


Rearrangement algorithm and maximum entropy

Carole Bernard^{1,2} · Oleg Bondarenko³ ·
Steven Vanduffel² 

Published online: 6 September 2017

© The Author(s) 2017. This article is an open access publication

Abstract We study properties of the block rearrangement algorithm (BRA) in the context of inferring dependence among variables given their marginal distributions and the distribution of their sum. We show that when all distributions are Gaussian the BRA yields solutions that are “close to each other” and exhibit almost maximum entropy, i.e., the inferred dependence is Gaussian with a correlation matrix that has maximum possible determinant. We provide evidence that, when the distributions are no longer Gaussian, the property of maximum determinant continues to hold. The consequences of these findings are that the BRA can be used as a stable algorithm for inferring a dependence that is economically meaningful.

Keywords Entropy · Block rearrangement algorithm · Inferring dependence

The authors gratefully acknowledge funding from IFSID and the Global Risk Institute (GRI) for the related project “Inferring tail correlations from option prices”. Part of this research was performed while C. Bernard was visiting the Institute for Pure and Applied Mathematics (IPAM). C. Bernard acknowledges support from the project on Systemic Risk funded by the GRI and the Louis Bachelier Institute. We would like to thank Edgars Jakobsons from ETH Zürich for various interesting comments.

✉ Steven Vanduffel
steven.vanduffel@vub.ac.be

Carole Bernard
carole.bernard@grenoble-em.com

Oleg Bondarenko
olegb@uic.edu

¹ Department of Accounting, Law and Finance, Grenoble Ecole de Management, Grenoble, France

² Department of Economics and Political Sciences, Vrije Universiteit Brussel (VUB), Brussels, Belgium

³ Department of Finance, University of Illinois at Chicago, Chicago, IL, USA

1 Introduction

In this paper, we study properties of the Block Rearrangement Algorithm (BRA), which generalizes the standard Rearrangement Algorithm (RA) introduced by [Puccetti and Rüschendorf \(2012\)](#) and further developed in [Embrechts et al. \(2013\)](#). The RA has applications in various disciplines. [Embrechts et al. \(2013\)](#) use the RA in quantitative risk management to obtain sharp approximations for the maximum possible Value-at-Risk (VaR) of a portfolio sum of d dependent risks when all marginal distributions are known but no information on dependence is available; see [Aas and Puccetti \(2014\)](#) for a case study. The RA can also be used in situations under which partial information on dependence is available (see [Bernard et al. 2017a, b](#); [Bernard and Vanduffel 2015](#)). The algorithm has also important applications in operations research. Consider the production of some goods requiring d different specialized operations and assume that there are n identical assembly lines. For each operation j there are n workers available; the i -th worker requires an amount of time equal to x_{ij} to execute the task. The problem of optimal Assembly Line Crew Scheduling (ALCS) is then to (re)allocate the nd workers so that all assembly lines are operating and the maximum production time is minimized (see e.g., [Coffman and Yannakakis 1984](#)). The RA attempts to achieve this situation.

We study properties of the RA in the context of the problem of inferring dependence among risks X_j ($j = 1, 2, \dots, d$) under the sole knowledge of their marginal distributions as well as of the distribution of their sum $S = \sum_{i=1}^d X_i$. Besides some theoretical appeal, the solution to this problem is also of practical relevance in finance. Using the price information of options (written) on individual stocks as well as on the index that comprises these stocks, it is indeed possible to infer the risk-neutral distribution F_j of future stock returns X_j ($j = 1, 2, \dots, d$) as well as the distribution F_S of the future index return $S := X_1 + X_2 + \dots + X_d$ ([Ross 1976](#); [Breedon and Litzenberger 1978](#); [Banz and Miller 1978](#); [Ait-Sahalia and Lo 2000](#); [Bondarenko 2003](#)). The joint distribution is not observable directly. We develop an algorithm to infer dependence among variables with given marginal distributions F_j and for which the distribution of the sum F_S is known. In other words, our algorithm makes it possible to infer in a model-free and forward-looking way the dependence structure of stock returns consistent with the current market prices of options. There are various potential applications. For instance, the inferred dependence can be used to price multivariate derivatives non-parametrically. It is then straightforward to price arbitrarily weighted basket options, exchange options or spread options such that their prices are consistent with existing options written on individual assets and on a given index. Another possible application is to develop an indicator of herd behavior among stock returns using information on option prices only. Such an index can then be used as a signal of systemic risk.

The method that we propose to infer dependence is solely based on the Rearrangement Algorithm (RA) and its recent generalization, the Block Rearrangement Algorithm (BRA) ([Bernard et al. 2017a](#); [Bernard and McLeish 2016](#); [Boudt et al. 2017](#)). In general, there might be several compatible dependence structures¹ and one could reasonably expect that by randomizing its initialization, BRA allows one to infer all possible dependence structures. However, this is not the case. When all distributions are Gaussian, we find that among all possible candidate dependence structures the BRA yields a solution with (almost) maximum entropy. That is, the inferred dependence among the variables X_j ($j = 1, 2, \dots, d$) is approximately Gaussian and its correlation matrix has maximum determinant.

¹ “Compatibility” will be defined formally later in Definition 2.1. For example, there can be many ways to join normally distributed random variables such that their sum is normally distributed.

This feature is interesting for the following two reasons: First, the fact that application of the BRA yields solutions that do not really depend on its initialization ensures that the BRA can be used as a stable algorithm for inferring dependence among stock prices. Second, the stable solution is economically meaningful, as the maximum entropy (statistically proved when all distributions are Gaussian) corresponds to the intuition that in well-functioning markets maximum uncertainty must prevail. A dependence with lower entropy would mean that we use information that we do not possess. This insight is also explored in Rubinstein (1994), Jackwerth and Rubinstein (1996), and Stutzer (1996) in the context of inference of marginal risk neutral probabilities from observed option prices. In this regard, we note that the principle of maximum entropy appears in many other disciplines. For example, the second law of thermodynamics asserts that the entropy of irreversible processes can never decrease. Another example pertains to the field of statistics where the principle of maximum entropy is used to obtain prior probability distributions for Bayesian inference; see Jaynes (1957).

The paper is organized as follows. In Sect. 2 we recall the standard Rearrangement Algorithm and the Block Rearrangement Algorithm and present its use to infer dependence among random variables. In Sect. 3, we discuss the principle of maximum entropy. In Sect. 5 we show that for Gaussian margins the algorithm identifies the dependence with maximum entropy using the methodology and notation presented in Sect. 4. In Sect. 6 we perform a robustness analysis and assess the extent to which our main result remains valid in a non-Gaussian framework or when the standard RA is used. We provide evidence that when the distributions are skewed the inferred dependence is still unique and exhibits a correlation matrix with maximum determinant. Section 7 concludes.

2 Rearrangement algorithm to infer dependence

We first present the standard RA of Puccetti and Rüschendorf (2012) and Embrechts et al. (2013) and show how to modify it to infer the dependence among variables for which the marginal distributions and the distribution of the sum are known. We include the presentation of the algorithms for completeness.

2.1 Rearrangement algorithm

The RA has been originally proposed in the context of assessing the impact of model uncertainty on portfolio models and is considered as a breakthrough in this regard. Technically, it amounts to finding a dependence among risks X_j with known distributions F_j ($j = 1, 2, \dots, d$) such that the variance of their sum $S = X_1 + X_2 + \dots + X_d$ becomes as low as possible. As the algorithm requires random variables that are discretely distributed² we further assume that each risk X_j ($j = 1, 2, \dots, d$) is sampled into n equiprobable values, i.e., we consider the realizations $x_{ij} := F_j^{-1}(\frac{i-0.5}{n})$ ($i = 1, \dots, n; j = 1, 2, \dots, d$). We obtain an $n \times d$ matrix \mathbf{X} :

$$\mathbf{X} = [X_1, \dots, X_d] = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1d} \\ x_{21} & x_{22} & \dots & x_{2d} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nd} \end{bmatrix}.$$

² Note that continuous distributions can be approximated to any degree of accuracy by discrete ones.

The matrix can be readily seen as the representation of a d -dimensional random vector; each column corresponds to a random variable and each row is a possible joint outcome that occurs with probability $1/n$. This interpretation allows us to also apply probabilistic operators (such as the mean, the variance and the correlation) onto columns of matrices.

We consider the problem of rearranging the elements x_{ij} of the matrix \mathbf{X} such that after rearrangement the variance of S is as small as possible. We allow for rearrangements within columns, as doing so affects the dependence among the $X_j, j = 1, 2, \dots, d$, but not their marginal distributions. By contrast, swapping values among columns will affect the marginal distributions and is not allowed. Clearly, in order for S to have smallest possible variance it must hold that for all $\ell = 1, 2, \dots, d, X_\ell$ is anti-monotonic with $\sum_{k=1, k \neq \ell}^d X_k$ (Puccetti and Rüschendorf 2012, Theorem 2.1); this observation lies at the core of their rearrangement method.

2.1.1 Standard rearrangement algorithm (RA)

1. For $j = 1, 2, \dots, d$, make the j th column anti-monotonic with the sum of the other columns, i.e., rearrange their values such that they have the property of being oppositely ordered to the row sums across the other columns.
2. If there is no improvement in $\text{var}\left(\sum_{k=1}^d X_k\right)$, output the current matrix \mathbf{X} , otherwise return to step 1.

At each step of this algorithm, we ensure that the j th column is anti-monotonic with the sum of the remaining columns, so that the columns, say X_j before rearranging and \tilde{X}_j after rearranging, verify obviously³

$$\text{var}\left(\sum_{k=1}^d X_k\right) \geq \text{var}\left(\sum_{k=1}^d \tilde{X}_k\right).$$

Hence, at each step of the algorithm the variance decreases. If the lower bound of zero is attained then we have found the situation of perfect mixability⁴ among the variables, i.e., a dependence that makes the sum constant, and a global optimum is obtained. Otherwise, the algorithm will converge to a local minimum.⁵ Indeed, there is no guarantee that the local minimum yields the minimum variance of the sum as another initialization of \mathbf{X} (i.e., a random permutation of the elements in each column) is likely to lead to another local optimum. Precisely, finding the dependence structure that minimizes the variance of the sum is a NP-complete problem (Haus 2015), which means that, most likely, there exists no deterministic algorithm with polynomial complexity. Thus, the above RA can only be an heuristic.

The standard RA has been recently generalized by considering blocks (Bernard and McLeish 2016; Bernard et al. 2017a; Boudt et al. 2017). The idea is as follows. If the distribution of $S = X_1 + X_2 + \dots + X_d$ has minimum variance it must actually hold that

³ Indeed $\text{var}\left(\sum_{k=1}^d X_k\right) = \text{var}\left(X_j + \sum_{k \neq j} X_k\right)$, and a necessary condition for $\text{var}\left(\sum_{k=1}^d X_k\right)$ to become minimum is that each X_j is anti-monotonic with $\sum_{k \neq j} X_k$.

⁴ The concept of mixability has been introduced in Wang and Wang (2011) and further extended in Wang and Wang (2016); see also Gaffke and Rüschendorf (1981) for some early results that are connected.

⁵ Each time a column is effectively rearranged, the variance of the sum strictly decreases. Since there is a finite number of possible rearranged matrices, the algorithm always terminates in a finite number of steps. At this point, all columns satisfy the anti-monotonicity property and a local optimum for the variance has been reached.

for any decomposition of $\{1, 2, \dots, d\} = I_1 \cup I_2$ into two disjoint sets I_1 and I_2 , the sums $S_1 := \sum_{k \in I_1} X_k$ and $S_2 := \sum_{k \in I_2} X_k$ are anti-monotonic (and not only for sets of the form $I_1 = \{j\}$ and $I_2 = \{1, 2, \dots, d\} \setminus I_1$, as in the standard RA). This observation makes it possible to generalize the standard RA by rearranging “blocks of columns” instead of one column at a time: The columns in the first set I_1 are stacked into a matrix (block) \mathbf{X}_1 and one rearranges its rows (i.e., one swaps entire rows) such that the row sums of \mathbf{X}_1 (reflecting S_1) are in increasing order. As for the matrix \mathbf{X}_2 that is formed by stacking the remaining columns, the rows are rearranged such that the row sums (reflecting S_2) are in decreasing order.

2.1.2 Block rearrangement algorithm (BRA)

1. Select a random sample of n_{sim} possible partitions of the columns $\{1, 2, \dots, d\}$ into two non-empty subsets $\{I_1, I_2\}$. When $d \leq 10$, we take $n_{sim} = 2^{d-1} - 1$ so that all non-trivial partitions are considered.⁶
2. For each of the n_{sim} partitions, create the matrices (blocks) \mathbf{X}_1 and \mathbf{X}_2 with corresponding row sums S_1 and S_2 and rearrange rows of \mathbf{X}_2 so that S_2 is anti-monotonic to S_1 .
3. If there is no improvement in $\text{var} \left(\sum_{k=1}^d X_k \right)$, output the current matrix \mathbf{X} , otherwise, return to step 1.

