Submitted to *Operations Research*
manuscript OPRE-2013-05-231

Robustness to Dependency in Portfolio Optimization using Overlapping Marginals

Xuan Vinh Doan
DIMAP and ORMS Group, Warwick Business School, University of Warwick, Coventry, CV4 7AL, United Kingdom, xuan.doan@wbs.ac.uk

Xiaobo Li
Industrial and Systems Engineering, University of Minnesota, Minneapolis, MN 55455, lixx3195@umn.edu

Karthik Natarajan
Engineering Systems and Design, Singapore University of Technology and Design, 8 Somapah Road, Singapore 487372, natarajan_karthik@sutd.edu.sg

In this paper, we develop a distributionally robust portfolio optimization model where the robustness is across different dependency structures among the random losses. For a Fréchet class of discrete distributions with overlapping marginals, we show that the distributionally robust portfolio optimization problem is efficiently solvable with linear programming. To guarantee the existence of a joint multivariate distribution consistent with the overlapping marginal information, we make use of a graph theoretic property known as the running intersection property. Building on this property, we develop a tight linear programming formulation to find the optimal portfolio that minimizes the worst-case Conditional Value-at-Risk measure. Lastly, we use a data-driven approach with financial return data to identify the Fréchet class of distributions satisfying the running intersection property and then optimize the portfolio over this class of distributions. Numerical results in two different datasets show that the distributionally robust portfolio optimization model improves on the sample-based approach

Key words: distributionally robust optimization, portfolio optimization, overlapping marginals

History: This paper was first submitted in May 2013, revised twice in April 2014 and December 2014, and finally accepted in August 2015.

1. Introduction

Optimization under uncertainty is an active research area with several interesting applications in the area of risk management. An example of a risk management problem is to choose a portfolio of assets with random returns such that the joint portfolio risk is minimized while a fixed level of expected return is guaranteed. Markowitz (1952) was the first to model this problem using variance as the risk measure. Several alternative risk measures have been proposed since for this problem. Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR) are two such popular risk measures (see, for example, Jorion (2001) and Rockafellar and Uryasev (2000)). However even assuming that the joint distribution of the random returns is known, the calculation of VaR and CVaR for a given portfolio involves multidimensional integrals, which can be computationally challenging. For discrete distributions, the computation of these risk measures requires the consideration of all the support points of the joint distribution. The number of support points of the joint distribution can however be exponentially large in comparison to the number of support points of the marginal distributions. For example, the problem of computing the probability that a sum of independent but not necessarily identically distributed integer random variables is less than a given number is known to be #P-hard (see Kleinberg et al. (2000)). Furthermore, if the assumed joint distribution does not match the actual distribution, the optimal portfolio allocation might perform poorly in out of sample realizations.

One popular approach to address this issue is that instead of assuming a complete joint distribution for the random returns of the risky assets, only reliable partial distributional information is used. Given limited distributional information, it is then natural to calculate the worst case bounds for the VaR and CVaR measures and optimize the worst case bounds. Several models have been proposed to capture the uncertainty (or ambiguity) in distributions. This includes the class of distributions with information on the first and second moments (see El Ghaoui et al. (2003), Delage and Ye (2010), Natarajan et al. (2009a), Zymler et al. (2013)), the Fréchet class of distributions with information on the univariate marginal distributions (see Meilijson and Nadas

(1979), Denuit et al. (1999)) and the Fréchet class of distributions with information on the multivariate marginals of non-overlapping subsets of asset returns (see Doan and Natarajan (2012), Garlappi et al. (2007), Rüschendorf (1991)). In this paper, we adopt a more general representation of distributional uncertainty where the multivariate marginals possibly overlap with each other.

1.1. Problem Setup

Throughout this paper, we use standard letters such as x to denote scalars, bold letters such as $\boldsymbol{x}$ to denote vectors, tilde notation such as $\tilde{c}$ to denote random variables and the calligraphic notation such as $\mathcal{C}$ to denote sets with $C = \text{size}(\mathcal{C})$.

Let $\tilde{\boldsymbol{c}}$ be a N-dimensional random vector where $\mathcal{N} = \{1, \ldots, N\}$ is the set of indices of the random vector. Consider a convex piecewise linear function of the random vector defined as:

$$\varphi(\tilde{\boldsymbol{c}}) \triangleq \max_{j \in \mathcal{M}} \left(\tilde{\boldsymbol{c}}^T \boldsymbol{a}_j + b_j \right), \tag{1}$$

where the maximum is over the set of affine functions of the random vector indexed by $\mathcal{M} = \{1, \ldots, M\}$ with $\boldsymbol{a}_j \in \Re^N$, $b_j \in \Re$ and $\tilde{\boldsymbol{c}}^T \boldsymbol{a}_j = \sum_{i \in \mathcal{N}} \tilde{c}_i a_{j_i}$. Let θ denote the N-dimensional joint distribution of $\tilde{\boldsymbol{c}}$. Associated with (1), is the evaluation of its expected value:

$$\mathbb{E}_\theta \left[\max_{j \in \mathcal{M}} \left(\tilde{\boldsymbol{c}}^T \boldsymbol{a}_j + b_j \right) \right], \tag{2}$$

and a stochastic optimization problem of the form:

$$\min_{\boldsymbol{x} \in \mathcal{X}} \mathbb{E}_\theta \left[\max_{j \in \mathcal{M}} \left(\tilde{\boldsymbol{c}}^T \boldsymbol{a}_j(\boldsymbol{x}) + b_j(\boldsymbol{x}) \right) \right], \tag{3}$$

where $\boldsymbol{a}_j(\boldsymbol{x})$ and $b_j(\boldsymbol{x})$ are assumed to be affine functions of the decision vector $\boldsymbol{x}$ that is chosen from a feasible region $\mathcal{X}$. Portfolio optimization with the CVaR measure lies within the scope of the stochastic optimization problem (3). To see this, let $\tilde{c}_i$ denote the random loss of the ith asset. The total portfolio loss is then $\tilde{\boldsymbol{c}}^T \boldsymbol{x}$ where x_i is the allocation in the ith asset. The CVaR of the portfolio for a given risk level $\alpha \in (0, 1)$ is defined as (see Rockafellar and Uryasev (2000, 2002)):

$$\text{CVaR}_\alpha^\theta \left(\tilde{\boldsymbol{c}}^T \boldsymbol{x} \right) \triangleq \min_{\beta \in \Re} \left(\beta + \frac{1}{1-\alpha} \mathbb{E}_\theta \left[\left(\tilde{\boldsymbol{c}}^T \boldsymbol{x} - \beta \right)^+ \right] \right),$$

where $y^+ = \max\{0, y\}$. The problem of finding the portfolio $\boldsymbol{x} \in \mathcal{X}$ that minimizes the CVaR is then formulated as:

$$\min_{\boldsymbol{x}\in\mathcal{X},\beta\in\Re} \left(\beta + \frac{1}{1-\alpha}\mathbb{E}_\theta\left[\left(\tilde{\boldsymbol{c}}^T\boldsymbol{x} - \beta\right)^+\right]\right).$$

Portfolio optimization with the VaR measure is however not an instance of problem (3) due to the inherent non-convexity of VaR. It is possible though to develop convex approximations to the VaR measure using the CVaR measure (see Nemirovski and Shapiro (2006)).

In the distributionally robust optimization setting, θ is not known exactly except that it lies in a set of distributions denoted by Θ. Then, it is natural to calculate upper and lower bounds on the expected value $\mathbb{E}_\theta\left[\varphi(\tilde{\boldsymbol{c}})\right]$. In this paper, we restrict our attention to finding the tightest upper bound on the expected value of the piecewise linear convex function in (2):

$$\sup_{\theta\in\Theta} \mathbb{E}_\theta\left[\max_{j\in\mathcal{M}}\left(\tilde{\boldsymbol{c}}^T\boldsymbol{a}_j + b_j\right)\right]. \tag{4}$$

The corresponding distributionally robust counterpart of (3) is:

$$\min_{\boldsymbol{x}\in\mathcal{X}} \sup_{\theta\in\Theta} \mathbb{E}_\theta\left[\max_{j\in\mathcal{M}}\left(\tilde{\boldsymbol{c}}^T\boldsymbol{a}_j(\boldsymbol{x}) + b_j(\boldsymbol{x})\right)\right], \tag{5}$$

where the optimal decision $\boldsymbol{x} \in \mathcal{X}$ is identified for the worst-case distribution in the set Θ.

1.2. Fréchet Class of Distributions

One of the early results in multivariate bounds is the work of Fréchet (1940), Fréchet (1951) who evaluated bounds on the cumulative distribution function of a random vector given only the marginal distributions of the random variables. The class of joint distributions with fixed marginal distributions is referred to as the Fréchet class of distributions and the bounds are referred to as Fréchet bounds. In this paper, we develop a new class of Fréchet bounds and apply it to to solve distributionally robust optimization problems.

Given a set $\mathcal{N}$, let $\mathcal{E} = \{\mathcal{J}_1, \ldots, \mathcal{J}_R\} \subseteq 2^{\mathcal{N}}$ be a cover of the set $\mathcal{N}$, i.e., $\bigcup_{r\in\mathcal{R}} \mathcal{J}_r = \mathcal{N}$, where $\mathcal{R} = \{1, \ldots, R\}$. Assume that there is no inclusion among the subsets, i.e., $\mathcal{J}_r \nsubseteq \mathcal{J}_s$ for any $r \neq s$. Typical examples of a cover include the partition, the star cover, and the series cover:

- **Partition:** $\mathcal{E}$ is a partition or a non-overlapping cover if for any $r \neq s$, $\mathcal{J}_r \cap \mathcal{J}_s = \emptyset$. The cover:

$$\mathcal{E} = \{\{1\}, \{2\}, \ldots, \{N\}\},$$

is called the *simple partition* and forms the basic Fréchet class of distributions. When some of the subsets consist of more than one random variable, the partition is referred to as a *non-overlapping multivariate marginal cover*. An example of such a partition is:

$$\mathcal{E} = \{\{1,2\}, \{3,4,\ldots,N\}\}.$$

- **Star cover:** Let $\{\mathcal{I}_0, \mathcal{I}_1, \ldots, \mathcal{I}_R\}$ be a partition of $\mathcal{N}$. Then $\mathcal{E}$ is a star cover if $\mathcal{J}_r = \mathcal{I}_r \cup \mathcal{I}_0$ for all $r \in \mathcal{R}$. The star cover:

$$\mathcal{E} = \{\{1,2\}, \{1,3\}, \ldots, \{1,N\}\},$$

is called the *simple star cover.*

- **Series cover:** Let $\{\mathcal{I}_0, \mathcal{I}_1, \ldots, \mathcal{I}_R\}$ be a partition of $\mathcal{N}$. Then $\mathcal{E}$ is a series cover if $\mathcal{J}_r = \mathcal{I}_{r-1} \cup \mathcal{I}_r$ for all $r \in \mathcal{R}$. The series cover:

$$\mathcal{E} = \{\{1,2\}, \{2,3\}, \ldots, \{N-1,N\}\},$$

is called the *simple series cover.*

Given a joint distribution θ of the random vector $\tilde{\boldsymbol{c}}$, let $\mathrm{proj}_{\mathcal{J}_r}(\theta)$ denote the marginal distribution of the sub-vector $\tilde{\boldsymbol{c}}_r$ formed from the components in the subset $\mathcal{J}_r$. This brings us to a general definition of the Fréchet class of distributions (see Rüschendorf (1991)).

Definition 1. The Fréchet class of distributions $\Theta_{\mathcal{E}}$ for a cover $\mathcal{E} = \{\mathcal{J}_1, \ldots, \mathcal{J}_R\}$ is defined as the set of all possible joint distributions of the random vector $\tilde{\boldsymbol{c}}$ with the given multivariate marginal distributions $\{\theta_r\}_{r\in\mathcal{R}}$ of the sub-vectors $\{\tilde{\boldsymbol{c}}_r\}_{r\in\mathcal{R}}$ as projections:

$$\Theta_{\mathcal{E}} \triangleq \{\theta \mid proj_{\mathcal{J}_r}(\theta) = \theta_r, \ \forall r \in \mathcal{R}\}.$$

Throughout the paper, we use the index r and the set $\mathcal{J}_r$ interchangeably. Given a real-valued function $\varphi(\cdot)$, and the Fréchet class of distributions $\Theta_{\mathcal{E}}$ for a given cover $\mathcal{E}$, the upper bound is computed as:

$$M_{\mathcal{E}}(\varphi) = \sup_{\theta \in \Theta_{\mathcal{E}}} \mathbb{E}_\theta \left[\varphi(\tilde{\boldsymbol{c}})\right]. \tag{6}$$

Several previous studies have focused on finding the Fréchet lower bound on the cumulative probability that the sum of random variables is strictly smaller than a given number z:

$$\inf_{\theta \in \Theta_{\mathcal{E}}} \mathbb{P}_\theta \left(\sum_{i \in \mathcal{N}} \tilde{c}_i < z \right).$$

Note that these bounds directly translate to upper bounds on the tail probability:

$$\sup_{\theta \in \Theta_{\mathcal{E}}} \mathbb{P}_\theta \left(\sum_{i \in \mathcal{N}} \tilde{c}_i \geq z \right).$$

The earliest known bounds for this problem were developed by Makarov (1982) and Rüschendorf (1982) for a simple partition with $N = 2$. This bound is given as:

$$\inf_{\theta \in \Theta_{\{\{1\},\{2\}\}}} \mathbb{P}_\theta \left(\tilde{c}_1 + \tilde{c}_2 < z\right) = \max \left\{ \sup_{\boldsymbol{d}: d_1 + d_2 = z} \left[F_1^-(d_1) + F_2(d_2)\right] - 1, 0 \right\}, \tag{7}$$

where $F_i(d_i) = \mathbb{P}\left(\tilde{c}_i \leq d_i\right)$ and $F_i^-(d_i) = \mathbb{P}\left(\tilde{c}_i < d_i\right)$. For $N \geq 3$, Kreinovich and Ferson (2006) showed that computing the tightest bound when $\mathcal{E}$ is a simple partition, is already NP-hard. Several weaker bounds have been proposed, among which is the *standard bound* of Embrechts et al. (2003) and Rüschendorf (2005):

$$\inf_{\theta \in \Theta_{\{\{1\},\dots,\{N\}\}}} \mathbb{P}_\theta \left(\sum_{i \in \mathcal{N}} \tilde{c}_i < z \right) \geq \max \left\{ \sup_{\boldsymbol{d}: \sum_{i \in \mathcal{N}} d_i = z} \left[F_1^-(d_1) + \sum_{i=2}^{N} F_i(d_i) \right] - (N-1), 0 \right\}. \tag{8}$$

While the standard bound is tight for $N = 2$, it is not tight for $N \geq 3$. For $N \geq 3$, Wang and Wang (2011) and Pucetti and Rüschendorf (2013) have developed tight bounds in special instances such as homogeneous univariate marginals with monotone densities and concave densities respectively. Embrechts and Puccetti (2006) have derived alternate dual bounds using results from mass transportation theory (see Rachev and Rüschendorf (1989)). For non-overlapping multivariate marginal covers, Puccetti and Rüschendorf (2012) showed that the Fréchet bound on the cumulative distribution function of a sum of random variables can be reduced to that of a simple partition. For the simple star cover, Rüschendorf (1991) introduced a conditioning method to derive Fréchet bounds using the bound for the simple partition. For the simple series cover, Embrechts and Puccetti (2010) proposed a variable splitting method to estimate Fréchet bounds. Puccetti and Rüschendorf (2012) have generalized this method to general overlapping covers. In general, given the hardness of computing the tightest lower bound on the cumulative distribution function of the sum of random variables, these methods generate tight bounds only in special instances.

In this paper, we compute a new Fréchet upper bound for the convex piecewise linear function of a random vector and solve the associated distributionally robust optimization problem. In Section 2, we review a graph theoretic condition on the structure of a cover referred to as the *running intersection property*. This property guarantees the Fréchet class of distributions is nonempty. Using the running intersection property, we show that the tightest upper bound for discrete distributions is efficiently computable with linear programming. Our proof is constructive and provides an explicit characterization of the distribution in the Fréchet class that attains the bound. The distribution is an alternative to the maximal entropy distribution which corresponds to a conditionally independent distribution in this setting. In Section 3, we compute new bounds on the worst-case VaR and CVaR measures. Simple examples are provided in this section to illustrate the quality of the bounds. In Section 4, we apply a data-driven approach with financial returns to identify the Fréchet class of distributions with the running intersection property and then optimize the portfolio over this class of distributions. We show that by combining simple heuristic algorithms to identify the cover with linear optimization to identify the optimal portfolio, superior out of sample performance is attainable. We finally conclude in Section 5.

2. Fréchet Bound for a Regular Cover

For a non-overlapping cover, the Fréchet class of distributions $\Theta_{\mathcal{E}}$ is nonempty if and only if each θ_r for $r \in \mathcal{R}$ is a valid distribution. Feasibility is ensured in this case by simply using a product measure on the marginal distributions. However for an arbitrary cover with overlap, the feasibility problem is itself known to be non-trivial. Honeyman et al. (1980) showed that for a general cover the problem of determining if there exists a multivariate joint distribution with the given marginals as projections is NP-complete. For an overlapping cover, the existence of a joint distribution clearly implies the pairwise consistency of the multivariate marginals:

$$\text{proj}_{\mathcal{J}_r \cap \mathcal{J}_s}(\theta_r) = \text{proj}_{\mathcal{J}_r \cap \mathcal{J}_s}(\theta_s), \ \forall r \neq s.$$

The reverse however need not be true, namely consistency does not imply the existence of joint distribution. Vorob'ev (1962) provided a simple counterexample to show this for the cover $\mathcal{E} = \{\{1,2\},\{2,3\},\{1,3\}\}$ (see Table 1). In this example, there is no possible joint distribution for

Table 1 Example where consistency of the bivariate marginals does not imply feasibility of a trivariate distribution.

$\tilde{c}_1$	$\tilde{c}_2$	Probability
0	0	0.5
1	1	0.5

$\tilde{c}_2$	$\tilde{c}_3$	Probability
0	0	0.5
1	1	0.5

$\tilde{c}_3$	$\tilde{c}_1$	Probability
0	1	0.5
1	0	0.5

$(\tilde{c}_1, \tilde{c}_2, \tilde{c}_3)$ even though the two-dimensional distributions of $(\tilde{c}_1, \tilde{c}_2)$, $(\tilde{c}_2, \tilde{c}_3)$ and $(\tilde{c}_3, \tilde{c}_1)$ are consistent on the overlapping elements. This brings us to the following definition of a *regular* or a *decomposable* cover (see Vorob'ev (1962), Kellerer (1964)).

