

# A Multivariate Distribution with Pareto Tails and Pareto Maxima

Costas Arkolakis  
Yale University

Andrés Rodríguez-Clare  
UC Berkeley and NBER

Jiun-Hua Su  
UC Berkeley \*

This Version: May, 2017

## Abstract

We present a new multivariate distribution with Pareto distributed tails and maxima. The distribution has a number of properties that make it useful for applied work. Compared to uncorrelated univariate Pareto distributions, the distribution features one additional parameter that governs the covariance of its realizations. We show that this distribution is indeed valid by proving a general result about  $n$ -increasing functions.

**Keywords:** Multivariate Pareto, Pareto tails, Pareto maxima.

---

\*We thank Barry Arnold, Xiaohong Chen, Roger Nelsen, Tianhao Wu for useful suggestions. All errors are, of course, our own.

## 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, firm sizes, stock returns, and city populations. Because of its empirical prevalence, but also its mathematical simplicity, the Pareto distribution has become an extremely important statistical tool for scientists across disciplines. 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.<sup>1</sup>

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 form. In particular, we show that the function

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

with support

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

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

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

Various multivariate Pareto distributions have been developed since the introduc-

---

<sup>1</sup>See Jupp and Mardia (1982) on incomes across different years, Chiragiev and Landsman (2007) on risk capital allocations, and Arkolakis *et al.* (2013) on multinational firms' sales to one destination from many possible production locations.

<sup>2</sup>The distribution (1) has finite mean if  $\theta > 1$  and finite variance if  $\theta > 2$ .

tion of multivariate Pareto distributions of the first and the second kind by Mardia (1962). Many of these examples and generalizations can be found in Arnold (1990), Arnold *et al.* (1992), and Kotz *et al.* (2002). Our distribution generalizes the bivariate distribution presented in Nelsen (2007) by combining an Archimedean copula that is not strict with the Pareto distribution of the first type.

To understand the properties of this distribution and the difference with past known examples we first study a known bivariate example. Consider the joint cumulative distribution

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

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  $F_i(z_i) = \mathbb{P}(Z_i \leq z_i) = 1 - T_i z_i^{-\theta}$  for  $i = 1, 2$ .<sup>3</sup>

In order to better understand the role of  $\rho$  in this bivariate example, we can study the upper tail dependence ( $\lambda_U$ ) of the copula. Roughly speaking,  $\lambda_U$  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  $\mathbb{P}(F_1(Z_1) > u \mid F_2(Z_2) > u)$ ,

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

As  $\rho \rightarrow 0$ ,  $\lambda_U \rightarrow 0$ , so that if  $Z_2$  is large then  $Z_1$  is large with probability 0, while as  $\rho \rightarrow 1$ ,  $\lambda_U \rightarrow 1$ , 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  $F(z) = 1 - \tilde{T} z^{-\theta}$  with support  $z \geq \tilde{T}^{1/\theta}$ . Moreover, the joint probability that  $j = \arg \max_i Z_i$  and that  $Z_j \geq z$  is simply

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

---

<sup>3</sup>This is copula 4.2.2 in Nelsen (2007).

This is a convenient property for many economic applications. For example, as explored in Arkolakis *et al.* (2013), 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  $G(z_1, z_2, \dots, z_n)$  is the distribution of productivity across locations, then we may need  $Z \equiv \max(Z_1, \dots, Z_n)$  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, 2007). Thus, the function

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

with domain defined implicitly by  $\sum_{i=1}^n (T_i z_i^{-\theta})^{1/(1-\rho)} \leq 1$  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.<sup>4</sup> In this paper we show that we can modify the support of the distribution to make it an  $N$ -box defined by  $z_i \geq \tilde{T}^{1/\theta}$  for all  $i$ , where  $\tilde{T} \equiv \left( \sum_{i=1}^n T_i^{1/(1-\rho)} \right)^{1-\rho}$ , as in the distribution introduced in (1). In the next section we discuss the properties of this distribution, while in Section 3 we prove that  $H(\mathbf{z})$  is a proper distribution.

## 2. Properties of the General 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.**  $Z \equiv \max(Z_1, \dots, Z_n)$  is distributed Pareto of type I (Arnold, 2015) with shape parameter  $\theta$  and scale parameter  $\tilde{T}^{1/\theta}$  – that is,  $\mathbb{P}(Z \leq z) = 1 - \tilde{T} z^{-\theta}$ .

