}\right]^{(\varrho-1)} u^{[-\theta(1-1/\varrho)]} \theta T_j^{1/\varrho} u^{[-(1+\theta/\varrho)]} \\ &= \theta T_j^{1/\varrho} \left[\sum_{i=1}^{n} T_i^{1/\varrho}\right]^{(\varrho-1)} \exp\left\{-\left(\sum_{i=1}^{n} T_i^{1/\varrho}\right)^\varrho u^{-\theta}\right\} u^{[-(\theta+1)]}.\end{aligned}$$

For ease of notation, let $A \equiv \left[\sum_{i=1}^{n} T_i^{1/\varrho}\right]^\varrho$. Consider the change of variables: $v = Au^{-\theta}$. We have

$$u^{[-(\theta+1)]} = A^{[-(1+1/\theta)]} v^{(1+1/\theta)} \text{ and } du = -A^{1/\theta} \theta^{-1} v^{[-(1+1/\theta)]} dv.$$

Therefore, we have

$$\begin{aligned}
&\Pr\left(\arg\max_i Z_i = j \cap \max_i Z_i \geq z\right) \\
&= \int_z^{\infty} \Pr(Z_i \leq u, \forall i \neq j | Z_j = u) h_j(u) du \\
&= \theta T_j^{1/\varrho} \left[\sum_{i=1}^{n} T_i^{1/\varrho}\right]^{(\varrho-1)} \int_z^{\infty} \exp\left\{-Au^{-\theta}\right\} u^{[-(\theta+1)]} du \\
&= \theta T_j^{1/\varrho} \left[\sum_{i=1}^{n} T_i^{1/\varrho}\right]^{(\varrho-1)} \int_0^{Az^{-\theta}} \exp\left\{-v\right\} A^{[-(1+1/\theta)]} v^{(1+1/\theta)} A^{1/\theta} \theta^{-1} v^{[-(1+1/\theta)]} dv \\
&= \theta T_j^{1/\varrho} \left[\sum_{i=1}^{n} T_i^{1/\varrho}\right]^{(\varrho-1)} A^{-1}\theta^{-1} \int_0^{Az^{-\theta}} \exp\left\{-v\right\} dv \\
&= \frac{T_j^{1/\varrho} \left[\sum_{i=1}^{n} T_i^{1/\varrho}\right]^{(\varrho-1)}}{\left[\sum_{i=1}^{n} T_i^{1/\varrho}\right]^{\varrho}} \left[1 - \exp\{-Az^{-\theta}\}\right] \\
&= \frac{T_j^{1/\varrho}}{\sum_{i=1}^{n} T_i^{1/\varrho}} \left[1 - \exp\left\{-\left[\sum_{i=1}^{n} T_i^{1/\varrho}\right]^{\varrho} z^{-\theta}\right\}\right].
\end{aligned}$$

Since $\Pr\left(\max_i Z_i \geq z\right) = 1 - \exp\left[-H_\theta(z, \ldots, z)\right] = 1 - \exp\left\{-\left(\sum_{i=1}^{n} T_i^{1/\varrho}\right)^{\varrho} z^{-\theta}\right\}$, we obtain

$$\Pr\left(\arg\max_i Z_i = j | \max_i Z_i \geq z\right) = \frac{T_j^{1/\varrho}}{\sum_{i=1}^{n} T_i^{1/\varrho}}.$$

5. Proof of Theorem 1

Consider an arbitrary n-box $B \equiv [\mathbf{a}, \mathbf{b}]$, where $\mathbf{a}, \mathbf{b} \in \overline{\mathbb{R}}^n$, with $\mathbf{a} \leq \mathbf{b}$.

Case 1 $b_i < 1$ for some i.

Then B lies entirely in the region where $G(\mathbf{z}) = 0$, so $V_G(B) = 0$.

Case 2 $\mathbf{a} \geq (1, \ldots, 1)$.

Then B lies in the region where $g(\mathbf{z}) = \frac{\partial^n G(\mathbf{z})}{\partial z_1 \partial z_2 \ldots \partial z_n} \geq 0$ almost everywhere so we can apply the fundamental theorem of calculus to show that

$$V_G(B) = \int_B g(\mathbf{z})\, dz_1\, dz_2 \cdots dz_n,$$

which is clearly non-negative.

Case 3 $\mathbf{b} \geq (1, \ldots, 1)$ and $a_i < 1$ for at least one i.

This implies that the set $\Omega \equiv \{k \in \mathbb{N}_n | a_k < 1\}$ is nonempty. To prove that the G-volume of B is non-negative, we first show that it can be expressed as the sum of the volume of sub-boxes that do not cross the 1-axis. That is, we show that

$$V_G(B) = V_G([\mathbf{a}^*, \mathbf{b}]) + \sum_{\xi \subseteq \Omega, \xi \neq \phi, \mathbb{N}_n} V_{G_\xi}([\mathbf{a}^*(\xi), \mathbf{b}(\xi)]). \tag{3}$$

where $\mathbf{a}^* \equiv \bar{\mathbf{a}}(\Omega)$. By case 2, we have that $V_G([\mathbf{a}^*, \mathbf{b}]) \geq 0$. In addition, $V_{G_\xi}([\mathbf{a}^*(\xi), \mathbf{b}(\xi)])$ is non-negative for all $\xi \subseteq \Omega, \xi \neq \phi, \mathbb{N}_n$ because for each $\xi \neq \mathbb{N}_n$, the density g_ξ of G_ξ on $(1, \infty]^{l(\xi)}$ is well defined and non-negative for all $\mathbf{x}(\xi)$ on $(1, \infty]^{l(\xi)}$, so

$$V_{G_\xi}([\mathbf{a}^*(\xi), \mathbf{b}(\xi)]) = \int_{[\mathbf{a}^*(\xi), \mathbf{b}(\xi)]} g_\xi(\mathbf{x}(\xi))\, d\mathbf{x}(\xi) \geq 0.$$

Thus it suffices to show that Equation (3) holds. The challenge here is that, by definition,