Definition 2. A cover structure for which the consistency of the multivariate marginals is sufficient to ensure that $\Theta_{\mathcal{E}}$ is nonempty is referred to as a regular cover (or decomposable cover). A cover that is not regular is referred to as an irregular cover.

While in general, there is no simple sufficient condition to test if $\Theta_{\mathcal{E}}$ is nonempty (see Kellerer (1991)), for regular covers the necessary and sufficient condition is to test the pairwise consistency of all the multivariate marginals.

2.1. Regular Cover

A regular cover is characterizable by several equivalent graph theoretic properties. We review some of the key properties using terminology from graphs and hypergraphs (see Berge (1976)) next. Associated with a cover is a hypergraph $\mathcal{H}(\mathcal{N},\mathcal{E})$ with a set of vertices $\mathcal{N}$ and a set of hyperedges $\mathcal{E}$ that form the set of nonempty subsets of $\mathcal{N}$. The graph $\mathcal{G}(\mathcal{H})$ associated with the hypergraph $\mathcal{H}$ is a graph with the same set of vertices as $\mathcal{H}$ and an edge between every vertex pair that lies in some hyperedge. Then the following four properties: (a), (b), (c) and (d) have shown to be equivalent to each other (see Beeri et al. (1983) and Lauritzen et al. (1984)).

(a) *The cover $\mathcal{E}$ on the set $\mathcal{N}$ is regular:*

For a regular cover, every pairwise consistent set of multivariate marginals is also globally consistent, namely there exists a joint distribution with the given multivariate marginals as projections.

(b) *The hypergraph $\mathcal{H}(\mathcal{N},\mathcal{E})$ is acyclic:*

A hypergraph $\mathcal{H}$ is acyclic if all the vertices of $\mathcal{H}$ can be deleted by repeatedly applying the following two operations: (1) Delete a vertex that occurs in only one hyperedge, (2) Delete an hyperedge that is contained in another hyperedge.

(c) *The hypergraph $\mathcal{H}(\mathcal{N},\mathcal{E})$ is conformal and chordal:*

A clique of a graph $\mathcal{G}$ is a set of pairwise adjacent vertices. A hypergraph $\mathcal{H}$ is conformal if every clique in $\mathcal{G}(\mathcal{H})$ is contained in an hyperedge of $\mathcal{H}$. A graph $\mathcal{G}$ is chordal if every cycle of four or more vertices has at least a chord (an edge connecting two non-adjacent vertices in the cycle). A hypergraph $\mathcal{H}$ is chordal if its graph $\mathcal{G}(\mathcal{H})$ is chordal.

(d) *The cover $\mathcal{E}$ on the set $\mathcal{N}$ satisfies the running intersection property (RIP):*

A cover $\mathcal{E}$ satisfies the RIP, if the elements of $\mathcal{E}$ can be ordered such that:

$$\forall r \in \mathcal{R} \setminus \{1\},\ \exists \sigma_r \in \mathcal{R}:\ 1 \leq \sigma_r < r \text{ and } \mathcal{J}_r \cap \left(\bigcup_{t=1}^{r-1} \mathcal{J}_t \right) \subseteq \mathcal{J}_{\sigma_r}. \tag{9}$$

Associated with the RIP in (9), define the parameters:

$$\begin{aligned} \sigma_r &\triangleq \min\left\{ i \in \mathcal{R} \;\middle|\; \mathcal{J}_r \cap \left(\bigcup_{t=1}^{r-1} \mathcal{J}_t \right) \subseteq \mathcal{J}_i \right\}, \forall r \in \mathcal{R} \setminus \{1\}, \\ \mathcal{K}_r &\triangleq \mathcal{J}_r \cap \left(\bigcup_{t=1}^{r-1} \mathcal{J}_t \right), \qquad\qquad\qquad\qquad \forall r \in \mathcal{R} \setminus \{1\}. \end{aligned} \tag{10}$$

A feasible joint distribution can be constructed in this case using conditional independence as follows (see Kellerer (1964) and Jiroušek (1991)):

$$\theta(\boldsymbol{c}) = \theta_1(\boldsymbol{c}_1) \times \frac{\theta_2(\boldsymbol{c}_2)}{\text{proj}_{\mathcal{K}_2}(\theta_2(\boldsymbol{c}_2))} \times \cdots \times \frac{\theta_R(\boldsymbol{c}_R)}{\text{proj}_{\mathcal{K}_R}(\theta_R(\boldsymbol{c}_R))}, \quad \forall \boldsymbol{c}. \tag{11}$$

Any one of the four properties: (a), (b), (c) or (d) is verifiable efficiently and in particular linear time algorithms have been developed to test for these properties (see Rose et al. (2004) and Tarjan and Yannakakis (1984)). Examples of irregular and regular covers are provided in Figure 1. Covers

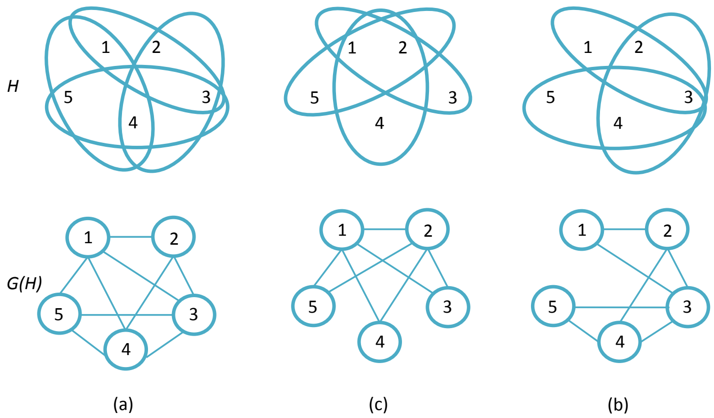

Figure 1 Examples of three different covers with their hypergraph and graph representations. (a) $\mathcal{E} = \{\{1,2,3\},\{2,3,4\},\{3,4,5\},\{4,5,1\}\}$ is an irregular cover. (b) $\mathcal{E} = \{\{1,2,3\},\{1,2,4\},\{1,2,5\}\}$ is a regular cover. (c) $\mathcal{E} = \{\{1,2,3\},\{2,3,4\},\{3,4,5\}\}$ is a regular cover. In (a), $\mathcal{H}$ is not acyclic. Though $\mathcal{G}(\mathcal{H})$ is chordal, $\mathcal{H}$ is not conformal since the clique $\{1,2,4\}$ is not contained in $\mathcal{E}$. In (b) and (c), $\mathcal{H}$ is acyclic, $\mathcal{H}$ is conformal and chordal and the RIP is satisfied. In (b), $\sigma_2 = 1$, $\sigma_3 = 1$, $\mathcal{K}_2 = \{1,2\}$, $\mathcal{K}_3 = \{1,2\}$. In (c), $\sigma_2 = 1$, $\sigma_3 = 2$, $\mathcal{K}_2 = \{2,3\}$, $\mathcal{K}_3 = \{3,4\}$.

(b) and (c) in Figure 1 correspond to star and series covers, both of which are regular covers. A joint distribution for the star cover in Figure 1(b) is constructed as follows:

$$\theta(\boldsymbol{c}) = \theta_{\{1,2,3\}}(c_1, c_2, c_3) \times \frac{\theta_{\{1,2,4\}}(c_1, c_2, c_4)}{\theta_{\{1,2\}}(c_1, c_2)} \times \frac{\theta_{\{1,2,5\}}(c_1, c_2, c_5)}{\theta_{\{1,2\}}(c_1, c_2)}, \quad \forall \boldsymbol{c}, \tag{12}$$

or equivalently:

$$\theta(\boldsymbol{c}) = \theta_{\{1,2\}}(c_1, c_2) \times \theta_{\{3|1,2\}}(c_3|c_1, c_2) \times \theta_{\{4|1,2\}}(c_4|c_1, c_2) \times \theta_{\{5|1,2\}}(c_5|c_1, c_2), \quad \forall \boldsymbol{c}, \tag{13}$$

where $\theta_{\{r|1,2\}}(c_r|c_1, c_2)$ is the conditional distribution of c_r given c_1 and c_2. Similarly a joint distribution for the series cover in Figure 1(c) is constructed as follows:

$$\theta(\boldsymbol{c}) = \theta_{\{1,2,3\}}(c_1, c_2, c_3) \times \frac{\theta_{\{2,3,4\}}(c_2, c_3, c_4)}{\theta_{\{2,3\}}(c_2, c_3)} \times \frac{\theta_{\{3,4,5\}}(c_3, c_4, c_5)}{\theta_{\{3,4\}}(c_3, c_4)}, \quad \forall \boldsymbol{c}, \tag{14}$$

or equivalently:

$$\theta(\boldsymbol{c}) = \theta_{\{1\}}(c_1) \times \theta_{\{2,3|1\}}(c_2, c_3|c_1) \times \theta_{\{4|2,3\}}(c_4|c_2, c_3) \times \theta_{\{5|3,4\}}(c_5|c_3, c_4), \quad \forall \boldsymbol{c}. \tag{15}$$

Properties of regular covers have been previously exploited in developing efficient database representation schemes (see Beeri et al. (1983)), in approximating high-dimensional probability distributions (see Chow and Liu (1968)), in developing tractable semidefinite relaxations in sparse polynomial optimization problems (see Lasserre (2006)) and in developing efficient inference algorithms in probabilistic graphical models (see Wainwright and Jordan (2008)). In this paper, we use the structure of regular covers in distributionally robust optimization problems. From this point onwards, we assume that the following condition is satisfied.

Assumption 1. *The cover $\mathcal{E}$ is regular with the elements satisfying the RIP property in (9) and the multivariate marginal distributions $\{\theta_r\}_{r\in\mathcal{R}}$ are consistent.*

In Section 4, we describe a data-driven approach that uses historical return data to generate regular covers with consistent marginal distributions. The following lemma provides a simpler condition to test the consistency of multivariate marginal distributions for regular covers.

Lemma 1. *Given a regular cover $\mathcal{E}$, the following condition is necessary and sufficient to ensure consistency among the marginal distributions:*

$$\mathrm{proj}_{\mathcal{K}_r}(\theta_r) = \mathrm{proj}_{\mathcal{K}_r}(\theta_{\sigma_r}),\ \forall r \in \mathcal{R} \setminus \{1\} : \mathcal{K}_r \neq \emptyset.$$

Proof. Using the RIP condition in (9), we have for all $r \in \mathcal{R} \setminus \{1\}$,

$$\begin{aligned} \mathcal{J}_r \cap \left(\bigcup_{t=1}^{r-1} \mathcal{J}_t \right) \subseteq \mathcal{J}_{\sigma_r} &\iff \mathcal{J}_r \cap \left(\bigcup_{t=1}^{r-1} \mathcal{J}_t \right) \subseteq \mathcal{J}_{\sigma_r} \cap \mathcal{J}_r, \\ &\iff \mathcal{J}_r \cap \mathcal{J}_t \subseteq \mathcal{J}_{\sigma_r} \cap \mathcal{J}_r, \qquad \forall t = 1, \ldots, r-1. \end{aligned}$$

This indicates that all the pairwise intersections for a regular cover are included in a set of $R-1$ intersections. Thus verifying consistency can be restricted to these pairs. Since, $\mathcal{J}_{\sigma_r} \cap \mathcal{J}_r = \left(\bigcup_{t=1}^{r-1} \mathcal{J}_t \right) \cap \mathcal{J}_r = \mathcal{K}_r$, the result is proved. $\square$

Lemma 1 implies that for regular covers, the feasibility of the Fréchet class of distributions can be ensured with $O(R)$ consistency requirements as opposed to $O(R^2)$ pairwise consistency requirements.

2.2. A New Fréchet Upper Bound

In this section, we compute the upper bound $M_{\mathcal{E}}(\varphi)$ for $\varphi(\tilde{\boldsymbol{c}}) = \max_{j \in \mathcal{M}} \left(\tilde{\boldsymbol{c}}^T \boldsymbol{a}_j + b_j \right)$ for regular covers with consistent multivariate marginals. Our main theorem provides a linear optimization formulation to compute the Fréchet upper bound for discrete distributions. Towards this, we define a N-dimensional vector $\boldsymbol{\eta}$ that allows us to express a linear function in $\tilde{\boldsymbol{c}}$ as separable functions with respect to the cover $\mathcal{E}$. Define the ith component of $\boldsymbol{\eta}$ as follows:

$$\eta_i \triangleq \frac{1}{\sum\limits_{r \in \mathcal{R}} \mathbb{I}\{i \in \mathcal{J}_r\}}, \quad \forall i \in \mathcal{N},$$

where $\mathbb{I}\{i \in \mathcal{J}_r\} = 1$ if $i \in \mathcal{J}_r$ and 0 otherwise. For example, in the simple star cover, this reduces to:

$$\boldsymbol{\eta} = \left(\frac{1}{N-1}, 1, \ldots, 1 \right)^T,$$

and in the simple series cover, this reduces to:

$$\boldsymbol{\eta} = \left(1, \frac{1}{2}, \frac{1}{2}, \ldots, \frac{1}{2}, 1\right)^T.$$

Then, $\tilde{\boldsymbol{c}}^T\boldsymbol{a} = \sum_{r\in\mathcal{R}} \tilde{\boldsymbol{c}}_r^T(\boldsymbol{\eta}_r \circ \boldsymbol{a}_r)$ where $\boldsymbol{\eta}_r$ and $\boldsymbol{a}_r$ are sub-vectors formed by the elements of $\boldsymbol{\eta}$ and $\boldsymbol{a}$ in $\mathcal{J}_r$ and $\circ$ is the Hadamard (entry-wise) product operator. Finally, let $\mathcal{C}_r$ denote the finite set of support values of the sub-vector $\tilde{\boldsymbol{c}}_r$ and $C_r = \text{size}(\mathcal{C}_r)$.

THEOREM 1. *Given a regular cover $\mathcal{E}$ and a consistent set $\{\theta_r\}_{r\in\mathcal{R}}$ of discrete marginal distributions with finite support, let $M_{\mathcal{E}}^{\mathrm{P}}(\varphi)$ be the optimal value to the primal linear program:*

$$\begin{aligned}
\max_{\vartheta_{j,r}(.),\lambda_j} \quad & \sum_{j\in\mathcal{M}}\sum_{r\in\mathcal{R}}\sum_{\boldsymbol{c}_r\in\mathcal{C}_r} \boldsymbol{c}_r^T(\boldsymbol{\eta}_r\circ\boldsymbol{a}_{jr})\cdot\vartheta_{j,r}(\boldsymbol{c}_r) + \sum_{j\in\mathcal{M}} b_j\lambda_j \\
\text{s.t.} \quad & \textit{(Nonnegativity of measure):} \\
& \vartheta_{j,r}(\boldsymbol{c}_r) \geq 0, && \forall\, \boldsymbol{c}_r\in\mathcal{C}_r, \forall\, r\in\mathcal{R}, \forall\, j\in\mathcal{M}, \\
& \textit{(Multivariate marginal requirement):} \\
& \sum_{j\in\mathcal{M}} \vartheta_{j,r}(\boldsymbol{c}_r) = \theta_r(\boldsymbol{c}_r), && \forall\, \boldsymbol{c}_r\in\mathcal{C}_r, \forall\, r\in\mathcal{R}, \\
& \textit{(Probability of } j\textit{th term attaining maximum):} \\
& \sum_{\boldsymbol{c}_r\in\mathcal{C}_r} \vartheta_{j,r}(\boldsymbol{c}_r) = \lambda_j, && \forall\, r\in\mathcal{R}, \forall\, j\in\mathcal{M}, \\
& \textit{(Consistency requirement):} \\
& \sum_{\boldsymbol{h}_r\in\mathcal{C}_r:\mathrm{proj}_{\mathcal{K}_r}(\boldsymbol{h}_r)=\boldsymbol{c}_{\mathcal{K}_r}} \vartheta_{j,r}(\boldsymbol{h}_r) = \sum_{\boldsymbol{h}_{\sigma_r}\in\mathcal{B}_{\sigma_r}:\mathrm{proj}_{\mathcal{K}_r}(\boldsymbol{h}_{\sigma_r})=\boldsymbol{c}_{\mathcal{K}_r}} \vartheta_{j,\sigma_r}(\boldsymbol{h}_{\sigma_r}), && \forall\, \boldsymbol{c}_{\mathcal{K}_r}\in\mathcal{C}_{\mathcal{K}_r}, \\
& && \forall\, r\in\mathcal{R}\setminus\{1\}:\mathcal{K}_r\neq\emptyset, \forall\, j\in\mathcal{M},
\end{aligned} \tag{16}$$

where the decision variables are the nonnegative measures $\vartheta_{j,r}(\boldsymbol{c}_r)$ and λ_j for $\boldsymbol{c}_r\in\mathcal{C}_r, r\in\mathcal{R}, j\in\mathcal{M}$. Then the Fréchet bound $M_{\mathcal{E}}(\varphi) = \max_{\theta\in\Theta_{\mathcal{E}}} \mathbb{E}_\theta\left[\max_{j\in\mathcal{M}}(\tilde{\boldsymbol{c}}^T\boldsymbol{a}_j + b_j)\right]$ is equal to $M_{\mathcal{E}}^{\mathrm{P}}(\varphi)$.

