JOURNAL OF INDUSTRIAL AND
MANAGEMENT OPTIMIZATION
Volume **10**, Number **4**, October **2014**

doi:10.3934/jimo.2014.10.1109

pp. **1109–1127**

CVAR PROXIES FOR MINIMIZING SCENARIO-BASED VALUE-AT-RISK

Helmut Mausser and Oleksandr Romanko

Quantitative Research, Risk Analytics, Business Analytics, IBM
185 Spadina Avenue, Toronto, ON M5T2C6, Canada

(Communicated by Hailiang Yang)

Abstract. Minimizing VaR, as estimated from a set of scenarios, is a difficult integer programming problem. Solving the problem to optimality may demand using only a small number of scenarios, which leads to poor out-of-sample performance. A simple alternative is to minimize CVaR for several different quantile levels and then to select the optimized portfolio with the best out-of-sample VaR. We show that this approach is both practical and effective, outperforming integer programming and an existing VaR minimization heuristic. The CVaR quantile level acts as a regularization parameter and, therefore, its ideal value depends on the number of scenarios and other problem characteristics.

1. **Introduction.** Minimizing a portfolio's Value-at-Risk (VaR) is a challenging optimization problem. One reason for this is that, aside from certain special cases, such as when losses are elliptically distributed, VaR is not a simple function of the positions in the portfolio. As a result, it is common practice to approximate the portfolio's loss distribution with a finite number of scenarios, and to optimize the VaR as estimated from this sample. Obtaining an accurate risk estimate, so that the optimized portfolio performs well on an out-of-sample basis, may demand an extremely large number of scenarios. This in itself is not necessarily problematic; optimizing other risk measures, such as the conditional VaR (CVaR), i.e., the average loss exceeding the VaR, also requires scenario approximation. However, VaR presents an additional challenge in that minimizing its estimator, the sample quantile, entails integer programming. This makes it increasingly difficult to find an optimal solution as the number of scenarios increases, so that this approach offers limited practical benefit. In contrast, CVaR optimization is a linear program, which can be solved much more readily. This computational advantage motivates using CVaR as a substitute, or proxy, for VaR in optimization problems.[1]

Further justification for optimizing CVaR in place of VaR stems from the fact that, for a given quantile level α, CVaR is an upper bound for VaR, i.e., $\mathrm{CVaR}_\alpha \geq$

2010 *Mathematics Subject Classification.* Primary: 91G10, 91G60, 91B30, 90C05; Secondary: 90C20, 90C11, 65K05.

Key words and phrases. Value-at-risk, conditional value-at-risk, optimization, regularization.

[1] Unlike CVaR, VaR generally is not subadditive, and thus only the former is a coherent risk measure [4]. While this makes CVaR preferable to VaR on theoretical grounds, herein we consider only their computational aspects, rather than their respective merits as risk measures per se. For discussion of the latter see, for example, Acerbi and Tasche [1] and Daníelsson et al. [6].

VaR_α.[2] Thus, one way to reduce VaR is simply to minimize its upper bound; intuitively, a portfolio that minimizes CVaR_α will also tend to have a low, if not minimal, VaR_α (see, for example, Rockafellar and Uryasev [17], Natarajan et al. [13]).

While minimizing an upper bound for VaR has a certain appeal, there is no reason to limit the CVaR quantile level only to α. A more general approach is to optimize $\mathrm{CVaR}_{\alpha'}$ for several quantile levels, $\alpha_{\min} \leq \alpha' \leq \alpha_{\max}$, and then to select the optimized portfolio having the lowest VaR_α. For example, Mausser and Rosen [12] used $95\% \leq \alpha' \leq 99.9\%$ in order to minimize VaR_α, for $\alpha = 99\%$ and 99.9%, with 20000 scenarios. However, they reported only in-sample results, making it impossible to assess the true performance of the CVaR proxies. Pagnoncelli et al. [15] considered a portfolio optimization problem with a chance-constraint, in the form of an upper bound for VaR. They used CVaR proxies with $80\% \leq \alpha' \leq 97\%$ to bound $\mathrm{VaR}_{0.90}$ with a scenario approximation of size 5000. Their results showed that, for α' in this range, the optimized portfolios were typically too conservative, i.e., on an out-of-sample basis, their $\mathrm{VaR}_{0.90}$ was usually below the upper bound.

Given these rather limited studies, questions remain about the effectiveness of using CVaR proxies in practice. Thus, one goal of this paper is to compare this approach, on both an in-sample and out-of-sample basis, with other scenario-based techniques for minimizing VaR. Specifically, we consider integer programming, subject to a computational time limit, as well as a heuristic procedure proposed by Larsen et al. [9]. The heuristic also makes use of CVaR proxies but allows certain scenarios to have arbitrarily large losses (by excluding them from the CVaR calculation). This is consistent with the fact that the VaR measure ignores the sizes of the losses that exceed the quantile while CVaR does not.

Additionally, this paper examines the relationship between the CVaR quantile level and the quality of the proxy. Since it is not possible, as far as we know, to determine the best quantile level for a given problem a priori, a trial-and-error approach that evaluates several different values of α' remains necessary in practice. A better understanding of the interplay between the CVaR quantile level, problem characteristics and performance, can help to identify good candidates for α'. In particular, we study how the portfolio loss distribution and the number of scenarios affect the best-performing CVaR quantile level. Similar investigations have been conducted previously in slightly different contexts. For example, when solving chance-constrained problems, the specified constraint violation probability needs to be chosen in accordance with the size of the scenario approximation (Luedtke and Ahmed [10], Pagnoncelli et al. [16]).[3] Recently, Mausser and Romanko [11] showed that $\mathrm{CVaR}_{\alpha'}$ is an effective proxy for minimizing CVaR_α, where $\alpha' \leq \alpha$ and α' approaches α as the number of scenarios increases. Note that in the case of CVaR_α, the proxy was chosen purely to improve out-of-sample performance, rather than for computational reasons.

The layout of this paper is as follows. Section 2 describes the estimators for VaR and CVaR. Section 3 formulates the VaR and CVaR optimization problems and

[2] Nemirovski and Shapiro [14] show that CVaR_α is, in fact, the best conservative convex approximation to VaR_α.

[3] This approach, which effectively uses $\mathrm{VaR}_{\alpha'}$ (rather than $\mathrm{CVaR}_{\alpha'}$) as a proxy for VaR_α, still requires solving integer programs. Thus, its applicability to financial risk management, where problems typically entail a large number of scenarios, is rather limited in practice.

explains the heuristic procedure of Larsen et al. [9]. The results of our computational experiments are reported in Section 4. Finally, Section 5 contains concluding remarks.

2. Estimators of VaR and CVaR.

Let the continuous random variable L, having distribution function F, denote the proportion of a portfolio's value that is lost over a given time horizon, i.e., L represents the "loss return", or the negative of the typical return.[4] VaR_α is the loss that is exceeded with probability $1-\alpha$, i.e.,

$$\mathbb{P}(L > \mathrm{VaR}_\alpha) = 1 - \alpha, \tag{1}$$

or, equivalently, $\mathrm{VaR}_\alpha = F^{-1}(\alpha)$. CVaR_α is the expected loss given that it exceeds VaR_α, i.e.,

$$\mathrm{CVaR}_\alpha = \mathbb{E}(L \,|\, L > \mathrm{VaR}_\alpha). \tag{2}$$

More generally, for continuous or discrete random variables, the respective definitions are

$$\mathrm{VaR}_\alpha := \inf\{x \in \mathbb{R} \,|\, \mathbb{P}(L > x) \le 1 - \alpha\} \tag{3}$$

and (see, for example, Acerbi and Tasche [2], Rockafellar and Uryasev [18])

$$\mathrm{CVaR}_\alpha := \frac{1}{1-\alpha}\left[\mathbb{E}(L \,|\, L \ge \mathrm{VaR}_\alpha) - \mathrm{VaR}_\alpha \cdot (\mathbb{P}(L \ge \mathrm{VaR}_\alpha) - 1 + \alpha)\right]. \tag{4}$$

Consider a sample of N losses, drawn randomly from F. Let $\ell_{(k)}$ denote the k^{th} order statistic of the sample losses, so that $\ell_{(1)} \le \ell_{(2)} \le \ldots \le \ell_{(N)}$. A common estimator of VaR_α, consistent with Equation 3, is the sample α-quantile

$$q_{\alpha,N} = \ell_{(\lceil N\alpha \rceil)}, \tag{5}$$

where $\lceil x \rceil$ is the smallest integer greater than or equal to x. It follows that in a sample of size N, the loss exceeds $q_{\alpha,N}$ in no more than $\lfloor N(1-\alpha) \rfloor$ scenarios, where $\lfloor x \rfloor$ is the integer part of x. From Equation 4, an estimator of CVaR_α is

$$h_{\alpha,N} = \frac{1}{N(1-\alpha)}\left[(\lceil N\alpha \rceil - N\alpha)\,\ell_{(\lceil N\alpha \rceil)} + \sum_{k=\lceil N\alpha \rceil + 1}^{N} \ell_{(k)}\right], \tag{6}$$

or equivalently,

$$h_{\alpha,N} = \frac{1}{1-\alpha}\int_\alpha^1 q_{p,N}\,dp. \tag{7}$$

For example, suppose $N = 100$ and that the three largest losses are $\ell_{(98)} = 0.42$, $\ell_{(99)} = 0.44$ and $\ell_{(100)} = 0.50$. The respective estimates of $\mathrm{VaR}_{0.98}$ and $\mathrm{CVaR}_{0.98}$ are $q_{0.98,100} = \ell_{(98)} = 0.42$ and $h_{0.98,100} = (\ell_{(99)} + \ell_{(100)})/2 = 0.47$. Conversely, estimates of $\mathrm{VaR}_{0.975}$ and $\mathrm{CVaR}_{0.975}$ are $q_{0.975,100} = \ell_{(98)} = 0.42$ and $h_{0.975,100} = (\ell_{(98)} + 2\ell_{(99)} + 2\ell_{(100)})/5 = 0.46$, respectively. Note that, under mild conditions, both $q_{\alpha,N}$ and $h_{\alpha,N}$ are consistent and asymptotically unbiased (Arnold et al. [3], Brazauskas et al. [5]).