---

<sup>4</sup>We have evaluated the multivariate Pareto distributions in Mardia (1962), Arnold, Hanagal (1996), Li (2006), Asimit *et al.* (2010), Su and Furman (2016), as well as the copulas in Table 4.1 on page 116 in Nelsen (2007). In none of these cases the maximum is distributed Pareto.

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

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

Combined with property (i), this implies that

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

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

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

with support  $z_i \geq \tilde{T}^{1/\theta}$  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  $z_i = \tilde{T}^{1/\theta}$  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  $z_i = \tilde{T}^{1/\theta}$  can be seen by noting that  $F_i(\tilde{T}^{1/\theta}) = 1 - T_i/\tilde{T} > 0$  while  $F_i(\tilde{T}^{1/\theta} - \varepsilon) = 0$  for any positive  $\varepsilon$ . The distribution  $H(\cdot)$  is also discontinuous at any point  $\mathbf{z}$  in which  $z_i = \tilde{T}^{1/\theta}$  for some  $i$  except if  $z_i = \tilde{T}^{1/\theta}$  for all  $i$  – that is,  $H(\cdot)$  is continuous in  $(\tilde{T}^{1/\theta}, \infty]^n \cup z^*$  where  $z^* \equiv (\tilde{T}^{1/\theta}, \dots, \tilde{T}^{1/\theta})$ .

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

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

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  $\rho$ .** The upper tail dependence parameter between  $Z_i$  and  $Z_j$  is

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

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  $\rho \rightarrow 1$ , we have  $H(\mathbf{z}) \rightarrow \min\{F_1(z_1), \dots, F_n(z_n)\}$ , where  $F_i(\cdot)$  is the one dimensional marginal for each  $i$ . This implies that as  $\rho \rightarrow 1$  our distribution  $H(\mathbf{z})$  converges to the Frechet-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  $\bar{\mathbb{R}}^n$  and the distribution is concentrated on the set  $\bigcup_{i=1}^n \{(z_1, \dots, z_n) : z_i \geq z_j = \tilde{T}^{1/\theta} \text{ for all } j \neq i\}$ , 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  $\mathbb{P}(Z_i \leq z_i) = 1 - \tilde{T}z_i^{-\theta}$ .

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  $\theta$ .** Let  $\theta' \geq \theta$  and  $\tilde{T}^{1/\theta'} \geq 1$ ; then  $H_{\theta'}(\mathbf{z})$  stochastically dominates  $H_{\theta}(\mathbf{z})$ :  $H_{\theta'}(\mathbf{z}) \geq H_{\theta}(\mathbf{z})$ .

### 3. A Multivariate Pareto Distribution

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

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

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 (2007) page 46).

**Condition 1** (Lower and upper limits)

$\lim_{\mathbf{z} \rightarrow +\infty} H(\mathbf{z}) = 1$  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  $B \equiv [\mathbf{a}, \mathbf{b}] = [a_1, b_1] \times \dots \times [a_n, b_n]$  the  $n$ -box formed by the vectors  $\mathbf{a}$  and  $\mathbf{b}$ ,  $H$  is  $n$ -increasing if and only if

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

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

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

Condition 1 holds because  $\lim_{\mathbf{z} \rightarrow \infty} H(\mathbf{z}) = 1$  follows directly from  $\lim_{z_i \rightarrow \infty} z_i^{-\theta/(1-\rho)} = 0$  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  $[\sum_{i=1}^n (T_i z_i^{-\theta})^{1/(1-\rho)}]^{1-\rho} \geq 0$  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  $H(1, 1, \dots, 1) = 0$ .

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  $\frac{\partial^n H(\mathbf{z})}{\partial z_1 \partial z_2 \dots \partial z_n}$  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 hence 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; instead, it does not depend on any specific Paretian structure of  $H(\cdot)$ .

With this observation, 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 Theorem 1 to show  $H(\cdot)$  is

$n$ -increasing.