When all partitions can be assessed, [Bernard and McLeish \(2016\)](#) show that the BRA outperforms the standard RA by orders of magnitude in terms of the distance between the obtained local minimum for the variance of the sum and the global minimum for the variance of the sum. When the different partitions cannot be assessed in a reasonable amount of time (i.e., for matrices with many columns), a stopping rule needs to be used and it is no longer clear cut that BRA dominates the standard RA. [Boudt et al. \(2017\)](#) conduct a horse race and show that for typical applications in quantitative risk management the standard RA and BRA are close competitors. By contrast, for some rearrangement problems that appear in operations research, a suitably designed BRA is outperforming.

2.2 Inferring dependence

To infer the dependence among variables X_1, \dots, X_d with known distributions and for which the distribution of the sum S is also known, we apply the BRA on the following $n \times (d + 1)$ matrix \mathbf{M} that is defined as

$$\mathbf{M} = [X_1, \dots, X_d, -S] = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1d} & -s_1 \\ x_{21} & x_{22} & \dots & x_{2d} & -s_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nd} & -s_n \end{bmatrix}. \tag{2.1}$$

Elements x_{ij} are defined as before and we sample the realizations s_i of the sum S using its distribution F_S , i.e., $s_i := F_S^{-1} \left(\frac{i-0.5}{n} \right)$.

In this regard, we note it is not guaranteed that after running the BRA the row sums of the rearranged matrix are all equal to zero and that a compatible candidate dependence has been found. Formally, “compatibility” is defined as follows.

⁶ Since all examples in this paper consider $d \leq 10$, we use the entire set of partitions and select them in a random order.

Definition 2.1 (*Compatibility*) A set of d distributions F_1, F_2, \dots, F_d is said to be compatible with the distribution F_S if there exists a random vector (X_1, X_2, \dots, X_d) and a random variable S such that the X_i and S are distributed as F_i ($i = 1, 2, \dots, d$) and F_S , respectively, and $S = \sum_{i=1}^d X_i$ almost surely. In this case, the dependence among the X_i ($i = 1, 2, \dots, d$) is called a compatible dependence.

“Compatibility” is closely related to joint mixability (Wang and Wang 2016). Specifically, F_1, F_2, \dots, F_d is compatible with the distribution F_S if and only if F_1, F_2, \dots, F_d and F_{-S} are jointly mixable with null center.

Clearly, a necessary requirement for joint mixability of F_1, F_2, \dots, F_d and F_{-S} is that $E(S) = E(X_1) + \dots + E(X_d)$ holds. Wang and Wang (2016) study joint mixability in some generality and provide several theoretical results that ensure existence of a compatible dependence in various cases of interest. Furthermore, Puccetti and Wang (2015) propose the Mixability Detection Procedure (MDP) as a suitable method to numerically assess joint mixability for an *arbitrarily* set of bounded distributions. In what follows, we assume that there is at least one compatible dependence for the given distributions F_i ($i = 1, 2, \dots, d$) and F_S and we merely aim at inferring any such dependence.

In the ideal situation, we observe that after running the BRA the row sums of the rearranged matrix are all equal to zero and a compatible dependence has been found. In practice, however, due to discretization errors and the fact that the algorithm is a heuristic, and the solution is only approximate. In the remainder of the paper, we study the properties of this approximate solution. To do so, we repeat the experiment K times. Each time, we randomize each of the $d + 1$ columns of \mathbf{M} given in (2.1) and next we apply the BRA. For each $k = 1, 2, \dots, K$ we denote by $\mathbf{M}^{(k)} = [X_1^{(k)}, \dots, X_d^{(k)}, -S^{(k)}]$ the matrix after rearrangement. Note that the vectors $[X_1^{(k)}, \dots, X_d^{(k)}]$ only differ with respect to their interdependence. As for each k we randomize the initial matrix (i.e., we randomize the initial condition for the algorithm), it appears reasonable to assert that the procedure provides a way to describe the set of *all* possible dependence structures (copulas) that are consistent with the given information. Nevertheless, we find that application of the procedure leads to “a thin set” of dependence structures: the solutions $\mathbf{X}^{(k)} = [X_1^{(k)}, \dots, X_d^{(k)}]$ all have (nearly) the same correlation matrix and this matrix has maximum determinant. The only exceptions come from the situations in which the initial dependence structure is such that the variance of the row sums is already zero. In this case, the BRA has converged up-front and stops at the initial configuration. However, our study will show that even a tiny perturbation of this initial structure will suffice for the algorithm to converge to a dependence structure with correlation matrix that has maximum determinant. Moreover, when distributions are Gaussian we show that this inferred dependence actually reflects maximum entropy. The following section introduces the notation and some theoretical results that are needed for this purpose.

3 Maximum entropy

In the remainder of the paper, all variables X_1, \dots, X_d and their sum S have known distributions F_1, \dots, F_d and F_S . Let μ_1, \dots, μ_d and μ_S denote their respective means and $\sigma_1, \dots, \sigma_d$ and σ_S denote the standard deviations.

3.1 Correlation matrices with maximum determinant

In general, there could be many compatible dependence structures (copulas) for the given F_1, \dots, F_d and F_S . To compare them, we will focus on the properties of their corresponding (Pearson) correlation matrices. In this regard, it is convenient to formally define several sets of correlation matrices.

Definition 3.1 (*All correlation matrices*) Let \mathcal{B} denote the set of all $d \times d$ symmetric semi-definite matrices, whose elements are equal to one on the main diagonal and range between -1 and 1 everywhere else.

Consider a correlation matrix $R \in \mathcal{B}$,

$$R = \begin{bmatrix} 1 & r_{12} & \dots & r_{1d} \\ r_{21} & 1 & \dots & r_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ r_{d1} & r_{d2} & \dots & 1 \end{bmatrix}. \tag{3.2}$$

Since the standard deviations $\sigma_1, \dots, \sigma_d$ and σ_S are fixed, R (corresponding to a compatible dependence) must satisfy the following additional restriction:

$$\sigma_S^2 = \sum_{i=1}^d \sigma_i^2 + 2 \sum_{i=1}^{d-1} \sum_{j>i} \sigma_i \sigma_j r_{ij}.$$

Equivalently, we can re-write the above condition as the constraint on average correlation:

$$\frac{\sum_{i=1}^{d-1} \sum_{j>i} \sigma_i \sigma_j r_{ij}}{\sum_{i=1}^{d-1} \sum_{j>i} \sigma_i \sigma_j} = \rho^{imp}, \tag{3.3}$$

where constant ρ^{imp} is defined as

$$\rho^{imp} := \frac{\sigma_S^2 - \sum_{i=1}^d \sigma_i^2}{2 \sum_{i=1}^{d-1} \sum_{j>i} \sigma_i \sigma_j}.$$

This motivates our next definition of the constrained set $\mathcal{C}(\rho^{imp})$.

Definition 3.2 (*Constrained correlation matrices*) Let $\mathcal{C}(\rho^{imp})$ denote the set of correlation matrices $R \in \mathcal{B}$ satisfying the constraint in (3.3).

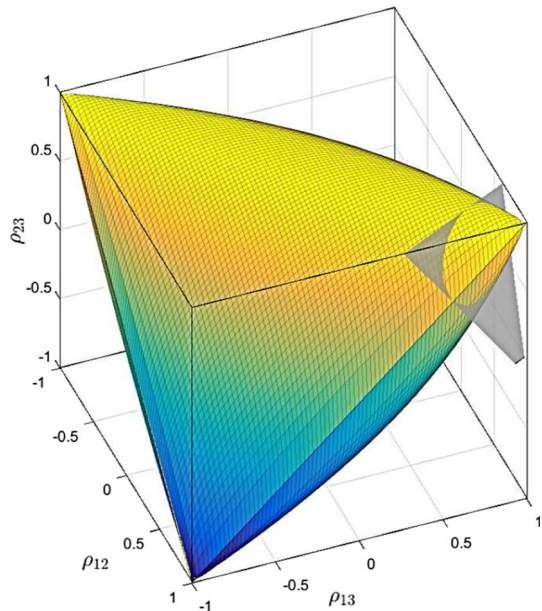
Figure 1 illustrates the two sets \mathcal{B} and $\mathcal{C}(\rho^{imp})$ for the special case when $d = 3$.

Proposition 3.3 *The set $\mathcal{C}(\rho^{imp})$ is convex and compact.*

Proof It is well-known that set \mathcal{B} is compact and convex. On the other hand, the hyperplane corresponding to the linear constraint (3.3) is clearly convex and closed. Thus, Proposition 3.3 holds. □

Assume that the F_1, \dots, F_d are compatible with F_S . Clearly, every compatible dependence must be such that its corresponding correlation matrix satisfies the constraint in (3.3). By contrast, there might be correlation matrices R that satisfy this constraint but for which no compatible dependence exists.

Fig. 1 The sets \mathcal{B} and $\mathcal{C}(\rho^{imp})$ in Definitions 3.1 and 3.2 for $d = 3$. The example uses $\sigma_1 = 1$, $\sigma_2 = \frac{2}{3}$, $\sigma_3 = \frac{1}{3}$, and $\rho^{imp} = 0.8$. The colored volume represents the set of all correlation matrices \mathcal{B} and the (gray) plane represents the linear constraint (3.3). Their intersection forms the constrained set $\mathcal{C}(\rho^{imp})$. (Color figure online)



Definition 3.4 (*Compatible correlation matrices*) Let \mathcal{D} denote the set of correlation matrices $R \in \mathcal{B}$ that result from all compatible dependence structures (copulas) for the given marginal distributions F_1, F_2, \dots, F_d and F_S .

It is clear that $\mathcal{D} \subset \mathcal{C}(\rho^{imp})$ but the opposite does not readily hold. The following proposition shows that in the Gaussian case \mathcal{D} and $\mathcal{C}(\rho^{imp})$ coincide.

Proposition 3.5 Assume that F_1, \dots, F_d and F_S are Gaussian distributions and that F_1, \dots, F_d are compatible with F_S , then $\mathcal{D} = \mathcal{C}(\rho^{imp})$.

Proof We only need to show that $\mathcal{C}(\rho^{imp}) \subset \mathcal{D}$. Consider $R \in \mathcal{C}(\rho^{imp})$: it is a well-defined correlation matrix satisfying (3.3). Then there exists a d -dimensional Gaussian joint distribution with a Gaussian copula G with correlation matrix R and marginal distributions F_1, F_2, \dots, F_d . Let (X_1, X_2, \dots, X_d) be a vector following this d -dimensional Gaussian joint distribution, then $S := X_1 + X_2 + \dots + X_d$ follows a Gaussian distribution identical to F_S . Therefore, the F_1, F_2, \dots, F_d are compatible with F_S . Thus there exists a compatible dependence structure among the X_i (distributed as F_i) having correlation matrix R . Thus R belongs to \mathcal{D} . \square

Remark 3.6 Note that under the framework of elliptical distributions (which includes Gaussian distributions as a special case), necessary and sufficient conditions for a set of marginals to be jointly mixable have been given in Wang and Wang (2016). In particular, Gaussian distributions $\mathcal{N}(0, \sigma_i^2)$, $i = 1, \dots, d + 1$ are jointly mixable if and only if $\sum_{i=1}^{d+1} \sigma_i \geq 2 \max_{i=1, \dots, d+1} \sigma_i$. In fact, the necessary part follows from subadditivity of standard deviation (triangle inequalities) and the sufficient part builds on closedness of the (multivariate) normal family under linear combinations. More details can be found in their paper.

In general, characterizing the set \mathcal{D} for any choice of marginal distributions is an open problem, as in general it will only be a strict subset of $\mathcal{C}(\rho^{imp})$.⁷

Proposition 3.7 (Maximum determinant for Gaussian margins) *There exists a unique matrix $R \in \mathcal{C}(\rho^{imp})$ that has maximum determinant among all matrices in $\mathcal{C}(\rho^{imp})$.*

Proof Since log determinant is continuous and $\mathcal{C}(\rho^{imp})$ is compact, there exists a matrix $R \in \mathcal{C}(\rho^{imp})$ which attains maximum determinant. Furthermore, since $\mathcal{C}(\rho^{imp})$ is a convex set and log determinant is strictly concave (see Boyd and Vandenberghe 2004, Chapter 3, Section 1, p. 74), the matrix that maximizes log determinant and thus also the determinant is unique. \square

Definition 3.8 Let $R_M(\rho^{imp})$ denote the matrix with maximum possible determinant among all $R \in \mathcal{C}(\rho^{imp})$ and let $\Delta_M(\rho^{imp})$ denote its determinant.