---

[4] We use this convention so that losses are positive rather than negative.

3. **Problem formulations.** Let $\boldsymbol{w} = \{w_j,\ j = 1, \ldots, J\}$ denote a portfolio of J assets, where w_j is the weight of asset j, and suppose that the asset returns, $\boldsymbol{r}$, follow the continuous multivariate distribution Ψ. If the return of asset j is r_j then the portfolio's loss is

$$L(\boldsymbol{w}) = \sum_{j=1}^{J} -r_j w_j. \tag{8}$$

The actual VaR of portfolio $\boldsymbol{w}$ is denoted $\mathrm{VaR}_\alpha(\boldsymbol{w}, \Psi)$, explicitly recognizing its dependence on the asset weights and the joint asset returns.

Let $\Omega \subseteq \mathbb{R}^J$ be the set of acceptable portfolios. For example, Ω may comprise only portfolios with no short positions, or those for which no stock represents more than 5% of the total value. We assume that Ω is convex, specifically, that it is defined by one or more linear constraints, i.e., $\Omega = \{\boldsymbol{w} \in \mathbb{R}^J \mid A\boldsymbol{w} \leq \boldsymbol{b}\}$ for some matrix A and vector $\boldsymbol{b}$. Our goal is to find $\boldsymbol{w}^* \in \Omega$ for which $\mathrm{VaR}_\alpha(\boldsymbol{w}^*, \Psi)$ is minimal, i.e., $\boldsymbol{w}^*$ is a solution to the optimization problem

$$\begin{array}{ll} \min\limits_{\boldsymbol{w}} & \mathrm{VaR}_\alpha(\boldsymbol{w}, \Psi) \\ \text{s.t.} & \boldsymbol{w} \in \Omega. \end{array} \tag{9}$$

If Ψ is a multivariate elliptical distribution, such as the normal, then $\mathrm{VaR}_\alpha(\boldsymbol{w}, \Psi)$ can be expressed in terms of the means and covariances of the asset returns. Specifically, if $\boldsymbol{r} \sim E_J(\overline{\boldsymbol{r}}, \Sigma, \phi)$ and each r_j has finite variance then Problem 9 becomes

$$\begin{array}{ll} \min\limits_{\boldsymbol{w}} & -\overline{\boldsymbol{r}}^T \boldsymbol{w} + Z_\alpha \sqrt{\boldsymbol{w}^T \Sigma \boldsymbol{w}} \\ \text{s.t.} & \boldsymbol{w} \in \Omega, \end{array} \tag{10}$$

where Z_α is the α-quantile of the standardized distribution. Problem 10 can be solved efficiently using second order conic programming (SOCP).

In practice, however, solving Problem 9 often is complicated by the fact that Ψ is unavailable in closed-form, or results in a computationally complex objective function. In light of this, suppose that we generate a sample S^o comprising N random observations (scenarios) from Ψ. If r_{ij} is the return of asset j in scenario i then the associated loss return of the portfolio $\boldsymbol{w}$ is

$$\ell_i(\boldsymbol{w}, S^o) = \sum_{j=1}^{J} -r_{ij} w_j. \tag{11}$$

We then minimize $q_{\alpha,N}(\boldsymbol{w}, S^o)$, an estimate of $\mathrm{VaR}_\alpha(\boldsymbol{w}, \Psi)$, by solving the following scenario approximation to Problem 9

$$\begin{array}{ll} \min\limits_{\boldsymbol{w},q,\boldsymbol{z}} & q \\ \text{s.t.} & \ell_i(\boldsymbol{w}, S^o) - M z_i - q \leq 0,\ i = 1, \ldots, N \\ & \sum\limits_{i=1}^{N} z_i \leq \lfloor N(1-\alpha) \rfloor \\ & z_i \in \{0, 1\},\ i = 1, \ldots, N \\ & \boldsymbol{w} \in \Omega. \end{array} \tag{12}$$

Observe that the binary variable $z_i = 1$ if the loss exceeds q in scenario i (M is a suitably large constant that bounds the excess loss). It is apparent that q will take on the smallest possible value that is exceeded in at most $\lfloor N(1-\alpha) \rfloor$ scenarios, which is consistent with minimizing the sample α-quantile. Problem 12, a mixed-integer program (MIP) with N binary variables, becomes increasingly difficult to solve as the number of tail scenarios, $\lfloor N(1-\alpha)\rfloor$, increases.

Let $\boldsymbol{w}^*(q_{\alpha,N}, S^o)$ identify the optimal solution to Problem 12. The in-sample risk of this portfolio is $q_{\alpha,N}(\boldsymbol{w}^*(q_{\alpha,N}, S^o), S^o)$ while its actual risk is $\mathrm{VaR}_\alpha(\boldsymbol{w}^*(q_{\alpha,N}, S^o), \Psi)$. Typically, instances of Problem 12 that can be solved optimally require N to be so small that the inherent estimation error limits the usefulness of the results, i.e., $q_{\alpha,N}(\boldsymbol{w}^*(q_{\alpha,N}, S^o), S^o)$ is extremely small but $\mathrm{VaR}_\alpha(\boldsymbol{w}^*(q_{\alpha,N}, S^o), \Psi)$ is much larger than $\mathrm{VaR}_\alpha(\boldsymbol{w}^*, \Psi)$.

When $\mathrm{CVaR}_{\alpha'}$ is used as a proxy for VaR_α, we minimize $h_{\alpha',N}(\boldsymbol{w}, S^o)$ by solving the following scenario approximation problem (Rockafellar and Uryasev [17])

$$\begin{array}{ll}\min\limits_{\boldsymbol{w},u,\boldsymbol{y}} & u + \dfrac{1}{N(1-\alpha')}\displaystyle\sum_{i=1}^{N} y_i \\ \text{s.t.} & \ell_i(\boldsymbol{w}, S^o) - u - y_i \le 0,\ i = 1,\ldots,N \\ & y_i \ge 0,\ i = 1,\ldots,N \\ & \boldsymbol{w} \in \Omega.\end{array} \tag{13}$$

Note that u is effectively an estimate of $\mathrm{VaR}_{\alpha'}$ while y_i is the amount, if any, by which the loss return exceeds the estimated $\mathrm{VaR}_{\alpha'}$ in scenario i. Problem 13 is a linear optimization problem. While the solution time increases with N, it does so much more slowly than for Problem 12, so that Problem 13 can be solved efficiently even for large sample sizes. The in-sample VaR of the resulting optimal portfolio $\boldsymbol{w}^*(h_{\alpha',N}, S^o)$ is $q_{\alpha,N}(\boldsymbol{w}^*(h_{\alpha',N}, S^o), S^o)$ and its actual VaR is $\mathrm{VaR}_\alpha(\boldsymbol{w}^*(h_{\alpha',N}, S^o), \Psi)$. The proxy is said to be effective if $\mathrm{VaR}_\alpha(\boldsymbol{w}^*(h_{\alpha',N}, S^o), \Psi) < \mathrm{VaR}_\alpha(\boldsymbol{w}^*(q_{\alpha,N}, S^o), \Psi)$.

To highlight the differences between the VaR and CVaR scenario approximation problems, suppose $N = 100$ and $\alpha = \alpha' = 95\%$. In this case, Problem 12 allows five losses to be arbitrarily large while minimizing the sixth-largest loss. In contrast, Problem 13 minimizes the average of the five largest losses. Clearly, reducing the five largest losses tends (indirectly) to lower the sixth-largest loss as well, so that $h_{0.95,100}$ is an intuitively appealing proxy for $q_{0.95,100}$.

Since the VaR measure is indifferent to the sizes of losses that exceed the quantile, one might expect further reductions in the VaR, i.e., the sixth-largest loss, to be obtained if the CVaR problem ignores the magnitudes of the five largest losses. This motivates the VaR minimization heuristic of Larsen et al. [9], which iteratively constructs a set of "inactive" scenarios from the initial sample S^o, whose losses are allowed to become arbitrarily large, while minimizing CVaR at a specified quantile level for the remaining "active" scenarios. In each iteration, a constant proportion, $0 < \xi \le 1$, of the active scenarios whose losses exceed the current VaR estimate is discarded, or made inactive. The process terminates when no more active scenarios can be discarded.

More precisely, to minimize VaR_α, the heuristic first solves Problem 13 with $\alpha' = \alpha$ to obtain an initial solution $\boldsymbol{w}_0^* \equiv \boldsymbol{w}^*(h_{\alpha,N}, S^o)$. In iteration $k > 0$, S^o is partitioned into sets of active (S_k^o) and inactive ($\bar{S}_k^o$) scenarios, of size N_k and $N - N_k$, respectively, where $N_k = \lfloor N(\alpha + (1-\alpha)(1-\xi)^k)\rfloor$. The inactive scenarios are those in which $\boldsymbol{w}_{k-1}^*$, the solution from iteration $k-1$, incurs the $N - N_k$ largest losses. The algorithm then minimizes $h_{\alpha_k,N_k}(\boldsymbol{w}, S_k^o)$ for some quantile level α_k by solving

$$
\begin{aligned}
\min_{\boldsymbol{w},u,\boldsymbol{y},\gamma} \quad & u + \frac{1}{N_k(1-\alpha_k)} \sum_{i=1}^{N_k} y_i \\
\text{s.t.} \quad & \ell_i(\boldsymbol{w}, S_k^o) - u - y_i \le 0, \; i = 1, \dots, N_k \\
& y_i \ge 0, \; i = 1, \dots, N_k \\
& \ell_i(\boldsymbol{w}, S_k^o) \le \gamma, \; i = 1, \dots, N_k \\
& \ell_i(\boldsymbol{w}, \bar{S}_k^o) \ge \gamma, \; i = N_k + 1, \dots, N \\
& \boldsymbol{w} \in \Omega,
\end{aligned}
\tag{14}
$$

where the variable γ is a loss threshold which ensures that no active scenario incurs a loss return greater than that of any inactive scenario.