Proof. We first show that $M^{\mathrm{P}}_{\mathcal{E}}(\varphi)$ is a valid upper bound of $M_{\mathcal{E}}(\varphi)$. For any joint distribution $\theta \in \Theta_{\mathcal{E}}$, the expected value in (2) can be expressed as follows:

$$\begin{aligned}
\mathbb{E}_\theta\left[\varphi(\tilde{\boldsymbol{c}})\right] &= \mathbb{E}_\theta\left[\max_{j\in\mathcal{M}}\left(\sum_{r\in\mathcal{R}}\tilde{\boldsymbol{c}}_r^T(\boldsymbol{\eta}_r\circ\boldsymbol{a}_{j_r})+b_j\right)\right]\\
&= \sum_{j\in\mathcal{M}}\mathbb{E}_\theta\left[\sum_{r\in\mathcal{R}}\tilde{\boldsymbol{c}}_r^T(\boldsymbol{\eta}_r\circ\boldsymbol{a}_{j_r})+b_j \;\middle|\; \text{the } j\text{th term is max}\right]\mathbb{P}_\theta\left(\text{the } j\text{th term is max}\right)\\
&= \sum_{j\in\mathcal{M}}\sum_{r\in\mathcal{R}}\sum_{\boldsymbol{c}_r\in\mathcal{C}_r}\boldsymbol{c}_r^T(\boldsymbol{\eta}_r\circ\boldsymbol{a}_{j_r})\cdot\mathbb{P}_\theta\left(\tilde{\boldsymbol{c}}_r=\boldsymbol{c}_r,\ \text{the } j\text{th term is max}\right)+\\
&\quad \sum_{j\in\mathcal{M}}b_j\mathbb{P}_\theta\left(\text{the } j\text{th term is max}\right)\\
&= \sum_{j\in\mathcal{M}}\sum_{r\in\mathcal{R}}\sum_{\boldsymbol{c}_r\in\mathcal{C}_r}\boldsymbol{c}_r^T(\boldsymbol{\eta}_r\circ\boldsymbol{a}_{j_r})\cdot\vartheta_{j,r}(\boldsymbol{c}_r)+\sum_{j\in\mathcal{M}}b_j\lambda_j,
\end{aligned}$$

where the decision variables are the measures $\vartheta_{j,r}(\boldsymbol{c}_r)$ and λ_j defined as follows:

$$\begin{aligned}
\vartheta_{j,r}(\boldsymbol{c}_r) &= \mathbb{P}_\theta\left(\tilde{\boldsymbol{c}}_r=\boldsymbol{c}_r,\ \text{the } j\text{th term is max}\right),\\
\lambda_j &= \mathbb{P}_\theta\left(\text{the } j\text{th term is max}\right).
\end{aligned}$$

Thus $\vartheta_{j,r}(\boldsymbol{c}_r)\geq 0$ which corresponds to the nonnegativity of measure requirement. Note that if the function value $\varphi(\boldsymbol{c})$ has multiple terms attaining the maximum for some value of $\boldsymbol{c}$, one can arbitrarily choose any one of them without changing the expected value, for example, the term with the minimum index. Hence, for all $\boldsymbol{c}_r\in\mathcal{C}_r$ and $r\in\mathcal{R}$,

$$\begin{aligned}
\sum_{j\in\mathcal{M}}\vartheta_{j,r}(\boldsymbol{c}_r) &= \mathbb{P}_\theta\left(\tilde{\boldsymbol{c}}_r=\boldsymbol{c}_r\right)\\
&= \theta_r(\boldsymbol{c}_r),
\end{aligned}$$

which corresponds to the given multivariate marginal requirement. In addition, for all $r\in\mathcal{R}$ and $j\in\mathcal{M}$,

$$\begin{aligned}
\sum_{\boldsymbol{c}_r\in\mathcal{C}_r}\vartheta_{j,r}(\boldsymbol{c}_r) &= \mathbb{P}_\theta\left(\text{the } j\text{th term is max}\right)\\
&= \lambda_j,
\end{aligned}$$

which corresponds to the probability of the jth term attaining the maximum. From Lemma 1, consistency of the measures $\vartheta_{j,r}(.)$ for a given term j is guaranteed by equality of the projections of the measures:

$$\mathrm{proj}_{\mathcal{K}_r}(\vartheta_{j,r})=\mathrm{proj}_{\mathcal{K}_r}(\vartheta_{j,\sigma_r}).$$

In terms of the decision variables, this corresponds to the last set of constraints in (16). Thus, for any distribution $\theta \in \Theta_{\mathcal{E}}$, all the constraints in (16) are satisfied, which implies $M_{\mathcal{E}}^{\mathrm{P}}(\varphi)$ is an upper bound:

$$M_{\mathcal{E}}^{\mathrm{P}}(\varphi) \geq M_{\mathcal{E}}(\varphi) = \max_{\theta \in \Theta_{\mathcal{E}}} \mathbb{E}_{\theta}\left[\varphi(\tilde{\boldsymbol{c}})\right].$$

We next prove that the bound is tight. Observe that the linear program (16) is bounded and feasible implying that the optimal objective value is attained. Consider an optimal solution of problem (16) denoted by $\vartheta_{j,r}^*(\boldsymbol{c}_r)$ and λ_j^*. We have:

$$\begin{aligned}
\sum_{j \in \mathcal{M}} \lambda_j^* &= \sum_{j \in \mathcal{M}} \sum_{\boldsymbol{c}_r \in \mathcal{C}_r} \vartheta_{j,r}^*(\boldsymbol{c}_r) \\
&= \sum_{\boldsymbol{c}_r \in \mathcal{C}_r} \theta_r(\boldsymbol{c}_r) \\
&= 1.
\end{aligned}$$

In addition, $\lambda_j^* \geq 0$ for all $j \in \mathcal{M}$. Thus $\boldsymbol{\lambda}^*$ is a probability vector. We now construct a joint distribution θ^* for $\tilde{\boldsymbol{c}}$ based on $\vartheta_{j,r}^*(\boldsymbol{c}_r)$ and λ_j^* as follows:

(a) Choose term $j \in \mathcal{M}$ with probability λ_j^*.

(b) For each $r \in \mathcal{R}$, assign a measure $\theta_{j,r}^*(\boldsymbol{c}_r)$ for $\boldsymbol{c}_r \in \mathcal{C}_r$ where $\theta_{j,r}^*(\boldsymbol{c}_r) = \vartheta_{j,r}^*(\boldsymbol{c}_r)/\lambda_j^*$. Note that if $\lambda_j^* = 0$, we simply drop that index.

(c) Choose a feasible joint distribution in the Fréchet class of distributions $\theta_j^* \in \Theta_{\mathcal{E}}(\theta_{j,1}^*, \ldots, \theta_{j,R}^*)$ and generate $\tilde{\boldsymbol{c}}$ with distribution θ_j^*.

It is clear that $\theta_{j,r}^*$ is a valid and consistent probability measure for $\tilde{\boldsymbol{c}}_r$, $r \in \mathcal{R}$, since $(\vartheta_{j,r}^*(\boldsymbol{c}_r), \lambda_j^*)$ is a feasible solution to problem (16). Hence, $\Theta_{\mathcal{E}}(\theta_{j,1}^*, \ldots, \theta_{j,R}^*) \neq \emptyset$ since the cover is regular, which implies the existence of a joint distribution θ_j^* for all $j \in \mathcal{M}$. For all $r \in \mathcal{R}$, the probability of $\tilde{\boldsymbol{c}}_r$ taking the value $\boldsymbol{c}_r$ is:

$$\begin{aligned}
\sum_{j \in \mathcal{M}} \lambda_j^* \cdot \frac{\vartheta_{j,r}^*(\boldsymbol{c}_r)}{\lambda_j^*} &= \sum_{j \in \mathcal{M}} \vartheta_{j,r}^*(\boldsymbol{c}_r) \\
&= \theta_r(\boldsymbol{c}_r).
\end{aligned}$$

Thus, we have $\theta^* \in \Theta_{\mathcal{E}}$. Hence the following inequality holds:

$$\begin{aligned}
\mathbb{E}_{\theta_j^*}\left[\max_{k\in\mathcal{M}}\left(\sum_{r\in\mathcal{R}}\tilde{\boldsymbol{c}}_r^T(\boldsymbol{\eta}_r\circ\boldsymbol{a}_{\boldsymbol{k}r})+b_k\right)\right] &\geq \mathbb{E}_{\theta_j^*}\left[\sum_{r\in\mathcal{R}}\tilde{\boldsymbol{c}}_r^T(\boldsymbol{\eta}_r\circ\boldsymbol{a}_{jr})+b_j\right]\\
&= \sum_{r\in\mathcal{R}}\mathbb{E}_{\theta_j^*}\left[\tilde{\boldsymbol{c}}_r^T(\boldsymbol{\eta}_r\circ\boldsymbol{a}_{jr})+b_j\right]\\
&= \sum_{r\in\mathcal{R}}\frac{1}{\lambda_j^*}\sum_{\boldsymbol{c}_r\in\mathcal{C}_r}\boldsymbol{c}_r^T(\boldsymbol{\eta}_r\circ\boldsymbol{a}_{jr})\cdot\vartheta_{j,r}^*(\boldsymbol{c}_r)+b_j,
\end{aligned}$$

where the first inequality is obtained by simple choosing the jth term in the function for θ_j^*. Then we have a lower bound since:

$$\begin{aligned}
\mathbb{E}_{\theta^*}\left[\max_{k\in\mathcal{M}}\left(\sum_{r\in\mathcal{R}}\tilde{\boldsymbol{c}}_r^T(\boldsymbol{\eta}_r\circ\boldsymbol{a}_{\boldsymbol{k}r})+b_k\right)\right] &= \sum_{j\in\mathcal{M}}\lambda_j^*\cdot\mathbb{E}_{\theta_j^*}\left[\max_{k\in\mathcal{M}}\left(\sum_{r\in\mathcal{R}}\tilde{\boldsymbol{c}}_r^T(\boldsymbol{\eta}_r\circ\boldsymbol{a}_{\boldsymbol{k}r})+b_k\right)\right]\\
&\geq \sum_{j\in\mathcal{M}}\lambda_j^*\left[\sum_{r\in\mathcal{R}}\frac{1}{\lambda_j^*}\sum_{\boldsymbol{c}_r\in\mathcal{C}_r}\boldsymbol{c}_r^T(\boldsymbol{\eta}_r\circ\boldsymbol{a}_{jr})\cdot\vartheta_{j,r}^*(\boldsymbol{c}_r)+b_j\right]\\
&= \sum_{j\in\mathcal{M}}\sum_{r\in\mathcal{R}}\sum_{\boldsymbol{c}_r\in\mathcal{C}_r}\boldsymbol{c}_r^T(\boldsymbol{\eta}_r\circ\boldsymbol{a}_{jr})\cdot\vartheta_{j,r}^*(\boldsymbol{c}_r)+b_j\lambda_j^*\\
&= M_{\mathcal{E}}^P(\varphi).
\end{aligned}$$

Hence

$$M_{\mathcal{E}}(\varphi)=\max_{\theta\in\Theta_{\mathcal{E}}}\mathbb{E}_\theta\left[\max_{j\in\mathcal{M}}\left(\boldsymbol{a}_j^T\tilde{\boldsymbol{c}}+b_j\right)\right]\geq\mathbb{E}_{\theta^*}\left[\max_{j\in\mathcal{M}}\left(\boldsymbol{a}_j^T\tilde{\boldsymbol{c}}+b_j\right)\right]\geq M_{\mathcal{E}}^P(\varphi).$$

Together, we have $M_{\mathcal{E}}(\varphi)=M_{\mathcal{E}}^{\mathrm{P}}(\varphi)$. □

Several remarks about the theorem and its proof are in order:

(a) The proof of Theorem 1 is inspired from the proofs in Bertsimas et al. (2006) and Natarajan et al. (2009b) for univariate marginals and Doan and Natarajan (2012) for non-overlapping multivariate marginals. Theorem 1 extends these results to overlapping multivariate marginals. The main generalization is that we incorporate a new set of linear constraints that guarantee the consistency of the distributions and thereby the existence of a joint distribution that attains the bound for overlapping regular covers.

(b) The conditionally independent distribution in (11) is a feasible distribution in the set $\Theta_{\mathcal{E}}$. This distribution maximizes the Shannon entropy among all the measures $\theta\in\Theta_{\mathcal{E}}$ (see Jiroušek

(1991)). Theorem 1 provides an alternate distribution in the set $\Theta_{\mathcal{E}}$ that maximizes the expected value of a piecewise linear convex function of the random vector.

(c) The representation of the split vector $\boldsymbol{\eta}$ is not unique. In particular, we can define values $\eta_i^r \geq 0$, such that $\sum_{r\in\mathcal{R}} \eta_i^r = 1$ and $\eta_i^r = 0$ if $i \notin \mathcal{J}_r$ for all $r \in \mathcal{R}$ and $i \in \mathcal{N}$. For example, instead of splitting variables equally among all the relevant subsets, we can set $\eta_i^{r(i)} = 1$ for all $i \in \mathcal{N}$, where $r(i) = \min\{r : i \in \mathcal{J}_r\}$. This does not affect the result of Theorem 1.

(d) A total of $\sum_{r\in R} C_r$ probabilities are specified as an input to the linear optimization problem where C_r is the support of each sub-vector $\tilde{\boldsymbol{c}}_r$. The total number of decision variables in the primal linear program in Theorem 1 is $M\sum_{r\in\mathcal{R}} C_r + M$, and the total number of constraints is $M\sum_{r\in\mathcal{R}} C_r + RM + \sum_{r\in\mathcal{R}} C_r + M\sum_{r\in\mathcal{R}\setminus\{1\}} C_{\mathcal{K}_r}$. Hence the size of the linear program is polynomially bounded in the parameters N, M, R and the maximum support size $\max_{r\in\mathcal{R}} C_r$. If the marginals are constructed from historical data, as in the numerical experiments in Section 4, the maximum support size is bounded by the number of data samples. With the number of data samples in the order of hundred, we will demonstrate in the numerical section that the linear program (16) can be solved efficiently.

(e) It is possible for the maximum support size $\max_{r\in\mathcal{R}} C_r$ to be exponential in the number of random variables N. For example, if the entire joint distribution is given, up to K^N probabilities might need to be specified where K is the maximum number of values that any individual random variable takes. In this case, if the size of the subsets are restricted to be $O(\log(N))$, then $\max_{r\in\mathcal{R}} C_r \leq K^{O(\log(N))}$ and is polynomially bounded in N. The size of the linear program is then polynomially bounded in the parameters N, M, R and K (the maximum number of support points of any univariate marginal). In general, if the size of the subsets is small enough as compared to N, solving the linear program (16) to compute Fréchet bounds would be more efficient than computing the expected value (2) given the complete joint distribution of up to K^N probabilities.

We conclude this section by showing that the result in Doan and Natarajan (2012) for general partitions can be derived from the result of Theorem 1 for general covers. By assigning dual variables $f_r(\boldsymbol{c}_r)$, $d_{j,r}$ and $g_{j,r}(\boldsymbol{c}_{\mathcal{K}_r})$ to the equalities in formulation (16), the dual linear program is formulated as follows:

$$\begin{aligned}
M_{\mathcal{E}}^{\mathrm{D}}(\varphi) = \min_{f_r(\cdot), g_{j,r}(\cdot), d_{j,r}} & \sum_{r\in\mathcal{R}} \sum_{\boldsymbol{c}_r\in\mathcal{C}_r} f_r(\boldsymbol{c}_r)\theta_r(\boldsymbol{c}_r) \\
\text{s.t. } & f_r(\boldsymbol{c}_r) \geq \boldsymbol{c}_r^T(\boldsymbol{\eta}_r \circ \boldsymbol{a}_{j_r}) - d_{j,r} - g_{j,r}(\boldsymbol{c}_{\mathcal{K}_r}) + \sum_{t>r:\sigma_t=r} g_{j,t}(\boldsymbol{c}_{\mathcal{K}_t}), \ \forall \boldsymbol{c}_r \in \mathcal{C}_r, \\
& \qquad \forall r\in\mathcal{R}, \forall j \in \mathcal{M}, \\
& \sum_{r\in\mathcal{R}} d_{j,r} + b_j = 0, \qquad \forall j\in\mathcal{M},
\end{aligned} \tag{17}$$

where we define $\mathcal{K}_1 = \emptyset$, and for $r\in\mathcal{R}$, if $\mathcal{K}_r = \emptyset$, we define $\tilde{\boldsymbol{c}}_{\mathcal{K}_r} = 0$, and $g_{j,r}(0) = 0$, for all $j \in \mathcal{M}$. Formulation (17) can be concisely rewritten as:

$$\begin{aligned}
M_{\mathcal{E}}^{\mathrm{D}}(\varphi) = \min_{g_{j,r}(\cdot), d_{j,r}} & \sum_{r\in\mathcal{R}} \mathbb{E}_{\theta_r}\left[\max_{j\in\mathcal{M}}\left(\tilde{\boldsymbol{c}}_r^T(\boldsymbol{\eta}_r \circ \boldsymbol{a}_{j_r}) - d_{j,r} - g_{j,r}(\tilde{\boldsymbol{c}}_{\mathcal{K}_r}) + \sum_{t>r:\sigma_t=r} g_{j,t}(\tilde{\boldsymbol{c}}_{\mathcal{K}_t})\right)\right] \\
\text{s.t. } & \sum_{r\in\mathcal{R}} d_{j,r} + b_j = 0, \quad \forall j\in\mathcal{M},
\end{aligned} \tag{18}$$

Linear programming duality implies that $M_{\mathcal{E}}(\varphi) = M_{\mathcal{E}}^{\mathrm{D}}(\varphi)$. For a general partition, the dual variables $g_{j,r}(\boldsymbol{c}_{\mathcal{K}_r})$ correspond to the marginal consistency constraints in the primal problem (16). When $\mathcal{E}$ is a partition, the marginal consistency constraints are not present and hence the corresponding dual variables are deleted. Thus formulation (17) for the partition case reduces to:

$$\begin{aligned}
\min_{d_{j,r}} & \sum_{r\in\mathcal{R}} \mathbb{E}_{\theta_r}\left[\max_{j\in\mathcal{M}}\left(\tilde{\boldsymbol{c}}_r^T \boldsymbol{a}_{j_r} - d_{j,r}\right)\right] \\
\text{s.t. } & \sum_{r\in\mathcal{R}} d_{j,r} + b_j = 0, \quad \forall j\in\mathcal{M},
\end{aligned} \tag{19}$$

which is equivalent to the non-overlapping marginal formulation in Doan and Natarajan (2012).

2.3. Connected Regular Covers: Star and Series Case

In this section, we simplify the Fréchet bound for a class of covers with a special structure that we term as *connected regular covers*.

Definition 3. A cover $\mathcal{E}$ is said to be connected if for any $s,t\in\mathcal{R}$, $s\neq t$, there exists a sequence $r_1 = s, r_2, \ldots, r_m = t \in \mathcal{R}$ such that $\mathcal{J}_{r_j} \cap \mathcal{J}_{r_{j+1}} \neq \emptyset$ for all $j = 1,\ldots,m-1$.

It is clear that partitions are not connected covers. The simple star and series covers are examples of connected covers. The next lemma characterizes the connectedness of regular covers.

LEMMA 2. *A regular cover $\mathcal{E}$ is connected if and only if $\mathcal{K}_r \neq \emptyset$ for all $r \in \mathcal{R} \setminus \{1\}$.*

The proof of the Lemma is provided in the Appendix. This characterization of connected regular covers allows us to slightly simplify the formulation to compute $M_{\mathcal{E}}(\varphi)$.