Without the linear constraint (3.3) on the correlation matrix, the maximum determinant (over the set \mathcal{B}) is equal to 1, which is achieved when all off-diagonal elements of R are zero. With this constraint, however, $\Delta_M(\rho^{imp})$ is typically strictly less than 1. To find $\Delta_M(\rho^{imp})$, we solve numerically the constrained optimization in Matlab. Specifically, we search for a symmetric correlation matrix R in (3.2) with $\frac{1}{2}d(d - 1)$ unknown parameters, which maximizes the determinant subject to the constraints that all eigenvalues are non-negative and the additional linear constraint in (3.3).

3.2 Maximum entropy dependence for Gaussian margins

Entropy refers to disorder or uncertainty of a system. It plays a fundamental role in various disciplines including information theory, computing and thermodynamics (e.g., the second law of thermodynamics). In his fundamental work (Shannon 1948) defines the entropy H of a discrete random variable X with probability mass function $P(X)$ as $H(X) = -E(\log(P(X)))$. There are natural extensions to measure the entropy of multivariate distributions. Let f be the density of a multivariate distribution of (X_1, \dots, X_d) , then the entropy is defined as

$$H(X_1, \dots, X_d) = -E(\log(f(X_1, \dots, X_d))).$$

The following result characterizes the multivariate distribution of a random vector (X_1, \dots, X_d) with Gaussian margins. Its proof follows from the fact that among all absolutely continuous distributions with unbounded support having a given invertible covariance matrix, the multivariate Gaussian distribution has maximum entropy.

Proposition 3.9 (Maximum entropy for Gaussian margins) *The entropy of the multivariate distribution of a random vector (X_1, \dots, X_d) with Gaussian margins and invertible correlation matrix R satisfies*

$$H(X_1, \dots, X_d) \leq \frac{d}{2} (1 + \ln(2\pi)) + \sum_{i=1}^d \ln(\sigma_i) + \frac{1}{2} \ln(\det(R)), \tag{3.4}$$

where the equality holds if and only if (X_1, \dots, X_d) is a multivariate Gaussian distribution.

⁷ See for example the range of possible correlation coefficients between two different lognormal distributions in McNeil et al. (2005) that is strictly included in $(-1, 1)$.

This proposition (and its proof) can be found as Theorem 8.6.5 in [Thomas and Cover \(2006\)](#), where it is formulated in a slightly different way. Specifically, these authors prove that

$$H(X_1, \dots, X_d) \leq \frac{d}{2} (1 + \ln(2\pi)) + \frac{1}{2} \ln(\det(\Sigma)),$$

in which Σ is the covariance matrix of the vector (X_1, \dots, X_d) . Since the determinant of the covariance matrix Σ writes as $\det[\Sigma] = \det[R] \cdot \left(\prod_{i=1}^d \sigma_i^2\right)$, we obtain (3.4).

As a direct consequence of Proposition 3.9, the maximum entropy distribution that is consistent with given Gaussian margins and a given average correlation ρ^{imp} is uniquely characterized by a Gaussian dependence with a correlation matrix $R_M(\rho^{imp})$.

Our numerical experiments show that there is a striking connection between the BRA and maximum entropy in that every run of BRA leads to

- (i) a dependence close to Gaussian,
- (ii) a correlation matrix with determinant close to the maximum possible determinant.

Given that we do not formally prove the result, but rather provide evidence based on extensive simulations, we need to introduce relevant concepts that make it possible to reach such conclusion. This will be the topic of the next section.

Remark 3.10 The conclusion regarding the maximum entropy property remains unchanged when the “sum” is not put in the last column. This can be seen by relabeling the columns. For every column it indeed holds in the situation of (joint) mixability that $X_k = -\sum_{i \neq k} X_i$. We have $d + 1$ columns for which the determinants of their correlation matrix is 0, but when we extract d columns, the determinant must be maximum. In addition, after we prove that the dependence is Gaussian, then if (X_1, \dots, X_d) is Gaussian, then $(X_1, \dots, X_{k-1}, X_{k+1}, \dots, X_d, -S)$ is also a Gaussian vector.

4 Maximum entropy and BRA

We run K times the BRA. For $k = 1, 2, \dots, K$, let $R^{(k)}$ denote the output correlation matrix of $\mathbf{X}^{(k)}$:

$$R^{(k)} := \begin{bmatrix} 1 & \rho_{12}^{(k)} & \dots & \rho_{1d}^{(k)} \\ \rho_{21}^{(k)} & 1 & \dots & \rho_{2d}^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{d1}^{(k)} & \rho_{d2}^{(k)} & \dots & 1 \end{bmatrix}.$$

and denote its determinant by $\Delta^{(k)}$. We also define the realized (or effective) average correlation $\varrho^{(k)}$ as

$$\varrho^{(k)} = \frac{\sum_{i=1}^{d-1} \sum_{j>i} \sigma_i \sigma_j \rho_{ij}^{(k)}}{\sum_{i=1}^{d-1} \sum_{j>i} \sigma_i \sigma_j}, \tag{4.5}$$

in which σ_i specifically denotes the standard deviation of $X_i^{(k)}$ and $\rho_{ij}^{(k)}$ is the pairwise correlation between $X_i^{(k)}$ and $X_j^{(k)}$, for all i and j .⁸ Finally, we define the deviation $V^{(k)}$ as

⁸ Note that the solutions $\mathbf{X}^{(k)}$ are obtained by K separate runs of the BRA on matrix \mathbf{M} in (2.1) with randomized columns. The elements in each column do not change from one run to the other, only their order changes.

the L^2 -norm of the $n \times 1$ residual vector $e^{(k)} := X_1^{(k)} + \dots + X_d^{(k)} - S^{(k)}$:

$$V^{(k)} = \left\| e^{(k)} \right\|_2. \tag{4.6}$$

Similarly to all other quantities above, $V^{(k)}$ depends on n , but, for the ease of presentation, we omit the explicit dependence on n .

Proposition 4.1 *The following inequality holds:*

$$\left| \varrho^{(k)} - \rho^{imp} \right| \leq C_1 V^{(k)} + C_2 \left(V^{(k)} \right)^2, \tag{4.7}$$

where C_1 and C_2 are constants depending on the standard deviations $\sigma_1, \dots, \sigma_d$ and σ_S .

Proof For simplicity, we drop the superscript in $X_i^{(k)}$, $S^{(k)}$, $\varrho^{(k)}$, $e^{(k)}$ and $V^{(k)}$. Denote $U = X_1 + \dots + X_d$, so that $e = U - S$. Using the definitions of ϱ and ρ^{imp} :

$$\varrho - \rho^{imp} = \frac{\sigma_U^2 - \sum_{i=1}^d \sigma_i^2}{2 \sum_{i=1}^{d-1} \sum_{j>i} \sigma_i \sigma_j} - \frac{\sigma_S^2 - \sum_{i=1}^d \sigma_i^2}{2 \sum_{i=1}^{d-1} \sum_{j>i} \sigma_i \sigma_j} = \frac{\sigma_U^2 - \sigma_S^2}{2 \sum_{i=1}^{d-1} \sum_{j>i} \sigma_i \sigma_j}.$$

Therefore,

$$\left| \varrho - \rho^{imp} \right| = \frac{|\sigma_U^2 - \sigma_S^2|}{2 \sum_{i=1}^{d-1} \sum_{j>i} \sigma_i \sigma_j}.$$

However,

$$|\sigma_U^2 - \sigma_S^2| \leq \sigma_e^2 + 2\sigma_e \sigma_S \leq V^2 + 2V\sigma_S,$$

which completes the proof of the proposition. □

Intuitively, Proposition 4.1 states that, as the accuracy of a candidate solution improves (i.e., the deviation $V^{(k)}$ decreases), the realized average correlation $\varrho^{(k)}$ generally gets closer to ρ^{imp} .

Each run of the algorithm leads to the realized determinant $\Delta^{(k)}$, average correlation $\varrho^{(k)}$, and deviation $V^{(k)}$ ($k = 1, 2, \dots, K$). We denote with $\Delta := (\Delta^{(1)}, \Delta^{(2)}, \dots, \Delta^{(K)})$, $\varrho := (\varrho^{(1)}, \varrho^{(2)}, \dots, \varrho^{(K)})$ and $V := (V^{(1)}, V^{(2)}, \dots, V^{(K)})$, respectively, and introduce measures of dispersion that make it possible to assess them in a meaningful way.

4.1 Distance to the maximum determinant matrix

To assess the similarity of the outputs of BRA, we make use of the following measure of dispersion.

Definition 4.2 (*Mean Percentage Error*) For any vector of interest Z taking values $z^{(k)}$ ($k = 1, 2, \dots, K$) and some “reference point” z , we denote by $MPE(Z)$ its average percentage error, where percentage error (PE) of a $z^{(k)}$ is simply given as

$$PE(z^{(k)}) := 100 \left[\frac{z^{(k)} - z}{z} \right].$$

Footnote 8 continued

Thus, they all have the same standard deviations $\sigma_1, \dots, \sigma_d$ and σ_S , independent of k . When a continuous Normal distribution is approximated with n discrete values, the resulting standard deviation might slightly differ from its continuous counterpart. In this paper, we use $n = 1000$ and $10,000$, for which the discrepancies are extremely small. Nevertheless, we re-scale the discretized values to ensure that all standard deviations match exactly their continuous counterparts $\sigma_1, \dots, \sigma_d$ and σ_S .

The main discrepancy measure that we consider concerns the normalized determinants $\sqrt[d]{\Delta^{(k)}}$. The determinant is a measure of volume in \mathbb{R}^d and it satisfies the scaling property that $\det[cR] = c^d \det[R]$ for constant c . Thus, $\sqrt[d]{\Delta}$ can be viewed as “dimension-invariant.” Specifically, we consider

$$MPE_1 \left(\sqrt[d]{\Delta} \right), \tag{4.8}$$

which measures by how much *on average* normalized determinants $\sqrt[d]{\Delta^{(k)}}$ differ from the reference value $\sqrt[d]{\Delta_M(\rho^{imp})}$. That is, for $k = 1, 2, \dots, K$ we first compute the percentage errors $PE \left(\sqrt[d]{\Delta^{(k)}} \right)$ with respect to $\sqrt[d]{\Delta_M(\rho^{imp})}$ and then take their average.

In this regard, one would expect that $\sqrt[d]{\Delta^{(k)}} \leq \sqrt[d]{\Delta_M(\rho^{imp})}$, implying that distance $MPE_1 \left(\sqrt[d]{\Delta} \right)$ would be a suitable measure of discrepancy (as the average of K errors that all have the same sign). However, the interpretation of $MPE_1 \left(\sqrt[d]{\Delta} \right)$ is somewhat ambiguous. Due to discretization, it is usually not even possible to obtain a dependence that yields a realized correlation equal to the target correlation ρ^{imp} . Typically, $V^{(k)} \neq 0$ and $\varrho^{(k)}$ is slightly smaller in absolute value than ρ^{imp} , which means that $\Delta_M(\varrho^{(k)}) > \Delta_M(\rho^{imp})$. Hence, it may happen for some k that $PE \left(\sqrt[d]{\Delta^{(k)}} \right) > 0$. To accommodate for this possibility, we consider an alternative measure of discrepancy where $PE \left(\sqrt[d]{\Delta^{(k)}} \right)$ is computed with respect to the reference point $\sqrt[d]{\Delta_M(\varrho^{(k)})}$ instead of the previous $\sqrt[d]{\Delta_M(\rho^{imp})}$. This choice of the reference point ensures that $PE \left(\sqrt[d]{\Delta^{(k)}} \right) \leq 0$. We denote the resulting percentage error as $PE_2 \left(\sqrt[d]{\Delta^{(k)}} \right)$ and its average as

$$MPE_2 \left(\sqrt[d]{\Delta} \right). \tag{4.9}$$

Finally, the average percentage error of ϱ to the target value ρ^{imp} is denoted by $MPE(\varrho)$.

It is easy to verify that these three measures are scale invariant. In the following examples, we observe that all these measures converge to 0 when the number of discretization steps n increases.

Property 4.3 (Scale invariance) *The measures $MPE_1 \left(\sqrt[d]{\Delta} \right)$, $MPE_2 \left(\sqrt[d]{\Delta} \right)$ and $MPE(\varrho)$ are scale-invariant.*

4.2 Testing the Gaussian dependence

We also test whether for $k = 1, 2, \dots, K$ the different $\mathbf{M}^{(k)} = [X_1^{(k)}, \dots, X_d^{(k)}, -S^{(k)}]$ can be seen as draws from a multivariate normal distribution. Very well-known tests in this regard are Mardias tests of multinormality or variants hereof. For general multivariate data, [Mardia \(1970\)](#) constructed two statistics for measuring multivariate skewness and kurtosis, which can be used for testing the hypothesis of normality ([Mardia 1974, 1975; Mardia et al. 1980](#)). Usually the tests whether kurtosis (MK) and skewness (MS) are consistent with a Gaussian model are performed separately, but there are also so-called omnibus tests that assess them simultaneously. In this paper, we use these two tests and perform them separately.