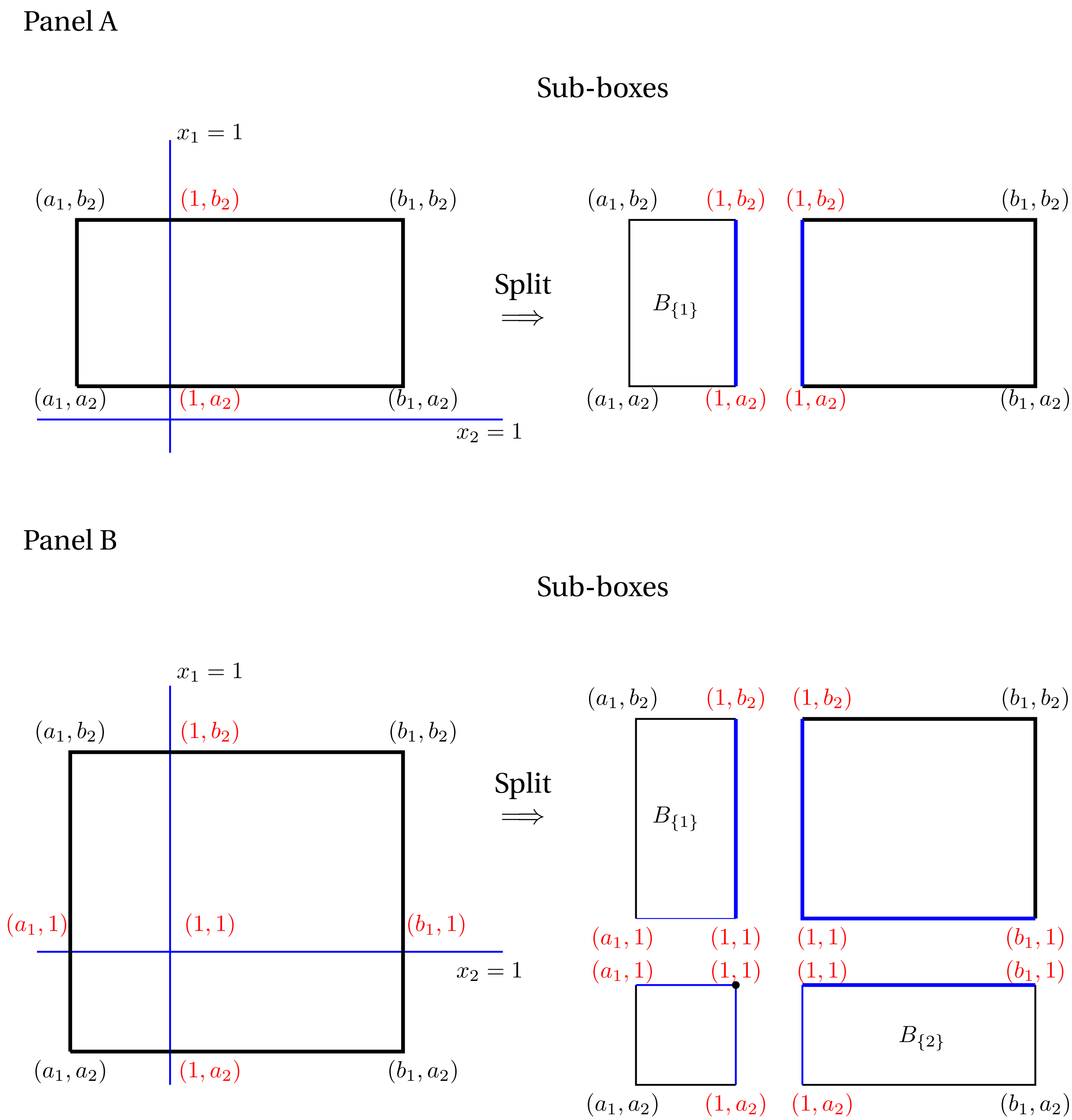
$$V_{G_\xi}\left([\mathbf{a}^*(\xi), \mathbf{b}(\xi)]\right) = \sum_{\mathbf{d} \in \Theta([\mathbf{a}^*(\xi), \mathbf{b}(\xi)])} (-1)^{m(A(\mathbf{d}; [\mathbf{a}^*(\xi), \mathbf{b}(\xi)]))} G_\xi(\mathbf{d}), \tag{4}$$

with $\boldsymbol{d} \in \overline{\mathbb{R}}^{l(\xi)}$ whereas $V_G(B)$ is defined as a sum of $G(\boldsymbol{c})$ across vertices $\boldsymbol{c} \in \overline{\mathbb{R}}^n$.

To proceed, let $\Gamma \equiv \{\mathbf{c} \in \overline{\mathbb{R}}^n | c_k \in \{a_k, a_k^*, b_k\}$ for all $k \in \mathbb{N}_n\}$ be the collection of all vertices after the decomposition of B into sub-boxes and let $\Psi \equiv \Gamma \backslash \Theta(B)$ be the set of "new" vertices generated by that decomposition. For any nonempty $\xi \subseteq \Omega$ that is not equal to $\mathbb{N}_n$, let $\Psi_\xi \subseteq \Psi$ be defined by $\Psi_\xi \equiv \{\mathbf{c} \in \Psi | c_k = a_k^*$ for all $k \in \xi\}$, let $B_\xi \equiv [\bar{\mathbf{a}}(\Omega \backslash \xi), \bar{\mathbf{b}}(\xi)]$ be the n-dimensional sub-box. Figure 2 shows two cases for the decomposition of a 2-box.