PROPOSITION 1. *Given a connected regular cover $\mathcal{E}$ and a consistent set $\{\theta_r\}_{r\in\mathcal{R}}$ of discrete marginal distributions with finite support, let $M_{\mathcal{E}}^{\mathrm{PC}}(\varphi)$ be the optimal value to the primal linear program:*

$$
\begin{aligned}
\max_{\vartheta_{j,r}(\cdot)} \; & \sum_{j\in\mathcal{M}}\sum_{r\in\mathcal{R}}\sum_{\boldsymbol{c}_r\in\mathcal{C}_r} \left(\boldsymbol{c}_r^T(\boldsymbol{\eta}_r \circ \boldsymbol{a}_{j_r}) + \rho_r b_j\right)\cdot \vartheta_{j,r}(\boldsymbol{c}_r) \\
\text{s.t. } & \textit{(Nonnegativity of measure):} \\
& \vartheta_{j,r}(\boldsymbol{c}_r) \geq 0, \qquad \forall\, \boldsymbol{c}_r \in \mathcal{C}_r, \forall\, r\in\mathcal{R}, \forall\, j\in\mathcal{M}, \\
& \textit{(Multivariate marginal requirement):} \\
& \sum_{j\in\mathcal{M}} \vartheta_{j,r}(\boldsymbol{c}_r) = \theta_r(\boldsymbol{c}_r), \qquad \forall\, \boldsymbol{c}_r \in \mathcal{C}_r, \forall\, r\in\mathcal{R}, \\
& \textit{(Consistency requirement):} \\
& \sum_{\boldsymbol{h}_r\in\mathcal{C}_r:\, \mathrm{proj}_{\mathcal{K}_r}(\boldsymbol{h}_r)=\boldsymbol{c}_{\mathcal{K}_r}} \vartheta_{j,r}(\boldsymbol{h}_r) = \sum_{\boldsymbol{h}_{\sigma_r}\in\mathcal{C}_{\sigma_r}:\, \mathrm{proj}_{\mathcal{K}_r}(\boldsymbol{h}_{\sigma_r})=\boldsymbol{c}_{\mathcal{K}_r}} \vartheta_{j,\sigma_r}(\boldsymbol{h}_{\sigma_r}), \; \forall\, \boldsymbol{c}_{\mathcal{K}_r} \in \mathcal{C}_{\mathcal{K}_r}, \\
& \qquad \forall\, r\in\mathcal{R}\setminus\{1\}, \forall\, j\in\mathcal{M},
\end{aligned}
\tag{20}
$$

where ρ_r are arbitrary constants that satisfy $\sum_{r\in\mathcal{R}}\rho_r = 1$ and the decision variables are the measures $\vartheta_{j,r}(\boldsymbol{c}_r)$ for $\mathbf{c}_r \in \mathcal{C}_r, r\in\mathcal{R}, j\in\mathcal{M}$. Then the Fréchet bound $M_{\mathcal{E}}(\varphi) = \max_{\theta\in\Theta_{\mathcal{E}}} \mathbb{E}_\theta\left[\max_{j\in\mathcal{M}}\left(\tilde{\boldsymbol{c}}^T\boldsymbol{a}_j + b_j\right)\right]$ is equal to $M_{\mathcal{E}}^{\mathrm{PC}}(\varphi)$.

Proof. We claim that the constraints:

$$\sum_{\boldsymbol{c}_r\in\mathcal{C}_r} \vartheta_{j,r}(\boldsymbol{c}_r) = \lambda_j, \quad \forall\, r\in\mathcal{R}, \forall\, j\in\mathcal{M},$$

in (16) are redundant if the regular cover $\mathcal{E}$ is connected. From Lemma 2, $\mathcal{K}_r \neq \emptyset$ for all $r\in\mathcal{R}$. Using the last set of constraints in (16) (or (20)), we obtain the following equalities:

$$\sum_{\boldsymbol{c}_r\in\mathcal{C}_r} \vartheta_{j,r}(\boldsymbol{c}_r) = \sum_{\boldsymbol{c}_s\in\mathcal{C}_s} \vartheta_{j,s}(\boldsymbol{c}_s), \quad \forall\, r,s\in\mathcal{R}, r\neq s, \forall\, j\in\mathcal{M}.$$

Thus, we can drop the decision variables λ_j, by replacing $\sum_{j\in\mathcal{M}} b_j\lambda_j$ by $\sum_{j\in\mathcal{M}}\sum_{r\in\mathcal{R}}\sum_{\boldsymbol{c}_r\in\mathcal{C}_r}\rho_r b_j\vartheta_{j,r}(\boldsymbol{c}_r)$ in the objective given that $\sum_{r\in\mathcal{R}}\rho_r = 1$. Thus for connected regular covers, (20) is equivalent to (16) and it implies that in this case, $M_{\mathcal{E}}(\varphi) = M_{\mathcal{E}}^{\mathrm{PC}}(\varphi)$. □

The problem (20) has fewer variables and lesser constraints as compared to (16), which also allows us to simplify its dual formulation by remove the corresponding set of dual variables. The dual formulation is written as follows:

$$M_{\mathcal{E}}^{\mathrm{DC}}(\varphi) = \min_{g_{j,r}(\cdot)}\sum_{r\in\mathcal{R}}\mathbb{E}_{\theta_r}\left[\max_{j\in\mathcal{M}}\left(\tilde{\boldsymbol{c}}_r^T(\boldsymbol{\eta}_r\circ\boldsymbol{a}_{jr}) - g_{j,r}\left(\tilde{\boldsymbol{c}}_{\mathcal{K}_r}\right) + \sum_{t>r:\sigma_t=r} g_{j,t}\left(\tilde{\boldsymbol{c}}_{\mathcal{K}_t}\right) + \rho_r b_j\right)\right]. \tag{21}$$

To illustrate this formulation, we consider two simple examples of the Fréchet bound for star and series covers.

Series cover: For the simple series cover, we have $R = N-1$ and $\sigma_r = r-1$ for all $r = 2,...,N-1$. Letting $\rho_r = 1/(N-1)$ for all $r\in\mathcal{R}$, we can reformulate (21) as:

$$\begin{aligned}\min_{g_{j,r}(.)}\ & \mathbb{E}_{\theta_{\{1,2\}}}\left[\max_{j\in\mathcal{M}}\left(\tilde{c}_1 a_{j_1} + \frac{\tilde{c}_2 a_{j_2}}{2} + g_{j,2}\left(\tilde{c}_2\right) + \frac{b_j}{N-1}\right)\right] + \\ & \sum_{r=2}^{N-2}\mathbb{E}_{\theta_{\{r,r+1\}}}\left[\max_{j\in\mathcal{M}}\left(\frac{\tilde{c}_r a_{j_r}}{2} + \frac{\tilde{c}_{r+1}a_{j_{r+1}}}{2} - g_{j,r}\left(\tilde{c}_r\right) + g_{j,r+1}\left(\tilde{c}_{r+1}\right) + \frac{b_j}{N-1}\right)\right] + \\ & \mathbb{E}_{\theta_{\{N-1,N\}}}\left[\max_{j\in\mathcal{M}}\left(\tilde{c}_N a_{j_N} + \frac{\tilde{c}_{N-1}a_{j_{N-1}}}{2} - g_{j,N-1}\left(\tilde{c}_{N-1}\right) + \frac{b_j}{N-1}\right)\right].\end{aligned} \tag{22}$$

Star cover: For the simple star cover, we have $R = N-1$ and $\sigma_r = 1$ for all $r = 2,\ldots,N-1$. The dual Fréchet bound for the star cover is reformulated from (21) as follows:

$$\min_{g_{j,r}(.)}\ \mathbb{E}_{\theta_{\{1,2\}}}\left[\max_{j\in\mathcal{M}}\left(\tilde{c}_2 a_{j2} + \tilde{c}_1 a_{j_1} + \sum_{r=2}^{N-1} g_{j,r}\left(\tilde{c}_1\right) + b_j\right)\right] + \sum_{r=2}^{N-1}\mathbb{E}_{\theta_{\{1,r+1\}}}\left[\max_{j\in\mathcal{M}}\left(\tilde{c}_{r+1}a_{j_{r+1}} - g_{j,r}\left(\tilde{c}_1\right)\right)\right].$$

We define a new set of decision variables as follows:

$$g_{j,1}\left(c_1\right) = -\sum_{r=2}^{N-1} g_{j,r}\left(c_1\right) - c_1 a_{j_1} - b_j, \quad \forall c_1\in\mathcal{C}_1, \forall j\in\mathcal{M}.$$

The formulation then reduces to:

$$\begin{aligned}\min_{g_{j,r}(.)}\ & \sum_{r=1}^{N-1}\mathbb{E}_{\theta_{\{1,r+1\}}}\left[\max_{j\in\mathcal{M}}\left(\tilde{c}_{r+1}a_{j_{r+1}} - g_{j,r}\left(\tilde{c}_1\right)\right)\right] \\ \text{s.t. } & \sum_{r=1}^{N-1} g_{j,r}\left(c_1\right) = -c_1 a_{j_1} - b_j, \quad \forall c_1\in\mathcal{C}_1, \forall j\in\mathcal{M}.\end{aligned} \tag{23}$$

Conditioning with respect to the marginal distribution of $\tilde{c}_1$ as in Puccetti and Rüschendorf (2012) and using the dual representation for the partition case in (19), we obtain an equivalent formulation:

$$\mathbb{E}_{\theta_{\{1\}}}\left\{\sup_{\theta\in\Theta_{\{\{2|1\},\ldots,\{N|1\}\}}}\mathbb{E}_{\theta}\left[\max_{j\in\mathcal{M}}\left(\tilde{\boldsymbol{c}}^T\boldsymbol{a}_j+b_j\right)\Big|\tilde{c}_1\right]\right\}.$$

For a fixed $\tilde{c}_1=c_1$, the inner problem is the Fréchet bound where θ belongs to the Fréchet class defined by the conditional marginal distributions. Thus, the Fréchet bound for the star cover is equivalent to:

$$\mathbb{E}_{\theta_{\{1\}}}\left\{\sup_{\theta\in\Theta_{\{\{2|1\},\ldots,\{N|1\}\}}}\mathbb{E}_{\theta}\left[\varphi(\tilde{\boldsymbol{c}})\Big|\tilde{c}_1\right]\right\},$$

which indicates that it is reduces to the computation of Fréchet bounds with a simple partition. A similar observation has been made by Puccetti and Rüschendorf (2012) for more general objective functions.

3. Bounds for CVaR and VaR

In this section, we evaluate new Fréchet bounds for the CVaR and VaR measures.

3.1. CVaR Bound

The worst case CVaR with respect to the Fréchet class of distributions for $\alpha\in(0,1)$ is defined as (see Natarajan et al. (2009a) and Zhu and Fukushima (2009)):

$$\begin{aligned}\mathrm{WCVaR}_{\alpha}^{\Theta_{\mathcal{E}}}\left(\tilde{\boldsymbol{c}}^T\boldsymbol{x}\right) &\triangleq \sup_{\theta\in\Theta_{\mathcal{E}}}\mathrm{CVaR}_{\alpha}^{\theta}\left(\tilde{\boldsymbol{c}}^T\boldsymbol{x}\right)\\ &=\sup_{\theta\in\Theta_{\mathcal{E}}}\min_{\beta\in\Re}\left(\beta+\frac{1}{1-\alpha}\mathbb{E}_{\theta}\left[\left(\tilde{\boldsymbol{c}}^T\boldsymbol{x}-\beta\right)^+\right]\right).\end{aligned}\tag{24}$$

Since the multivariate marginal distributions $\{\theta_r\}_{r\in\mathcal{R}}$ are assumed to be discrete with finite support, the expected value is finite and the supremum is attained by a joint distribution. Interchanging the minimum and maximum in the worst-case CVaR formulation and from the convexity of the objective function with respect to β and linearity with respect to the measure θ, we get:

$$\mathrm{WCVaR}_{\alpha}^{\Theta_{\mathcal{E}}}\left(\tilde{\boldsymbol{c}}^T\boldsymbol{x}\right)=\min_{\beta\in\Re}\left(\beta+\frac{1}{1-\alpha}\max_{\theta\in\Theta_{\mathcal{E}}}\mathbb{E}_{\theta}\left[\left(\tilde{\boldsymbol{c}}^T\boldsymbol{x}-\beta\right)^+\right]\right).\tag{25}$$

Thus, in order to compute the upper bound on the CVaR, we need to compute an upper bound on the expected value $\mathbb{E}_{\theta}\left[\left(\tilde{\boldsymbol{c}}^T\boldsymbol{x}-\beta\right)^+\right]$. We provide a simple example to illustrate the computation of the expected value next.

3.1.1. Example Consider a sum of N random variables. We compute the Fréchet upper bound:

$$\max_{\theta \in \Theta_{\mathcal{E}}} \mathbb{E}_\theta \left[\left(\sum_{i \in \mathcal{N}} \tilde{c}_i - \beta \right)^+ \right] \tag{26}$$

and compare it with expected value under the maximum entropy (ME) distribution in (11):

$$\mathbb{E}_{\mathrm{ME}} \left[\left(\sum_{i \in \mathcal{N}} \tilde{c}_i - \beta \right)^+ \right].$$

Consider the bivariate uniform discrete distributions provided in Table 2 for a simple series cover with $N = 4$. The maximum-entropy distribution in this case is the independent uniform distribution with $P_{\mathrm{ME}}(\boldsymbol{c}) = 1/16$ for all $\boldsymbol{c} \in \{0,1\}^4$.

Table 2 Consistent bivariate marginals for simple series cover with $\mathcal{E} = \{\{1,2\},\{2,3\},\{3,4\}\}$.

$\tilde{c}_1$	$\tilde{c}_2$	Probability	$\tilde{c}_2$	$\tilde{c}_3$	Probability	$\tilde{c}_3$	$\tilde{c}_4$	Probability
0	0	0.25	0	0	0.25	0	0	0.25
0	1	0.25	0	1	0.25	0	1	0.25
1	0	0.25	1	0	0.25	1	0	0.25
1	1	0.25	1	1	0.25	1	1	0.25

In order to compute the upper bound, we apply Proposition 1 with $M=2$, $\boldsymbol{a}_1=\boldsymbol{e}$, $b_1=-\beta$, $\boldsymbol{a}_2=\boldsymbol{0}$ and $b_2=0$. The primal linear program for (26) is formulated as follows:

$$
\begin{aligned}
\max_{\vartheta_r(\cdot)} \quad & \sum_{(v,w)\in\{0,1\}^2} \left(\left(v + \frac{w}{2} - \frac{\beta}{3} \right) \cdot \vartheta_{\{1,2\}}(v,w) + \left(\frac{v}{2} + \frac{w}{2} - \frac{\beta}{3} \right) \cdot \vartheta_{\{2,3\}}(v,w) \right) + \\
& \sum_{(v,w)\in\{0,1\}^2} \left(\frac{v}{2} + w - \frac{\beta}{3} \right) \cdot \vartheta_{\{3,4\}}(v,w) \\
\text{s.t.} \quad & \text{(Nonnegativity of measure):} \\
& \vartheta_{\{1,2\}}(v,w),\ \vartheta_{\{2,3\}}(v,w),\ \vartheta_{\{3,4\}}(v,w) \geq 0, && \forall (v,w)\in\{0,1\}^2, \\
& \text{(Multivariate marginal requirement):} \\
& \vartheta_{\{1,2\}}(v,w) \leq 0.25, && \forall (v,w)\in\{0,1\}^2, \\
& \vartheta_{\{2,3\}}(v,w) \leq 0.25, && \forall (v,w)\in\{0,1\}^2, \\
& \vartheta_{\{3,4\}}(v,w) \leq 0.25, && \forall (v,w)\in\{0,1\}^2, \\
& \text{(Consistency requirement):} \\
& \vartheta_{\{2,3\}}(0,0) + \vartheta_{\{2,3\}}(0,1) = \vartheta_{\{1,2\}}(0,0) + \vartheta_{\{1,2\}}(1,0), \\
& \vartheta_{\{2,3\}}(1,0) + \vartheta_{\{2,3\}}(1,1) = \vartheta_{\{1,2\}}(0,1) + \vartheta_{\{1,2\}}(1,1), \\
& \vartheta_{\{3,4\}}(0,0) + \vartheta_{\{3,4\}}(0,1) = \vartheta_{\{2,3\}}(0,0) + \vartheta_{\{2,3\}}(1,0), \\
& \vartheta_{\{3,4\}}(1,0) + \vartheta_{\{3,4\}}(1,1) = \vartheta_{\{2,3\}}(0,1) + \vartheta_{\{2,3\}}(1,1),
\end{aligned}
\tag{27}
$$

Since the support of $\sum_{i\in\mathcal{N}} \tilde{c}_i$ is restricted in $\{0,1,2,3,4\}$, we vary β in $[0,4]$. Figure 2 provides a comparison of the bounds. The Fréchet bound provides an upper bound on the expected value with respect to the maximum-entropy distribution.