In the next section, we show that for Gaussian marginal distributions, the BRA yields a Gaussian dependence with corresponding correlation matrix that has maximum determinant.

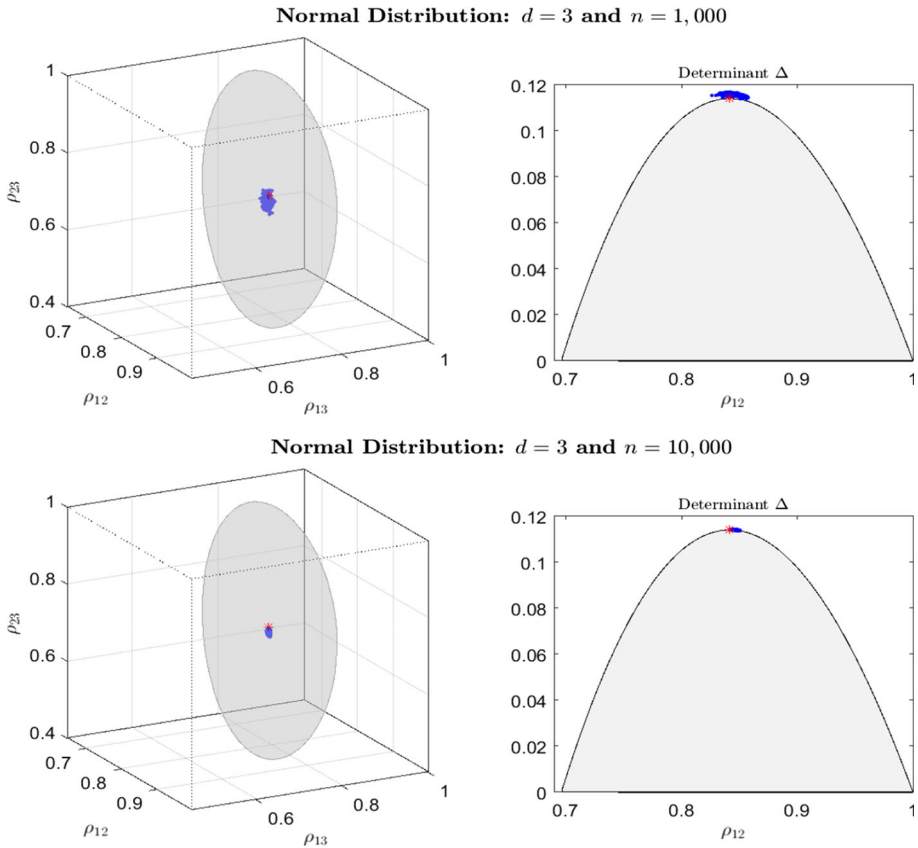


Fig. 2 Normal distribution with $d = 3$ and $n = 1000$. For each case, there are $K = 500$ blue dots corresponding to different runs of the BRA. The shaded gray area represents the constrained set $\mathcal{C}(\rho^{imp})$ and the red star indicates the maximal matrix R_M . The left panel shows $K = 500$ realizations of the correlations ρ_{12} , ρ_{13} , and ρ_{23} . The right panel shows the relation Δ versus ρ_{12} . (Color figure online)

5 Normal distribution

Assume that the components X_i ($i = 1, 2, \dots, d$) and their sum S are all normally distributed, and that they are “compatible” with the distribution of the sum (as per Definition 2.1). To simplify the exposition, all variables are centered, i.e., they have zero mean. The standard deviations $\sigma_1, \dots, \sigma_d$ are assumed to linearly decrease from 1 to $1/d$, whereas the standard deviation of the sum σ_S is chosen in such a way that $\rho^{imp} = 0.8$. The number of components d ranges from 3 to 10. To discretize the marginal distributions, we use either $n = 1000$ or $n = 10,000$. In each experiment, we run the BRA $K = 500$ times to obtain candidate approximate dependencies, as described in Sect. 2.2.

Our first experiment considers the case of $d = 3$ and is illustrated in Fig. 2. The left panel shows the relationships between the three pairwise correlations ρ_{12} , ρ_{13} , and ρ_{23} . On this plane, the gray “ellipse” represents the constrained set $\mathcal{C}(\rho^{imp})$.⁹ See also Fig. 1 for a

⁹ The shape only looks like an ellipse—but it is not an ellipse according to the mathematical definition of an ellipse.

Table 1 Normal distribution

d	$\sqrt[d]{\Delta_M(\rho_{imp})}$	$MPE_1(\sqrt[d]{\Delta})$	$MPE_2(\sqrt[d]{\Delta})$	$100 \times \bar{V}$	$MPE(\varrho)$	$MS(\%)$	$MK(\%)$
$n = 1000$							
3	0.485	0.511	-0.105	8.027	-0.254	0.8	1.2
4	0.430	-0.004	-0.213	3.480	-0.076	0.4	3.0
5	0.398	-0.199	-0.287	1.679	-0.030	0.4	2.6
7	0.360	-0.352	-0.373	0.449	-0.007	0.0	3.8
10	0.332	-0.474	-0.476	0.081	-0.001	2.0	5.6
$n = 10,000$							
3	0.485	0.007	-0.061	4.699	-0.028	0.4	1.8
4	0.430	-0.096	-0.118	1.685	-0.008	0.4	5.6
5	0.398	-0.118	-0.127	0.808	-0.003	1.0	4.2
7	0.360	-0.108	-0.110	0.198	-0.001	2.0	2.4
10	0.332	-0.069	-0.070	0.032	-0.000	2.2	2.6

For each run of BRA $k = 1, \dots, 500$, we obtain $\Delta^{(k)}, \varrho^{(k)}$ and $V^{(k)}$ and use these to compute sample averages. All MPEs are reported as percents. \bar{V} denotes the average of the K values $V^{(k)}$. The last two columns report the results of the two multivariate normality tests, MS and MK. For each test, we report the percent of rejections (with p -value less than 0.05). The table is reproduced for two levels of discretization $n = 1000$ and $n = 10,000$ and when $d = 3, 4, 5, 7$ and 10

graphical illustration of $\mathcal{C}(\rho^{imp})$. All the points in this gray area are candidate solutions and correspond to a global optimum for the variance minimization problem under consideration. With the help of Matlab, we solve the high-dimensional constrained optimization problem to find the correlation matrix $R_M(\rho^{imp}) \in \mathcal{C}(\rho^{imp})$ with maximum determinant $\Delta_M(\rho^{imp})$ (see also Sect. 4) and we represent it with a red star. We then run the BRA 500 times and show the corresponding solutions with blue dots. It is striking to observe how for all 500 runs the candidate solutions obtained with the BRA exhibit correlations which are very close to the ones of the maximum determinant matrix. This conclusion is also confirmed by the right panel which plots the values of $\Delta^{(k)}$ versus $\rho_{12}^{(k)}$. When the level of discretization n increases from 1000 to 10,000, the cluster of the blue dots becomes even tighter and closer to the maximum determinant matrix.

The representation of the range of possible correlation through an “ellipse” is convenient when $d = 3$, but such representation needs to be adapted in higher dimensions. Therefore, in Table 1 we provide several measures that allow to interpret the properties of the output matrix when d increases up to 10 dimensions. In particular, we report the measures of discrepancy $MPE_1(\sqrt[d]{\Delta^{(k)}})$ and $MPE_2(\sqrt[d]{\Delta^{(k)}})$. The last two columns of this table report the results of two statistical tests MS and MK that we performed to test for multivariate normality (see Sect. 4.2).

We make the following observations related to Table 1. First, we observe that the results illustrated in Fig. 2 for $d = 3$ carry over to higher dimensions. Indeed, $MPE_1(\sqrt[d]{\Delta})$ and $MPE_2(\sqrt[d]{\Delta})$ are both very small, which confirms the convergence of the BRA to a candidate dependence that has the correlation matrix with maximum determinant (MPEs are computed by averaging the values displayed in Figs. 4, 5). In addition, the rate of convergence is illustrated on Fig. 3.

Second, we observe that the algorithm converges better when n increases, as measured by the mean of the 500 squared errors $V^{(k)}$. Somewhat surprisingly, the convergence also improves when the dimension d increases. Third, we find that the realized correlation $\varrho^{(k)}$

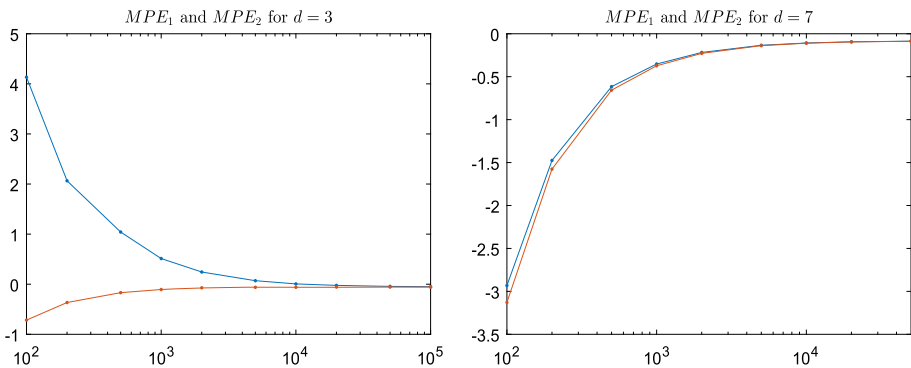


Fig. 3 Normal distribution with $d = 3$ and $d = 7$. For each case, mean pricing errors $MPE_1(\sqrt[d]{\Delta})$ (blue line) and $MPE_2(\sqrt[d]{\Delta})$ (red line) are shown for various value of n . (Color figure online)

on average gets closer and closer to ρ^{imp} when d increases and when n increases. This observation is consistent with Proposition 4.1. Third, consistently with intuition, we observe that $\sqrt[d]{\Delta}$ is much more stable when the discretization level n increases. Finally, for all runs the statistical tests leads to an acceptance of the null-hypothesis of multivariate normality at confidence level of 94%. From Table 1 we can thus conclude that the BRA yields a dependence structure that is Gaussian and with correlation matrix that has maximum determinant.

Figures 4 and 5 hereafter provide additional support for these findings by exploring further the structure of the candidate solutions generated by the BRA when the discretization level $n = 1000$.¹⁰ Specifically, for each case of dimensionality $d = 3, 5, 7$ or 10 , the first three panels plot the measures $PE_1(\sqrt[d]{\Delta^{(k)}})$, $PE_2(\sqrt[d]{\Delta^{(k)}})$ and $V^{(k)}$ with respect to $PE(\varrho^{(k)})$, for each of the 500 runs of the BRA. The last panel shows the distributions of pairwise correlations $\rho_{12}^{(k)}, \rho_{13}^{(k)}, \dots, \rho_{d-1d}^{(k)}$. The plots confirm that the different candidate solutions are all very similar and exhibit close to maximum entropy.

We make additional observations related to Figs. 4 and 5. First, we note that $\varrho^{(k)}$ is always less than ρ^{imp} . Intuitively, to increase the determinant, correlations should be as close to zero as possible. Second, since the constraint 3.3 must be respected, the optimizer increases those correlations $\rho_{ij}^{(k)}$, for which the weights $\sigma_i \sigma_j$ are relatively high, and reduces those correlations, for which these weights are relatively low. This particular pattern for correlations is evident from all figures. Note also that correlations $\rho_{ij}^{(k)}$ that correspond to higher products $\sigma_i \sigma_j$ are recovered by the BRA more accurately. Third, we observe that V and ϱ are negatively correlated. Intuitively, when the BRA converges to an approximate solution that is relatively far from a global solution, ϱ has more “room” to deviate from ρ^{imp} , which translates in a higher mean percentage error relative to ρ^{imp} .

Remark 5.1 Under the probabilistic interpretation of matrices given in Sect. 2, the output of RA and BRA algorithms represent discrete mutually complete dependences (see Lancaster 1963) and in particular so-called multivariate Shuffle-of-Mins (see Durante and Sánchez 2012). Nevertheless, we show that the Gaussian assumption cannot be rejected. This shows once moreagai interplay between discrete optimization method that can only provide discrete results and the fact that any dependence structure can be proxied by a sufficiently fine Shuffle-of-Min.

¹⁰ Note that we have also reproduced the Figs. 4 and 5 for $n = 10,000$ and, as expected, even higher accuracy is achieved. For brevity, these figures are omitted.