The quantile level α_k in Problem 14 is chosen so that the CVaR of $\boldsymbol{w}_{k-1}^*$, as estimated from the active scenarios in iteration k, is as close as possible to VaR_α of $\boldsymbol{w}_{k-1}^*$, as estimated from the full sample, i.e., $h_{\alpha_k, N_k}(\boldsymbol{w}_{k-1}^*, S_k^o) \approx q_{\alpha,N}(\boldsymbol{w}_{k-1}^*, S^o)$. Since $q_{\alpha,N}(\boldsymbol{w}_k^*, S^o)$ may increase as k increases (i.e., solutions are not guaranteed to improve monotonically), the heuristic simply returns the best solution encountered.[5]

Note that no iterations are performed if $N(1-\alpha) < 1$, as no scenarios can be discarded in this case. Otherwise, the heuristic performs a total of K iterations after solving the initial problem, where

$$
K = \begin{cases} 1 & \text{if } \xi = 1 \text{ or } N(1-\alpha) = 1, \\ \left\lceil \dfrac{\ln\left(\lceil N\alpha \rceil + 1 - N\alpha\right) - \ln\left(N(1-\alpha)\right)}{\ln(1-\xi)} \right\rceil & \text{otherwise.} \end{cases}
\tag{15}
$$

It follows that K increases with N and decreases with α and ξ. Larsen et al. [9] found that the in-sample solution quality improves as more iterations are performed, i.e., as ξ decreases.

While Problem 13 and Problem 14 both minimize CVaR, the former has some computational advantages. First, since there is no need to distinguish active and inactive scenarios, Problem 13 does not require the additional N constraints (involving γ), making it only about half the size of Problem 14. Second, both the heuristic and the CVaR proxy approaches typically require solving a sequence of optimization problems with different quantile levels. Since instances of Problem 13 differ only in terms of α', while successive iterations of the heuristic modify α_k as well as the sets S_k^o and $\bar{S}_k^o$ in Problem 14, it is somewhat easier to "warm start" an optimization with an existing solution in the former case.

4. **Numerical experiments.** For evaluation purposes, we optimized two portfolios of international stocks, selected based on historical data from Reuters Data Scope Select (DSS). Consideration was given only to those stocks having

- a complete set of monthly data from January 2003 to June 2010 (i.e., 90 months of data);
- a share price of at least 1 USD throughout the entire period;
- a market capitalization of at least 50 million USD throughout the entire period.

[5] The procedure described refers to Algorithm A2 in Larsen et al. [9]. A related procedure, Algorithm A1, was found to be less effective both in the original paper and in our computational experiments, and therefore is not discussed herein.

TABLE 1. Historical return data and simulation errors.

Returns		Mean	Std Dev	Skewness	Kurtosis	Correlation
Normal	Historical Average	0.0103	0.0897	0.0135	3.0249	0.1984
	Historical Minimum	-0.0073	0.0371	-0.1300	2.7374	-0.3054
	Historical Maximum	0.0397	0.1996	0.1263	3.2582	0.8009
	Largest Error	6.71E-09	9.74E-08	1.40E-06	3.18E-06	5.71E-04
Non-Normal	Historical Average	0.0102	0.1272	-0.9424	15.3818	0.1672
	Historical Minimum	-0.0106	0.0455	-3.0507	13.5686	-0.4194
	Historical Maximum	0.0397	0.3134	2.7327	17.6680	0.9022
	Largest Error	2.44E-08	1.59E-07	1.67E-06	2.58E-05	1.16E-01

We first computed monthly USD returns for all candidates, thereby capturing the effects of both equity and currency risks. Two sets of 100 stocks then were selected based on the degree of normality of their returns:

- "normal" stocks have Jarque-Bera statistics between 0 and 0.27 (p-values, as tabulated by Wuertz and Katzgraber [19], above 0.85);
- "non-normal" stocks have Jarque-Bera statistics between 500 and 822 (p-values below 0.0001).

We generated scenarios of joint returns for each set of 100 stocks, consistent with the historical moments and correlations, using the procedure of Høyland et al. [8].[6] For normal stocks, root mean squared error tolerances were 0.001 for the moments and 0.001 for the correlations. In the non-normal case, we found that the procedure did not converge (i.e., did not produce a valid sample) if the correlation tolerance was 0.001. Therefore, we specified error tolerances of 0.001 for the moments and 0.1 for the correlations, and then simply discarded samples whose correlation error exceeded 0.02. A total of 1.5 million scenarios were generated in each case; 25 samples of size 20000 were designated scenario approximations for optimization purposes while the other one million scenarios were taken to represent the "true" (out-of-sample) asset return distribution Ψ, i.e., $\mathrm{VaR}_\alpha(\boldsymbol{w}, \Psi)$ is given by the sample α-quantile from the one million scenarios.

Table 1 shows the statistical properties of the historical returns along with the largest simulation errors, computed as the maximum difference between a historical statistic and the corresponding out-of-sample statistic, for any stock. The results indicate that the scenarios are consistent with historical observations. Clearly, any portfolio of normal stocks will have a loss distribution that is close to normal, while this will not necessarily be true for non-normal stock portfolios.

Our goal is to find portfolios that minimize the out-of-sample VaR at quantile levels $\alpha = 90\%$, 95%, 99% and 99.9%, subject to there being no short positions, i.e.,

$$\Omega := \left\{ \boldsymbol{w} \in \mathbb{R}^{100} \, : \, \sum_{j=1}^{100} w_j = 1, \; w_j \geq 0 \text{ for } j = 1, \ldots, 100 \right\}. \tag{16}$$

We considered scenario approximations of size $N = 1000$, 5000, 10000 and 20000. Thus, each sample of 20000 scenarios (recall that there are 25 such samples for both normal and non-normal returns), gives rise to 16 VaR minimization problems, corresponding to all possible combinations of sample sizes and quantile levels. Note

[6] The source code was downloaded from http://work.michalkaut.net.

Table 2. Normal and out-of-sample VaR in basis points.

	α (%)			
	90	95	99	99.9
Normal VaR	189.45	269.72	416.59	578.58
Out-of-sample VaR	189.98	270.02	415.50	574.12

that a scenario approximation of size N was obtained by selecting the first N scenarios from the sample.

For each α and N, we performed the following optimizations:[7]

1. Minimize $q_{\alpha,N}(\boldsymbol{w}, S^o)$ by solving Problem 12 subject to a 30 minute time limit;
2. Minimize $q_{\alpha,N}(\boldsymbol{w}, S^o)$ by applying the heuristic with $\xi = 1.0$, 0.5 and 0.1;
3. Minimize $h_{\alpha',N}(\boldsymbol{w}, S^o)$ by solving Problem 13 for $\alpha' = 50\%, 50.1\%, \ldots, 99.9\%$, i.e., 500 quantile levels in total.

The in-sample and out-of-sample VaR were recorded for each optimized portfolio, and the results were averaged over 25 trials.

4.1. **Normal returns.** First, consider the portfolio of stocks with (almost) normal returns. Since the deviation from multivariate normality is small, in addition to using scenario approximation we also solved Problem 10 with $\overline{\boldsymbol{r}}$ and Σ computed from the historical returns and Z_α given by the standard normal α-quantile. Table 2 shows that for the resulting optimal portfolios, the normal, historical VaR agrees well with the true VaR as computed from the one million out-of-sample scenarios.

Table 3 reports the average in-sample VaR of portfolios obtained by minimizing $q_{\alpha,N}(\boldsymbol{w}, S^o)$ using the MIP formulation and the heuristic, as well as minimizing the CVaR proxy $h_{\alpha',N}(\boldsymbol{w}, S^o)$ at quantile levels $\alpha' = \alpha$, α^+ and α^*, where

- α corresponds to the standard CVaR upper bound for VaR;
- α^+ yields the lowest average in-sample VaR of all α' considered;
- α^* yields the lowest average out-of-sample (actual) VaR of all α' considered.

By definition, solving the MIP must yield the lowest in-sample VaR for a given problem (since this approach minimizes the sample α-quantile directly). As seen in Table 3, this is the case whenever a near-optimal MIP solution can be obtained within the 30 minute time limit, i.e., generally, when the number of tail scenarios, $N(1-\alpha)$, is sufficiently small. When the MIP solution is of poor quality, the best in-sample result is obtained by the heuristic procedure with $\xi = 0.1$.

Minimizing the upper bound $h_{\alpha,N}(\boldsymbol{w}, S^o)$ consistently fails to match the performance of the heuristic on an in-sample basis. Even the best in-sample CVaR proxy, $h_{\alpha^+,N}(\boldsymbol{w}, S^o)$, yields solutions whose quality is comparable only to that of the poorest heuristic solutions ($\xi = 1.0$). Evidently, iteratively discarding scenarios is critical for reducing the in-sample VaR.