### 3.1. Main Result

We introduce some additional definitions. Let  $\mathbb{N}_n \equiv \{1, 2, \dots, n\}$  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

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

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  $\mathbf{x}$  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  $G_\xi$  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  $\phi$  to be the empty set.<sup>5</sup> 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.<sup>6</sup>

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  $G_\xi$ -volume of these sub-boxes and that each  $G_\xi$ -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:

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

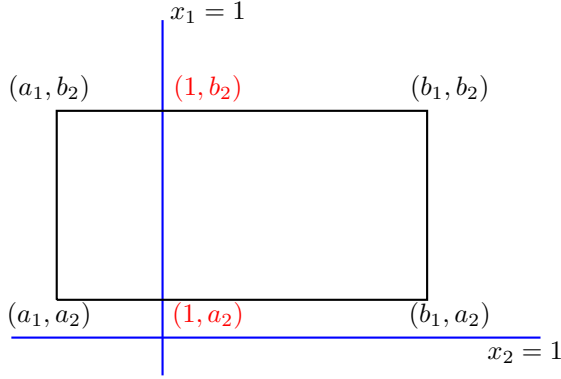
Similarly, in Panel B, we consider a 2-box  $B \equiv [a_1, b_1] \times [a_2, b_2]$  with  $a_1 < 1 < b_1$  and  $a_2 < 1 < b_2$  and let  $\mathbf{a}^* = (1, 1)$ . The  $G$ -volume of this box can similarly be decomposed

<sup>5</sup>For example, for  $n = 2$  and  $\xi = \{1\}$ , we have  $\mathbf{x}(\xi) = x_2$ ,  $\bar{\mathbf{x}}(\xi) = (1, x_2)$  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  $G_\xi = G_\phi = G$ .

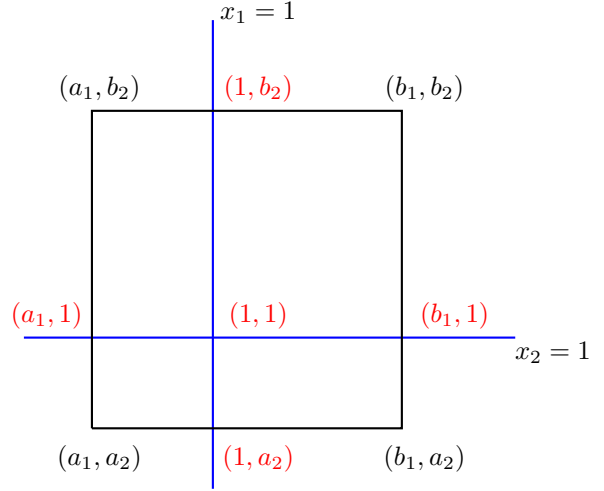
<sup>6</sup>In the example above with  $n = 2$  and  $\xi = \{1\}$ ,  $V_{G_\xi}([\mathbf{a}(\xi), \mathbf{b}(\xi)]) = G(1, b_2) - G(1, a_2)$ .



Panel A



Panel B

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

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 don't need to 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  $G_\xi$  volume is non-negative as long as its density  $g_\xi$  is non-negative on  $(1, \infty]^{l(\xi)}$ . It is thus obvious that the function  $G(\cdot)$  is  $n$ -increasing.

**Theorem 1.** Suppose a function  $G : \overline{\mathbb{R}}^n \rightarrow \mathbb{R}$  satisfies (i)  $G(\mathbf{z}) = 0$  for  $\mathbf{z} \notin [1, \infty]^n$ , (ii)  $g(\mathbf{z}) = \frac{\partial^n G(\mathbf{z})}{\partial z_1 \partial z_2 \dots \partial z_n} \geq 0$  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.

*Proof.* 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**  $\mathbf{a} \geq (1, \dots, 1)$ .

Then  $B$  lies in the region where  $g(\mathbf{z}) = \frac{\partial^n G(\mathbf{z})}{\partial z_1 \partial z_2 \dots \partial z_n} \geq 0$  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**  $\mathbf{b} \geq (1, \dots, 1)$  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

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

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  $g_\xi$  of  $G_\xi$  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,