We next incorporate the bound to provide an explicit formulation for the Fréchet bound of CVaR. Applying the dual formulation, the worst-case CVaR bound is computed as follows:

$$
\mathrm{WCVaR}_{\alpha}^{\Theta_{\mathcal{E}}} \left(\sum_{i\in\mathcal{N}} \tilde{c}_i \right) = \min_{g_r(.),\beta} \left(\beta + \frac{1}{1-\alpha} \sum_{r\in\mathcal{R}} \mathbb{E}_{\theta_r} \left[\left(\tilde{\boldsymbol{c}}_r^T \boldsymbol{\eta}_r - g_r\left(\tilde{\boldsymbol{c}}_{\mathcal{K}_r}\right) + \sum_{t>r:\sigma_t=r} g_t\left(\tilde{\boldsymbol{c}}_{\mathcal{K}_t}\right) - \frac{\beta}{R} \right)^+ \right] \right). \tag{28}
$$

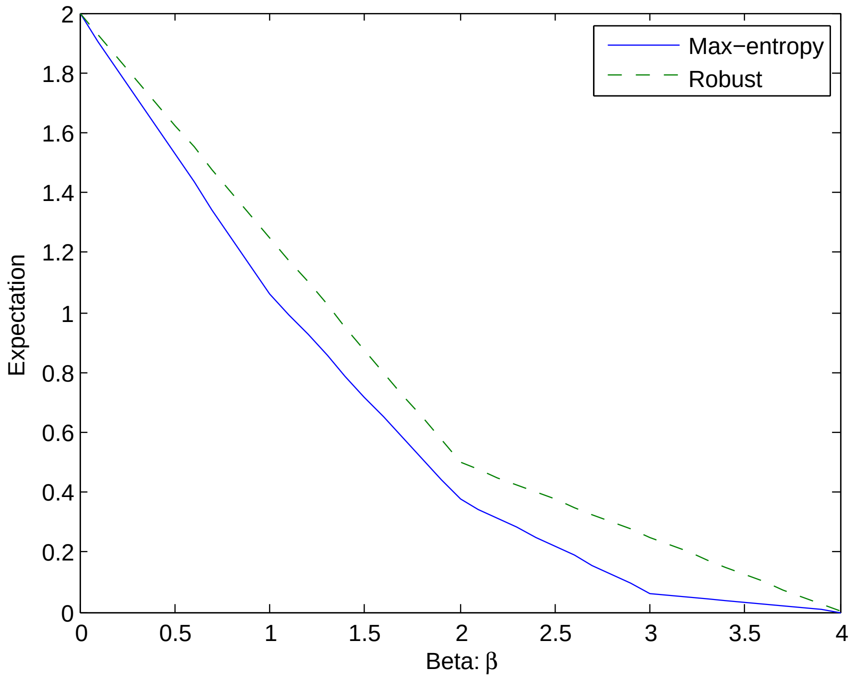

Figure 2 Comparison between the Fréchet bound and the maximum entropy distribution

For the distributional information in Table 2, we can reformulate (28) with additional decision variables as a linear program:

$$\begin{aligned}
\min_{g_r(.),\beta,z_r(.)} \ & \beta + \frac{1}{4(1-\alpha)} \sum_{(v,w)\in\{0,1\}^2} \left(z_{\{1,2\}}(v,w) + z_{\{2,3\}}(v,w) + z_{\{3,4\}}(v,w)\right) \\
\text{s.t. } & z_{\{1,2\}}(v,w) - g_{\{2,3\}}(w) \geq v + \frac{w}{2} - \frac{\beta}{3}, && \forall\,(v,w)\in\{0,1\}^2, \\
& z_{\{2,3\}}(v,w) + g_{\{2,3\}}(v) - g_{\{3,4\}}(w) \geq \frac{v}{2} + \frac{w}{2} - \frac{\beta}{3}, && \forall\,(v,w)\in\{0,1\}^2, \\
& z_{\{3,4\}}(v,w) + g_{\{3,4\}}(v) \geq \frac{v}{2} + w - \frac{\beta}{3}, && \forall\,(v,w)\in\{0,1\}^2, \\
& z_{\{1,2\}}(v,w),\, z_{\{2,3\}}(v,w),\, z_{\{3,4\}}(v,w) \geq 0, && \forall\ (v,w)\in\{0,1\}^2.
\end{aligned} \tag{29}$$

3.2. VaR Bound

The VaR for a portfolio for $\alpha \in (0,1)$ is defined as:

$$\text{VaR}_\alpha^\theta\left(\tilde{\boldsymbol{c}}^T\boldsymbol{x}\right) \triangleq \inf\left\{z\in\Re : \mathbb{P}_\theta\left(\tilde{\boldsymbol{c}}^T\boldsymbol{x}\leq z\right)\geq\alpha\right\}.$$

Hence, we have the following equivalence:

$$\text{VaR}_\alpha^\theta\left(\tilde{\boldsymbol{c}}^T\boldsymbol{x}\right)\leq z \ \Leftrightarrow\ \mathbb{P}_\theta\left(\tilde{\boldsymbol{c}}^T\boldsymbol{x}\leq z\right)\geq\alpha,$$

which implies that z is an upper bound of $\mathrm{VaR}_{\alpha}^{\theta}\left(\tilde{\boldsymbol{c}}^T \boldsymbol{x}\right)$, if and only if α is a lower bound of the cumulative distribution function $\mathbb{P}_{\theta}\left(\tilde{\boldsymbol{c}}^T \boldsymbol{x} \leq z\right)$. Since CVaR dominates VaR (see Rockafellar and Uryasev (2002)), we can use CVaR to derive lower bounds of the cumulative distribution function value. Given a Fréchet class $\Theta_{\mathcal{E}}$ of distributions, the worst-case VaR is defined as follows:

$$\mathrm{WVaR}_{\alpha}^{\Theta_{\mathcal{E}}}(\tilde{\boldsymbol{c}}^T \boldsymbol{x}) \triangleq \inf\left\{z \in \Re : \inf_{\theta \in \Theta_{\mathcal{E}}} \mathbb{P}_{\theta}\left(\tilde{\boldsymbol{c}}^T \boldsymbol{x} \leq z\right) \geq \alpha\right\}. \tag{30}$$

Since $\mathrm{WVaR}_{\alpha}^{\Theta_{\mathcal{E}}}(\tilde{\boldsymbol{c}}^T \boldsymbol{x}) = \sup_{\theta \in \Theta_{\mathcal{E}}} \mathrm{VaR}_{\alpha}^{\theta}(\boldsymbol{x})$, this implies $\mathrm{WVaR}_{\alpha}^{\Theta_{\mathcal{E}}}(\tilde{\boldsymbol{c}}^T \boldsymbol{x}) \leq \mathrm{WCVaR}_{\alpha}^{\Theta_{\mathcal{E}}}(\tilde{\boldsymbol{c}}^T \boldsymbol{x})$.

3.2.1. Example In the following example, we provide a lower bound on the cumulative distribution function of the sum of random variables $\mathbb{P}_{\theta}\left(\sum_{i \in \mathcal{N}} \tilde{c}_i \leq z\right)$ using CVaR approximations. Observe that,

$$\mathrm{WVaR}_{\alpha}^{\Theta_{\mathcal{E}}}\left(\sum_{i \in \mathcal{N}} \tilde{c}_i\right) \leq z \;\Leftrightarrow\; \inf_{\theta \in \Theta_{\mathcal{E}}} \mathbb{P}_{\theta}\left(\sum_{i \in \mathcal{N}} \tilde{c}_i \leq z\right) \geq \alpha,$$

The Fréchet bounds are related as follows:

$$\inf_{\theta \in \Theta_{\mathcal{E}}} \mathbb{P}_{\theta}\left(\sum_{i \in \mathcal{N}} \tilde{c}_i < z\right) \leq \inf_{\theta \in \Theta_{\mathcal{E}}} \mathbb{P}_{\theta}\left(\sum_{i \in \mathcal{N}} \tilde{c}_i \leq z\right) \leq \inf_{\theta \in \Theta_{\mathcal{E}}} \mathbb{P}_{\theta}\left(\sum_{i \in \mathcal{N}} \tilde{c}_i \leq z + \epsilon\right) \text{ for all } \epsilon > 0.$$

Since WCVaR is an upper bound on WVaR, we first compute the WCVaR bound and then use numerical inversion to find a lower bound on the cumulative distribution function. For the series cover, we compute the worst-case CVaR as follows:

$$\begin{aligned}
\mathrm{WCVaR}_{\alpha}^{\Theta_{\mathcal{E}}}\left(\sum_{i \in \mathcal{N}} \tilde{c}_i\right) = \min_{g_i(.),\beta}\; & \beta + \frac{1}{1-\alpha}\left(\mathbb{E}_{\theta_{\{1,2\}}}\left[\left(\tilde{c}_1 + \frac{\tilde{c}_2}{2} + g_2(\tilde{c}_2) - \frac{\beta}{N-1}\right)^+\right] + \right. \\
& \sum_{i=2}^{N-2} \mathbb{E}_{\theta_{\{i,i+1\}}}\left[\left(\frac{\tilde{c}_i + \tilde{c}_{i+1}}{2} - g_i(\tilde{c}_i) + g_{i+1}(\tilde{c}_{i+1}) - \frac{\beta}{N-1}\right)^+\right] + \\
& \left. \mathbb{E}_{\theta_{\{N-1,N\}}}\left[\left(\frac{\tilde{c}_{N-1}}{2} + \tilde{c}_N - g_{N-1}(\tilde{c}_{N-1}) - \frac{\beta}{N-1}\right)^+\right]\right).
\end{aligned} \tag{31}$$

For the numerical experiment, we construct bivariate marginal distributions for the simple series cover by using the independent copula and identical uniform univariate marginals in $[0,1]$. Then $F_i(c_i) = c_i$ for all $c_i \in [0,1]$ and the joint distribution for two random variables is is $F_{i,i+1}(c_i, c_{i+1}) = c_i c_{i+1}$ for all $(c_i, c_{i+1}) \in [0,1]^2$ with $i = 1, \ldots, N-1$. Clearly the set of these bivariate marginals is consistent. Given these continuous marginals, the problem in (31) is an infinite-dimensional linear optimization problem. To compute WCVaR, we use a discretization of the distribution function to compute upper and lower bounds. Consider a discrete distribution $\hat{F}_\omega$ approximation of F with M-points as in Embrechts and Puccetti (2010):

$$\hat{F}_{\boldsymbol{\omega}} \triangleq \frac{1}{M}\sum_{j\in\mathcal{M}} \mathbb{I}\{x \geq \omega_j\},$$

where $\boldsymbol{\omega} = \{\omega_1, ..., \omega_M\}$ is the set of M jump points. Let $q_j = \dfrac{j}{M}$ for $j = 0, \ldots, M$ and define two sets of jump points $\overline{\boldsymbol{\omega}} = \{q_1, \ldots, q_M\}$ and $\underline{\boldsymbol{\omega}} = \{q_0, \ldots, q_{M-1}\}$. Clearly, $\hat{F}_{\overline{\boldsymbol{\omega}}}$ and $\hat{F}_{\underline{\boldsymbol{\omega}}}$ provide lower and upper bounds for F. The discretized bivariate marginal distributions are constructed from the corresponding discretized univariate marginals using the independent copula. Let $\overline{\text{WCVaR}}_\alpha^{\Theta_\mathcal{E}}(\sum_{i\in\mathcal{N}} \tilde{c}_i)$ and $\underline{\text{WCVaR}}_\alpha^{\Theta_\mathcal{E}}(\sum_{i\in\mathcal{N}} \tilde{c}_i)$ denote the worst case CVaR bounds with respect to the discretized marginals respectively. Then, $\underline{\text{WCVaR}}_\alpha^{\Theta_\mathcal{E}}(\sum_{i\in\mathcal{N}} \tilde{c}_i) \leq \text{WCVaR}_\alpha^{\Theta_\mathcal{E}}\left(\sum_{i\in\mathcal{N}} \tilde{c}_i\right) \leq \overline{\text{WCVaR}}_\alpha^{\Theta_\mathcal{E}}(\sum_{i\in\mathcal{N}} \tilde{c}_i)$. The upper and lower bounds are computed from the linear optimization problem in (31).

Embrechts and Puccetti (2010) proposed a lower bound of the cumulative distribution function of the sum of random variables using the standard bound in (8) with variable splitting. In order to compute this bound, one needs to calculate $F_y(d) = \mathbb{P}\left(\tilde{c}_1 + \frac{\tilde{c}_2}{2} \leq d\right)$ and $F_z(d) = \mathbb{P}\left(\frac{\tilde{c}_1}{2} + \frac{\tilde{c}_2}{2} \leq d\right)$. In our example, this reduces to:

$$F_y(d) = \begin{cases} 0, & d < 0, \\ d^2, & 0 \leq d < 1/2, \\ d - 1/4, & 1/2 \leq d < 1, \\ -d^2 + 3d - 5/4, & 1 \leq d < 3/2, \\ 1, & d \geq 3/2, \end{cases} \quad \text{and} \quad F_z(d) = \begin{cases} 0, & d < 0, \\ 2d^2, & 0 \leq d < 1/2, \\ -2d^2 + 4d - 1, & 1/2 \leq d < 1, \\ 1, & d \geq 1. \end{cases}$$

Note that $F_y^-(d) = F_y(d)$ and $F_z^-(d) = F_z(d)$ given continuous distributions. The lower bound in Embrechts and Puccetti (2010) (referred to as the *reduced standard bound* (RSB)) is computed as follows:

$$\text{RSB}(x) = \max\left\{\sup_{\boldsymbol{d}\in\Re^{N-2}} \left[F_y(d_1) + F_y\left(x - \sum_{i=1}^{N-2} d_i\right) + \sum_{i=2}^{N-2} F_z(d_i)\right] - (N-2), 0\right\}. \tag{32}$$

The objective function in the inner maximization problem in (32) is unfortunately not concave, making it challenging to find optimal solutions for the optimization problem. To solve this problem, we use a numerical procedure outlined in the Appendix. In our computations, we set $M = 50$ and compute the new *series CVaR based bounds* (SECB). The two bounds on the cumulative distribution function $\text{SECB}^+(x)$ and $\text{SECB}^-(x)$ are evaluated by taking the inverse of $\overline{\text{WCVaR}}_\alpha^{\Theta\varepsilon}(\sum_{i\in\mathcal{N}}\tilde{c}_i)$ and $\underline{\text{WCVaR}}_\alpha^{\Theta\varepsilon}(\sum_{i\in\mathcal{N}}\tilde{c}_i)$, respectively. Figure 3 shows the three bounds, $\text{RSB}(x)$, $\text{SECB}^+(x)$, and $\text{SECB}^-(x)$, for $N = 4$ and $N = 6$. Observe that $\text{SECB}^+(x)$ and $\text{SECB}^-(x)$ are fairly close to each other. Since the actual CVaR bound lies the between the two curves, the discrete approximation with $M = 50$ for CVaR bounds is reasonably good in this example. It is also clear that our proposed approximation significantly improves on the existing reduced standard bound.

4. Robust Portfolio Optimization

In this section, we implement the distributionally robust portfolio optimization approach in two financial datasets and compare it with the sample based approach. Consider a portfolio of N assets and let $\tilde{\boldsymbol{\xi}}$ be the random return vector of the assets. The random loss of ith asset is then simply $\tilde{c}_i = -\tilde{\xi}_i$. Given a feasible asset allocation $\boldsymbol{x} \in \mathcal{X}$, the computation of CVaR of the joint portfolio requires the distribution of the random return vector $\tilde{\boldsymbol{c}}$. Assume that we have access to historical data of a finite set of samples from the financial market denoted by the set $\mathcal{C}$. The sample distribution θ assigns a probability of $1/C$ to each sample vector in $\mathcal{C}$. The optimal sample-based allocation with the minimum CVaR is obtained by solving:

$$\min_{\beta\in\Re,\boldsymbol{x}\in\mathcal{X}} \left(\beta + \frac{1}{(1-\alpha)C}\sum_{\boldsymbol{c}\in\mathcal{C}}\left[\left(\boldsymbol{c}^T\boldsymbol{x} - \beta\right)^+\right]\right), \tag{33}$$

which is representable as a linear program. However, the out of sample performance of such a approach is not necessarily good due to the possibility that the out of sample distribution is different from the in-sample distribution. Using simulated data, Lim et al. (2011) have shown that the CVaR measure with sample-based optimization results in fragile portfolios that are often unreliable due to estimation errors. The approach we adopt in this paper is to use historical data to extract the

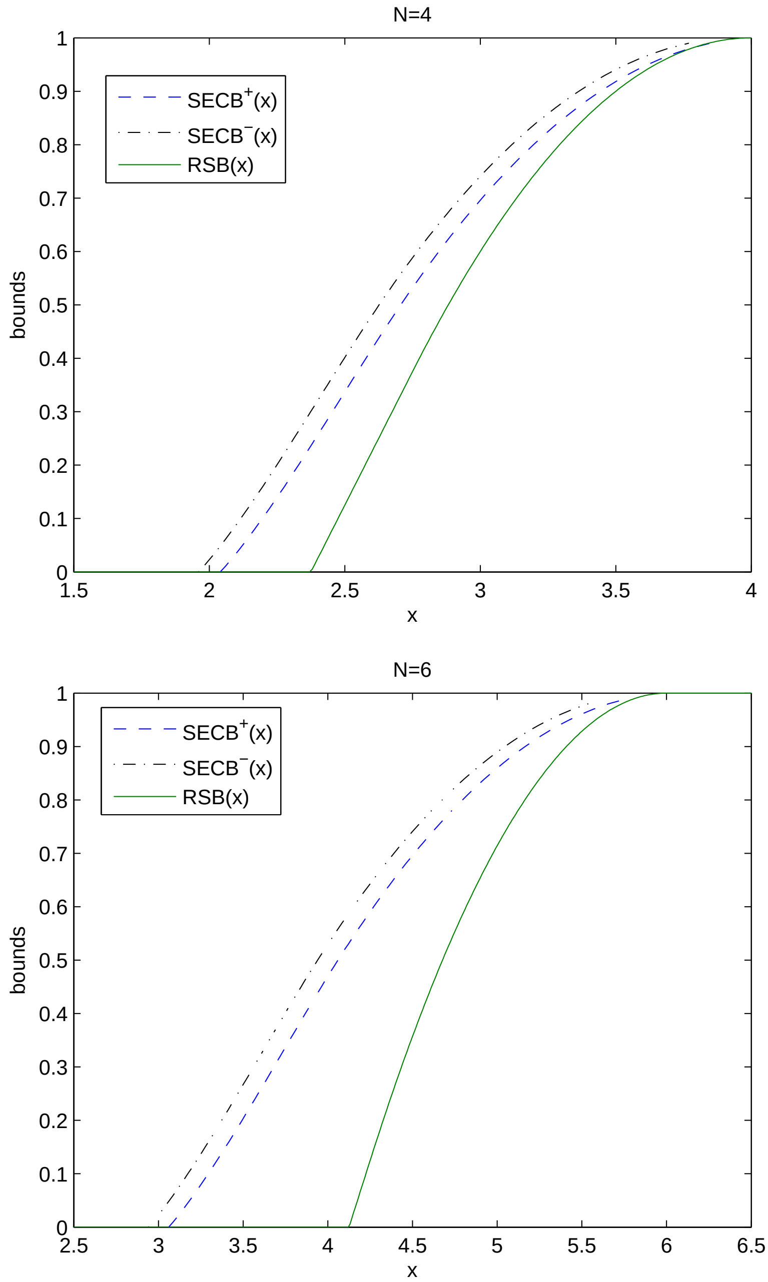

Figure 3 Different bounds of the cumulative distribution function of the sum of random variables with the simple series cover

stable dependencies among the random losses and only incorporate this reliable information into the optimization model. Given historical data, we construct a Fréchet class of distributions $\Theta_{\mathcal{E}}$ and solve the following distributional robust optimization problem:

$$\min_{\boldsymbol{x}\in\mathcal{X}}\max_{\theta\in\Theta_{\mathcal{E}}}\mathrm{CVaR}_{\alpha}^{\theta}(\tilde{\boldsymbol{c}}^T\boldsymbol{x}). \tag{34}$$

Using the dual representation in (28), the distributional robust portfolio optimization problem is formulated as:

$$\begin{aligned}\min_{g_r(.),\beta,\boldsymbol{x}}\ & \left(\beta+\frac{1}{1-\alpha}\sum_{r\in\mathcal{R}}\mathbb{E}_{\theta_r}\left[\left(\tilde{\boldsymbol{c}}_r^T(\boldsymbol{\eta}_r\circ\boldsymbol{x}_r)-g_r\left(\tilde{\boldsymbol{c}}_{\mathcal{K}_r}\right)+\sum_{t>r:\sigma_t=r}g_t\left(\tilde{\boldsymbol{c}}_{\mathcal{K}_t}\right)-\frac{\beta}{R}\right)^+\right]\right)\\ \text{s.t.}\quad & \boldsymbol{x}\in\mathcal{X}.\end{aligned}\tag{35}$$

If $\mathcal{X}$ is a polyhedron and the multivariate marginal support set of $\tilde{\boldsymbol{c}}_r$ is $\mathcal{C}_r$ with each sub-vector in the set equally likely, problem (35) is solvable as a linear optimization problem:

$$\begin{aligned}\min_{g_r(.),\beta,\boldsymbol{x},z_r(.)}\ & \beta+\frac{1}{1-\alpha}\sum_{r\in\mathcal{R}}\sum_{\boldsymbol{c}_r\in\mathcal{C}_r}\frac{z_r(\boldsymbol{c}_r)}{C_r}\\ \text{s.t.}\ & z_r(\boldsymbol{c}_r)+g_r\left(\boldsymbol{c}_{\mathcal{K}_r}\right)-\sum_{t>r:\sigma_t=r}g_t\left(\boldsymbol{c}_{\mathcal{K}_t}\right)\geq\boldsymbol{c}_r^T(\boldsymbol{\eta}_r\circ\boldsymbol{x}_r)-\frac{\beta}{R},\ \forall\boldsymbol{c}_r\in\mathcal{C}_r,\forall r\in\mathcal{R},\\ & z_r(\boldsymbol{c}_r)\geq 0,\qquad\qquad\forall\boldsymbol{c}_r\in\mathcal{C}_r,\forall r\in\mathcal{R},\\ & \boldsymbol{x}\in\mathcal{X}.\end{aligned}\tag{36}$$

Next, we discuss a data-driven approach to construct the Fréchet class of distributions of asset returns $\Theta_{\mathcal{E}}$.