On an out-of-sample basis, however, the results are markedly different (Table 4). Now, the best CVaR proxy, $h_{\alpha^*,N}(\boldsymbol{w}, S^o)$, always outperforms the MIP and the heuristic methods (note that no single value of ξ consistently yields the best heuristic result). Moreover, simply minimizing the upper bound also outperforms these methods unless $N(1-\alpha)$ is small. When $N = 20000$, the best proxy results almost

[7] The optimizations were performed using the CPLEX® solver version 12.3 on a server with 8 Quad-Core AMD Opteron Processors 8356 (32 cores in total) and 256 Gb of RAM. Four threads were used for solving all problems and CPLEX parameters remained at their default values.

TABLE 3. In-sample VaR in basis points (percentage difference from upper bound minimization in parenthesis) for normal returns. Shading identifies the best results.

α (%)	Method	N = 1000		5000		10000		20000	
90	MIP	168.27	*(-10.2)*	212.51	*(11.5)*	217.12	*(13.6)*	217.06	*(13.6)*
	Heur ($\xi = 0.1$)	143.81	*(-23.3)*	173.47	*(-9.0)*	179.65	*(-6.0)*	183.63	*(-3.9)*
	Heur ($\xi = 0.5$)	150.57	*(-19.7)*	178.77	*(-6.2)*	184.33	*(-3.5)*	187.15	*(-2.0)*
	Heur ($\xi = 1.0$)	174.09	*(-7.1)*	187.55	*(-1.6)*	189.74	*(-0.7)*	190.17	*(-0.4)*
	CVaR(α)	187.43	*(0.0)*	190.64	*(0.0)*	191.08	*(0.0)*	191.01	*(0.0)*
	CVaR(α^+)	174.29	*(-7.0)*	187.65	*(-1.6)*	189.26	*(-1.0)*	189.46	*(-0.8)*
	CVaR(α^*)	182.96	*(-2.4)*	188.28	*(-1.2)*	189.41	*(-0.9)*	189.86	*(-0.6)*
95	MIP	214.22	*(-18.9)*	285.04	*(5.2)*	296.32	*(9.8)*	298.97	*(10.5)*
	Heur ($\xi = 0.1$)	216.58	*(-18.0)*	250.02	*(-7.7)*	256.33	*(-5.1)*	261.36	*(-3.4)*
	Heur ($\xi = 0.5$)	223.71	*(-15.3)*	254.35	*(-6.1)*	259.94	*(-3.7)*	264.60	*(-2.2)*
	Heur ($\xi = 1.0$)	245.86	*(-6.9)*	266.15	*(-1.8)*	267.41	*(-1.0)*	269.43	*(-0.4)*
	CVaR(α)	264.01	*(0.0)*	270.98	*(0.0)*	269.99	*(0.0)*	270.56	*(0.0)*
	CVaR(α^+)	244.28	*(-7.5)*	265.40	*(-2.1)*	267.21	*(-1.0)*	268.90	*(-0.6)*
	CVaR(α^*)	261.13	*(-1.1)*	267.81	*(-1.2)*	268.22	*(-0.7)*	269.51	*(-0.4)*
99	MIP	314.74	*(-17.5)*	389.45	*(-5.2)*	434.89	*(4.7)*	451.78	*(8.6)*
	Heur ($\xi = 0.1$)	320.15	*(-16.0)*	376.62	*(-8.3)*	390.40	*(-6.1)*	398.80	*(-4.1)*
	Heur ($\xi = 0.5$)	337.21	*(-11.6)*	381.31	*(-7.1)*	394.02	*(-5.2)*	402.50	*(-3.3)*
	Heur ($\xi = 1.0$)	351.43	*(-7.8)*	398.57	*(-2.9)*	407.75	*(-1.9)*	411.99	*(-1.0)*
	CVaR(α)	381.30	*(0.0)*	410.67	*(0.0)*	415.54	*(0.0)*	416.06	*(0.0)*
	CVaR(α^+)	348.11	*(-8.7)*	398.19	*(-3.0)*	407.31	*(-2.0)*	411.11	*(-1.2)*
	CVaR(α^*)	403.40	*(5.8)*	413.98	*(0.8)*	416.23	*(0.2)*	415.46	*(-0.1)*
99.9	MIP	383.01	*(-5.9)*	475.55	*(-9.7)*	503.84	*(-9.1)*	535.63	*(-4.8)*
	Heur ($\xi = 0.1$)	402.52	*(-1.1)*	490.90	*(-6.8)*	505.58	*(-8.8)*	522.91	*(-7.1)*
	Heur ($\xi = 0.5$)	402.52	*(-1.1)*	502.76	*(-4.6)*	520.40	*(-6.2)*	528.57	*(-6.1)*
	Heur ($\xi = 1.0$)	402.52	*(-1.1)*	511.70	*(-2.9)*	530.68	*(-4.3)*	546.99	*(-2.8)*
	CVaR(α)	406.89	*(0.0)*	526.91	*(0.0)*	554.51	*(0.0)*	562.91	*(0.0)*
	CVaR(α^+)	400.66	*(-1.5)*	501.66	*(-4.8)*	527.03	*(-5.0)*	547.15	*(-2.8)*
	CVaR(α^*)	536.20	*(31.8)*	570.09	*(8.2)*	568.94	*(2.6)*	570.53	*(1.4)*

match those obtained by minimizing the normal, historical VaR (see Table 2); the discrepancy ranges from 0.07% for $\alpha = 90\%$ to 0.59% for $\alpha = 99.9\%$.

To clarify these results, Figure 1 plots the average in-sample and out-of-sample VaR at the 99% quantile level against the number of scenarios for the various methods (to improve readability, we exclude the results of the heuristic with $\xi = 0.5$ and 1.0). For all methods, the difference between the in-sample and out-of-sample VaR declines as N increases; intuitively, the approximation to the true loss return distribution improves with more scenarios.

Perhaps surprisingly, a lower in-sample VaR generally implies a higher out-of-sample VaR (with the exception of the MIP solutions for large N). This can be attributed to the optimization exploiting the idiosyncracies of the scenarios, essentially overfitting the data in order to obtain the best possible objective function value. In other words, when $N(1-\alpha)$ is relatively small, an in-sample VaR that is "too good" indicates that the optimization has seized upon sampling errors in the scenario approximation, which degrades out-of-sample performance.

TABLE 4. Out-of-sample VaR in basis points (percentage difference from upper bound minimization in parenthesis) for normal returns. Shading identifies the best results.

α (%)	Method	N							
		1000		5000		10000		20000	
90	MIP	214.80	*(7.3)*	212.38	*(10.0)*	216.19	*(12.4)*	216.80	*(13.1)*
	Heur ($\xi = 0.1$)	208.66	*(4.2)*	195.62	*(1.4)*	193.27	*(0.5)*	191.65	*(0.0)*
	Heur ($\xi = 0.5$)	205.28	*(2.5)*	194.04	*(0.5)*	192.31	*(0.0)*	191.45	*(-0.1)*
	Heur ($\xi = 1.0$)	203.04	*(1.4)*	193.81	*(0.4)*	192.58	*(0.2)*	191.91	*(0.1)*
	CVaR(α)	200.26	*(0.0)*	193.00	*(0.0)*	192.25	*(0.0)*	191.66	*(0.0)*
	CVaR(α^+)	199.39	*(-0.4)*	192.08	*(-0.5)*	190.51	*(-0.9)*	190.36	*(-0.7)*
	CVaR(α^*)	196.67	*(-1.8)*	191.25	*(-0.9)*	190.50	*(-0.9)*	190.11	*(-0.8)*
95	MIP	296.68	*(3.4)*	290.73	*(5.9)*	295.84	*(8.6)*	299.16	*(10.1)*
	Heur ($\xi = 0.1$)	294.43	*(2.6)*	277.83	*(1.2)*	274.45	*(0.7)*	272.55	*(0.3)*
	Heur ($\xi = 0.5$)	290.53	*(1.2)*	275.63	*(0.4)*	273.18	*(0.2)*	271.74	*(0.0)*
	Heur ($\xi = 1.0$)	290.15	*(1.1)*	275.62	*(0.4)*	273.22	*(0.3)*	272.00	*(0.1)*
	CVaR(α)	286.97	*(0.0)*	274.49	*(0.0)*	272.52	*(0.0)*	271.63	*(0.0)*
	CVaR(α^+)	284.88	*(-0.7)*	273.85	*(-0.2)*	272.31	*(-0.1)*	270.87	*(-0.3)*
	CVaR(α^*)	278.97	*(-2.8)*	272.04	*(-0.9)*	270.79	*(-0.6)*	270.27	*(-0.5)*
99	MIP	458.56	*(-1.3)*	437.49	*(2.1)*	438.48	*(3.7)*	451.62	*(7.5)*
	Heur ($\xi = 0.1$)	457.00	*(-1.6)*	431.11	*(0.6)*	426.30	*(0.8)*	422.59	*(0.6)*
	Heur ($\xi = 0.5$)	462.04	*(-0.5)*	430.07	*(0.4)*	425.04	*(0.5)*	421.02	*(0.2)*
	Heur ($\xi = 1.0$)	465.53	*(0.2)*	430.87	*(0.6)*	424.47	*(0.3)*	421.05	*(0.2)*
	CVaR(α)	464.45	*(0.0)*	428.47	*(0.0)*	423.02	*(0.0)*	420.12	*(0.0)*
	CVaR(α^+)	455.26	*(-2.0)*	427.22	*(-0.3)*	422.30	*(-0.2)*	419.82	*(-0.1)*
	CVaR(α^*)	431.59	*(-7.1)*	420.21	*(-1.9)*	418.23	*(-1.1)*	417.02	*(-0.7)*
99.9	MIP	644.85	*(-1.8)*	610.26	*(-3.3)*	602.72	*(-2.0)*	596.17	*(-0.9)*
	Heur ($\xi = 0.1$)	651.80	*(-0.7)*	614.87	*(-2.6)*	602.78	*(-2.0)*	595.50	*(-1.0)*
	Heur ($\xi = 0.5$)	651.80	*(-0.7)*	626.68	*(-0.7)*	609.72	*(-0.8)*	596.48	*(-0.8)*
	Heur ($\xi = 1.0$)	651.80	*(-0.7)*	633.80	*(0.4)*	612.57	*(-0.4)*	600.75	*(-0.1)*
	CVaR(α)	656.66	*(0.0)*	631.34	*(0.0)*	614.83	*(0.0)*	601.46	*(0.0)*
	CVaR(α^+)	647.90	*(-1.3)*	615.49	*(-2.5)*	601.23	*(-2.2)*	590.84	*(-1.8)*
	CVaR(α^*)	600.54	*(-8.5)*	583.23	*(-7.6)*	579.64	*(-5.7)*	577.49	*(-4.0)*

Conversely, a solution with an extremely poor in-sample VaR is also undesirable, as this suggests that the optimization has failed to structure the portfolio in a beneficial manner. This is apparent in the MIP solutions for $N = 10000$ and $N = 20000$, where the 30 minute time limit prevents finding a solution of sufficiently high quality.