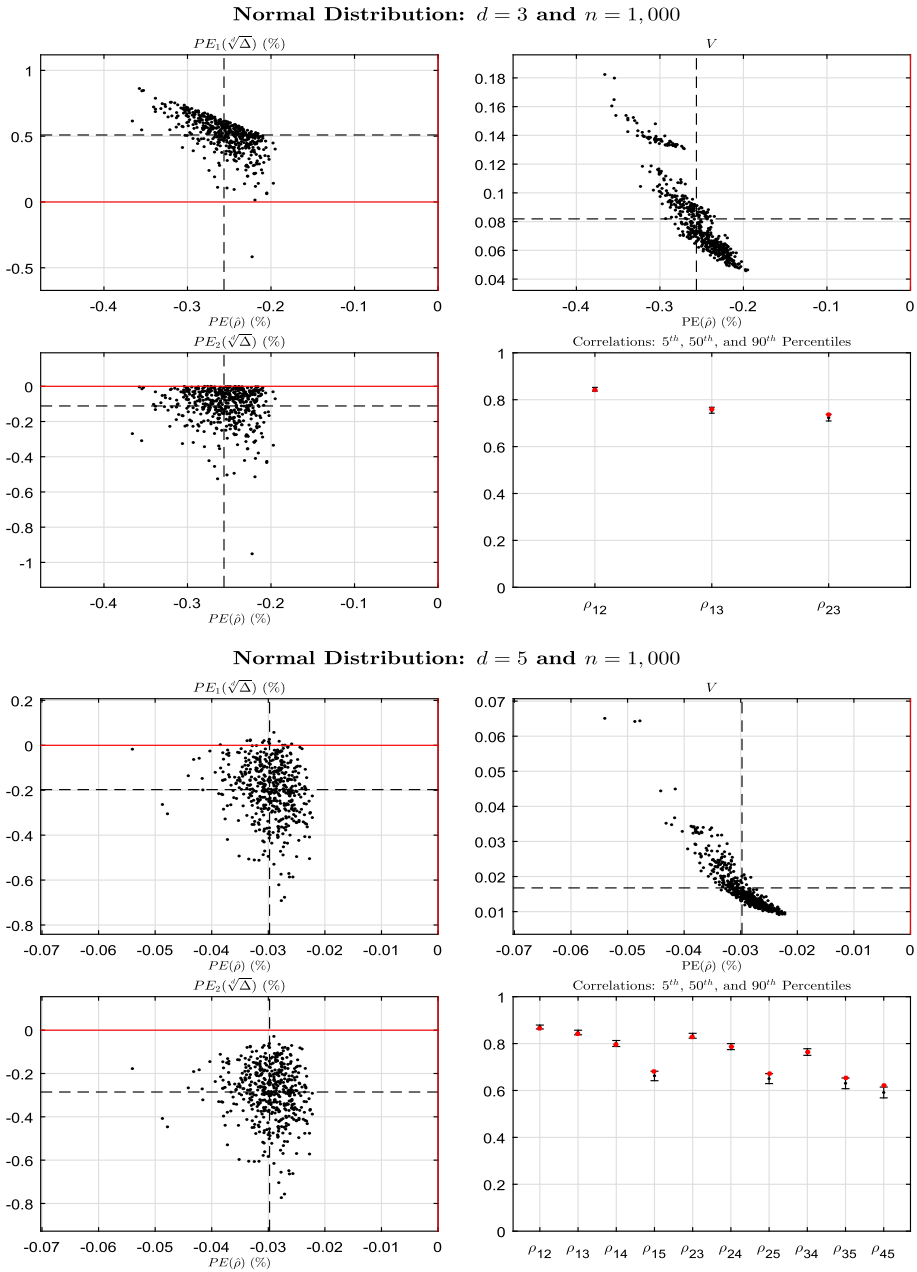


Fig. 4 Normal distribution with $d = 3$ and $d = 5$. Each *black dot* corresponds to one of the $K = 500$ BRA runs. For each case, the first three panels show $PE_1(\sqrt{\Delta}^{(k)})$, $V^{(k)}$, and $PE_2(\sqrt{\Delta}^{(k)})$ with respect to $PE(\varrho^{(k)})$. The *black dashed lines* indicate sample averages. The *right bottom panel* shows the 5th, 50th and 95th percentiles of distributions of correlations $\rho_{ij}^{(k)}$ generated by K runs. The *percentiles* are shown along $\frac{1}{2}d(d - 1)$ vertical lines, corresponding to correlations $\rho_{12}, \rho_{13}, \dots, \rho_{d-1d}$. The *red dots* are the corresponding correlations for the maximal matrix R_M . (Color figure online)

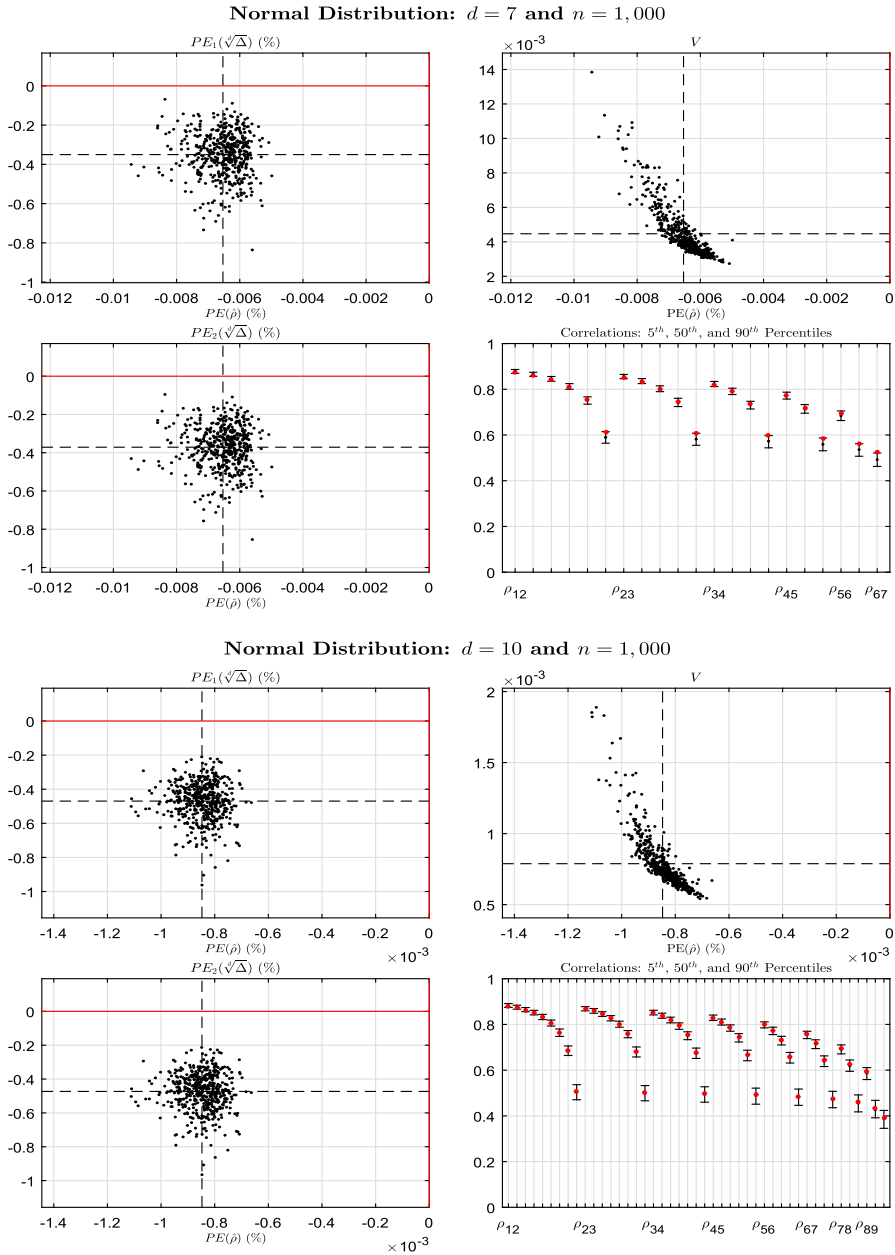


Fig. 5 Normal distribution with $d = 7$ and $d = 10$. Each *black dot* corresponds to one of the $K = 500$ BRA runs. For each case, the first three panels show $PE_1(\sqrt{\Delta}^{(k)})$, $V^{(k)}$, and $PE_2(\sqrt{\Delta}^{(k)})$ with respect to $PE(\rho^{(k)})$. The *black dashed lines* indicate sample averages. The *right bottom panel* shows the 5th, 50th and 95th percentiles of distributions of correlations $\rho_{ij}^{(k)}$ generated by K runs. The *percentiles* are shown along $\frac{1}{2}d(d - 1)$ vertical lines, corresponding to correlations $\rho_{12}, \rho_{13}, \dots, \rho_{d-1d}$. The *red dots* are the corresponding correlations for the maximal matrix R_M . (Color figure online)

6 Robustness

In this section, we study the robustness of our results. First, we investigate the sensitivity to the initial condition. Second, we investigate marginal distributions that are non-Gaussian. Finally, we investigate the impact when using the standard RA instead of the BRA.

6.1 Sensitivity to initial conditions

Recall that when the algorithm starts from a randomized matrix \mathbf{M} for which the row sums are all already equal to zero [as in (2.1)], then the algorithm has converged up-front and the output matrix is the same as the initial one (which is then just a global optimum). In this particular instance, the BRA does not converge to the maximum determinant correlation matrix. However, if we perturb slightly this initial (optimal) matrix, so that the row sums are now not exactly equal to zero, then we observe that the algorithm will converge to the maximum determinant correlation matrix.

This phenomenon is illustrated in Fig. 6, for the level of discretization $n = 1000, 3000,$ and $10,000$. We start with a particular global solution, marked by the green star, but with 0.2% of rows (of the matrix \mathbf{M}) swapped to create a small random perturbation. That is, 2 rows out of 1000, 6 rows out of 3000, and 20 rows out of 10,000 are randomly selected and swapped. Without the perturbation, the BRA would terminate where it starts (at the particular solution, marked by the green star). With the perturbation, however, the BRA typically converges to a solution close the maximum determinant matrix, marked by the red star. The “attraction” to the maximum determinant matrix gets progressively stronger and more striking as n increases. When $n = 1000$, one can see a “path” of blue dots between the green and red stars. When $n = 10,000$, all blue dots are now clustered very tightly around the red star.

6.2 Sensitivity to marginal distributions

When the marginal distributions are not Gaussian, it is not clear that the Gaussian copula is the copula that maximizes entropy. Therefore, we merely assess to which extent the property of maximum determinant for the correlation matrix, which we observed for Gaussian margins, remains valid when other distributional assumptions are made. In order to do so we resort to skewed t-distributions. To this end, we recall that a d -dimensional random vector \mathbf{X} is said to follow a normal mean-variance mixture, if

$$\mathbf{X} \sim \boldsymbol{\mu} + Y\boldsymbol{\gamma} + \sqrt{Y}\mathbf{Z} \quad (6.10)$$

where $\mathbf{Z} \sim N_d(0, \mathbf{W})$ is a d -dimensional multivariate normal with $E[\mathbf{Z}] = 0$ and $\text{var}(\mathbf{Z}) = \mathbf{W}$, where $\mathbf{W} \in \mathbb{R}^{d \times d}$ is positive semi-definite, $Y \geq 0$ is a non-negative scalar random variable independent of \mathbf{Z} , and $\boldsymbol{\gamma} \in \mathbb{R}^d$ and $\boldsymbol{\mu} \in \mathbb{R}^d$ are constants.

Specification (6.10) implies that $\mathbf{X}|_{Y=y} \sim N_d(\boldsymbol{\mu} + y\boldsymbol{\gamma}, y\mathbf{W})$. Furthermore,

$$E[\mathbf{X}] = \boldsymbol{\mu} + E[Y]\boldsymbol{\gamma}, \quad \text{and} \quad \text{cov}(\mathbf{X}) = E[Y]\mathbf{W} + \text{var}(Y)\boldsymbol{\gamma}\boldsymbol{\gamma}^t.$$

The multivariate Generalized Hyperbolic (GH) distribution is obtained when the scalar factor in the mixture Y follows a generalized inverse Gaussian distribution with parameters λ , χ and ψ (Chapter 3.2 in McNeil et al. 2010) and is denoted as $\mathbf{X} \sim GH_d(\lambda, \chi, \psi, \boldsymbol{\mu}, \mathbf{W}, \boldsymbol{\gamma})$.

