Scenario reliability assessment for CVaR minimization in two-stage stochastic programs

Ghazal Shah Abadi[1] and Sarah M. Ryan[2]
Department of Industrial and Manufacturing Systems Engineering
Iowa State University, Ames, IA 50011-2030

Abstract
Reliable scenarios are needed to obtain a high-quality solution to a stochastic program. Considering sets of scenarios and corresponding observed values of the uncertain parameters over a collection of historical instances, reliability is defined loosely as goodness of the scenarios' fit to the observations. For two-stage, risk-neutral models, a statistical tool was developed previously to assess the reliability of any given scenario generation method. This tool can diagnose over- or under-dispersion and/or bias in the scenario sets. For risk-averse decision makers who aim to minimize conditional value-at-risk (CVaR), only the scenarios that define the upper tail of the optimal cost distribution at the optimal solution are important. We develop a tool to assess the reliability of these so-called effective scenarios for CVaR minimization. Simulation studies of a financial investment problem demonstrate the ability of the tool to detect mismatches in mean, variance, or kurtosis between scenarios and the corresponding observations.

Keywords Risk-averse stochastic programming, scenario reliability assessment, Conditional-Value-at-Risk (CVaR), goodness of fit

1. Introduction

Stochastic programming is used to optimize a problem of decision making under uncertainty by modeling uncertain parameters in a mathematical program as random variables. In two-stage stochastic programs, decisions are divided into two sets. The first-stage decisions must be made before the realizations of the random variables are known, while second-stage decisions can provide recourse to the realizations observed. Approximating the underlying probability distribution of uncertainty by generating probabilistic scenarios for the uncertain parameter values has long presented a challenge in modeling stochastic optimization problems (Rahimian et al., 2018). If the generated set of scenarios well represents the joint distribution of the uncertain parameters, then a good solution for the stochastic program is obtained. In situations where

[1] Ghazal Shah Abadi (ghazals@iastate.edu) is a PhD student in Industrial Engineering at Iowa State University.
[2] Sarah M. Ryan (smryan@iastate.edu) is the C. G. "Turk" and Joyce A. Therkildsen Department Chair and Professor of Industrial and Manufacturing Systems Engineering at Iowa State University. With expertise in stochastic modeling and optimization under uncertainty, her research has been supported by the National Science Foundation, including a CAREER Award, the US Department of Energy, the Iowa Energy Center, an AT&T Industrial Ecology Faculty Fellowship, and industry consortia. She is a Fellow of the Institute of Industrial and Systems Engineers.

instances of the same problem are solved continually, it may be possible to re-enact the process of generating scenarios and solving the resulting optimization problem. Then the optimal objective values can be compared with the cost of implementing the solutions under the corresponding observations of the uncertain parameters. This procedure is very time consuming (Sari Ay & Ryan, 2019). To avoid it, Sari et al. (2016) developed a statistical tool to assess scenario reliability directly. For daily power generation scheduling instances, Sari and Ryan (2018) demonstrated that reliable scenarios, as identified by this tool, produce high quality solutions, as measured by the average cost of their implementation.

Scenario reduction methods have been studied intensively for identifying a moderate-sized subset of scenarios to approximate the joint distribution of the random variables. Although most of them were developed for classical risk-neutral stochastic optimization problems, they could also be applied to models that incorporate risk-aversion; specifically, when minimizing Conditional Value at Risk (CVaR), since this risk measure can be expressed as an expectation. Recently, however, a more efficient scenario reduction technique was developed for CVaR minimization problems by taking the definition of *effective scenarios* into account (Arpón et al., 2018). If the optimal objective value of a problem changes by removing a scenario, that would be an effective scenario; otherwise, it is considered as ineffective (Rahimian et al., 2018).

If historical observations of the random variables are available, the reliability of a scenario generation and/or reduction method can be assessed. In ensemble forecasting, if the forecasts and observations are consistent statistically, they are interchangeable, and the forecasts can be considered as reliable (Gneiting & Raftery, 2007). Pinson and Girard (2012) studied reliability assessment of probabilistic scenarios. They considered equally likely wind power scenarios and developed some statistical metrics to assess them without taking account what effect those scenarios have on the stochastic programming problems. The reliability of equally likely scenarios (Pinson & Girard, 2012) can be checked by a minimum spanning tree (MST) rank histogram studied by Wilks (2004). The minimum spanning tree lengths are used as a pre-ranking function in this method. For considering unequally likely wind power scenarios for use in stochastic unit commitment, Sari et al. (2016) modified some of the statistical metrics used by Pinson and Girard (2012). In particular, motivated by the use of the mass transportation distance (MTD), also known as Weierstrass distance, they developed the MTD rank histogram, in which the MST length was replaced by the MTD as a pre-ranking function.