Equation (4) combined with the fact that $G_\xi(\mathbf{c}(\xi)) = G(\bar{\mathbf{c}}(\xi))$ and the fact that that $\boldsymbol{d} \in \Theta([\mathbf{a}^*(\xi), \mathbf{b}(\xi)])$ if and only if $\boldsymbol{d} = \boldsymbol{c}(\xi)$ with $\boldsymbol{c} \in \Theta(B_\xi) \cap \Psi_\xi$ implies that

$$V_{G_\xi}\left([\mathbf{a}^*(\xi), \mathbf{b}(\xi)]\right) = \sum_{\mathbf{d} = \mathbf{c}(\xi):\ \mathbf{c} \in \Theta(B_\xi) \cap \Psi_\xi} (-1)^{m(A(\mathbf{d}; [\mathbf{a}^*(\xi), \mathbf{b}(\xi)]))} G_\xi(\mathbf{d})$$



In Panel A, $\Omega = \{1\}$, $\mathbf{a}^* = (1, a_2)$, $\Gamma = \{(a_1, a_2), (1, a_2), (b_1, a_2), (a_1, b_2), (1, b_2), (b_1, a_2)\}$, $\Psi = \{(1, a_2), (1, b_2)\}$, *and* $B_{\{1\}} = [a_1, 1] \times [a_2, b_2]$;
in Panel B, $\Omega = \{1, 2\}$, $\mathbf{a}^* = (1, 1)$, $\Gamma = \{(a_1, a_2), (1, a_2), (b_1, a_2), (a_1, 1), (1, 1), (b_1, 1), (a_1, b_2), (1, b_2), (b_1, b_2)\}$, $\Psi = \{(a_1, 1), (1, a_2), (1, 1), (1, b_2), (b_1, 1)\}$, $B_{\{1\}} = [a_1, 1] \times [1, b_2]$, *and* $B_{\{2\}} = [1, b_1] \times [a_2, 1]$.

Figure 2: Sub-boxes

$$= \sum_{\mathbf{c}\in\Theta(B_\xi)\cap\Psi_\xi} (-1)^{m(A(\mathbf{c}(\xi);[\mathbf{a}^*(\xi),\mathbf{b}(\xi)]))} G(\bar{\mathbf{c}}(\xi)).$$

Since $\bar{\mathbf{c}}(\xi) = \mathbf{c}$ for all $\mathbf{c} \in \Psi_\xi$ and $A(\mathbf{c}(\xi); [\mathbf{a}^*(\xi), \mathbf{b}(\xi)]) = \{k \in \mathbb{N}_n \setminus \xi | c_k = a_k^*\}$, we can then write

$$V_{G_\xi}([\mathbf{a}^*(\xi), \mathbf{b}(\xi)]) = \sum_{\mathbf{c}\in\Theta(B_\xi)\cap\Psi_\xi} (-1)^{m\left(\{k\in\mathbb{N}_n\setminus\xi|c_k=a_k^*\}\right)} G(\mathbf{c}).$$

But note that, for any $\mathbf{c} \in \Psi_\xi$ we must have

$$\begin{aligned} m(\{k \in \mathbb{N}_n\backslash\xi|c_k = a_k^*\}) =& m(\{k \in \mathbb{N}_n|c_k = a_k^*\}) - m(\{k \in \xi|c_k = a_k^*\}) \\ =& m(\{k \in \Omega|c_k = a_k^*\}) + m(\{k \in \mathbb{N}_n\backslash\Omega|c_k = a_k^*\}) - m(\xi). \end{aligned}$$

For any $\mathbf{c} \in \Gamma$, let $\Omega(\mathbf{c}) \equiv \{k \in \Omega | c_k = 1\}$ and $\tilde{\Omega}(\mathbf{c}) \equiv \{k \in \mathbb{N}_n \setminus \Omega | c_k = a_k\}$. Combining these definitions with the previous equation we can then write

$$m(\{k \in \mathbb{N}_n\backslash\xi|c_k = a_k^*\}) = m(\Omega(\mathbf{c})) - m(\xi) + m(\tilde{\Omega}(\mathbf{c}))$$

and hence

$$V_{G_\xi}([\mathbf{a}^*(\xi), \mathbf{b}(\xi)] = \sum_{\mathbf{c}\in\Theta(B_\xi)\cap\Psi_\xi} (-1)^{[m(\Omega(\mathbf{c}))-m(\xi)+m(\tilde{\Omega}(\mathbf{c}))]} G(\mathbf{c}).$$

But if $\mathbf{c} \in \Theta(B_\xi)\backslash\Psi_\xi$ then $c_k = a_k < 1$ for some $k \in \xi$ and $G(\mathbf{c}) = 0$, so we can write

$$V_{G_\xi}([\mathbf{a}^*(\xi), \mathbf{b}(\xi)] = \sum_{\mathbf{c}\in\Theta(B_\xi)} (-1)^{[m(\Omega(\mathbf{c}))-m(\xi)+m(\tilde{\Omega}(\mathbf{c}))]} G(\mathbf{c}),$$

and noting that $B_\xi = [\bar{\mathbf{a}}(\Omega\backslash\xi), \bar{\mathbf{b}}(\xi)]$, we then have

$$V_{G_\xi}([\mathbf{a}^*(\xi), \mathbf{b}(\xi)] = \sum_{\mathbf{c}\in\Theta([\bar{\mathbf{a}}(\Omega\backslash\xi),\bar{\mathbf{b}}(\xi)])} (-1)^{[m(\Omega(\mathbf{c}))-m(\xi)+m(\tilde{\Omega}(\mathbf{c}))]} G(\mathbf{c}). \tag{5}$$

Note that

$$\begin{aligned} V_G([\mathbf{a}^*,\mathbf{b}]) &= \sum_{\mathbf{c}\in\Theta([\mathbf{a}^*,\mathbf{b}])} (-1)^{m(\{k|c_k=a_k^*\})}G(\mathbf{c}) \\ &= \sum_{\mathbf{c}\in\Theta([\bar{\mathbf{a}}(\Omega\backslash\phi),\bar{\mathbf{b}}(\phi)])} (-1)^{[m(\Omega(\mathbf{c}))-m(\phi)+m(\tilde{\Omega}(\mathbf{c}))]}G(\mathbf{c}) \end{aligned} \tag{6}$$

because $[\mathbf{a}^*,\mathbf{b}] = [\bar{\mathbf{a}}(\Omega\backslash\phi),\bar{\mathbf{b}}(\phi)]$ and $\{k|c_k = a_k^*\} = \Omega(\mathbf{c})\cup\tilde{\Omega}(\mathbf{c})$ by definition. Combining equalities (5) and (6), we obtain