4.1. Construction of Regular Covers

In the context of distributionally robust optimization, the dependency structure of the random variables is often incorporated using moment information. Some of the common classes of distributions employed in the financial literature are distributions with first and second moment information (see for example, El Ghaoui et al. (2003), Natarajan et al. (2009a), and Delage and Ye (2010)) and multivariate normal distributions with parameter uncertainty in the mean and covariance matrix (see Garlappi et al. (2007)). The resulting optimization formulations are tractable conic programs. Our approach is to use a Fréchet class of distributions with possibly overlapping marginals to capture information of dependencies among the random parameters. An important aspect of such an approach is to identify the cover structure $\mathcal{E}$ to balance over-fitting the data and getting overly conservative solutions due to lack of information. In order to construct the cover $\mathcal{E}$, we use time-dependent correlation information of the asset returns.

Our underlying assumption in identifying the cover is that we include pairs of assets in the same subset if the changes in correlation between the two assets over time is minimal. We propose the following two step data-driven approach to identify the regular cover:

Step 1: Split the historical data into two sets of equal size and construct the sampling distributions $\mathbb{P}_1$ and $\mathbb{P}_2$ for the asset returns from these two sets. For each pair of assets (i,j), we compute $\Delta\rho_{i,j} = |\rho_{i,j}^{\mathbb{P}_1} - \rho_{i,j}^{\mathbb{P}_2}|$, where $\rho_{i,j}^{\mathbb{Q}}$ is the correlation of the losses of two assets i and j for the distribution $\mathbb{Q}$. The simplest choice of correlation is the common Pearson correlation coefficient but other correlation measures could also be chosen. The objective is to identify the pairs (i,j) of assets with small values of $\Delta\rho_{i,j}$. Under the assumption that the stable dependency structure is captured by pairs with minimal change in correlation over time, these pairs of assets should be included in the same subset. There are different ways to identify such pairs of assets. In this paper, we implement two such approaches:

(a) **Minimum spanning tree approach (MST):** Use $\Delta\rho_{i,j}$ as the weight for the edge (i,j) in a complete graph of N vertices. Find the minimum spanning tree in this graph and keep all $N-1$ pairs of assets which define the tree. The choice of the minimum spanning tree implies that pairs with small changes in correlation coefficients over time are likely to be selected. The MST approach is inspired from the Chow-Liu method (Chow and Liu (1968)) which provides a second-order product approximation of a joint probability distribution using the mutual information measure and the spanning tree algorithm. We employ a similar method but use changes in the correlation coefficients as the cost terms.

(b) **Edge budgeting based approach (EB):** Remove all the pairs except for a fraction $r_a \in [0,1]$ of the total number of $N(N-1)/2$ pairs with the smallest values of $\Delta\rho_{i,j}$. Clearly, if $r_a = 0$, no pair will be selected. On the other hand, if $r_a = 1$, we keep all the pairs. The parameter r_a allows us to control the number of pairs of assets to be selected.

Step 2: Construct an undirected graph $\mathcal{G}$ where the set of nodes is the set of assets, and the set of edges is the pairs of assets selected in Step 1. If the graph $\mathcal{G}$ is chordal then one can construct a regular cover $\mathcal{E}$ efficiently. A linear time lexicographic breadth-first search (L-BFS) algorithm is used to determine whether a graph is chordal and to construct the regular cover $\mathcal{E}$ (see Rose et al. (2004) and Tarjan and Yannakakis (1984)).

For MST approach, the graph is a tree and hence chordal. The resulting cover $\mathcal{E}$ has $N-1$ two-element subsets corresponding to individual selected pairs. The MST cover can be viewed as a generalization of the simple star and simple series covers. In general, the resulting graph from EB approach is not chordal. If the graph $\mathcal{G}$ is not chordal, one adds in a set of additional edges, which are called *fill-in* edges, to make the graph chordal. Even though the problem of finding the fill-in with the minimum number of edges is NP-complete (see Yannakakis (1981)), there are efficient algorithms to find fill-ins with reasonably small number of edges (see for example, Huang and Darwiche (1996) and Natanzon et al. (2000)). In our experiments, we use the MINWEIGHTELIMORDER function in PMTK3, a Matlab toolkit for probabilistic modeling (see Dunham and Murphy (2012)), which is based on a fill-in algorithm developed by Huang and Darwiche (1996). One can then construct a regular cover $\mathcal{E}$ from the modified chordal graph using the L-BFS algorithm. Note that this data-driven approach of identify regular covers is only heuristic. When we use fill-in edges, there will pairs of assets with larger change in correlation over time included in the cover. We will show in Section 4 some examples of how many fill-in edges are needed for the regular covers and how changes in correlation over time of these additional pairs of assets are compared with those of original ones.

Given a regular cover $\mathcal{E}$, the marginals are constructed from historical data. Financial data is however non-stationary. While one could use the original sampling marginals if sufficient stationary historical data is available, in our experiments we found that it was difficult to get a nontrivial Fréchet class of distribution with the data of a few hundred days. To tackle this issue, we round the samples to ensure that the Fréchet class of distributions is non-trivial. The approach we adopt is to cluster the historical data of each asset return into several clusters and to replace the data within each cluster by the respective cluster mean. The marginals with the rounded samples are then used in the optimization approach. Under this construction, the mean of the rounded samples remains the same as that of original ones. Another benefit is that the size of supports of marginal distributions is reduced and this reduces the computational time to solve the problems. In the next section, we investigate the effects of the rounding procedure as well as the effects of using dependence structures in numerical experiments with real financial market data.

4.2. Dataset 1: Fama-French Portfolio

Given the volatility of financial markets, investors re-balance their portfolios periodically. We solve the portfolio optimization problem in each period under the assumption that historical daily return data of assets from the last two periods are available and are used to estimate distributions of daily returns in the current period. We allow for short selling and consider the following set of allowable allocations:

$$\mathcal{X} = \left\{ \boldsymbol{x} \in \Re^n : \boldsymbol{e}^T \boldsymbol{x} = 1, \boldsymbol{\mu}^T \boldsymbol{x} \geq \mu_t \right\},$$

where $\boldsymbol{e}$ is the vector of all ones, μ_t is the target return and $\boldsymbol{\mu}$ is the expected return vector of all assets. Given the Fréchet class of distributions obtained either from the MST or EB approach, we solve (36) to find the portfolio allocation in each period. In order to evaluate results obtained from the distributionally robust optimization approach, we compare the results with two other approaches:

1. **Sample-based approach (SB):** The original sample distribution is used and the allocation is computed by solving the problem (33).

2. **Rounded-sampled-based approach (RSB):** The rounded sample distribution is used in (33) instead of the original sample distribution. This strategy serves as a control to validate that the effect of the rounding procedure is not drastic.

The first data set we analyze consists of historical daily returns of an industry portfolio obtained from the Fama & French data library (French and Fama (2013)). The portfolio consists of NYSE, AMEX and NASDAQ stocks classified by industry. This include industries such as finance, health, textiles, food and machinery. A total of 4400 observations of daily return data were available in a period spanning approximately 15 years before the financial crisis, from August 18, 1989 to February 1, 2007. Consider an investor who plans to invest in the portfolio with $N = 49$ risky assets. He would like to minimize the risk of his investment, while guaranteeing a certain level of average return by choosing an appropriate trading strategy. The investor re-balances his portfolio every 200 days. We divide the 4400 samples into 22 periods, with each period consisting of 200 days. The investor starts his investment from the beginning of the third period. From then on, at the beginning of each period, the investor uses the portfolio return data of the last two periods to make the decision on the portfolio allocation for the current period.

In the experiments, we cluster the return data into 10 clusters. We use the R package Ckmeans.1d.1p, which is based on a k-means clustering dynamic programming algorithm in one dimension (see Wang and Song (2011)). The target return μ_t is varied between 0.04% and 0.08%. For each target return, we apply the four trading strategies for 20 periods. We then compute the aggregate out of sample mean and out of sample CVaR. The out of sample efficient frontier is constructed by varying the target return. The numerical tests were conducted in 64-bit Matlab 2011a with the CVX solver (see CVX Research Inc. (2012)).

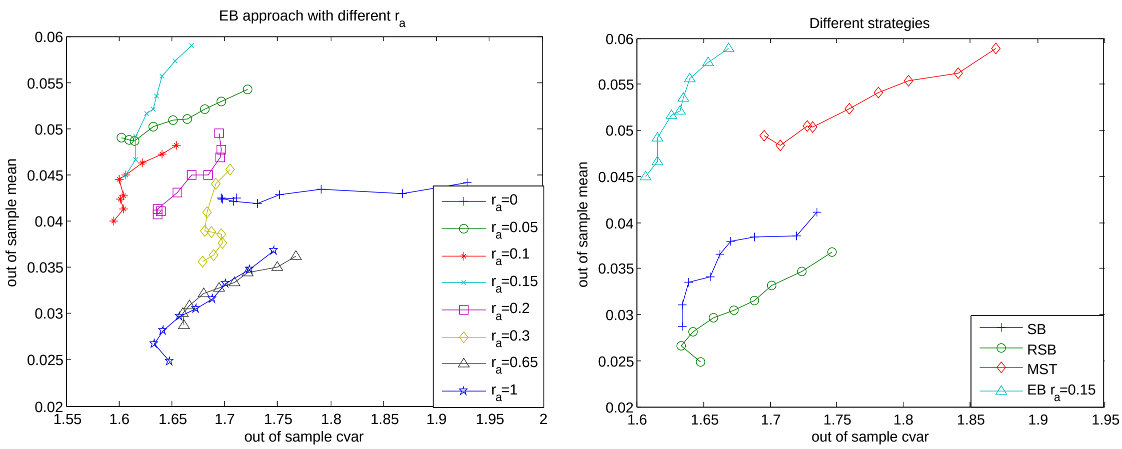

Figure 4 Out-of-sample efficient frontiers of different strategies with $\alpha = 0.95$

The graph on the left in Figure 4 shows the out of sample efficient frontiers of the EB trading strategy for different values of the parameter r_a when $\alpha = 0.95$. If $r_a = 0$, the EB strategy uses only univariate marginals and in this example, its efficient frontier is worse than that of the EB strategy for other small values of r_a. This is to be expected since we use no dependency information from the financial market in this case. As r_a increases, the performance of the EB strategy improves, and the best efficient frontier is achieved around $r_a = 0.15$. The performance then gradually deteriorates as r_a continues to increase to 1. This result indicates that by using only partial dependency information it is possible to enhance the performance of the trading strategy in the out of sample data. The graph on the right in Figure 4 shows the efficient frontiers of the four different strategies. The EB strategy is plotted for the optimum value $r_a = 0.15$. We can see that optimal EB and MST strategies are the better performing strategies in comparison to SB and RSB.

A well-known phenomenon in financial data is that the estimation of the out of sample mean is inaccurate (see Merton (1980)). The out of sample means are between 0.03% and 0.055%, while the target returns are between 0.04% and 0.08%. We conduct an experiment directly using the out of sample mean data in the optimization formulation. While clearly impractical, this serves to check the effect of the inaccuracies in the estimation of the mean return on the comparative performance of the different strategies. Figure 5 shows that in this case the EB strategy is the best performing strategy while the MST strategy does not perform as well. From these experiments, we conclude that the optimal EB strategy achieves the best performance in this dataset. Note that our approach is completely data-driven from identifying the cover to computing the optimal portfolio.

4.2.1. Robustness Tests In this section, we test the robustness of the results, by implementing the distributional robust optimization model with a few modifications.

1. In the first test, we vary the CVaR parameter α. The results with $\alpha = 0.9$ are displayed in Figure 6. The best EB strategy is obtained around $r_a = 0.1$. Similar to the results obtained with $\alpha = 0.95$, the performance of EB and MST are better than sample-based approaches and the efficient frontiers of SB and RSB are fairly close to each other.

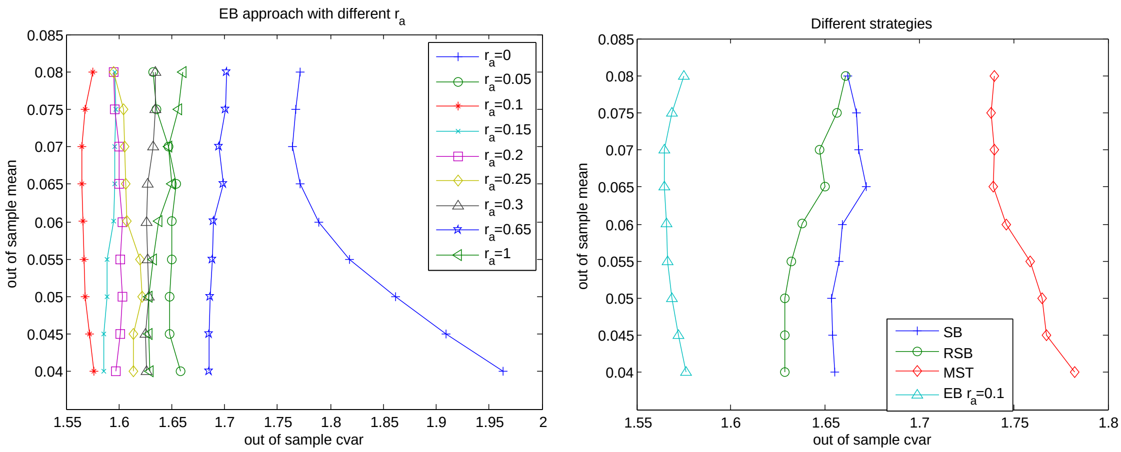

Figure 5 Out-of-sample efficient frontiers of different strategies with $\alpha = 0.95$ and out of sample mean fixed

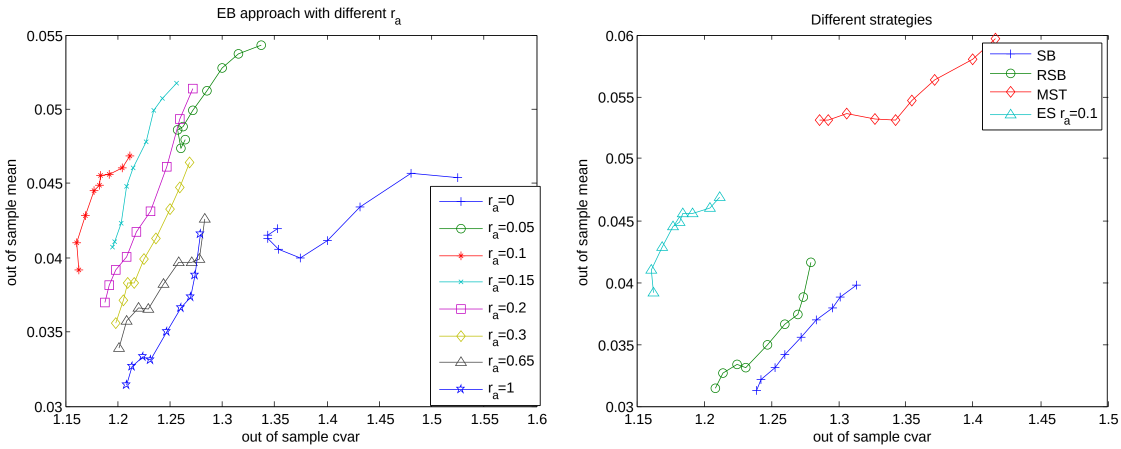

Figure 6 Out-of-sample efficient frontiers of different trading strategies with $\alpha = 0.9$

2. In the second test, we evaluate the effects of the rounding procedure by repeating the experiment with 20 support points. The results are displayed in Figure 7. From the figure, it is clear that one reaches a similar conclusion regarding the effectiveness of the optimal EB and MST approach and the closeness between efficient frontiers of SB and RSB strategies.