Evidently, the best out-of-sample results are associated with solutions whose in-sample VaR is “good, but not too good”, i.e., the optimized portfolio is structured to have a low VaR, without being overly tuned to a particular scenario approximation. This suggests that the optimization should be restrained, or dampened, as necessary to account for the level of sampling error in the scenario approximation. The standard way to do this is through regularization, which may involve shrinkage estimators or adding constraints to the problem (see, for example, DeMiguel et al. [7]). Our experiments indicate that a similar effect can be obtained by modifying parameters of the solution algorithm (e.g., decreasing the time limit for the MIP solver, increasing ξ for the heuristic) or by using a CVaR proxy.

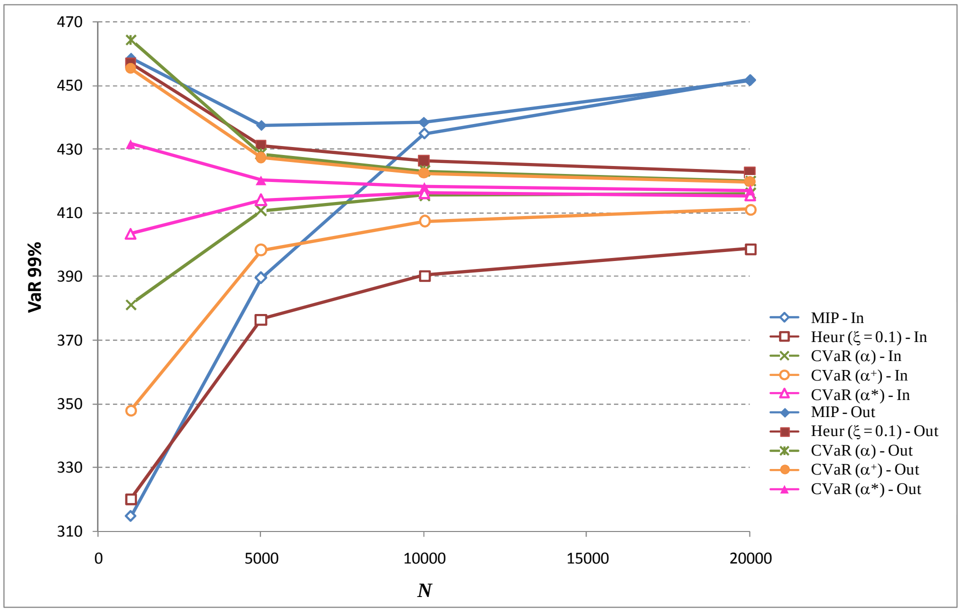

FIGURE 1. In- and out-of-sample VaR (basis points) as a function of sample size.

TABLE 5. α^+ and α^* (%) for normal returns.

α	α^+				α^*			
	$N = 1000$	$N = 5000$	$N = 10000$	$N = 20000$	$N = 1000$	$N = 5000$	$N = 10000$	$N = 20000$
90	88.5	84.9	75.8	80.4	72.6	76.6	73.6	73.4
95	93.7	94.0	94.4	92.3	81.0	85.5	85.5	86.3
99	98.0	98.7	98.8	98.9	85.4	91.0	93.5	94.6
99.9	99.3	99.7	99.8	99.8	86.7	92.2	97.2	96.9

When using a CVaR proxy, the quantile level α' effectively controls the amount of dampening. Thus, the proxy quantile level that performs best on an out-of-sample basis, α^*, depends on the amount of sampling error in the scenario approximation. As shown in Table 5, α^* is always less than α, the VaR quantile level, and it tends to increase with N, the number of scenarios (i.e., α^* becomes larger as the sampling error declines).[8]

For normal distributions, it is possible to compute the limiting value of α^* as $N \to \infty$. Recall that if losses are normally distributed then both VaR and CVaR can be expressed as a weighted sum of the mean and the standard deviation. Specifically, suppose that losses are distributed $\mathcal{N}(\mu, \sigma)$ and denote the normal distribution and density functions by Φ and ϕ, respectively. Then

$$\mathrm{VaR}_\alpha = \mu + Z_\alpha \sigma \tag{17}$$

and

$$\mathrm{CVaR}_\alpha = \mu + K_\alpha \sigma \tag{18}$$

[8] In contrast, α^+ remains more or less constant for a given α.

TABLE 6. $\tilde{\alpha}$ for selected VaR quantile levels α (%).

α	90	95	99	99.9
$\tilde{\alpha}$	75.44	87.45	97.42	99.74

where $Z_\alpha = \Phi^{-1}(\alpha)$ and $K_\alpha = \phi(Z_\alpha)(1-\alpha)^{-1}$. As such, minimizing VaR_α is equivalent to minimizing $\mathrm{CVaR}_{\tilde{\alpha}}$, where $\tilde{\alpha}$ satisfies $\phi(Z_{\tilde{\alpha}}) = Z_\alpha \cdot (1-\tilde{\alpha})$. Table 6 gives $\tilde{\alpha}$ for the quantile levels considered in this paper.

In our experiments, the optimized out-of-sample loss distributions are close to normal; all loss distributions referenced in Table 4, regardless of the sample, have skewness between -0.03 and 0.03, and excess kurtosis between -0.04 and 0.02. Thus, α^* should approach $\tilde{\alpha}$ as N increases and, in fact, the results in Table 5 exhibit this trend. Essentially, the difference $(\tilde{\alpha} - \alpha^*)$ reflects the amount of dampening. Note that α^* approaches $\tilde{\alpha}$ more quickly as α decreases. This is consistent with the fact that the volatility of VaR and CVaR estimates increases with α, i.e., for a given N, more dampening is required as $\alpha \to 1$.

To show the effect of the CVaR proxy level in more detail, Figure 2 plots the in- and out-of-sample VaR against α' for $\alpha = 99\%$ and $N = 1000$, 5000 and 20000 (graphs for other values of α and N exhibit similar characteristics). While the in-sample VaR attains a sharp minimum, the out-of-sample VaR profile is much flatter (see Table 5 for the quantile levels, α^+ and α^*, corresponding to the respective minima). This lack of curvature is attractive from a practical perspective; a trial-and-error approach that evaluates only a limited number of α' values is likely to produce a good CVaR proxy, even if it does not find α^* precisely. For example, while $\alpha^* = 85.4\%$ for $N = 1000$, any $56.5\% \leq \alpha' \leq 98.4\%$ yields an out-of-sample VaR of less than 457 basis points, thereby improving on the best result obtained by the heuristic or MIP methods. As shown in Table 7, for any combination of α and N, it is relatively easy to find a good value of α' by trial-and-error.

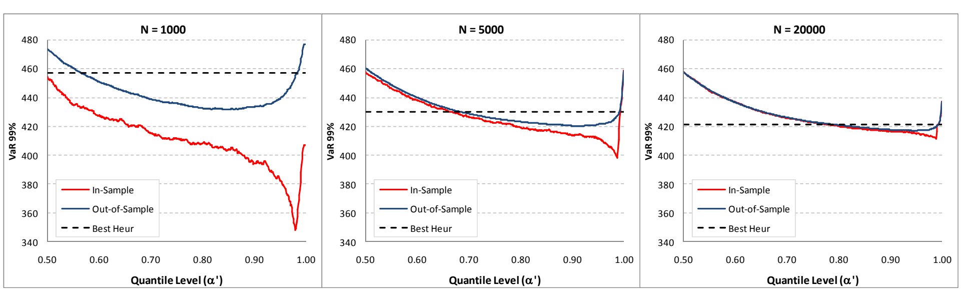

FIGURE 2. In- and out-of-sample VaR (basis points) as a function of CVaR quantile level for normal returns.

4.2. **Non-normal returns.** We now consider stocks with non-normal returns. The methods that deliver the best in-sample (Table 8) and out-of-sample (Table 9) performance are identical to those for normal asset returns. However, the differences between methods are larger in this case.

Figure 3 plots the average skewness and excess kurtosis of the out-of-sample loss distributions of portfolios obtained by minimizing $h_{\alpha',N}(\boldsymbol{w}, S^o)$ for $50\% \leq \alpha' \leq$

TABLE 7. Range of α' (%) for which the CVaR proxy outperforms MIP and heuristic methods for normal returns.