$$\begin{aligned} &V_G([\mathbf{a}^*,\mathbf{b}]) + \sum_{\xi\subseteq\Omega,\xi\neq\phi,\mathbb{N}_n} V_{G_\xi}([\mathbf{a}^*(\xi),\mathbf{b}(\xi)]) \\ &= \sum_{\xi\subseteq\Omega,\xi\neq\mathbb{N}_n}\ \sum_{\mathbf{c}\in\Theta([\bar{\mathbf{a}}(\Omega\backslash\xi),\bar{\mathbf{b}}(\xi)])} (-1)^{[m(\Omega(\mathbf{c}))-m(\xi)+m(\tilde{\Omega}(\mathbf{c}))]}G(\mathbf{c}) \\ &= \sum_{\xi\subseteq\Omega}\ \sum_{\mathbf{c}\in\Theta([\bar{\mathbf{a}}(\Omega\backslash\xi),\bar{\mathbf{b}}(\xi)])} (-1)^{[m(\Omega(\mathbf{c}))-m(\xi)+m(\tilde{\Omega}(\mathbf{c}))]}G(\mathbf{c}) \end{aligned}$$

because if $\Omega = \mathbb{N}_n$ then $G(\mathbf{c}) = 0$ for all $\mathbf{c}\in\Theta([\bar{\mathbf{a}}(\Omega\backslash\mathbb{N}_n),\bar{\mathbf{b}}(\mathbb{N}_n)]) = \Theta([\mathbf{a},\mathbf{a}^*])$. Therefore we have

$$\begin{aligned} &V_G([\mathbf{a}^*,\mathbf{b}]) + \sum_{\xi\subseteq\Omega,\xi\neq\phi,\mathbb{N}_n} V_{G_\xi}([\mathbf{a}^*(\xi),\mathbf{b}(\xi)]) \\ &= \sum_{\xi\subseteq\Omega}\ \sum_{\mathbf{c}\in\Theta([\bar{\mathbf{a}}(\Omega\backslash\xi),\bar{\mathbf{b}}(\xi)])} (-1)^{[m(\mathit{\Omega}(\mathbf{c}))-m(\xi)+m(\tilde{\mathit{\Omega}}(\mathbf{c}))]}G(\mathbf{c}) \\ &= \sum_{\xi\subseteq\Omega}\ \sum_{\mathbf{c}\in\Theta([\bar{\mathbf{a}}(\Omega\backslash\xi),\bar{\mathbf{b}}(\xi)])\backslash\Psi} (-1)^{[m(\mathit{\Omega}(\mathbf{c}))-m(\xi)+m(\tilde{\mathit{\Omega}}(\mathbf{c}))]}G(\mathbf{c}) \\ &\quad + \sum_{\xi\subseteq\Omega}\ \sum_{\mathbf{c}\in\Theta([\bar{\mathbf{a}}(\Omega\backslash\xi),\bar{\mathbf{b}}(\xi)])\cap\Psi} (-1)^{[m(\mathit{\Omega}(\mathbf{c}))-m(\xi)+m(\tilde{\mathit{\Omega}}(\mathbf{c}))]}G(\mathbf{c}) \\ &\equiv \text{Term } 1 + \text{Term } 2. \end{aligned}$$

The rest of the proof consists in showing that (i) $\text{Term } 1 = V_G(B)$ and that (ii) $\text{Term } 2 = 0$. To show (i), note first that for all $\xi\subseteq\Omega$ and $\mathbf{c}\in\Theta([\bar{\mathbf{a}}(\Omega\setminus\xi),\bar{\mathbf{b}}(\xi)])\setminus\Psi$, we have $m(\mathit{\Omega}(\mathbf{c})) = 0$ by definition of Ψ, and $m(\{k\in\mathit{\Omega}|c_k = a_k\}) = m(\xi)$. So,

$$\begin{aligned} &m(\mathit{\Omega}(\mathbf{c})) - m(\xi) + m(\tilde{\mathit{\Omega}}(\mathbf{c})) \\ &= -2m(\xi) + m(\xi) + m(\{k\in\mathbb{N}_n\setminus\mathit{\Omega}|c_k = a_k^*\}) \end{aligned}$$

$$
\begin{aligned}
&= -2m(\xi) + m(\{k \in \mathit{\Omega} | c_k = a_k\}) + m(\{k \in \mathbb{N}_n \setminus \mathit{\Omega} | c_k = a_k\}) \\
&= -2m(\xi) + m(\{k \in \mathbb{N}_n | c_k = a_k\}) \\
&= -2m(\xi) + m(\mathrm{A}(\mathbf{c};[\mathbf{a},\mathbf{b}])),
\end{aligned}
$$

where the second equality holds because $a_k^* = a_k$ if $k \notin \Omega$. Given that $\Theta([\mathbf{a},\mathbf{b}]) = \{\mathbf{c} \in \Theta([\bar{\mathbf{a}}(\Omega \setminus \xi), \bar{\mathbf{b}}(\xi)]) \setminus \Psi : \xi \subseteq \Omega\}$ we then have

$$
\begin{aligned}
\text{Term 1} &= \sum_{\xi \subseteq \Omega} \sum_{\mathbf{c} \in \Theta([\bar{\mathbf{a}}(\Omega \setminus \xi), \bar{\mathbf{b}}(\xi)]) \setminus \Psi} (-1)^{[m(\mathit{\Omega}(\mathbf{c})) - m(\xi) + m(\tilde{\mathit{\Omega}}(\mathbf{c}))]} G(\mathbf{c}) \\
&= \sum_{\xi \subseteq \Omega} \sum_{\mathbf{c} \in \Theta([\bar{\mathbf{a}}(\Omega \setminus \xi), \bar{\mathbf{b}}(\xi)]) \setminus \Psi} (-1)^{[-2m(\xi)]} (-1)^{[m(\mathrm{A}(\mathbf{c};[\mathbf{a},\mathbf{b}]))]} G(\mathbf{c}) \\
&= \sum_{\mathbf{c} \in \Theta([\mathbf{a},\mathbf{b}])} (-1)^{m(\mathrm{A}(\mathbf{c};[\mathbf{a},\mathbf{b}]))} G\left(\mathbf{c}\right) \\
&= V_G\left(B\right).
\end{aligned}
$$