We also ran the numerical experiments with 40 support points and average computational times of all models are reported in Table 3. Given the fixed number of historical data ($M = 400$), computational times of the sample-based approaches does not depend on the number K of support points of (univariate) marginal distributions. The effect of K on computational time is prominent

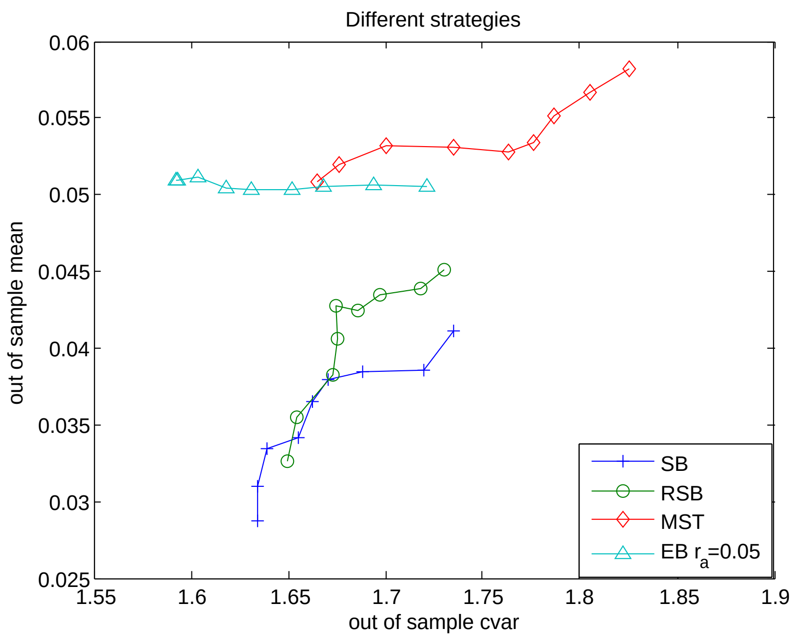

Figure 7 Out-of-sample efficient frontiers of different trading strategies with 20 supports and $\alpha = 0.95$

for the univariate marginal (EB with $r_a = 0.00$) and MST approach while it is much less for the general EB approach.

3. In the third test, we verify the results by using a nonparametric correlation measure in generating the cover. We make use of the Kendall tau rank correlation measure (see Kendall (1938) and Embrechts et al. (2002)). Given a set of n observations

$$(v_1, w_1), (v_2, w_2), \ldots, (v_T, w_T),$$

we call a pair (i, j) concordant if $v_i \geq v_j, w_i \geq w_j$ or $v_i \leq v_j, w_i \leq w_j$; otherwise, the pair is called disconcordant. The Kendall tau is defined as:

$$\tau = \frac{2(T_{con} - T_{dis})}{T(T-1)},$$

where T_{con} is the number of concordant pairs and T_{dis} is the number of disconcordant pairs. Then $-1 \leq \tau \leq 1$ with $\tau = 1$ if the agreement between the rankings is perfect and $\tau = -1$ if the disagreement between the two rankings is perfect. The results displayed in Figure 8 where *coeff* corresponds to the Pearson correlation coefficient and *tau* corresponds to Kendall tau correlation.

Approach	K	Variable number	Constraint number	Computational time (sec)
SB/RSB (EB $r_a = 1.00$)	10	850	402	1.12
	20	-	-	-
	40	-	-	-
Univariate marginal (EB $r_a = 0.00$)	10	1078	492	1.32
	20	2058	982	2.64
	40	4018	1962	5.20
EB $r_a = 0.15$	10	31879	11302	82.84
	20	36535	12854	102.07
	40	39215	13702	106.11
MST	10	6373	2905	15.12
	20	17665	8316	49.00
	40	32391	15209	98.95

Table 3 Computational times with different numbers of support points

We observe that the insights are similar as before, namely the optimal EB and MST strategies cases outperform the SB and RSB approaches significantly.

4. In the final test, we ran the numerical experiments for two additional time intervals: the time interval from August 21, 2006 to August 10,2010, which covers the recent financial crisis, and the after-crisis time interval from October 23, 2009 to July 31, 2014. Note that the first two periods of 200 days each in these time intervals are only used as historical data while decisions are made from the third period onwards. The efficient frontiers of the four different strategies, which include the EB approach with the best value of the parameter r_a, are plotted in Figure 9. The EB approach again performs better than other approaches in both time interval, which is consistent with other settings. The MST approach is very good in terms of out-of-sample return for the after-crisis time interval but not for the crisis one. The best values of r_a for the EB approach are 0.75 and 0.55 for

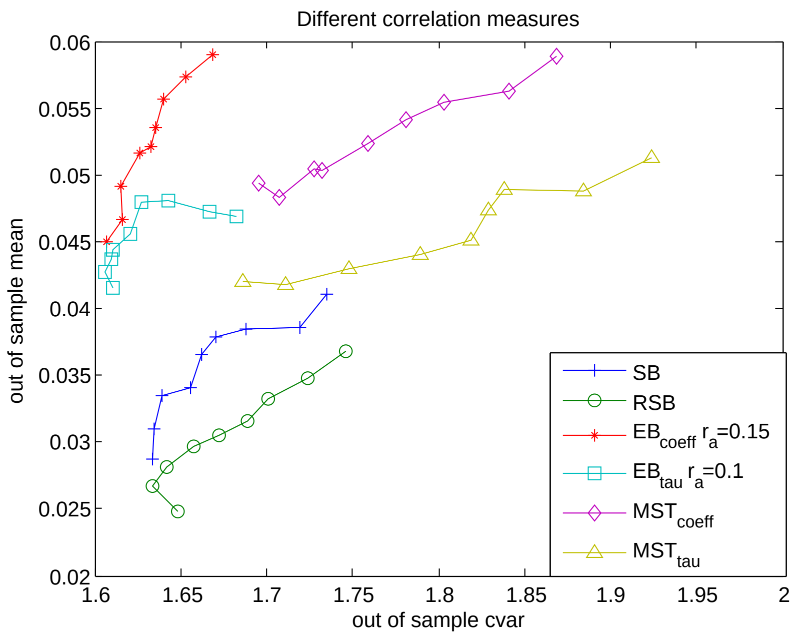

Figure 8 Out-of-sample efficiency frontier of different correlation measures with 10 supports and $\alpha = 0.95$

the crisis and after-crisis periods, respectively. It seems more dependence information is needed in these periods, especially the crisis periods.

We conclude this section by showing an example of resulting cover from the EB strategy. In this example, almost all of 48 two-element subsets of MST cover appears in the subsets of EB cover. This implies that there is greater dependence information assumed in the EB approach in comparison to the MST approach. The EB cover in this example has 37 subsets, with the largest subsets consisting of 13 elements. The top diagram in Figure 10 shows the first four subsets of this particular cover while the bottom diagram shows the fourth subset and five additional ones. Note that the cover structure is already much more complicated than the tree structure of the MST cover. In this example, there are original 176 edges obtained from Step 1 of the EB approach. Using the fill-in algorithm developed by Huang and Darwiche (1996), we add 84 additional edges in Step 2 to generate a regular cover. The distribution of $\Delta\rho_{i,j}$ of all edges (i, j) is shows in Figure 11 as well that of additional edges. When we vary the value of r_a, the number of additional edges needed is also varied with respect to the number of original edges and the number of remaining edges. Figure 12 shows the relationship between these numbers. With respect to the number of original

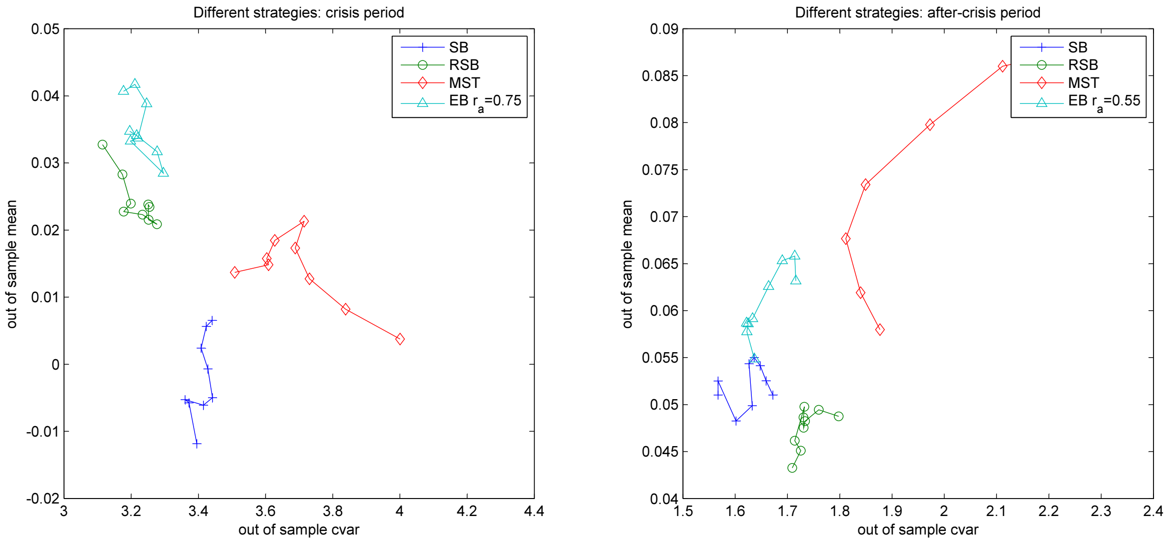

Figure 9 Out-of-sample efficient frontiers of different trading strategies for crisis and after-crisis periods

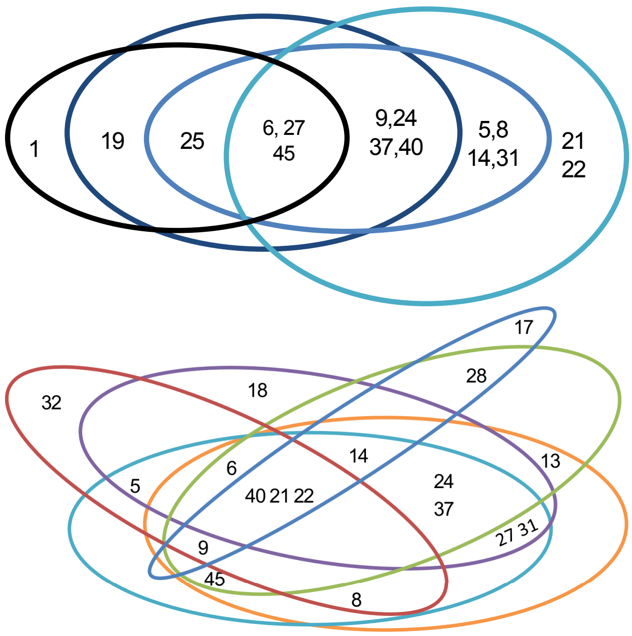

Figure 10 Partial Venn diagrams of an EB cover

edges, we need to add more edges when r_a is small. Figure 13 shows the distributions of difference of correlation when $r_a = 0.15$ for two cases with smallest and largest numbers of additional edges. It demonstrates the fill-in heuristic is reasonable in several cases while there are still cases in which we need to add a large number of additional edges with larger difference in correlation to make the cover regular.

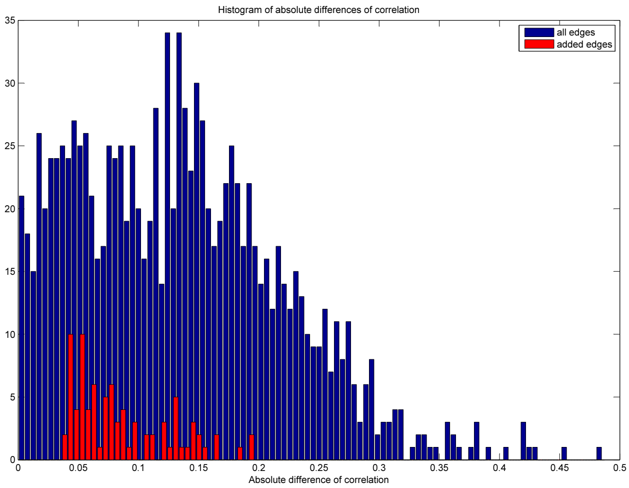

Figure 11 Distributions of difference in correlation

4.3. Dataset 2: Index Fund Portfolio

The second data set we analyze is from the OR-library (see Beasley (1990)). This data set was originally used for index tracking with the S&P100 index and 98 stocks. A detailed description of the data set can be found in Canakgoz and Beasley (2009). The data provides weekly stock prices, together with the index price, from March 1992 to September 1997, collected from DATASTREAM. The price data is transformed into return data using $\frac{p_{i+1}-p_i}{p_i}$ where p_i is the price on day i. Since the portfolio consists of an index and the individual stocks, in addition to the previous approaches we make use of the simple star cover for comparison purposes where the index forms the common star element. We group the 280 data points into 4 periods. The decision maker makes a decision

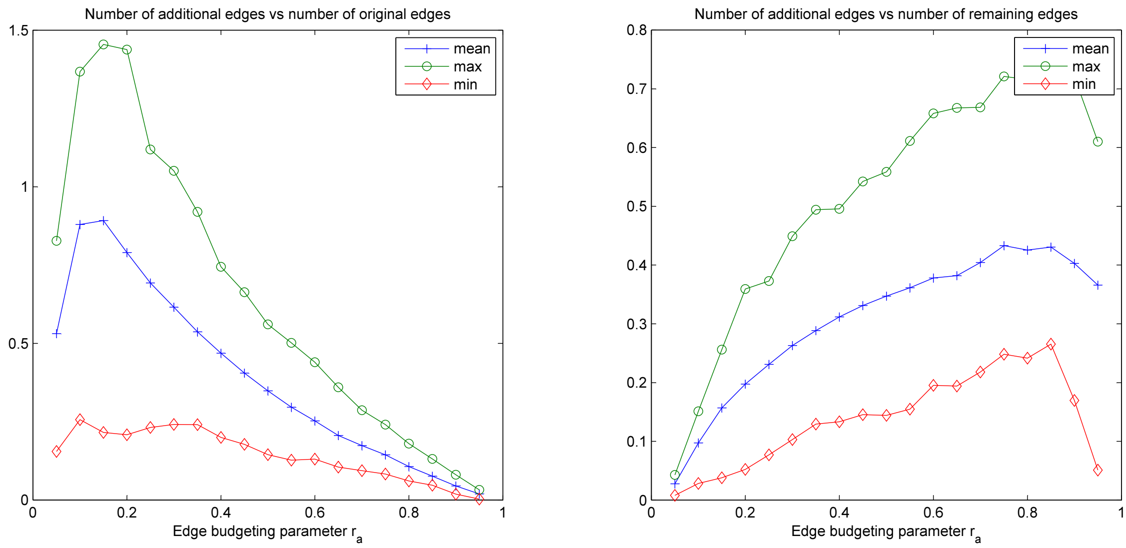

Figure 12 Comparison between the number of additional edges with that of original edges and remaining edges

at each period based on last two periods.The results are displayed in Figure 14. From the results, we find that the star cover (STAR) dominates the univariate marginal model, i.e., EB with $r_a = 0$, as well as SB and RSB significantly. However the best EB strategy (with $r_a = 0.02$ in this case) is even better than the STAR strategy. This suggests that the additional effort of finding a regular cover is useful in obtaining better out of sample performance. Note that the value of r_a is very small for the best EB strategy in this case which implies a weak dependence structure is assumed.

5. Conclusion

In this paper, we make use of the graph theoretic - running intersection property to develop a linear program to compute Fréchet bounds on random portfolio risks. The formulation is shown to be efficiently solvable for the discrete distribution case. New robust bounds on CVaR and VaR of the joint portfolio with overlapping multivariate marginal distribution information are provided.

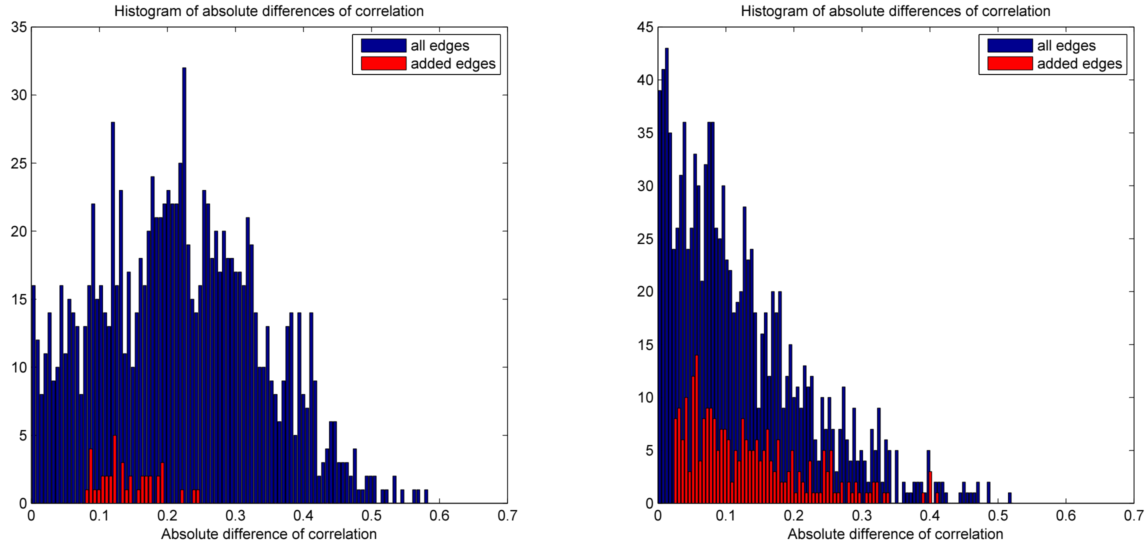

Figure 13 Distributions of difference in correlation with smallest and largest numbers of additional edges

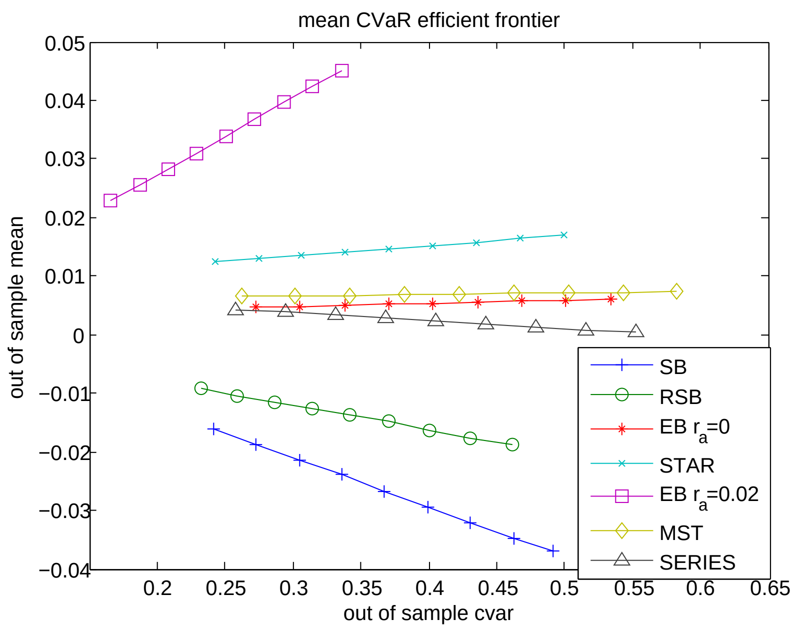

Figure 14 Out-of-sample efficiency frontier of different trading strategies

Based on the tight and efficiently solvable bounds, we propose a novel data-driven robust portfolio optimization model. This model identifies the overlapping cover structure by computing the changes in correlation over time. In conjunction with a linear optimization model, we show that the results help improve on the performance of sample based approaches.