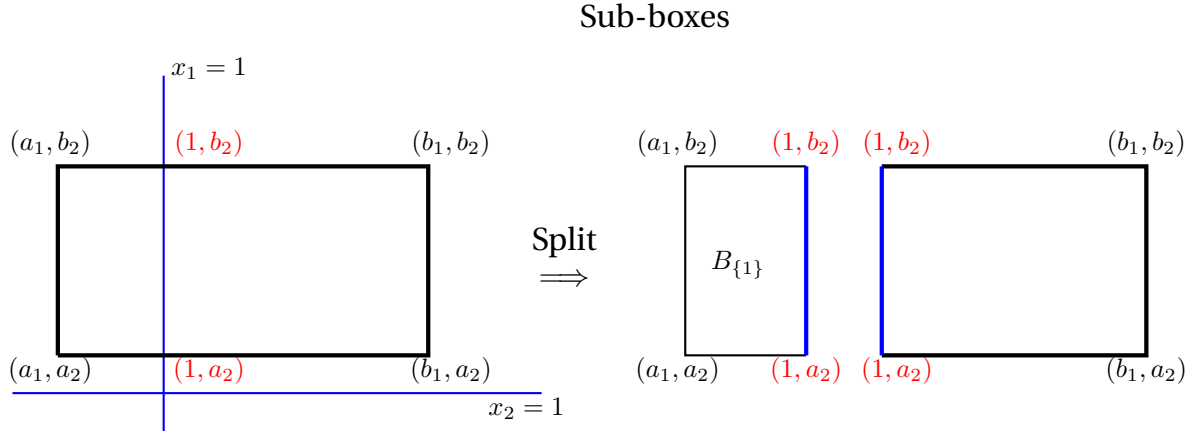
$$V_{G_\xi}([\mathbf{a}^*(\xi), \mathbf{b}(\xi)]) = \sum_{\mathbf{d} \in \Theta([\mathbf{a}^*(\xi), \mathbf{b}(\xi)])} (-1)^{m(A(\mathbf{d}; [\mathbf{a}^*(\xi), \mathbf{b}(\xi)]))} G_\xi(\mathbf{d}), \quad (4)$$

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

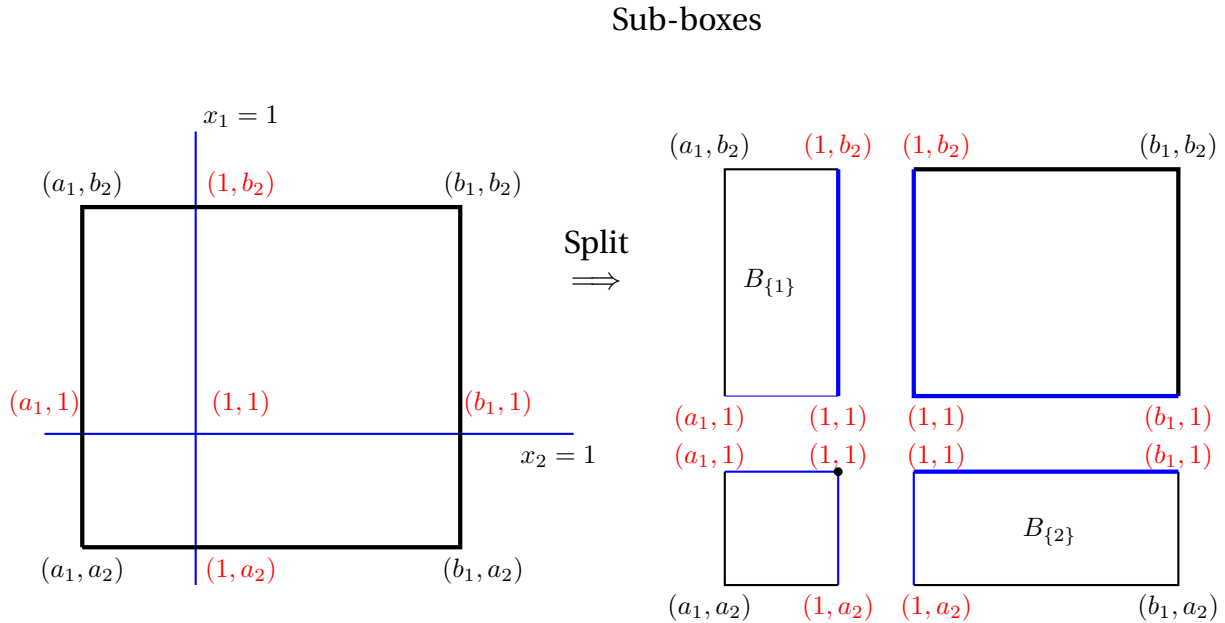
To proceed, let  $\Gamma \equiv \{\mathbf{c} \in \bar{\mathbb{R}}^n | c_k \in \{a_k, a_k^*, b_k\} \text{ for all } k \in \mathbb{N}_n\}$  be the collection of all vertices after the decomposition of  $B$  into sub-boxes and let  $\Psi \equiv \Gamma \setminus \Theta(B)$  be the set of “new” vertices generated 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  $\Psi_\xi \equiv \{\mathbf{c} \in \Psi | c_k = a_k^* \text{ for all } k \in \xi\}$ , let  $B_\xi \equiv [\bar{\mathbf{a}}(\Omega \setminus \xi), \bar{\mathbf{b}}(\xi)]$  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