To show (ii), observe that $\Psi = \cup_{\xi \subseteq \Omega} [\Theta([\bar{\mathbf{a}}(\Omega \setminus \xi), \bar{\mathbf{b}}(\xi)]) \cap \Psi]$ and that in fact any vertex $\mathbf{c} \in \Psi$ will be repeated $\binom{m(\Omega(\mathbf{c}))}{v}$ times (although not necessarily with the same sign) in the set of vertices $\Theta([\bar{\mathbf{a}}(\Omega \setminus \xi), \bar{\mathbf{b}}(\xi)]) \cap \Psi$ across different $\xi \subseteq \Omega(\mathbf{c})$ with $m(\xi) = v$ and hence will repeated $\sum_{v=0}^{m(\Omega(\mathbf{c}))} \binom{m(\Omega(\mathbf{c}))}{v}$ across all possible $\xi \subseteq \Omega(\mathbf{c})$. Hence we have

$$
\begin{aligned}
\text{Term 2} &= \sum_{\xi \subseteq \Omega} \sum_{\mathbf{c} \in \Theta([\bar{\mathbf{a}}(\Omega \setminus \xi), \bar{\mathbf{b}}(\xi)]) \cap \Psi} (-1)^{[m(\mathit{\Omega}(\mathbf{c})) - m(\xi) + m(\tilde{\mathit{\Omega}}(\mathbf{c}))]} G(\mathbf{c}) \\
&= \sum_{\mathbf{c} \in \Psi} \left[\sum_{v=0}^{m(\Omega(\mathbf{c}))} \binom{m\left(\Omega(\mathbf{c})\right)}{v} (-1)^{m(\Omega(\mathbf{c})) - v + m\left(\tilde{\Omega}(\mathbf{c})\right)} \right] G(\mathbf{c}) \\
&= \sum_{\mathbf{c} \in \Psi} \left[(-1)^{m(\Omega(\mathbf{c})) + m\left(\tilde{\Omega}(\mathbf{c})\right)} \sum_{v=0}^{m(\Omega(\mathbf{c}))} \binom{m\left(\Omega(\mathbf{c})\right)}{v} (-1)^v \right] G(\mathbf{c}) \\
&= \sum_{\mathbf{c} \in \Psi} \left[(-1)^{m(\Omega(\mathbf{c})) + m\left(\tilde{\Omega}(\mathbf{c})\right)} (1-1)^{m(\Omega(\mathbf{c}))} \right] G(\mathbf{c}) \\
&= 0,
\end{aligned}
$$

where the second to last equality holds because we apply the binomial theorem, which states that $\sum_{i=0}^{p} \binom{n}{i} x^i = (1+x)^p$ for any positive integer p and real number

x. This completes the proof.

6. Derivation of Results for the Roy Model in Section 5.

For comparison, let $\varrho \equiv 1-\rho$. Let c be a nonnegative constant and $(w_1,\ldots,w_K)\in\mathbb{R}_+^K$ be a K-dimensional vector of positive real numbers. For each $i=1,\ldots,K$, let

$$\pi_i \equiv \frac{\left(T_i w_i^{\theta}\right)^{(1/\varrho)}}{\sum_{j=1}^{K}\left[T_j w_j^{\theta}\right]^{(1/\varrho)}}$$

and

$$\begin{aligned}\Omega_i(c) &\equiv \left\{(z_1,\ldots,z_K)\in\mathbb{R}^K : w_i z_i \geq \max w_j z_j \cap \max w_j z_j \geq c\right\}\\ &= \left\{(z_1,\ldots,z_K)\in\mathbb{R}^K : w_i z_i \geq \max w_j z_j \cap z_i \geq c/w_i\right\}.\end{aligned}$$

We are interested in $\frac{\mathcal{E}_i}{L} \equiv \mathbb{E}\left[Z_i \mathbb{1}_{[\boldsymbol{Z}\in\Omega_i(c)]}\right]$, where

$$\mathbb{1}_{[\boldsymbol{Z}\in\Omega_i(c)]} = \begin{cases} 1, & \text{if } \boldsymbol{Z}\equiv(Z_1,\ldots,Z_K)\in\Omega_i(c),\\ 0, & \text{otherwise}.\end{cases}$$

Note that

$$\begin{aligned}\frac{\mathcal{E}_i}{L} &= \mathbb{E}\left[\mathbb{E}[Z_i \mathbb{1}_{[\boldsymbol{Z}\in\Omega_i(c)]}|Z_i]\right]\\ &= \mathbb{E}\left[Z_i \mathbb{1}_{[Z_i\geq c/w_i]}\,\mathbb{E}[\mathbb{1}_{[w_i z_i\geq \max w_j z_j]}|Z_i]\right]\\ &= \mathbb{E}\left[Z_i \mathbb{1}_{[Z_i\geq c/w_i]}\,\Pr(w_i Z_i \geq \max w_j Z_j|Z_i)\right]\\ &= \int_{c/w_i}^{\infty} u\cdot \Pr(w_i Z_i \geq \max w_j Z_j|Z_i=u)h_i(u)du.\end{aligned}$$

Roy Model with the multivariate Fréchet distribution

Consider the multivariate Fréchet distribution

$$F_\theta(z_1,\ldots,z_K;\varrho) = \exp\left\{-\left[\sum_{i=1}^{K}\left(T_i z_i^{-\theta}\right)^{1/\varrho}\right]^{\varrho}\right\}.$$

Suppose that $\boldsymbol{Z} = (Z_1, \ldots, Z_K)$ follows the multivariate Fréchet distribution above. Let h_i denote the marginal density of Z_i. Note that for $u > 0$,

$$\begin{aligned}
& \Pr(Z_j \leq uw_i/w_j, \forall j \neq i | Z_i = u) h_i(u) \\
= & \left. \frac{\partial F_\theta(z_1, \ldots, z_n; \varrho)}{\partial z_i} \right|_{z_j = uw_i/w_j \text{ for each } j} \\
= & \exp\left\{ - \left[\sum_{j=1}^{K} \left(T_j (uw_i/w_j)^{-\theta} \right)^{1/\varrho} \right]^{\varrho} \right\} \left\{ \sum_{j=1}^{K} \left[T_j \left(uw_i/w_j \right)^{-\theta} \right]^{1/\varrho} \right\}^{(\varrho-1)} \theta T_i^{1/\varrho} u^{(-1-\theta/\varrho)} \\
= & \exp\left\{ - \left[\sum_{j=1}^{K} \left(T_j w_j^\theta \right)^{1/\varrho} \right]^{\varrho} (uw_i)^{-\theta} \right\} \left\{ \sum_{j=1}^{K} \left[T_j w_j^\theta \right]^{1/\varrho} \right\}^{(\varrho-1)} \theta T_i^{1/\varrho} w_i^{[-\theta(1-1/\varrho)]} u^{-(\theta+1)}.
\end{aligned}$$