α	N 1000	5000	10000	20000
90	[54.5, 93.4]	[59.8, 91.9]	[60.8, 90.1]	[62.3, 88.9]
95	[57.2, 96.2]	[66.2, 96.5]	[69.3, 96.1]	[72.8, 95.2]
99	[56.5, 98.4]	[68.3, 99.1]	[73.6, 99.2]	[79.6, 99.3]
99.9	[53.6, 99.4]	[61.6, 99.6]	[66.3, 99.8]	[70.7, 99.8]

TABLE 8. In-sample VaR in basis points (percentage difference from upper bound minimization in parenthesis) for non-normal returns. Shading identifies the best results.

α (%)	Method	N 1000		5000		10000		20000	
90	MIP	156.74	*(-17.8)*	233.06	*(17.8)*	252.38	*(27.6)*	283.25	*(42.7)*
	Heur ($\xi = 0.1$)	132.89	*(-30.3)*	168.97	*(-14.6)*	175.83	*(-11.1)*	181.94	*(-8.3)*
	Heur ($\xi = 0.5$)	147.04	*(-22.9)*	181.08	*(-8.4)*	187.61	*(-5.1)*	192.46	*(-3.0)*
	Heur ($\xi = 1.0$)	173.94	*(-8.8)*	194.30	*(-1.8)*	195.92	*(-0.9)*	197.62	*(-0.4)*
	CVaR(α)	190.66	*(0.0)*	197.79	*(0.0)*	197.72	*(0.0)*	198.47	*(0.0)*
	CVaR(α^+)	177.01	*(-7.2)*	187.54	*(-5.2)*	186.91	*(-5.5)*	187.69	*(-5.4)*
	CVaR(α^*)	181.31	*(-4.9)*	187.67	*(-5.1)*	187.08	*(-5.4)*	187.77	*(-5.4)*
95	MIP	217.35	*(-24.8)*	313.43	*(6.0)*	343.00	*(15.7)*	376.14	*(26.2)*
	Heur ($\xi = 0.1$)	224.02	*(-22.5)*	265.05	*(-10.3)*	273.91	*(-7.6)*	280.22	*(-6.0)*
	Heur ($\xi = 0.5$)	236.08	*(-18.3)*	273.95	*(-7.3)*	282.81	*(-4.6)*	289.63	*(-2.9)*
	Heur ($\xi = 1.0$)	263.52	*(-8.8)*	289.97	*(-1.9)*	293.63	*(-0.9)*	296.59	*(-0.5)*
	CVaR(α)	289.10	*(0.0)*	295.62	*(0.0)*	296.38	*(0.0)*	298.15	*(0.0)*
	CVaR(α^+)	265.34	*(-8.2)*	287.09	*(-2.9)*	289.27	*(-2.4)*	289.85	*(-2.8)*
	CVaR(α^*)	275.71	*(-4.6)*	288.73	*(-2.3)*	289.73	*(-2.2)*	290.39	*(-2.6)*
99	MIP	348.83	*(-20.5)*	445.15	*(-7.5)*	514.18	*(5.9)*	546.75	*(11.7)*
	Heur ($\xi = 0.1$)	360.10	*(-17.9)*	429.70	*(-10.7)*	449.68	*(-7.4)*	463.11	*(-5.4)*
	Heur ($\xi = 0.5$)	375.90	*(-14.3)*	440.31	*(-8.5)*	457.26	*(-5.9)*	469.22	*(-4.1)*
	Heur ($\xi = 1.0$)	398.72	*(-9.1)*	463.56	*(-3.7)*	476.06	*(-2.0)*	484.25	*(-1.0)*
	CVaR(α)	438.59	*(0.0)*	481.20	*(0.0)*	485.76	*(0.0)*	489.29	*(0.0)*
	CVaR(α^+)	395.75	*(-9.8)*	463.10	*(-3.8)*	475.74	*(-2.1)*	481.03	*(-1.7)*
	CVaR(α^*)	442.09	*(0.8)*	475.42	*(-1.2)*	478.80	*(-1.4)*	483.00	*(-1.3)*
99.9	MIP	440.05	*(-6.4)*	565.87	*(-10.1)*	614.90	*(-10.0)*	673.99	*(-4.7)*
	Heur ($\xi = 0.1$)	465.81	*(-1.0)*	583.19	*(-7.3)*	618.29	*(-9.5)*	646.00	*(-8.6)*
	Heur ($\xi = 0.5$)	465.81	*(-1.0)*	599.73	*(-4.7)*	634.39	*(-7.2)*	659.97	*(-6.6)*
	Heur ($\xi = 1.0$)	465.81	*(-1.0)*	606.29	*(-3.7)*	651.97	*(-4.6)*	679.16	*(-3.9)*
	CVaR(α)	470.30	*(0.0)*	629.28	*(0.0)*	683.40	*(0.0)*	706.91	*(0.0)*
	CVaR(α^+)	467.00	*(-0.7)*	608.63	*(-3.3)*	645.90	*(-5.5)*	682.77	*(-3.4)*
	CVaR(α^*)	635.96	*(35.2)*	696.35	*(10.7)*	709.38	*(3.8)*	720.18	*(1.9)*

99.9%. All distributions exhibit positive excess kurtosis while negative skewness is apparent when $\alpha' > 65\%$ ($\alpha' > 70\%$ if $N = 1000$). Intuitively, as α' increases, $\text{CVaR}_{\alpha'}$ is minimized by lengthening the left tail relative to the right tail and making the distribution more peaked.

The lack of normality affects the values of α^+ and α^* (Table 10). When $\alpha = 90\%$ or 95%, both α^+ and α^* are generally lower than they are for normal returns. In

Table 9. Out-of-sample VaR in basis points (percentage difference from upper bound minimization in parenthesis) for non-normal returns. Shading identifies the best results.

α (%)	Method	N							
		1000		5000		10000		20000	
90	MIP	229.42	*(10.3)*	233.97	*(15.5)*	250.74	*(24.6)*	280.63	*(39.9)*
	Heur ($\xi = 0.1$)	214.90	*(3.3)*	196.90	*(-2.8)*	193.55	*(-3.8)*	192.13	*(-4.2)*
	Heur ($\xi = 0.5$)	210.10	*(1.0)*	198.92	*(-1.8)*	198.15	*(-1.5)*	198.37	*(-1.1)*
	Heur ($\xi = 1.0$)	210.92	*(1.4)*	203.18	*(0.3)*	201.44	*(0.1)*	200.72	*(0.1)*
	CVaR(α)	208.01	*(0.0)*	202.61	*(0.0)*	201.23	*(0.0)*	200.59	*(0.0)*
	CVaR(α^+)	203.07	*(-2.4)*	192.34	*(-5.1)*	190.98	*(-5.1)*	190.47	*(-5.0)*
	CVaR(α^*)	199.23	*(-4.2)*	192.16	*(-5.2)*	190.96	*(-5.1)*	190.44	*(-5.1)*
95	MIP	333.61	*(5.0)*	319.81	*(4.5)*	345.45	*(13.7)*	383.99	*(26.7)*
	Heur ($\xi = 0.1$)	324.23	*(2.1)*	303.84	*(-0.7)*	299.42	*(-1.5)*	296.82	*(-2.0)*
	Heur ($\xi = 0.5$)	319.47	*(0.6)*	302.73	*(-1.1)*	301.08	*(-0.9)*	300.25	*(-0.9)*
	Heur ($\xi = 1.0$)	322.01	*(1.4)*	306.57	*(0.2)*	304.31	*(0.1)*	303.15	*(0.1)*
	CVaR(α)	317.60	*(0.0)*	306.00	*(0.0)*	303.86	*(0.0)*	302.97	*(0.0)*
	CVaR(α^+)	314.06	*(-1.1)*	296.36	*(-3.1)*	294.36	*(-3.1)*	294.04	*(-2.9)*
	CVaR(α^*)	306.02	*(-3.6)*	295.81	*(-3.3)*	294.27	*(-3.2)*	293.43	*(-3.1)*
99	MIP	561.20	*(-0.6)*	522.92	*(0.7)*	530.58	*(3.2)*	588.24	*(15.1)*
	Heur ($\xi = 0.1$)	553.82	*(-1.9)*	517.09	*(-0.4)*	508.97	*(-1.0)*	503.65	*(-1.5)*
	Heur ($\xi = 0.5$)	558.61	*(-1.1)*	515.14	*(-0.8)*	509.55	*(-0.9)*	504.98	*(-1.2)*
	Heur ($\xi = 1.0$)	567.51	*(0.5)*	520.40	*(0.2)*	512.58	*(-0.3)*	510.93	*(0.0)*
	CVaR(α)	564.70	*(0.0)*	519.29	*(0.0)*	514.00	*(0.0)*	511.12	*(0.0)*
	CVaR(α^+)	552.02	*(-2.2)*	515.56	*(-0.7)*	503.61	*(-2.0)*	499.43	*(-2.3)*
	CVaR(α^*)	529.06	*(-6.3)*	501.85	*(-3.4)*	497.82	*(-3.1)*	494.90	*(-3.2)*
99.9	MIP	912.10	*(-2.0)*	829.02	*(-2.6)*	802.73	*(-6.0)*	799.32	*(-6.2)*
	Heur ($\xi = 0.1$)	942.59	*(1.3)*	839.44	*(-1.4)*	815.61	*(-4.5)*	809.44	*(-5.0)*
	Heur ($\xi = 0.5$)	942.59	*(1.3)*	854.55	*(0.3)*	828.19	*(-3.0)*	811.47	*(-4.8)*
	Heur ($\xi = 1.0$)	942.59	*(1.3)*	853.42	*(0.2)*	840.59	*(-1.6)*	851.30	*(-0.1)*
	CVaR(α)	930.42	*(0.0)*	851.59	*(0.0)*	853.94	*(0.0)*	852.45	*(0.0)*
	CVaR(α^+)	920.72	*(-1.0)*	846.17	*(-0.6)*	824.06	*(-3.5)*	823.62	*(-3.4)*
	CVaR(α^*)	856.22	*(-8.0)*	777.32	*(-8.7)*	766.03	*(-10.3)*	759.03	*(-11.0)*