We mention a couple of areas of possible future research. Firstly, while we restrict our attention to applications in robust portfolio optimization, the bounds proposed in this paper are much more general. Studying the implication of these bounds in areas such as queueing and inventory models is a natural extension. Secondly, we restrict our attention in this paper to regular covers with consistent set of marginals. Finding bounds when the cover is irregular or the marginals are inconsistent are open questions.

Appendix A: Proof of Lemma 2

$\Rightarrow$. Since the cover is connected, for all $r \in \mathcal{R} \setminus \{1\}$, there exists a sequence $s_1 = r, s_2, \ldots, s_m = r-1$ that links r to $r-1$. If $s_2 < r$, we have: $\mathcal{K}_r = \mathcal{J}_r \cap \left(\bigcup_{t=1}^{r-1} \mathcal{J}_t \right) \supseteq \mathcal{J}_r \cap \mathcal{J}_{s_2} \neq \emptyset$. If $s_2 > r$, there exist three consecutive indices in the sequence such that $s_{j-1} < s_j$ and $s_j > s_{j+1}$. We have: $\mathcal{J}_{s_{j-1}} \cap \mathcal{J}_{s_j} \neq \emptyset$ and $\mathcal{J}_{s_{j+1}} \cap \mathcal{J}_{s_j} \neq \emptyset$. Since $\mathcal{E}$ satisfies the RIP, we have $\mathcal{J}_{\sigma_{s_j}} \supseteq \mathcal{J}_{s_j} \cap \mathcal{J}_{\sigma_{s_j}} = \mathcal{K}_{s_j} = \mathcal{J}_{s_j} \cap \left(\bigcup_{t=1}^{s_j-1} \mathcal{J}_t \right) \supseteq \mathcal{J}_{s_j} \cap \left(\mathcal{J}_{s_{j-1}} \cup \mathcal{J}_{s_{j+1}} \right)$, thus:

$$\mathcal{J}_{\sigma_{s_j}} \supseteq \mathcal{J}_{s_j} \cap \mathcal{J}_{s_{j-1}}$$
$$\mathcal{J}_{\sigma_{s_j}} \supseteq \mathcal{J}_{s_j} \cap \mathcal{J}_{s_{j+1}}.$$

Taking the intersection of $\mathcal{J}_{s_{j-1}}$ and $\mathcal{J}_{s_{j+1}}$ on both sides respectively, we have:

$$\mathcal{J}_{\sigma_{s_j}} \cap \mathcal{J}_{s_{j-1}} \supseteq \mathcal{J}_{s_j} \cap \mathcal{J}_{s_{j-1}} \neq \emptyset$$
$$\mathcal{J}_{\sigma_{s_j}} \cap \mathcal{J}_{s_{j+1}} \supseteq \mathcal{J}_{s_j} \cap \mathcal{J}_{s_{j+1}} \neq \emptyset.$$

Thus we can replace s_j with σ_{s_j} in the sequence, with $\sigma_{s_j} < s_j$. Continuing on doing the process, we can find a sequence with $s_2 < r$.

$\Leftarrow$. We shall prove by induction on R. When $R = 2$, if $\mathcal{K}_2 = \mathcal{J}_1 \cap \mathcal{J}_2 \neq \emptyset$, the cover is obviously connected. Suppose the statement is true for $R = k$, let us consider $R = k+1$. Since $\mathcal{K}_r \neq \emptyset$ for all $r = 2, \ldots, k$, the subsets $\mathcal{J}_1, \ldots, \mathcal{J}_k$ are connected. Since $\mathcal{K}_{k+1} \neq 0$, $\mathcal{J}_{k+1} \cap \mathcal{J}_{\sigma_{k+1}} \neq \emptyset$. Thus for all $r = 2, \ldots, k$, there exists a sequence linking r with $k+1$ via σ_{k+1}.

Appendix B: Algorithm for Numerical Example in Section 3.2.1

Note that if there exists $i=1,\dots,N-2$ such that $d_i \le 0$, the inequality $H(\boldsymbol{d};x)\le 0$ always holds since F_y and F_z are cumulative distribution functions. Thus we can restrict the feasible region to the positive orthant, $d_i>0$ for all $i=1,\dots,N-2$. Similarly, we only need to consider solutions that satisfies $d_{N-1}=x-\sum_{i=1}^{N-2} d_i>0$. Note that F_y and F_z are both continuously differentiable in $(0,+\infty)$. Consider the first-order necessary optimality conditions, $\nabla H(\boldsymbol{d};x)=\mathbf{0}$:

$$F_y'(d_1)=F_y'\left(x-\sum_{i=1}^{N-2} d_i\right)=F_z'(d_i),\qquad i=2,\dots,N-2. \tag{37}$$

In order to solve this system of equations, we need the derivative F_y' and F_z':

$$F_y'(d)=\begin{cases}2d, & 0<d<1/2,\\ 1, & 1/2\le d<1,\\ -2d+3, & 1\le d<3/2,\\ 0, & d\ge 3/2,\end{cases}\qquad \text{and}\qquad F_z'(d)=\begin{cases}4d, & 0<d<1/2,\\ -4d+4, & 1/2\le d<1,\\ 0, & d\ge 1.\end{cases}$$

We then need to consider (37) three distinct cases with $F_y'(d_1)=0$, $0<F_y'(d_1)<1$, and $F_y'(d_1)=1$.

1. $F_y'(d_1)=0$: Since $d_1>0$, we have $d_1\ge 3/2$. Similarly, we have $d_i\ge 1$ for all $i=2,\dots,N-2$ and finally, $d_1=x-\sum_{i=1}^{N-2} d_i>3/2$. This case happens only when $x=\sum_{i=1}^{N-1} d_1\ge (N-3)+2\times\frac{3}{2}=N$, which is the trivial case with $\text{RBS}(x)=1$ since $\sum_{i\in\mathcal{N}}\tilde{c}_i<N$ almost surely.

2. $0<F_y'(d_1)<1$: Let $z=F_y'(d_1)/4$, we have $z\in(0,1/4)$ and $F_y'(2z)=F_y'(d_1)$. Given the formulation of $F_y'(d)$, we can easily show that $d_1\in\{2z,3/2-2z\}$. Similarly, we have: $d_{N-1}\in\{2z,3/2-2z\}$. Now consider d_i, $i=2,\dots,N-2$, we have $F_z'(z)=F_z'(1-z)=F_z'(d_i)$. Thus $d_i\in\{z,1-z\}$ for all $i=2,\dots,N-2$. Let k be the number of decision variables among d_i, $i=2,\dots,N-2$ that take the value of z. We have: k can take any value from 0 to $N-3$. Similarly, let l be the number of decision variables among $\{d_1,d_{N-1}\}$ that take the value of $2z$, $l=0,1,2$. Using the constraint $\sum_{i=1}^{N-1} d_i=x$, we obtain the following equation on z:

$$(N-k-3l/2)-(N+1-2k-4l)z=x.$$

If this equation results in a solution $z\in(0,1/4)$, we achieve a set of solutions $\boldsymbol{d}$ of the original problem which satisfy the first-order optimality condition (37). It means we would need to consider $3(N-2)$ possible value pairs of (k,l) and check the feasibility of z to find all potential candidates of the optimal solution for this case.

3. $F_y'(d_1)=1$: In this case, we have $d_1\in[1/2,1]$ and so is d_{N-1}. For $i=2,\dots,N-2$, we have $d_i\in\{1/4,3/4\}$. Similarly to the previous case, we let k be the number of decision variables among d_i, $i=2,\dots,N-2$ that take the value of $1/4$, we then have:

$$(3N-9-2k)/4+d_1+d_{N-1}=x.$$

In addition, the objective value in this case can be computed as

$$H(\boldsymbol{d};x) = d_1 + d_{N-1} - 1/2 + kF_z(1/4) + (N-3-k)F_z(3/4) - (N-2).$$

Thus, we just need to check whether $y = d_1 + d_{N-1} = x - (3N-9-2k)/4$ belong to the interval $[1,2]$. We need to perform this feasibility check for $N-2$ different values of k.
Following the analysis of these cases, we can find the optimal solution among all potential candidates, which will help us compute the standard bound RSB(x).

Acknowledgments

The authors would like to thank the Associate Editor and two anonymous reviewers for their valuable comments and detailed suggestions on improving the manuscript. The authors would also like to thank Zheng Zhichao for useful discussions regarding the paper. The research for the third author was partly funded by the SUTD-MIT International Design Center grant number IDG31300105 on 'Optimization for Complex Discrete Choice' and the MOE Tier 2 grant number MOE2013-T2-2-168 on 'Distributional Robust Optimization for Consumer Choice in Transportation Systems'.

References

Beasley, J. E. 1990. OR-Library: Distributing test problems by electronic mail. *Journal of the Operational Research Society* **41**(11) 1069–1072.

Beeri, C., R. Fagin, D. Maier, M. Yannakakis. 1983. On the desirability of acyclic database schemes. *Journal of the Association of Computing Machinery* **30**(3) 479–513.

Berge, C. 1976. *Graphs and Hypergraphs*. North-Holland, Amsterdam.

Bertsimas, D., K. Natarajan, C. P. Teo. 2006. Persistence in discrete optimization under data uncertainty. *Mathematical Programming, Series B* **108**(2-3) 251–274.

Canakgoz, N. A., J. E. Beasley. 2009. Mixed-integer programming approaches for index tracking and enhanced indexation. *European Journal of Operational Research* **196**(1) 384–399. URL `http://dblp.uni-trier.de/db/journals/eor/eor196.html#CanakgozB09`.

Chow, C. K., C. N. Liu. 1968. Approximating discrete probability distributions with dependence trees. *IEEE Transactions on Information Theory* **14**(3) 462–467.

CVX Research Inc. 2012. CVX: Matlab software for disciplined convex programming, version 2.0 beta. `http://cvxr.com/cvx`.

Delage, E., Y. Ye. 2010. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research* **58**(3) 595–612.

Denuit, M., C. Genest, É. Marceau. 1999. Stochastic bounds on sums of dependent risks. *Insurance: Mathematics and Economics* **25**(1) 85–104.

Doan, X. V., K. Natarajan. 2012. On the complexity of non-overlapping multivariate marginal bounds for probabilistic combinatorial optimization problems. *Operations Research* **60**(1) 138–149.

Dunham, M., K. Murphy. 2012. PMTK3: Probabilistic modeling toolkit for Matlab/Octave, version 3. `https://code.google.com/p/pmtk3/`.

El Ghaoui, L., M. Oks, F. Oustry. 2003. Worst-case value-at-risk and robust portfolio optimization: a conic programming approach. *Operations Research* **51**(4) 543–556.

Embrechts, P., A. Höing, A. Juri. 2003. Using copulae to bound the Value-at-Risk for functions of dependent risks. *Finance and Stochastics* **7**(2) 145–167.

Embrechts, P., A. McNeil, D. Straumann. 2002. Correlation and dependence in risk management: properties and pitfalls. *Risk Management: Value at Risk and Beyond* 176–223.

Embrechts, P., G. Puccetti. 2006. Bounds for functions of dependent risks. *Finance and Stochastics* **10**(3) 341–352.

Embrechts, P., G. Puccetti. 2010. Bounds for the sum of dependent risks having overlapping marginals. *Journal of Multivariate Analysis* **101**(1) 177–190.

Fréchet, M. 1940. *Les probabilités associées à un système d'événements compatibles et dépendants. I. Evenements en nombre fini fixe*. Actualités Sci. Ind., n. 859. Hermann et Cie, Paris.

Fréchet, M. 1951. Sur les tableaux de corrélation dont les marges sont données. *Annales de l'Université de Lyon, Section A, Series 3* **14** 53–77.

French, K., E. Fama. 2013. Fama & French data library. `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data\textunderscorelibrary.html`.

Garlappi, L., R. Uppal, T. Wang. 2007. Portfolio selection with parameter and model uncertainty: A multi-prior approach. *The Review of Financial Studies* **20**(7) 41–81.

Honeyman, P., R. E. Lander, M. Yannakakis. 1980. Testing the universal instance assumption. *Information Processing Letters* **10**(1) 14–19.

Huang, C., A. Darwiche. 1996. Inference in belief networks: A procedural guide. *International Journal of Approximate Reasoning* **15**(3) 225–263.

Jiroušek, R. 1991. Solution of the marginal problem and decomposable distribution. *Kybernetika* **27**(5) 403–412.

Jorion, P. 2001. *Value at Risk: The New Benchmark for Managing Financial Risk*. McGraw-Hill.

Kellerer, H. G. 1964. Verteilungsfunktionen mit gegebenen marginalverteilungen. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **3** 247–270.

Kellerer, H. G. 1991. Indecomposable marginal problems. G. D. Aglio, S. I. Kotz, G. Salinetti, eds., *Advances in Probability Distributions with Given Marginals*. Mathematics and Its Applications Volume 67, 139–149.

Kendall, M. G. 1938. A new measure of rank correlation. *Biometrika* **30**(1-2) 81–93. doi:10.1093/biomet/30.1-2.81. URL `http://biomet.oxfordjournals.org/content/30/1-2/81.short`.

Kleinberg, J., Y. Rabani, E. Tardos. 2000. Allocating bandwidth for bursty connections. *SIAM Journal of Computing* **30**(1) 191–217.

Kreinovich, V., S. Ferson. 2006. Computing best-possible bounds for the distribution of a sum of several variables is NP-hard. *International Journal of Approximate Reasoning* **41**(3) 331–342.

Lasserre, J. B. 2006. Convergent sdp-relaxations in polynomial optimization with sparsity. *SIAM Journal of Optimization* **17** 833–843.

Lauritzen, S. L., T. P. Speed, K. Vijayan. 1984. Decomposable graphs and hypergraphs. *Journal of the Australian Mathematical Society* **36** 12–29.

Lim, A. E. B., J. G. Shanthikumar, G-Y. Vahn. 2011. Conditional value-at-risk in portfolio optimization: Coherent but fragile. *Operations Research Letters* **39** 163–171.

Makarov, G. D. 1982. Estimates for the distribution function of the sum of two random variables with given marginal distributions. *Theory of Probability and Its Applications* **26** 803–806.

Markowitz, H. M. 1952. Portfolio selection. *Journal of Finance* **7** 77–91.

Meilijson, I., A. Nadas. 1979. Convex majorization with an application to the length of critical path. *Journal of Applied Probability* **16** 671–677.

Merton, R. C. 1980. On estimating the expected return on the market: an exploratory investigation. *Journal of Financial Economics* **8** 323–361.

Natanzon, A., R. Shamir, R. Sharan. 2000. A polynomial approximation algorithm for the minimum fill-in problem. *SIAM Journal on Computing* **30**(4) 1067–1079.

Natarajan, K., M. Sim, J. Uichanco. 2009a. Tractable robust expected utility and risk models for portfolio optimization. *Operations Research* **20**(4) 695–731.

Natarajan, K., M. Song, C. P. Teo. 2009b. Peristency model and its applications in choice modeling. *Management Science* **55**(3) 453–469.

Nemirovski, A., A. Shapiro. 2006. Convex approximations of chance constrained programs. *SIAM Journal of Optimization* **17** 969–996.

Puccetti, G., L. Rüschendorf. 2012. Bounds for joint portfolios of dependent risks. *Statistics and Risk Modelling* **29** 107–132.

Pucetti, G., L. Rüschendorf. 2013. Sharp bounds for sums of dependent risks. *Journal of Applied Probabililty* **50**(1) 42–53.

Rachev, S. T., L. Rüschendorf. 1989. *Mass Transportation Problems*, vol. I-II. Berlin Heidelberg New York: Springer.

Rockafellar, R. T., S. Uryasev. 2000. Optimization of conditional value-at-risk. *Journal of Risk* **2** 21–41.

Rockafellar, R. T., S. Uryasev. 2002. Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance* **26** 1443–1471.

Rose, D. J., R. E. Tarjan, G. S. Lueker. 2004. Algorithmic aspects of vertex elimination on graphs. *SIAM Journal on Computing* **5**(2) 266–283.

Rüschendorf, L. 1982. Random variables with maximum sums. *Advances in Applied Probability* **14**(3) 695–731.

Rüschendorf, L. 1991. Bounds for distributions with multivariate marginals. K. Mosler, M. Scarsini, eds., *Stochastic Orders and Decision under Risk*. IMS Lecture Notes - Monograph Series 19, 285–310.

Rüschendorf, L. 2005. Stochastic ordering of risks, influence of dependence and a.s. constructions. N. Balakrishnan, I. G. Bairamov, O. L. Gebizlioglu, eds., *Advances on Models, Characterizations and Applications*. CRC Press, 19–56.

Tarjan, R. E., M. Yannakakis. 1984. Simple linear-time algorithms to test chordality of graph, test acyclicity of hypergraphs, and selectively reduce acyclic hypergraphs. *SIAM Journal on Computing* **13** 566–579.

Vorob'ev, N. N. 1962. Consistent families of measures and their extensions. *Theory of Probability and Its Applications* **7** 147–163.

Wainwright, M. J., M. I. Jordan. 2008. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning* **1** 1–305.

Wang, B., R. Wang. 2011. The complete mixability and convex minimization problems with monotone marginal densities. *Journal of Multivariate Analysis* **102**(10) 1344–1360.

Wang, H., M. Song. 2011. Ckmeans.1d.dp: optimal k-means clustering in one dimension by dynamic programming. *The R Journal* **3**(2) 29–33.

Yannakakis, M. 1981. Computing the minimum fill-in is NP-complete. *SIAM Journal on Algebraic and Discrete Methods* **2**(1) 77–79.

Zhu, S., M. Fukushima. 2009. Worst-case conditional value-at-risk with application to robust portfolio management. *Operations Research* **57**(5) 1155–1168.

Zymler, S., D. Kuhn, B. Rustem. 2013. Worst-case value at risk of nonlinear portfolios. *Management Science* **59** 172–188.