It follows that

$$\begin{aligned}
\frac{\mathcal{E}_i}{L} &= \int_{c/w_i}^{\infty} u \cdot \Pr(w_i Z_i \geq \max w_j Z_j | Z_i = u) h_i(u) du \\
&= \int_{c/w_i}^{\infty} u \cdot \exp\left\{ - \left[\sum_{j=1}^{K} \left(T_j w_j^\theta \right)^{1/\varrho} \right]^{\varrho} (uw_i)^{-\theta} \right\} \\
& \qquad \cdot \left\{ \sum_{j=1}^{K} \left[T_j w_j^\theta \right]^{1/\varrho} \right\}^{(\varrho-1)} \theta T_i^{1/\varrho} w_i^{[-\theta(1-1/\varrho)]} u^{-(\theta+1)} du \\
&= \left\{ \sum_{j=1}^{K} \left[T_j w_j^\theta \right]^{1/\varrho} \right\}^{(\varrho-1)} \theta T_i^{1/\varrho} w_i^{[-\theta(1-1/\varrho)]} \\
& \qquad \cdot \int_{c/w_i}^{\infty} \exp\left\{ - \left[\sum_{j=1}^{K} \left(T_j w_j^\theta \right)^{1/\varrho} \right]^{\varrho} (uw_i)^{-\theta} \right\} u^{-\theta} du.
\end{aligned}$$

For ease of notation, let $A_{\boldsymbol{w}} \equiv \left[\sum_{j=1}^{K} \left(T_j w_j^\theta \right)^{1/\varrho} \right]^{\varrho}$, $B_{\boldsymbol{w}} \equiv A_{\boldsymbol{w}} w_i^{-\theta}$, and

$$\Gamma\left(1 - \frac{1}{\theta}; 0 \to B_{\boldsymbol{w}} (c/w_i)^{-\theta} \right) \equiv \int_0^{B_{\boldsymbol{w}} (c/w_i)^{-\theta}} \exp\{-s\} s^{[(1-1/\theta)-1]} ds.$$

Consider the change of variables: $v = B_{\boldsymbol{w}} u^{-\theta}$. We have

$$u^{-\theta} = B_{\boldsymbol{w}}^{-1} v \text{ and } du = -B_{\boldsymbol{w}}^{1/\theta} \theta^{-1} v^{[-(1+1/\theta)]} dv.$$

It follows that

$$
\begin{aligned}
\frac{\mathcal{E}_i}{L} &= A_{\boldsymbol{w}}^{(1-1/\varrho)}\theta T_i^{1/\varrho} w_i^{[-\theta(1-1/\varrho)]} \int_{c/w_i}^{\infty} \exp\left\{-B_{\boldsymbol{w}} u^{-\theta}\right\} u^{-\theta} du \\
&= A_{\boldsymbol{w}}^{(1-1/\varrho)}\theta T_i^{1/\varrho} w_i^{[-\theta(1-1/\varrho)]} \int_{0}^{B_{\boldsymbol{w}}(c/w_i)^{-\theta}} \exp\left\{-v\right\} B_{\boldsymbol{w}}^{-1} v B_{\boldsymbol{w}}^{1/\theta}\theta^{-1} v^{[-(1+1/\theta)]} dv \\
&= A_{\boldsymbol{w}}^{(1-1/\varrho)} B_{\boldsymbol{w}}^{[-(1-1/\theta)]} T_i^{1/\varrho} w_i^{[-\theta(1-1/\varrho)]} \int_{0}^{B_{\boldsymbol{w}}(c/w_i)^{-\theta}} \exp\left\{-v\right\} v^{[(1-1/\theta)-1]} dv \\
&= A_{\boldsymbol{w}}^{(1-1/\varrho)} B_{\boldsymbol{w}}^{[-(1-1/\theta)]} T_i^{1/\varrho} w_i^{[-\theta(1-1/\varrho)]} \cdot \Gamma\left(1-\frac{1}{\theta}; 0 \to B_{\boldsymbol{w}}(c/w_i)^{-\theta}\right) \\
&= A_{\boldsymbol{w}}^{(1-1/\varrho)} \left[A_{\boldsymbol{w}} w_i^{-\theta}\right]^{[-(1-1/\theta)]} T_i^{1/\varrho} w_i^{[-\theta(1-1/\varrho)]} \cdot \Gamma\left(1-\frac{1}{\theta}; 0 \to B_{\boldsymbol{w}}(c/w_i)^{-\theta}\right) \\
&= A_{\boldsymbol{w}}^{(1/\theta-1/\varrho)} w_i^{(\theta-1)} T_i^{1/\varrho} w_i^{[-\theta(1-1/\varrho)]} \cdot \Gamma\left(1-\frac{1}{\theta}; 0 \to B_{\boldsymbol{w}}(c/w_i)^{-\theta}\right) \\
&= A_{\boldsymbol{w}}^{(1/\theta-1/\varrho)} w_i^{-1} \left[T_i w_i^{\theta}\right]^{1/\varrho} \cdot \Gamma\left(1-\frac{1}{\theta}; 0 \to B_{\boldsymbol{w}}(c/w_i)^{-\theta}\right) \\
&= \left\{\sum_{j=1}^{K} \left[T_j w_j^{\theta}\right]^{1/\varrho}\right\}^{\varrho/\theta} \frac{\left[T_i w_i^{\theta}\right]^{1/\varrho}}{\sum_{j=1}^{K}\left[T_j w_j^{\theta}\right]^{1/\varrho}} \cdot \frac{1}{w_i} \cdot \Gamma\left(1-\frac{1}{\theta}; 0 \to B_{\boldsymbol{w}}(c/w_i)^{-\theta}\right) \\
&= \frac{\pi_i}{w_i} \cdot \Gamma\left(1-\frac{1}{\theta}; 0 \to B_{\boldsymbol{w}}(c/w_i)^{-\theta}\right) \cdot \left\{\sum_{j=1}^{K} \left[T_j w_j^{\theta}\right]^{1/\varrho}\right\}^{\varrho/\theta}.
\end{aligned}
$$