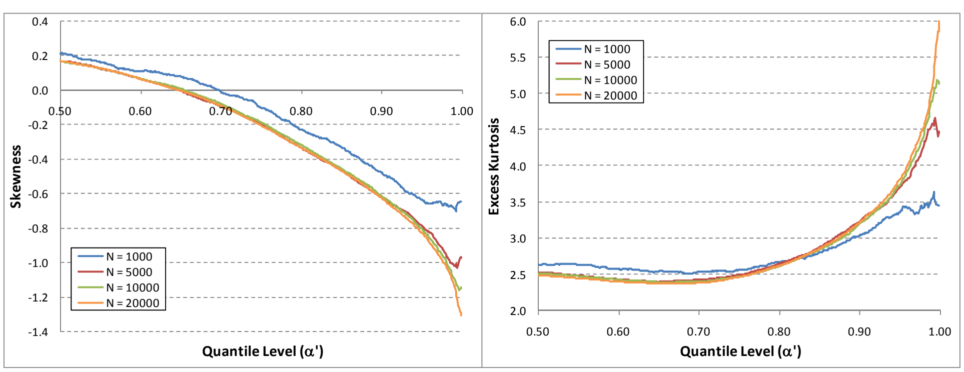

Figure 3. Average skewness and excess kurtosis of out-of-sample portfolio loss distributions when minimizing CVaR at quantile level α' for non-normal returns.

TABLE 10. α^+ and α^* (%) for non-normal returns.

α	α^+				α^*			
	$N = 1000$	$N = 5000$	$N = 10000$	$N = 20000$	$N = 1000$	$N = 5000$	$N = 10000$	$N = 20000$
90	84.2	71.5	68.0	67.5	69.1	68.5	68.1	68.2
95	93.5	84.4	83.0	85.2	81.5	80.8	81.1	81.4
99	98.1	98.8	98.1	97.8	92.4	94.0	95.5	95.3
99.9	99.4	99.8	99.8	99.8	95.8	97.5	98.0	97.9

TABLE 11. $\tilde{\alpha}$ and α^* for selected VaR quantile levels α (%) for one million non-normal return scenarios.

α	90	95	99	99.9
$\tilde{\alpha}$	73.8	86.7	97.3	99.7
α^*	69.6	82.6	96.8	99.7

contrast, when $\alpha = 99\%$ or 99.9%, α^* is higher than for normal returns while α^+ is about the same.

Since neither the limiting value of α^* nor the quantile level $\tilde{\alpha}$, for which $\mathrm{CVaR}_{\tilde{\alpha}}$ equals VaR_α, can be found analytically, we obtain their approximate values by minimizing $\mathrm{CVaR}_{\alpha'}$ using the one million out-of-sample scenarios (Table 11). The values of $\tilde{\alpha}$ are lower than those for normally distributed losses (see Table 6), consistent with positive excess kurtosis, i.e., heavier-than-normal tails. While the limiting value of α^* equals $\tilde{\alpha}$ for normal distributions, in this case $\alpha^* \leq \tilde{\alpha}$.

The out-of-sample VaR profile (Figure 4) exhibits greater curvature than when losses are normally distributed (Figure 2), indicating that VaR is more sensitive to the quantile level used when optimizing CVaR in this case. Intuitively, this is because the skewness and the kurtosis of the optimized portfolio loss distributions vary with α' (see Figure 3), unlike when the loss distributions are normal. As such, the ranges of α' values for which the CVaR proxy outperforms the MIP and heuristics methods tend to be smaller for non-normal loss distributions (Table 12). Nevertheless, the ranges suggest that it is still relatively easy to find a good value for α' by trial-and-error.

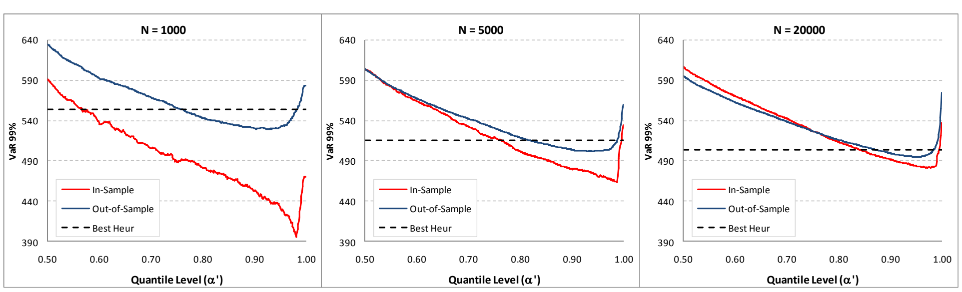

FIGURE 4. In- and out-of-sample VaR (basis points) as a function of CVaR quantile level for non-normal returns.

To assess the relative performance of the heuristic and CVaR proxy methods in stressed conditions, we repeated the experiments with non-normal stocks while

Table 12. Range of α' (%) for which the CVaR proxy outperforms MIP and heuristic methods for non-normal returns.

α	N			
	1000	5000	10000	20000
90	[50.0, 91.4]	[52.9, 83.3]	[55.7, 80.3]	[57.7, 77.9]
95	[59.7, 95.5]	[64.5, 93.4]	[67.4, 92.2]	[70.5, 90.5]
99	[76.0, 98.2]	[82.0, 98.7]	[85.0, 98.6]	[87.3, 98.4]
99.9	[82.4, 99.2]	[87.0, 99.6]	[91.1, 99.6]	[90.6, 99.5]

increasing their historical correlations by approximately 25% (yielding average, minimum and maximum correlations of 0.20, -0.34 and 0.96, respectively). The results were practically identical to those with the historical correlations, and are therefore not reported here.

4.3. **Computational performance.** Table 13 and Table 14 show the average solution times for the normal and non-normal stock portfolios, respectively. To aid in comparing the various methods, we also report the number of CVaR proxy problems (i.e., instances of Problem 13) that can be solved in an equivalent amount of time. Evidently, the ability of the CVaR proxy approach to leverage warm starts is advantageous, allowing a relatively large number of CVaR problems to be solved in the time taken by other methods. Note, however, that such comparisons depend on the range and the granularity of the quantile levels for the CVaR proxy. For example, our experimental approach, which starts with $\alpha' = 99.9\%$ and then decreases α' by 0.1% in subsequent iterations, allows each CVaR problem to be solved extremely quickly but may entail a relatively large number of iterations. Alternatively, starting at a lower quantile level and/or using a larger step size may prove to be more computationally effective, even if the individual CVaR problems take longer to solve.

Indeed, while we evaluated a total of 500 CVaR-optimized portfolios for each problem in order to accurately relate VaR and the CVaR quantile level, our experiments suggest that this level of precision is unnecessary in practice. Rather, finding a quantile level that produces a good CVaR proxy (i.e., yielding an out-of-sample VaR better than both the MIP and heuristic methods) requires only a small number of trials. For example, trying only the six values $\alpha' = 70\%$, 75%, 80%, 85%, 90% and 95% is sufficient to outperform the MIP and heuristic methods for all 32 problems considered in this paper.

Certainly, further tuning of the methods may improve their respective results. Algorithms for solving the MIP, for instance, typically have a large number of parameters that affect the solution quality, particularly when a time limit is imposed. Likewise, the performance of the heuristic depends on the parameter ξ, for which we considered only three possible values. However, we maintain that from a practical standpoint, finding suitable parameter values for these methods takes longer than tuning the CVaR proxy by varying α'. While both the CVaR proxy and the heuristic solve similar problems (respectively, Problem 13 and Problem 14), evaluating a single ξ parameter value requires iteratively solving many such problems. For each problem considered in our experiments (aside from the case of $\alpha = 99.9\%$ and $N = 1000$, where all of $\xi = 0.1$, 0.5 and 1.0 perform only one iteration), obtaining the three heuristic solutions entailed solving at least 21 instances of Problem 14. As noted above, the CVaR proxy yields better results while solving only six instances of

TABLE 13. Solution time in seconds (equivalent number of CVaR proxy problems) for normal returns.