A useful property of the GH family is that it is closed under linear operations, meaning, in particular, that $S = \sum_i X_i$ has also a GH distribution. In this paper, we are interested in a special case of the GH family where $\lambda = -\nu/2$, $\chi = \nu$ and $\psi = 0$, so that Y follows an inverse Gamma distribution, $Y \sim IG(\nu/2, \nu/2)$. This special case corresponds to a Skewed- t

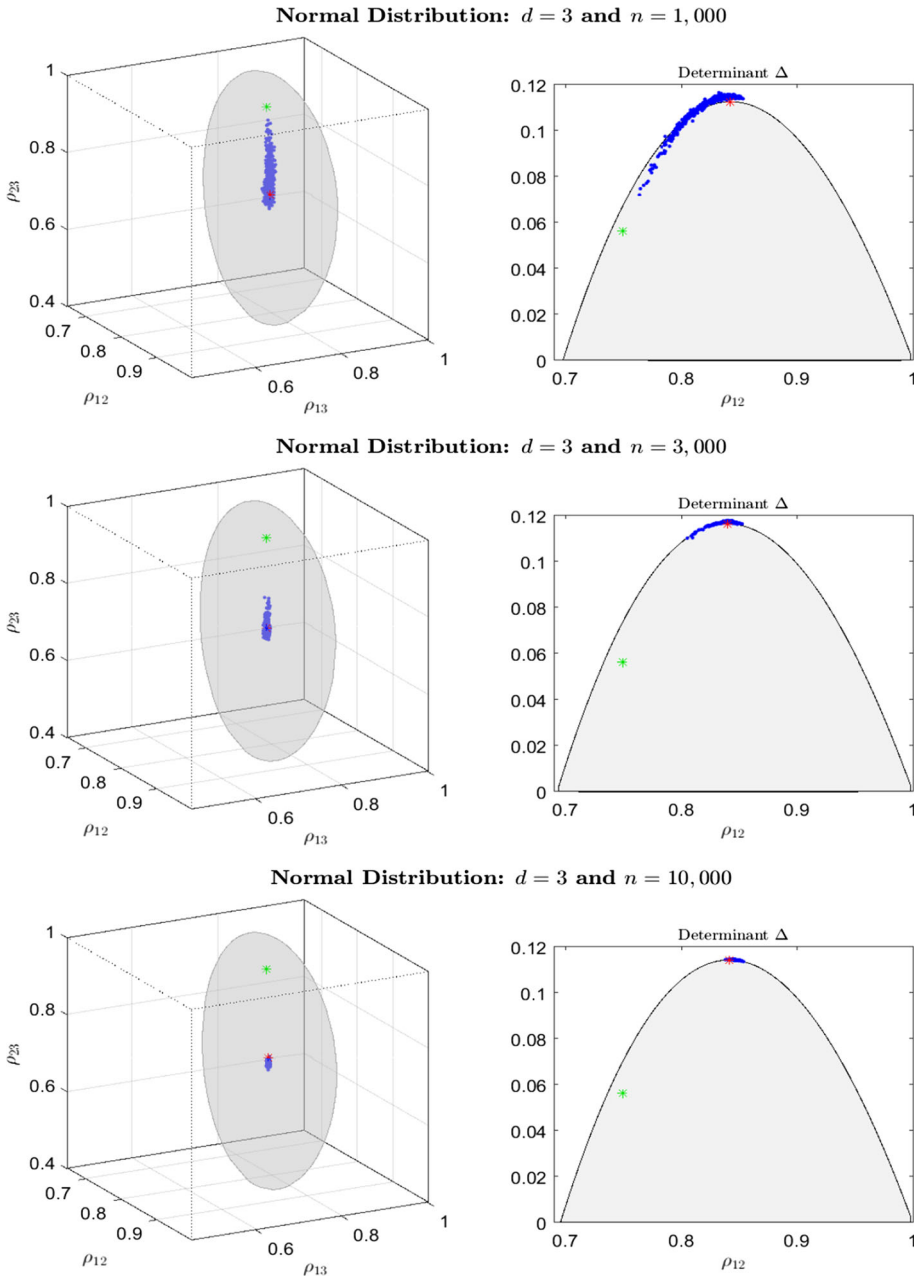


Fig. 6 Normal distribution with $d = 3$ and $n = 1000, 3000$, or $10,000$. For each case, there are $K = 500$ blue dots corresponding to different runs of BRA. Each run of BRA starts at a particular solution (marked by the green star), but with $2/6/20$ random rows swapped when $n = 1000/3000/10,000$. The shaded gray area represents the constrained set $C(\rho^{imp})$ and the red star indicates the maximal matrix R_M . The left panel shows realizations of the correlations ρ_{12} , ρ_{13} , and ρ_{23} . The right panel shows the relation Δ versus ρ_{12} . (Color figure online)

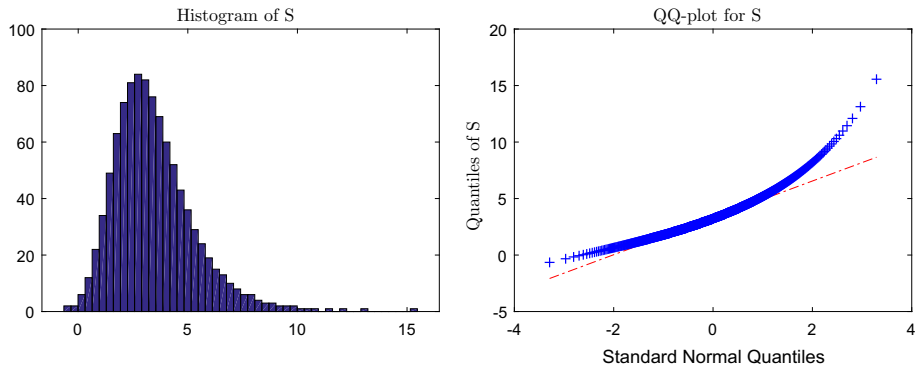


Fig. 7 Histogram and QQ-plot for sum S generated with multivariate Skewed- t distribution when $d = 3$ and $n = 1000$

distribution, for which $\mathbf{X} \sim Skew_d(\nu, \boldsymbol{\mu}, \mathbf{W}, \boldsymbol{\gamma})$ and

$$\text{cov}(\mathbf{X}) = \frac{\nu}{\nu - 2} \mathbf{W} + \frac{2\nu^2}{(\nu - 2)^2(\nu - 4)} \boldsymbol{\gamma} \boldsymbol{\gamma}^t, \tag{6.11}$$

where the covariance matrix is defined as long as $\nu > 4$. The sum S as well as the components X_i ($i = 1, 2, \dots, d$) now follow one-dimensional Skewed- t distribution. In particular,

$$S \sim Skew_1 \left(\nu, \sum_i \mu_i, \mathbf{1W1}^t, \sum_i \gamma_i \right).$$

We simulate from the multivariate Skewed- t distribution to obtain realizations for the the components X_i ($i = 1, 2, \dots, d$) and their sum S . Next, we use this information to construct the matrix \mathbf{M} on which we repeatedly apply the BRA. As before, we assume that the number of components d ranges from 3 to 10 and the level of discretization is either $n = 1000$ or

Table 2 Skewed- t distribution

d	$\sqrt[d]{\Delta M(\rho_{imp})}$	$MPE_1(\sqrt[d]{\Delta})$	$MPE_2(\sqrt[d]{\Delta})$	100Mean (V)	$MPE(\varrho)$
$n = 1000$					
3	0.485	1.907	-0.115	26.241	-0.840
4	0.430	0.828	-0.212	11.875	-0.378
5	0.398	0.630	-0.282	9.034	-0.308
7	0.360	0.184	-0.379	2.931	-0.177
10	0.332	0.215	-0.517	0.636	-0.220
$n = 10,000$					
3	0.485	0.390	-0.048	33.098	-0.181
4	0.430	0.131	-0.093	18.495	-0.081
5	0.398	0.087	-0.102	14.457	-0.064
7	0.360	0.035	-0.085	3.956	-0.037
10	0.332	0.087	-0.056	1.189	-0.043

For each run of BRA $k = 1, \dots, K$, we use $\Delta^{(k)}$, $\rho^{(k)}$, $\xi^{(k)}$, and $V^{(k)}$ to compute sample averages. MPEs are in percents. The table is reproduced for two levels of discretization $n = 1000$ and $n = 10,000$ and when $d = 3, 4, 5, 7$ and 10

$n = 10,000$. For the ease of comparison, we set the parameters of the Skewed- t distribution so that the covariance matrix $\text{cov}(\mathbf{X})$ is the same as in the normal case. Specifically, as in the normal case, the standard deviations $\sigma_1, \dots, \sigma_d$ of the marginals linearly decrease from 1 to $1/d$ and all pairwise correlations are equal to $\rho^{imp} = 0.8$. For each component X_i , the location parameter is zero, $\mu_i = 0$, and the skewness parameter is proportional to σ_i , that is, $\gamma_i = c\sigma_i = 1.5\sigma_i$. We set $\nu = 15$ and use the relation in (6.11) to deduce the covariance matrix \mathbf{W} for the normal factors \mathbf{Z} . Our choice of the parameters guarantees that \mathbf{W} is positive semi-definite.

Figure 7 illustrates the shape of the Skewed- t distribution used in the simulations. It plots the histogram and QQ-plot for the sum S when $d = 3$ and $n = 1000$. Table 2 reports the results for $K = 500$ runs of the BRA when d ranges from 3 to 10 and for two levels of discretization, $n = 1000$ and $n = 10,000$.

In Fig. 8, we analyze the case $d = 3$. As before, the “ellipse” in the left panel represents the constrained set $\mathcal{C}(\rho^{imp})$. For all 500 runs, the candidate solutions exhibit correlations that match closely the ones of the maximum determinant matrix. This conclusion only strengthens when the level of discretization increases from 1000 to 10,000. The candidate solutions

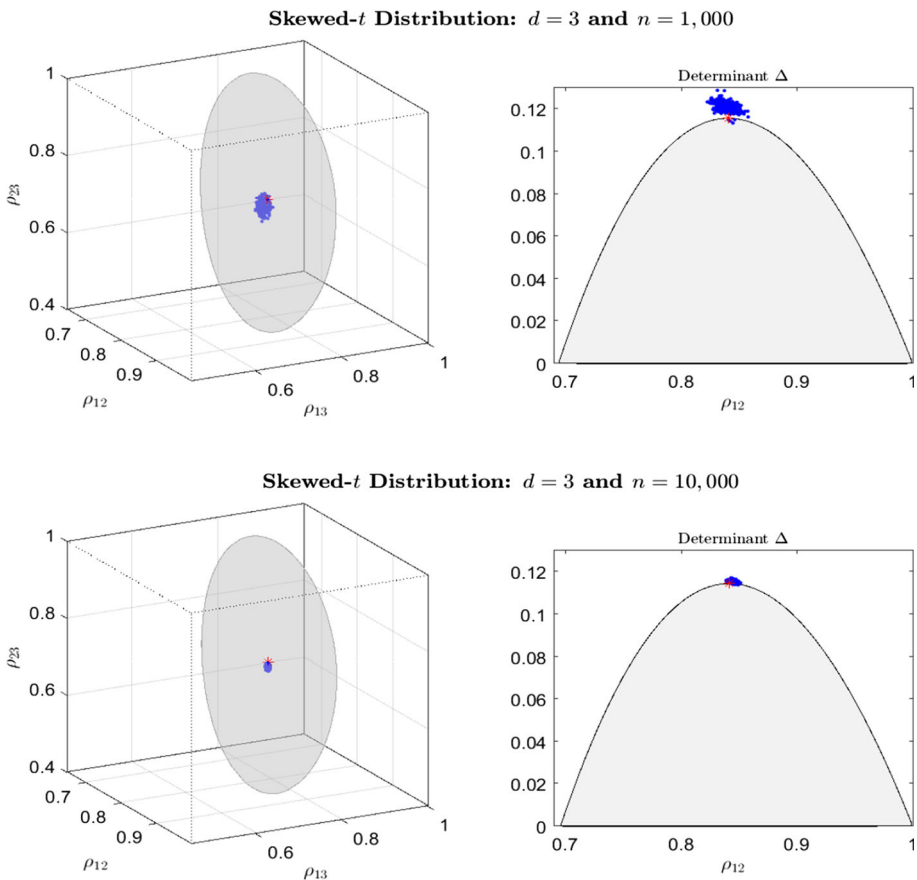


Fig. 8 Skewed- t distribution with $d = 3$ and $n = 1000$ or $n = 10,000$. For each case, there are $K = 500$ blue dots corresponding to different runs of BRA. The shaded gray area represents the constrained set $\mathcal{C}(\rho^{imp})$ and the red star indicates the maximal matrix R_M . The left panel shows $K = 500$ realizations of the correlations ρ_{12} , ρ_{13} , and ρ_{23} . The right panel shows the relation Δ versus ρ_{12} . (Color figure online)

generated by the BRA exhibit more variation compared to the normal distribution. This is to be expected, as components X_i now have considerable skewness and heavier tails. As a result, outliers are more frequent and the overall precision of the BRA declines noticeably. Nevertheless, the solutions remain tightly clustered around the one corresponding to the maximum determinant.