Taking $c = 0$ yields

$$
\frac{\mathcal{E}_i}{L} = \frac{\pi_i}{w_i}\Gamma\left(1-\frac{1}{\theta}\right)\left\{\sum_{j=1}^{K}\left[T_j w_j^{\theta}\right]^{1/\varrho}\right\}^{\varrho/\theta}.
$$

Roy Model with the MVP

Let c be a constant with $c > \max\{w_1, \ldots, w_K\}\tilde{T}^{1/\theta}$. Let $\tilde{T} \equiv \left(\sum_{j=1}^{K} T_j^{1/\varrho}\right)^{\varrho}$. Consider the MVP

$$
F_\theta(z_1, \ldots, z_K; \varrho) = 1 - \left[\sum_{j=1}^{K}\left(T_j z_j^{-\theta}\right)^{1/\varrho}\right]^{\varrho}
$$

with the support $z_j \geq \tilde{T}^{1/\theta}$ for all $j = 1, \ldots, K$ and $\theta > 1$. Suppose that $\boldsymbol{Z} = (Z_1, \ldots, Z_K)$ follows the MVP above. Let h_i denote the marginal density of Z_i on $(\tilde{T}^{1/\theta}, \infty)$. Note that for $u > \tilde{T}^{1/\theta}$,

$$
\begin{aligned}
&\Pr(Z_j \leq uw_i/w_j, \forall j \neq i | Z_i = u) h_i(u) \\
= &\left.\frac{\partial F_\theta(z_1, \ldots, z_n; \varrho)}{\partial z_i}\right|_{z_j = uw_i/w_j \text{ for each } j} \\
= &\,\theta \left\{ \sum_{j=1}^{K} \left[T_j \left(uw_i/w_j \right)^{(-\theta)} \right]^{(1/\varrho)} \right\}^{-(1-\varrho)} T_i^{(1/\varrho)} u^{(-1-\theta/\varrho)} \\
= &\,\theta \left\{ \sum_{j=1}^{K} \left[T_j \left(w_i/w_j \right)^{(-\theta)} \right]^{(1/\varrho)} \right\}^{-(1-\varrho)} T_i^{(1/\varrho)} u^{(\theta(1-\varrho)/\varrho)} u^{(-1-\theta/\varrho)} \\
= &\,\theta \left\{ \sum_{j=1}^{K} \left[T_j \left(w_i/w_j \right)^{(-\theta)} \right]^{(1/\varrho)} \right\}^{-(1-\varrho)} T_i^{(1/\varrho)} u^{-(\theta+1)}.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
\frac{\mathcal{E}_i}{L} &= \int_{c/w_i}^{\infty} u \cdot \Pr(w_i Z_i \geq \max w_j Z_j | Z_i = u) h_i(u) du \\
&= \int_{c/w_i}^{\infty} u \cdot \theta \left\{ \sum_{j=1}^{K} \left[T_j \left(w_i/w_j \right)^{(-\theta)} \right]^{(1/\varrho)} \right\}^{-(1-\varrho)} T_i^{(1/\varrho)} u^{-(\theta+1)} du \\
&= \theta \left\{ \sum_{j=1}^{K} \left[T_j \left(w_i/w_j \right)^{(-\theta)} \right]^{(1/\varrho)} \right\}^{-(1-\varrho)} T_i^{(1/\varrho)} \int_{c/w_i}^{\infty} u^{-\theta} du \\
&= \theta \left\{ \sum_{j=1}^{K} \left[T_j w_j^{\theta} \right]^{(1/\varrho)} \right\}^{-(1-\varrho)} w_i^{\left(\frac{\theta(1-\varrho)}{\varrho}\right)} T_i^{(1/\varrho)} \frac{1}{\theta - 1} (w_i/c)^{(\theta-1)} \\
&= \frac{\theta}{\theta - 1} \left\{ \sum_{j=1}^{K} \left[T_j w_j^{\theta} \right]^{(1/\varrho)} \right\}^{-(1-\varrho)} w_i^{-1} \left(T_i w_i^{\theta} \right)^{(1/\varrho)} c^{(1-\theta)} \\
&= \frac{\theta c^{(1-\theta)}}{\theta - 1} \cdot \left\{ \sum_{j=1}^{K} \left[T_j w_j^{\theta} \right]^{(1/\varrho)} \right\}^{\varrho} \cdot \frac{\left(T_i w_i^{\theta} \right)^{(1/\varrho)}}{\sum_{j=1}^{K} \left[T_j w_j^{\theta} \right]^{(1/\varrho)}} \cdot \frac{1}{w_i} \\
&= \frac{\pi_i}{w_i} \cdot \frac{\theta c^{(1-\theta)}}{\theta - 1} \cdot \left\{ \sum_{j=1}^{K} \left[T_j w_j^{\theta} \right]^{(1/\varrho)} \right\}^{\varrho}.
\end{aligned}
$$

References

AGUIAR, M., BILS, M., CHARLES, K. K. and HURST, E. (2018). Leisure luxuries and the labor supply of young men. *Journal of Political Economy*, **129** (2), 337–382. https://doi.org/10.1086/711916

ARKOLAKIS, C., RAMONDO, N., RODRÍGUEZ-CLARE, A. and YEAPLE, S. (2018). Innovation and production in the global economy. *American Economic Review*, **108** (8), 2128–2173. https://doi.org/10.1257/aer.20141743

ARNOLD, B. C. (1983). *Pareto distributions*. International Cooperative Publishing House, Fairland, MD.