α (%)	Method	N 1000		5000		10000		20000	
90	MIP	1805.67	*(41,039)*	1800.36	*(3,887)*	1801.08	*(656)*	1823.65	*(573)*
	Heur ($\xi = 0.1$)	9.69	*(220)*	155.87	*(337)*	300.36	*(109)*	793.22	*(249)*
	Heur ($\xi = 0.5$)	3.33	*(76)*	54.83	*(118)*	95.71	*(35)*	258.09	*(81)*
	Heur ($\xi = 1.0$)	1.10	*(25)*	8.17	*(18)*	21.73	*(8)*	44.88	*(14)*
	CVaR(α)	0.35	*(8)*	6.43	*(14)*	17.80	*(6)*	36.45	*(11)*
95	MIP	1807.25	*(41,075)*	1800.39	*(3,887)*	1800.90	*(656)*	1847.37	*(581)*
	Heur ($\xi = 0.1$)	7.71	*(175)*	128.40	*(277)*	241.77	*(88)*	613.29	*(193)*
	Heur ($\xi = 0.5$)	2.43	*(55)*	43.03	*(93)*	72.90	*(27)*	193.79	*(61)*
	Heur ($\xi = 1.0$)	1.09	*(25)*	6.21	*(13)*	16.13	*(6)*	42.57	*(13)*
	CVaR(α)	0.29	*(7)*	4.52	*(10)*	12.10	*(4)*	33.81	*(11)*
99	MIP	1812.67	*(41,198)*	1800.32	*(3,887)*	1800.66	*(656)*	1802.31	*(567)*
	Heur ($\xi = 0.1$)	4.63	*(105)*	84.35	*(182)*	145.47	*(53)*	365.68	*(115)*
	Heur ($\xi = 0.5$)	1.66	*(38)*	25.88	*(56)*	47.75	*(17)*	106.47	*(33)*
	Heur ($\xi = 1.0$)	0.93	*(21)*	4.88	*(11)*	15.08	*(5)*	23.91	*(8)*
	CVaR(α)	0.24	*(6)*	3.14	*(7)*	11.04	*(4)*	15.46	*(5)*
99.9	MIP	13.13	*(298)*	1395.99	*(3,014)*	1801.06	*(656)*	1801.27	*(566)*
	Heur ($\xi = 0.1$)	0.75	*(17)*	32.30	*(70)*	65.10	*(24)*	170.08	*(53)*
	Heur ($\xi = 0.5$)	0.76	*(17)*	10.58	*(23)*	20.91	*(8)*	52.95	*(17)*
	Heur ($\xi = 1.0$)	0.81	*(18)*	3.98	*(9)*	7.36	*(3)*	14.03	*(4)*
	CVaR(α)	0.23	*(5)*	2.32	*(5)*	3.34	*(1)*	6.30	*(2)*
CVaR(α'), $50 \leq \alpha' \leq 99.9$		22.00	*(500)*	231.57	*(500)*	1372.60	*(500)*	1590.50	*(500)*

TABLE 14. Solution time in seconds (equivalent number of CVaR proxy problems) for non-normal returns.

α (%)	Method	N 1000		5000		10000		20000	
90	MIP	1806.49	*(39,944)*	1800.47	*(3,466)*	1801.54	*(635)*	1837.13	*(541)*
	Heur ($\xi = 0.1$)	9.46	*(209)*	171.46	*(330)*	322.46	*(114)*	787.94	*(232)*
	Heur ($\xi = 0.5$)	2.90	*(64)*	53.76	*(103)*	97.75	*(34)*	258.04	*(76)*
	Heur ($\xi = 1.0$)	1.10	*(24)*	8.91	*(17)*	22.41	*(8)*	49.80	*(15)*
	CVaR(α)	0.34	*(7)*	7.23	*(14)*	18.32	*(6)*	41.32	*(12)*
95	MIP	1806.07	*(39,934)*	1800.35	*(3,466)*	1801.14	*(635)*	1820.54	*(536)*
	Heur ($\xi = 0.1$)	7.18	*(159)*	127.25	*(245)*	232.00	*(82)*	629.80	*(185)*
	Heur ($\xi = 0.5$)	2.45	*(54)*	43.09	*(83)*	70.18	*(25)*	192.52	*(57)*
	Heur ($\xi = 1.0$)	1.03	*(23)*	6.99	*(13)*	15.92	*(6)*	43.25	*(13)*
	CVaR(α)	0.27	*(6)*	5.27	*(10)*	11.87	*(4)*	34.83	*(10)*
99	MIP	1812.21	*(40,070)*	1800.34	*(3,466)*	1800.85	*(635)*	1802.97	*(531)*
	Heur ($\xi = 0.1$)	4.32	*(95)*	79.27	*(153)*	143.25	*(51)*	356.54	*(105)*
	Heur ($\xi = 0.5$)	1.64	*(36)*	24.06	*(46)*	45.03	*(16)*	104.55	*(31)*
	Heur ($\xi = 1.0$)	0.79	*(18)*	4.92	*(9)*	14.31	*(5)*	24.62	*(7)*
	CVaR(α)	0.23	*(5)*	3.19	*(6)*	10.38	*(4)*	16.20	*(5)*
99.9	MIP	9.31	*(206)*	1768.25	*(3,404)*	1800.45	*(635)*	1800.81	*(530)*
	Heur ($\xi = 0.1$)	0.77	*(17)*	32.80	*(63)*	63.78	*(22)*	169.22	*(50)*
	Heur ($\xi = 0.5$)	0.80	*(18)*	10.32	*(20)*	21.33	*(8)*	51.45	*(15)*
	Heur ($\xi = 1.0$)	0.71	*(16)*	3.90	*(8)*	7.39	*(3)*	14.19	*(4)*
	CVaR(α)	0.22	*(5)*	2.23	*(4)*	3.38	*(1)*	6.27	*(2)*
CVaR(α'), $50 \leq \alpha' \leq 99.9$		22.61	*(500)*	259.74	*(500)*	1418.20	*(500)*	1697.70	*(500)*

Problem 13. Of course, a more refined search for α^*, such as bisection, may further improve the efficiency and/or the solution quality of the CVaR proxy approach.

5. **Conclusions.**

When using scenario approximation, minimizing CVaR is an effective way to obtain portfolios with a low out-of-sample VaR. Optimizing CVaR for a number of different quantile levels and then selecting the portfolio with the lowest VaR works well in practice for several reasons. First, the procedure is robust, i.e., good results are obtained for a wide range of quantile levels. Second, changing the quantile level in the objective function allows a series of CVaR problems to be solved efficiently by warm start methods. Our experiments show that it is possible to outperform MIP and heuristic methods for minimizing VaR by trying only a small number of different CVaR quantile levels.

Minimizing CVaR in place of VaR is a form of regularization, as it tends to limit overfitting. While CVaR proxies typically cannot match the in-sample results of MIP and heuristic methods, their out-of-sample performance is often much better, particularly when the number of tail scenarios is small. The CVaR quantile level effectively controls the amount of regularization. Thus, the quantile level that yields the best CVaR proxy depends on the size of the scenario approximation, as well as on problem characteristics like the loss distribution. All else being equal, the best CVaR quantile level for a problem increases as the number of scenarios increases.

REFERENCES

[1] C. Acerbi and D. Tasche, Expected Shortfall: A natural coherent alternative to Value at Risk, *Economic Notes*, **31** (2002), 379–388.

[2] C. Acerbi and D. Tasche, On the coherence of expected shortfall, *Journal of Banking & Finance*, **26** (2002), 1487–1503.

[3] B. C. Arnold, N. Balakrishnan and H. N. Nagaraja, *A First Course in Order Statistics*, SIAM Publishers, Philadelphia, PA, 2008.

[4] P. Artzner, F. Delbaen, J.-M. Eber and D. Heath, Coherent measures of risk, *Mathematical Finance*, **9** (1999), 203–228.

[5] V. Brazauskas, B. L. Jones, M. L. Puri and R. Zitikis, Estimating conditional tail expectation with actuarial applications in view, *Journal of Statistical Planning and Inference*, **138** (2008), 3590–3604.

[6] J. Daníelsson, B. N. Jorgensen, G. Samorodnitsky, M. Sarma and C. G. de Vries, Subadditivity re-examined: The case for Value-at-Risk, preprint, London School of Economics, 2005.

[7] V. DeMiguel, L. Garlappi, F. J. Nogales and R. Uppal, A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms, *Management Science*, **55** (2009), 798–812.

[8] K. Høyland, M. Kaut and S. W. Wallace, A heuristic for moment-matching scenario generation, *Computational Optimization and Applications*, **24** (2003), 169–185.

[9] N. Larsen, H. Mausser and S. Uryasev, Algorithms for optimization of Value-at-Risk, in *Financial Engineering, E-commerce and Supply Chain* (eds. P. Pardalos and V. K. Tsitsiringos), Kluwer Academic Publishers, Norwell, MA, (2002), 19–46.

[10] J. Luedtke and S. Ahmed, A sample approximation approach for optimization with probabilistic constraints, *SIAM Journal on Optimization*, **19** (2008), 674–699.

[11] H. Mausser and O. Romanko, Bias, exploitation and proxies in scenario-based risk minimization, *Optimization*, **61** (2012), 1191–1219.

[12] H. Mausser and D. Rosen, Efficient risk/return frontiers for credit risk, *Journal of Risk Finance*, **2** (2000), 66–78.

[13] K. Natarajan, D. Pachamanova and M. Sim, Incorporating asymmetric distribution information in robust value-at-risk optimization, *Management Science*, **54** (2008), 573–585.

[14] A. Nemirovski and A. Shapiro, Convex approximations of chance constrained programs, *SIAM Journal on Optimization*, **17** (2006), 969–996.

[15] B. K. Pagnoncelli, S. Ahmed and A. Shapiro, *Computational study of a chance constrained portfolio selection problem*, Optimization Online, 2008. Available from: http://www.optimization-online.org/DB_HTML/2008/02/1899.html.

[16] B. K. Pagnoncelli, S. Ahmed and A. Shapiro, Sample average approximation method for chance constrained programming: Theory and applications, *Journal of Optimization Theory and Applications*, **142** (2009), 399–416.

[17] R. T. Rockafellar and S. Uryasev, Optimization of conditional value-at-risk, *Journal of Risk*, **2** (2000), 21–41.

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

[19] D. Wuertz and H. Katzgraber, *Precise Finite-Sample Quantiles of the Jarque-Bera Adjusted Lagrange Multiplier Test*, MPRA Paper No. 19155, University of Munich, 2009. Available from: http://mpra.ub.uni-muenchen.de/19155/.

Received September 2012; 1st revision February 2013; 2nd revision October 2013.

E-mail address: Helmut.Mausser@ca.ibm.com
E-mail address: Oleksandr.Romanko@ca.ibm.com