Panel A



Panel B



*In Panel A*,  $\Omega = \{1\}$ ,  $\mathbf{a}^* = (1, a_2)$ ,  $\Gamma = \{(a_1, a_2), (1, a_2), (b_1, a_2), (a_1, b_2), (1, b_2), (b_1, b_2)\}$ ,  $\Psi = \{(1, a_2), (1, b_2)\}$ , 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

$\mathbf{d} \in \Theta([\mathbf{a}^*(\xi), \mathbf{b}(\xi)])$  if and only if  $\mathbf{d} = \mathbf{c}(\xi)$  with  $\mathbf{c} \in \Theta(B_\xi) \cap \Psi_\xi$  implies that

$$\begin{aligned} V_{G_\xi}([\mathbf{a}^*(\xi), \mathbf{b}(\xi)]) &= \sum_{\mathbf{d}=\mathbf{c}(\xi): \mathbf{c} \in \Theta(B_\xi) \cap \Psi_\xi} (-1)^{m(A(\mathbf{d}; [\mathbf{a}^*(\xi), \mathbf{b}(\xi)]))} G_\xi(\mathbf{d}) \\ &= \sum_{\mathbf{c} \in \Theta(B_\xi) \cap \Psi_\xi} (-1)^{m(A(\mathbf{c}(\xi); [\mathbf{a}^*(\xi), \mathbf{b}(\xi)]))} G(\bar{\mathbf{c}}(\xi)). \end{aligned}$$

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

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

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

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

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

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

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  $\mathbf{c} \in \Theta(B_\xi) \setminus \Psi_\xi$  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  $B_\xi = [\bar{\mathbf{a}}(\Omega \setminus \xi), \bar{\mathbf{b}}(\xi)]$ , we then have

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

Note that

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

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

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

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

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

The rest of the proof consists in showing that (i) Term 1 =  $V_G(B)$  and that (ii) 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  $m(\Omega(\mathbf{c})) = 0$  by definition of  $\Psi$ , and  $m(\{k \in \Omega | c_k = a_k\}) = m(\xi)$ . So,

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

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

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

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 be 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

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

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

**Remark.** Condition (iii) in Theorem 1 ensures that functions  $G_\xi$  are smooth on  $(1, \infty]^{l(\xi)}$  so that the  $G_\xi$ -volumes are non-negative by the fundamental theorem of calculus. The functions  $g_\xi$  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

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

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

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

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

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

*Proof.* 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

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

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

$$\begin{aligned} h_\xi(\mathbf{x}(\xi)) &= \frac{\partial H_\xi(\mathbf{x}(\xi))}{\partial \mathbf{x}(\xi)} \\ &= (1 - \rho) \left( \frac{\theta}{1 - \rho} \right)^{l(\xi)} \left[ \prod_{i=1}^{l(\xi)-1} (i - (1 - \rho)) \right] \\ &\quad \cdot \left[ \prod_{i \in \mathbb{N}_n \setminus \xi} \mathbf{x}_i^{-1} (T_i \mathbf{x}_i^{-\theta})^{\left(\frac{1}{1-\rho}\right)} \right] \left[ \sum_{i \in \xi} T_i^{\left(\frac{1}{1-\rho}\right)} + \sum_{i \in \mathbb{N}_n \setminus \xi} (T_i \mathbf{x}_i^{-\theta})^{\left(\frac{1}{1-\rho}\right)} \right]^{(1-\rho-l(\xi))}, \end{aligned}$$

which is non-negative for all  $\mathbf{x}(\xi) \in (1, \infty]^{l(\xi)}$ .<sup>7</sup> This completes the proof.  $\square$

## Appendix

### The maximum is distributed Pareto.

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

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

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

---

<sup>7</sup>By convention, we denote  $\prod_{i=1}^0 u_i = 1$  for any sequence  $\{u_i\}$  of real numbers.