— (1990). A flexible family of multivariate pareto distributions. *Journal of Statistical Planning and Inference*, **24** (2), 249–258. https://doi.org/10.1016/0378-3758(90)90045-V

— (2015). *Pareto distributions*. Chapman and Hall/CRC, 2nd edn. https://doi.org/10.1201/b18141

—, CASTILLO, E. and SARABIA, J.-M. (1992). *Conditionally specified distributions*. Springer. https://doi.org/10.1007/978-1-4612-2912-4

ASIMIT, A. V., FURMAN, E. and VERNIC, R. (2010). On a multivariate pareto distribution. *Insurance: Mathematics and Economics*, **46** (2), 308–316. https://doi.org/10.1016/j.insmatheco.2009.11.004

BHAT, C. R. and GUO, J. (2004). A mixed spatially correlated logit model: Formulation and application to residential choice modeling. *Transportation Research Part B: Methodological*, **38** (2), 147–168. https://doi.org/10.1016/S0191-2615(03)00005-5

BINDER, A. J. and BOUND, J. (2019). The declining labor market prospects of less-educated men. *Journal of Economic Perspectives*, **33** (2), 163–190. https://doi.org/10.1257/jep.33.2.163

BIRKNER, M., SCHEUER, N. and WÄLDE, K. (2023). The dynamics of Pareto distributed wealth in a small open economy. *Economic Theory*, **76**, 607–644. https://doi.org/10.1007/s00199-022-01471-z

CHIRAGIEV, A. and LANDSMAN, Z. (2007). Multivariate pareto portfolios: TCE-based capital allocation and divided differences. *Scandinavian Actuarial Journal*, **2007** (4), 261–280. https://doi.org/10.1080/03461230701554007

DHAENE, J., DENUIT, M., GOOVAERTS, M., KAAS, R. and VYNCKE, D. (2002). The concept of comonotonicity in actuarial science and finance: Theory. *Insurance: Mathematics and Economics*, **31** (1), 3–33. https://doi.org/10.1016/S0167-6687(02)00134-8

EATON, J. and KORTUM, S. (2002). Technology, geography, and trade. *Econometrica*, **70** (5), 1741–1779. https://doi.org/10.1111/1468-0262.00352

— and — (2018). Trade in goods and trade in services. In *World Trade Evolution*, Routledge, pp. 82–125. https://doi.org/10.4324/9781351061544-4

FAN, J. and LUO, W. (2026). Financing multinationals. *Review of Economics and Statistics*, Accepted. https://doi.org/10.1162/REST.a.1701

GABAIX, X. (2009). Power laws in economics and finance. *Annual Review of Economics*, **1**, 255–294. https://doi.org/10.1146/annurev.economics.050708.142940

GABAIX, X. (2016). Power laws in economics: An introduction. *Journal of Economic Perspectives*, **90** (1), 185–206. https://doi.org/10.1257/jep.30.1.185

GALLE, S., RODRÍGUEZ-CLARE, A. and YI, M. (2022). Slicing the pie: Quantifying the aggregate and distributional effects of trade. *Review of Economic Studies*, **90** (1), 331–375. https://doi.org/10.1093/restud/rdac020

HANAGAL, D. D. (1996). A multivariate pareto distribution. *Communications in Statistics - Theory and Methods*, **25** (7), 1471–1488. https://doi.org/10.1080/03610929608831779

HSIEH, C.-T., HURST, E., JONES, C. I. and KLENOW, P. J. (2019). The allocation of talent and U.S. economic growth. *Econometrica*, **87** (5), 1439–1474. https://doi.org/10.3982/ECTA11427

HSU, W.-T., LU, L. and PICARD, P. M. (2023). Income inequality, productivity, and international trade. *Economic Theory*, **76**, 203–249. https://doi.org/10.1007/s00199-022-01456-y

JUPP, P. E. and MARDIA, K. V. (1982). A characterization of the multivariate pareto distribution. *Annals of Statistics*, **10** (3), 1021–1024. https://doi.org/10.1214/aos/1176345894

KOTZ, S., BALAKRISHNAN, N. and JOHNSON, N. L. (2000). *Continuous multivariate distributions: Models and applications*. Wiley-Interscience, 2nd edn. https://doi.org/10.1002/0471722065

LAGAKOS, D. and WAUGH, M. E. (2013). Selection, agriculture, and cross-country productivity differences. *American Economic Review*, **103** (2), 948–980. https://doi.org/10.1257/aer.103.2.948

LAN, T. (2019). *The coagglomeration of innovation and production*. Tech. rep.

LI, E. Z. (2025). *E-commerce and regional inequality: A trade framework and evidence from Amazon's expansion*. Tech. rep.

LI, H. (2006). *Tail dependence of multivariate Pareto distributions*. Tech. rep.

LIND, N. and RAMONDO, N. (2023). Trade with correlation. *American Economic Review*, **113** (2), 317–353. https://doi.org/10.1257/aer.20190781

LIU, C. and MA, X. (2023). Migration, tariffs, and china's export surge. *Journal of International Economics*, **140**, 103696. https://doi.org/10.1016/j.jinteco.2022.103696

MARDIA, K. V. (1962). Multivariate pareto distributions. *Annals of Mathematical Statistics*, **33** (3), 1008–1015. https://doi.org/10.1214/aoms/1177704468

MCFADDEN, D. (1978). Modelling the choice of residential location. In A. Karlqvist, L. Lundqvist, F. Snickars and J. W. Weibull (eds.), *Spatial Interaction Theory and Planning Models*, *3*, Elsevier, pp. 75–96.

MELITZ, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, **71** (6), 1695–1725. https://doi.org/10.1111/1468-0262.00467

NELSEN, R. B. (2006). *An introduction to copulas*. Springer New York, 2nd edn. https://doi.org/10.1007/0-387-28678-0

RAMONDO, N. and RODRÍGUEZ-CLARE, A. (2013). Trade, multinational production, and the gains from openness. *Journal of Political Economy*, **121** (2), 273–322. https://doi.org/10.1086/670136

SHEN, Y. (2024). Corporate income taxation and multinational production. *Review of International Economics*, **32** (3), 1039–1070. https://doi.org/10.1111/roie.12718

SU, J. and FURMAN, E. (2017). A form of multivariate pareto distribution with applications to financial risk measurement. *ASTIN Bulletin*, **47** (1), 331–357. https://doi.org/10.1017/asb.2016.22

WANG, Z. (2017). *Multinational production and external economies of scale*. Tech. rep.

XIANG, W. (2023). *Clean growth and environmental policies in the global economy*. Tech. rep.