Two additional figures in “Appendix”, Figs. 11 and 12, provide further insights into the structure of the output for $n = 1000$. For each case of $d = 3, 5, 7$ or 10 , the first three panels plot the measures $PE_1(\sqrt[d]{\Delta^{(k)}})$, $PE_2(\sqrt[d]{\Delta^{(k)}})$, and $V^{(k)}$ with respect to $PE(\varrho^{(k)})$, for each of the 500 runs of the BRA. The last panel shows the distributions of pairwise correlations $\rho_{12}^{(k)}, \rho_{13}^{(k)}, \dots, \rho_{(d-1)d}^{(k)}$. The plots confirm that the candidate solutions are all similar and attain close to maximum determinant.

6.3 Sensitivity to rearrangement algorithm

Finally, we assess to which extent our main findings change when using the standard RA, which is described in Sect. 2. Specifically, we follow the same Gaussian setup as in Sect. 5,

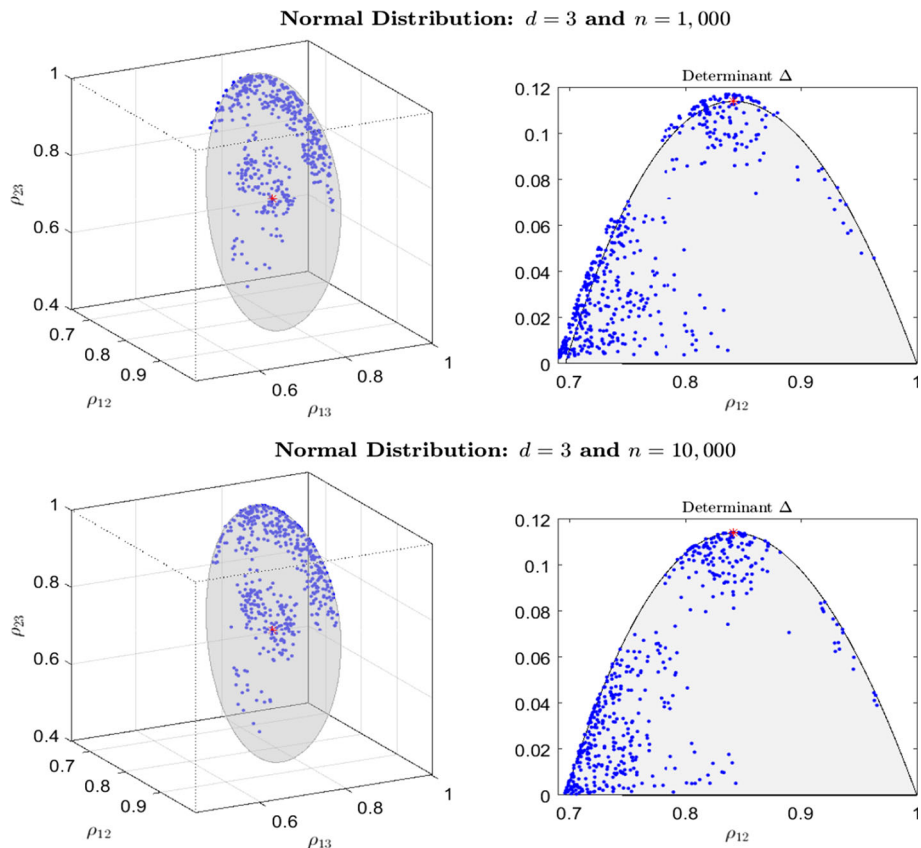


Fig. 9 Normal distribution with $d = 3$ and $n = 1000$ or $n = 10,000$, using the standard RA instead of the BRA. As before, there are $K = 500$ blue dots corresponding to different runs of the RA. The shaded gray area represents the constrained set $C(\rho^{imp})$ and the red star indicates the maximal matrix R_M . The left panel shows $K = 500$ realizations of the correlations ρ_{12}, ρ_{13} , and ρ_{23} . The right panel shows the relation Δ versus ρ_{12} . (Color figure online)

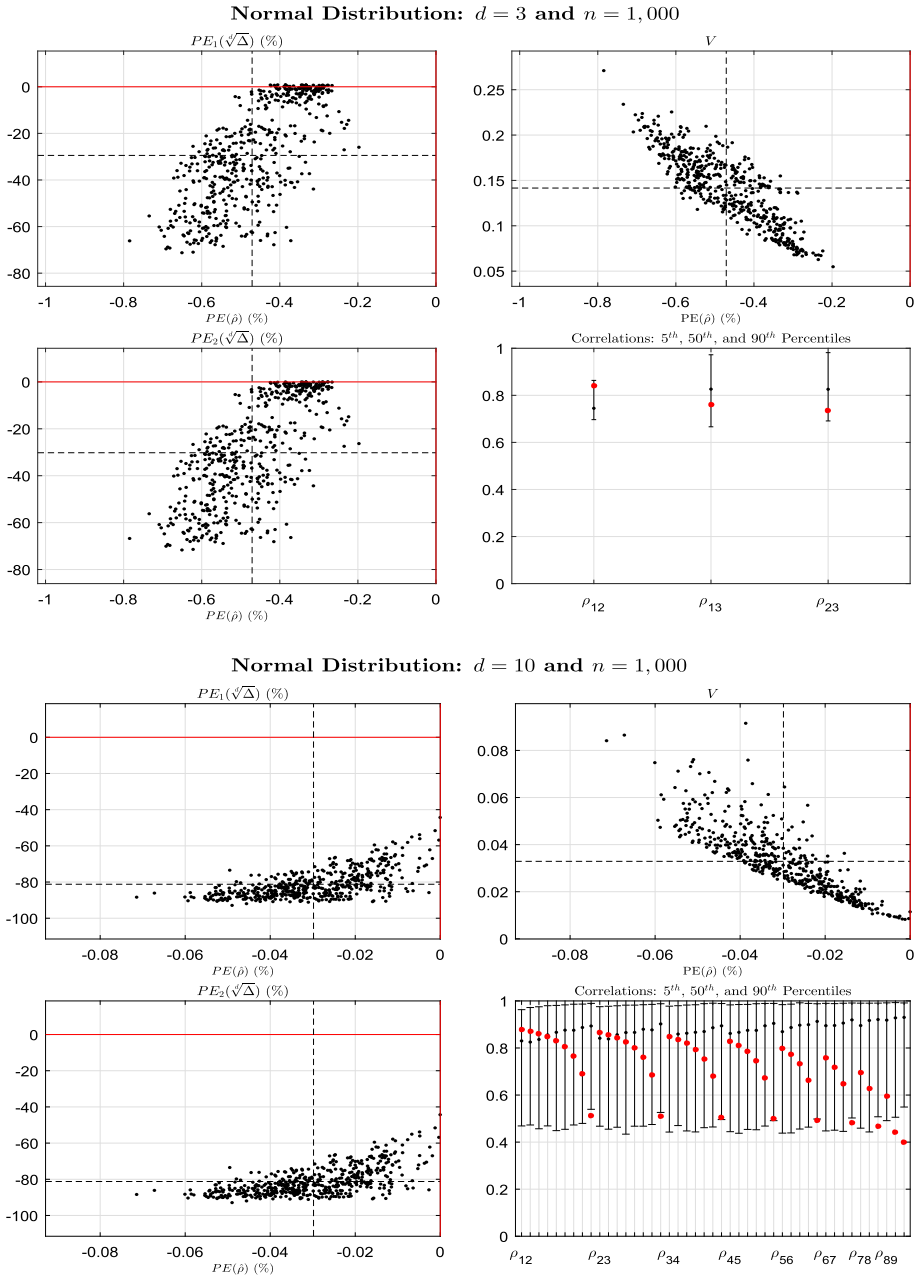


Fig. 10 Normal distribution with $d = 3$ and $d = 10$, using the standard RA instead of the BRA. Each black dot corresponds to one of the $K = 500$ BRA runs. For each case, the first three panels show $PE_1(\sqrt{\Delta^{(k)}})$, $V^{(k)}$, and $PE_2(\sqrt{\Delta^{(k)}})$ with respect to $PE(\rho^{(k)})$. The black dashed lines indicate sample averages. The right bottom panel shows the 5th, 50th and 95th percentiles of distributions of correlations $\rho_{ij}^{(k)}$ generated by K runs. The percentiles are shown along $\frac{1}{2}d(d-1)$ vertical lines, corresponding to correlations $\rho_{12}, \rho_{13}, \dots, \rho_{d-1d}$. The red dots are the corresponding correlations for the maximal matrix R_M . (Color figure online)

Table 3 Normal distribution, using the standard RA instead of the BRA

d	$\sqrt[d]{\Delta_M(\rho_{imp})}$	$MPE_1(\sqrt[d]{\Delta})$	$MPE_2(\sqrt[d]{\Delta})$	$100 \times \bar{V}$	$MPE(\varrho)$	$MS(\%)$	$MK(\%)$
$n = 1000$							
3	0.485	-29.458	-30.210	14.158	-0.471	15.2	38.6
4	0.430	-50.694	-50.996	10.894	-0.244	37.8	64.4
5	0.398	-59.430	-59.587	8.389	-0.144	47.8	78.0
7	0.360	-72.132	-72.185	5.334	-0.066	75.8	92.0
10	0.332	-81.216	-81.231	3.290	-0.030	89.8	97.4
$n = 10,000$							
3	0.485	-31.255	-31.334	7.885	-0.051	11.2	28.2
4	0.430	-50.353	-50.383	4.902	-0.024	25.4	51.2
5	0.398	-62.438	-62.452	3.911	-0.014	39.0	68.2
7	0.360	-75.895	-75.899	2.272	-0.007	65.8	86.2
10	0.332	-85.108	-85.108	1.347	-0.003	87.2	96.6

For each run of RA $k = 1, \dots, 500$, we obtain $\Delta^{(k)}, \varrho^{(k)}$ and $V^{(k)}$ and use these to compute sample averages. All MPEs are reported as percents. The last two columns report the results of the two multivariate normality tests, MS and MK. For each test, shown are the the percent of rejections (with p -value less than 0.05). The table is reproduced for two levels of discretization $n = 1000$ and $n = 10,000$ and when $d = 3, 4, 5, 7$ and 10

except now we use the standard RA to infer the dependence among the variables instead of the BRA. We report the results in Figs. 9, 10, and Table 3.

It is clear that the standard RA does not recover the dependence with maximum entropy. From Fig. 10, we observe that the 500 experiments lead to correlation matrices that may have determinants very far from the maximum determinant matrix. Figure 10 illustrates the magnitude of the distance to the maximum determinant matrix and shows that pairwise correlation can be in a very wide range (fourth panel for $d = 3$ and eighth panel for $d = 10$) compared to the results found in Figs. 4 and 5 using the BRA. Furthermore, one observes from Table 3 that both multivariate Gaussian tests fail so that the hypothesis that the dependence is a Gaussian copula is now rejected.

7 Conclusion

We perform an extensive study of the Block Rearrangement Algorithm (BRA) when used to infer dependence structures among variables that are consistent with information on the marginal distributions of the variables and the distribution of their sum. We find that when the variables are normally distributed the BRA always yields a very similar solution which moreover has a property of maximum entropy. This stylized fact is remarkable as we would expect a wide range of compatible dependence structures to be attained given that we start from a random initial structure and here we merely obtain just one dependence structure exhibiting maximum entropy. These findings are robust in that for non-normal distributions, BRA still yields a unique dependence that exhibits a correlation matrix with maximum determinant. Hence, we obtain strong support for using the BRA as a stable algorithm to infer an economically meaningful dependence. Providing a formal proof for the property of maximum entropy seems very challenging (as we do not observe a systematic increase of entropy at each step of the algorithm) and is left for future research.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix: Additional figures

See Figs. 11 and 12.