To assess the reliability of scenarios for CVaR minimization, we develop a method to evaluate the cost of the observation with respect to the cost distribution near optimality of the scenarios generated for that instance. To focus on effective scenarios, we adapt the scenario reduction approach of (Arpón et al., 2018) for risk-averse problems formulated as:

$$\min_{x\in X} CVaR_\alpha[G(x,\xi)], \tag{1}$$

where x is the set of (first-stage) decisions and ξ is the random vector. This method considers only those scenarios that define the upper tail of $G(x,.)$, which compose the subset of scenarios that is important for our problem.

This work is applicable to problems where similar instances are solved repeatedly, and we have observational data for a collection of past instances. A statistical tool based on the probability of the observation objective values, evaluated according to the cumulative distribution functions of the corresponding scenario costs, is developed to assess the reliability of a scenario generation process for a two-stage, risk-neutral stochastic program. Similar to the MTD or MST rank histograms, this tool can diagnose bias or a dispersion mismatch in the scenario sets. If the risk-averse objective is instead to minimize conditional value-at-risk (CVaR) of cost, we should consider only the scenarios that produce cost in the upper tail of the distribution at the optimal solution. We perform simulation studies to test the effectiveness of the proposed reliability assessment method over a set of randomly generated instances. We apply our method to a financial investment problem for a tradeoff between risk and return (Guo & Ryan, 2021a). To test our method, we generate a set of random scenarios for each of a set of synthetic instances. We can find the optimal solution and objective function of the investment problem considering the generated scenarios set for each instance. Then, we keep those scenarios whose cost falls in the upper tail of the cost distribution and modify their probabilities to form an empirical cumulative distribution function (cdf). The value of this function, evaluated at the observation, is a quantitative measure of the observation's location within the cost distribution. A histogram of these values provides a visual assessment of scenario reliability, or its lack. Similar to the MTD rank histogram, an upward slope from left to right can be seen in the histograms when the ensemble of scenarios is under-dispersed. On the contrary, an over-dispersed ensemble of scenarios results in a downward sloping histogram. Reliable scenario sets produce flat histograms. Because heavy-tailed distributions for investment returns are observed frequently, we also test the ability of our tool to detect kurtosis mismatches.

The paper that motivates this work (Arpón et al., 2018) falls in a stream of progress on scenario reduction for two-stage stochastic programming. In the classical, distribution-oriented approach, the goal is to approximate the "true" distribution of the random vector, ξ, as closely as possible. Recently, Kaut (2021) studied methods to approximate a set of historical data for ξ with subsets selected in various ways, while Rujeerapaiboon et al. (2022) expanded the set of candidate scenarios beyond the atoms of the true distribution. Increasingly, attention has turned to proble-oriented scenario reduction methods, in which the optimization context is considered. (Fairbrother et al., 2022) developed an approach based on sampling from risk regions, whereas Arpón et al. (2018) based their method on probability metrics. In a parallel effort, Henrion and Römisch (2022) approximated the recourse function by formulating a semi-infinite program. Prochazka and Wallace (2020) discussed a conceptual approach to constructing a scenario tree that minimizes the discrepancy between in-sample and out-of-sample performance. Bertsimas and Mundru (2022) minimized a problem-dependent divergence, defined as the loss in decision quality when a reduced set of scenarios is used to optimize. Most recently, Zhang et al. (2023) selected a subset of scenarios from among a large finite set to minimize the error in approximating the recourse function.