## 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  $u > \tilde{T}^{1/\theta}$ ,

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

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

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

## Marginal distributions.

For any nonempty proper subset  $\xi$  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  $\infty$  otherwise. The lower-dimension marginal corresponding to  $\xi$  is

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

## Pareto tails.

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

$$\mathbb{P}(Z_i \geq z | Z_i \geq a) = \frac{\mathbb{P}(Z_i \geq z)}{\mathbb{P}(Z_i \geq a)} = \left(\frac{z}{a}\right)^{-\theta}.$$

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

## The role of $\rho$ .

Fix a point  $\mathbf{z} = (z_1, \dots, z_n) \in \overline{\mathbb{R}}^n$ . If  $z_i < \tilde{T}^{1/\theta}$  for some  $i$ , then  $H_\rho(\mathbf{z}) = 0 = \min_j F_j(z_j)$  because  $F_i(z_i) = 0$ . Suppose  $z_i \geq \tilde{T}^{1/\theta}$  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  $\rho \rightarrow 1$ ,

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

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

It also follows that for every pair  $(i, j)$ ,  $H(\mathbf{z}; \{i, j\}) \rightarrow 1 - \max\{T_i z_i^{-\theta}, T_j z_j^{-\theta}\}$  as  $\rho \rightarrow 1$ .

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  $\rho \rightarrow 1$ .

## 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  $z_i < (\tilde{T}')^{1/\theta}$  for some  $i$ .

Suppose  $z_i \geq (\tilde{T}')^{1/\theta}$  for all  $i$ . We have

$$\left( \sum_{i=1}^n (T_i z_i^{-\theta})^{1/(1-\rho)} \right)^{1-\rho} \leq \left( \sum_{i=1}^n (T'_i z_i^{-\theta})^{1/(1-\rho)} \right)^{1-\rho}$$

and thus

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

by definition of  $H(\cdot)$ .

### Stochastic dominance with respect to $\theta$ .

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

## References

- ARKOLAKIS, C., RAMONDO, N., RODRÍGUEZ-CLARE, A. and YEAPLE, S. (2013). *Innovation and production in the global economy*. Tech. rep., National Bureau of Economic Research.
- ARNOLD, B. (). Pareto distributions. 1983. *International Cooperative Publ. House, Fairland, MD*.
- , CASTILLO, E. and SARABIA, J. M. (1992). *Conditionally specified distributions*. Springer.
- ARNOLD, B. C. (1990). A flexible family of multivariate pareto distributions. *Journal of statistical planning and inference*, **24**, 249–258.
- (2015). *Pareto distributions*. CRC Press.
- ASIMIT, A. V., FURMAN, E. and VERNIC, R. (2010). On a multivariate pareto distribution. *Insurance: Mathematics and Economics*, **46** (2), 308–316.
- CHIRAGIEV, A. and LANDSMAN, Z. (2007). Multivariate pareto portfolios: Tce-based capital allocation and divided differences. *Scandinavian Actuarial Journal*, **4**, 261–280.
- DHAENE, J., DENUIT, M., GOOVAERTS, M. J., KAAS, R. and VYNCKE, D. (2002). The concept of comonotonicity in actuarial science and finance: theory. *Insurance: Mathematics and Economics*, **31**, 3–33.
- HANAGAL, D. D. (1996). A multivariate pareto distribution. *Communications in Statistics-Theory and Methods*, **25** (7), 1471–1488.
- JUPP, P. E. and MARDIA, K. V. (1982). A characterization of the multivariate pareto distribution. *The Annals of Statistics*, **10**, 1021–1024.
- KOTZ, S., BALAKRISHNAN, N. and JOHNSON, N. L. (2002). *Continuous Multivariate Distributions*. New York: John Wiley & Sons.
- LI, H. (2006). *Tail dependence of multivariate Pareto distributions*. Tech. rep., Technical Report 2006.

MARDIA, K. V. (1962). Multivariate pareto distributions. *The Annals of Mathematical Statistics*, **33**, 1008–1015.

NELSEN, R. B. (2007). *An introduction to copulas*. Springer Science & Business Media.

SU, J. and FURMAN, E. (2016). A form of multivariate pareto distribution with applications to financial risk measurement. *ASTIN Bulletin*, pp. 1–27.