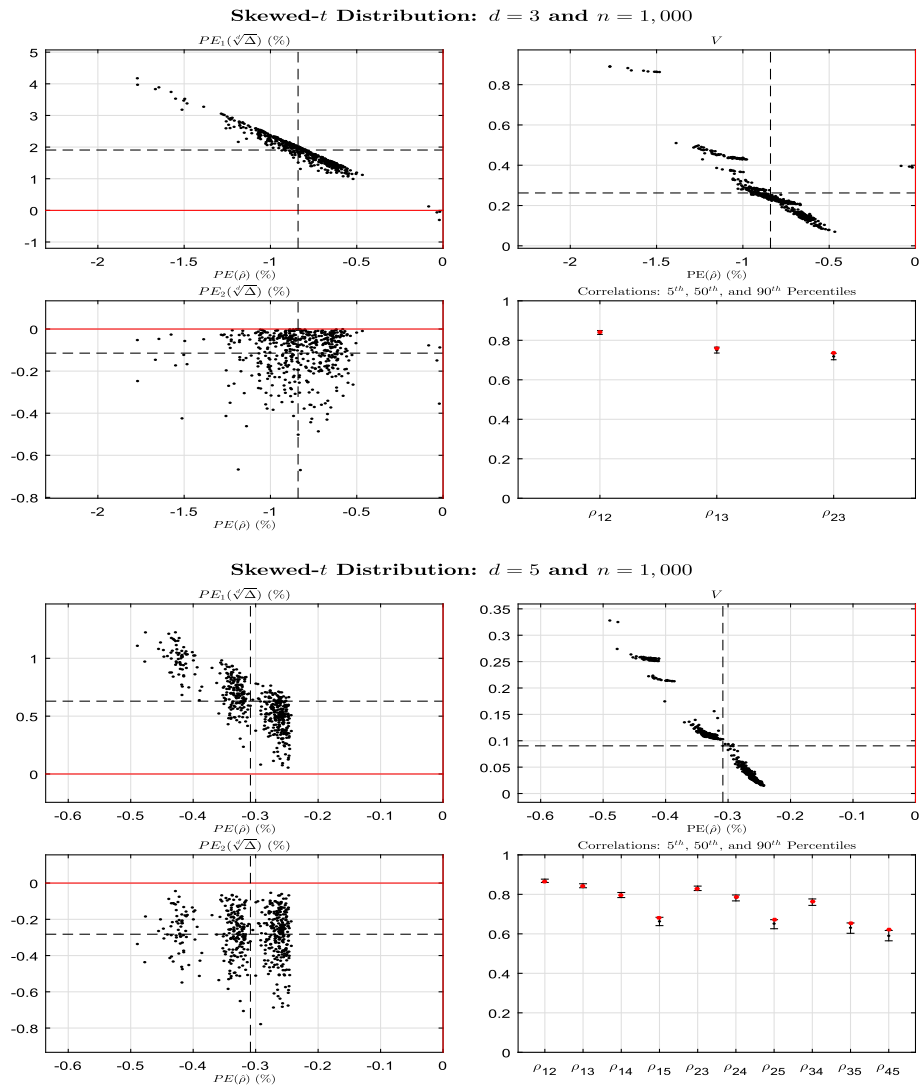


Fig. 11 Skewed-*t* distribution with $d = 3$ and $d = 5$. Each black dot represents one of $K = 500$ BRA runs. For each case, the first three panels show $PE_1(\sqrt{d/\Delta^{(k)}})$, $V^{(k)}$, and $PE_2(\sqrt{d/\Delta^{(k)}})$ with respect to $PE(\varrho^{(k)})$. The black dashed lines indicate sample averages. The right bottom panel shows the 5th, 50th and 95th percentiles of distributions of correlations $\rho_{ij}^{(k)}$ generated by K runs. The percentiles are shown along $\frac{1}{2}d(d-1)$ vertical lines, corresponding to correlations $\rho_{12}, \rho_{13}, \dots, \rho_{d-1d}$. The red dots are the corresponding correlations for the maximal matrix R_M . (Color figure online)

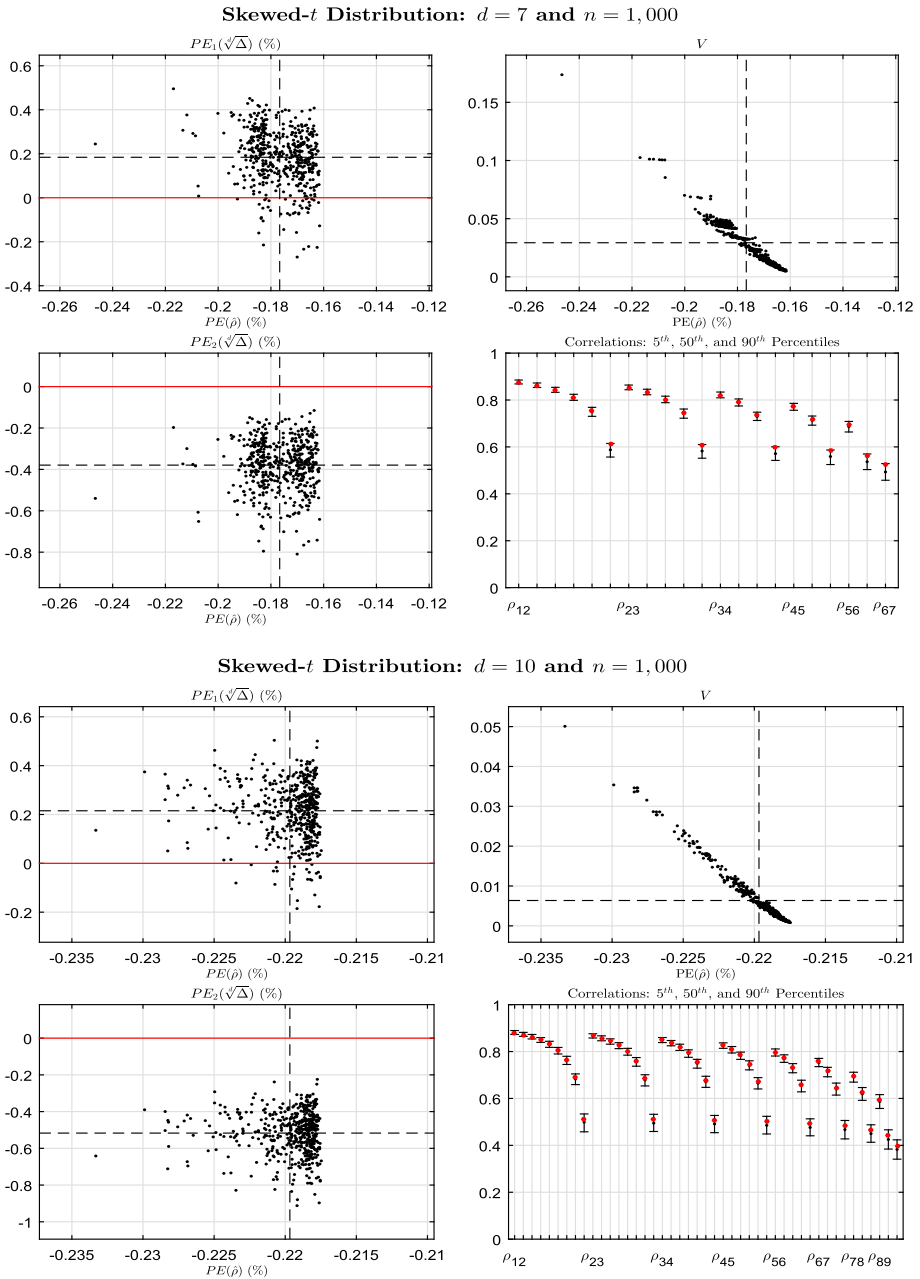


Fig. 12 Skewed-*t* distribution with $d = 7$ and $d = 10$. Each *black dot* represents one of $K = 500$ BRA runs. For each case, the first three panels show $PE_1(\sqrt{d\Delta}^{(k)})$, $V^{(k)}$, and $PE_2(\sqrt{d\Delta}^{(k)})$ with respect to $PE(\rho^{(k)})$. The *black dashed lines* indicate sample averages. The *right bottom panel* shows the 5th, 50th and 95th percentiles of distributions of correlations $\rho_{ij}^{(k)}$ generated by K runs. The *percentiles* are shown along $\frac{1}{2}d(d - 1)$ vertical lines, corresponding to correlations $\rho_{12}, \rho_{13}, \dots, \rho_{d-1d}$. The *red dots* are the corresponding correlations for the maximal matrix R_M . (Color figure online)

References

- Aas, K., & Puccetti, G. (2014). Bounds on total economic capital: The DNB case study. *Extremes*, 17(4), 693–715.
- Ait-Sahalia, Y., & Lo, A. W. (2000). Nonparametric risk management and implied risk aversion. *Journal of Econometrics*, 94(1), 9–51.
- Banz, R., & Miller, M. (1978). Prices for state-contingent claims: Some estimates and applications. *Journal of Business*, 51, 653–672.
- Bernard, C., & McLeish, D. (2016). Algorithms for finding copulas minimizing convex functions of sums. *Asia-Pacific Journal of Operational Research*, 33, 1650040–26.
- Bernard, C., Rüschendorf, L., & Vanduffel, S. (2017a). Value-at-risk bounds with variance constraints. *Journal of Risk and Insurance*, 84(2), 923–959.
- Bernard, C., Rüschendorf, L., Vanduffel, S., & Yao, J. (2017b). How robust is the value-at-risk of credit risk portfolios? *The European Journal of Finance*, 23(6), 507–534.
- Bernard, C., & Vanduffel, S. (2015). A new approach to assessing model risk in high dimensions. *Journal of Banking and Finance*, 58, 166–178.
- Bondarenko, O. (2003). Estimation of risk-neutral densities using positive convolution approximation. *Journal of Econometrics*, 116(1), 85–112.
- Boudt, K., Jakobsons, E., & Vanduffel, S. (2017). Block rearranging elements within matrix columns to minimize the variability of the row sums. *4OR: A Quarterly Journal of Operations Research* doi:10.1007/s10288-017-0344-4.
- Boyd, S., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge: Cambridge University Press.
- Breedon, D. T., & Litzenberger, R. H. (1978). Prices of state-contingent claims implicit in option prices. *Journal of Business*, 51(4), 621–651.
- Coffman, E., & Yannakakis, M. (1984). Permuting elements within columns of a matrix in order to minimize maximum row sum. *Mathematics of Operations Research*, 9(3), 384–390.
- Durante, F., & Sánchez, J. F. (2012). On the approximation of copulas via shuffles of Min. *Statistics & Probability Letters*, 82(10), 1761–1767.
- Embrechts, P., Puccetti, G., & Rüschendorf, L. (2013). Model uncertainty and VaR aggregation. *Journal of Banking & Finance*, 37(8), 2750–2764.
- Gaffke, N., & Rüschendorf, L. (1981). On a class of extremal problems in statistics. *Optimization*, 12(1), 123–135.
- Haus, U.-U. (2015). Bounding stochastic dependence, complete mixability of matrices, and multidimensional bottleneck assignment problems. *Operations Research Letters*, 43, 74–79.
- Jackwerth, J. C., & Rubinstein, M. (1996). Recovering probability distributions from option prices. *Journal of Finance*, 51(5), 1611–1631.
- Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical Review*, 106(4), 620.
- Lancaster, H. (1963). Correlation and complete dependence of random variables. *The Annals of Mathematical Statistics*, 34(4), 1315–1321.
- Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57(3), 519–530.
- Mardia, K. V. (1974). Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhyā: The Indian Journal of Statistics, Series B*, 51(5), 115–128.
- Mardia, K. V. (1975). Assessment of multinormality and the robustness of Hotelling’s T2 test. *Applied Statistics*, 24(2), 163–171.
- Mardia, K. V., Kent, J. T., & Bibby, J. M. (1980). *Multivariate analysis*. London: Academic Press.
- McNeil, A. J., Frey, R., & Embrechts, P. (2005). *Quantitative risk management: Concepts, techniques, and tools*. Princeton: Princeton University Press.
- McNeil, A. J., Frey, R., & Embrechts, P. (2010). *Quantitative risk management: Concepts, techniques, and tools*. Princeton: Princeton University Press.
- Puccetti, G., & Rüschendorf, L. (2012). Computation of sharp bounds on the distribution of a function of dependent risks. *Journal of Computational and Applied Mathematics*, 236(7), 1833–1840.
- Puccetti, G., & Wang, R. (2015). Detecting complete and joint mixability. *Journal of Computational and Applied Mathematics*, 280, 174–187.
- Ross, S. (1976). Options and efficiency. *Quarterly Journal of Economic*, 90, 75–89.
- Rubinstein, M. (1994). Implied binomial trees. *The Journal of Finance*, 49(3), 771–818.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379–423.
- Stutzer, M. (1996). A simple nonparametric approach to derivative security valuation. *Journal of Finance*, 51, 1633–1652.

- Thomas, J. A., & Cover, T. M. (2006). *Elements of information theory*. New York: Wiley.
- Wang, B., & Wang, R. (2011). The complete mixability and convex minimization problems with monotone marginal densities. *Journal of Multivariate Analysis*, *102*(10), 1344–1360.
- Wang, B., & Wang, R. (2016). Joint mixability. *Mathematics of Operations Research*, *41*, 808–826.