Our work is more closely related to the evaluation of methods that have been developed to generate scenarios that accurately represent the stochastic processes. Defining the cost of the stochastic programming problem considering the scenarios produced by a scenario generation method is a way to evaluate the considered method (Wang et al., 2011). However, this approach is computationally very demanding. Kaut and Wallace (2007) defined in-sample and out-of-sample stability as two measures of the quality of scenarios. If the expected costs of solutions produced by different scenarios sets are similar, then in-sample stability exists. Out-of-sample stability occurs when alternative scenario sets produce solutions with similar true expected cost. Simulation studies, similar to those in this paper, were performed by Sari et al. (2016) to demonstrate that the MTD rank histogram has similar diagnostic abilities as the MST rank histogram. A new version of the mass transportation distance (MTD) rank histogram was used by Sari Ay and Ryan (2019) to assess the reliability of the unequally likely scenarios used for unit commitment and server location problems. They used past instances to assess the reliability of scenarios to avoid solving stochastic programming instances. They consider the impact that each scenario has on the problem. Using MTD rank histograms to assess the reliability of scenarios guided the selection of hyperparameters used to estimate parameters of the distribution of stock index or momentum portfolio returns (Guo & Ryan, 2021b, 2023). Emirhüseyinoğlu et al. (2023) used them to validate machine learning-generated scenarios for crop yield. (Ryan & Shah Abadi, 2022) adapted the MTD rank histogram to assess the reliability of scenarios with respect to the upper tail of the cost distribution. Over- or under-dispersion and/or bias in the scenario sets can be diagnosed with this graphical tool. For a risk-averse newsvendor model, they verified that non-flat shapes of the rank histogram are the result of the mismatch between scenarios and observations.

The multi-dimensional character of ensemble weather forecasts motivated the use of the MST length as a pre-rank function. The MTD played a similar role for multi-dimensional scenarios needed in many applications of stochastic programming. In this paper, we aim to assess the reliability of scenarios whose costs fall in the upper tail of the cost distribution. Because cost has a single dimension, we modify the pre-rank function to more simply evaluate the cost of the observation with respect to the cost distribution generated by scenarios. For the special case of CVaR that equals expectation, we compare the results of our new rank histogram with those of the MTDRh applied to the costs of scenarios and observations and also to the scenarios and observations directly.

This rest of the paper is organized as follows. The research problem is described in Section 2. In Section 3, the proposed reliability assessment method is described. An investment problem is introduced in Section 4. Section 5 describes of the simulation studies performed for two distributions of rate of return, including the normal distribution and a heavy tailed distribution. Finally, conclusions are summarized in Section 6.

2. Problem Description

The inherent uncertainty in many real systems motivates the use of stochastic programming for decision making. For computational tractability, a discrete set of probabilistic scenarios is generated to approximate the joint probability distribution of the uncertain parameters of the optimization problem. The question that arises after generating scenarios is how well they match the true distribution of the observed data. However, the objective function represents the decision-maker's aim. When decision-maker's risk-aversion can be expressed as in (1), the reliability of those scenarios whose costs define the upper $(1-\alpha)$-probability tail of the cost distribution is important, and we can ignore the other scenarios. Obviously, assessing all scenarios is a time-consuming process. The goal of scenario reliability assessment considering risk-aversion is to help decision-makers efficiently find out how well the interesting part of the cost distribution is approximated.

The verification rank histogram is a tool to assess how the scenario sets generated for past instances compare with the corresponding observations. The mass transportation distance (MTD) rank histogram is an example of the verification rank histograms using MTD as a pre-ranking function used for assessing the reliability of unequally likely scenarios. If the resulting rank histograms are flat, we can conclude that the generated scenario sets, and their corresponding observations are drawn from the same distribution. MTD rank histograms can be applied to scenarios directly. To access the reliability of those scenarios whose cost falls in the tail, we develop a simpler histogram that works in the single dimension of the cost of scenarios.

Most of the scenario reduction methods are designed for the following form of optimization problem:

$$\min_{x \in X} E[G(x, \xi)], \tag{2}$$

where ξ is an uncertain random vector, x is a vector of decision variables, $G(\cdot)$ is an objective function that depends on ξ and x, and X is a convex closed set. Decision-making considering such a formulation is risk-neutral. However, decision-makers are sometimes risk-averse, which motivates the need to optimize a risk metric instead of expectation. Some risk measures can be expressed as expectations. One of those risk measures is Conditional-Value-at-Risk (CVaR), which is a coherent (i.e., translation-invariant, positively homogeneous, and convex) risk metric.

To define the CVaR, we should first review the Value at Risk (VaR).The Value at Risk (VaR) and the Conditional-Value-at-Risk (CVaR) of a random variable Y are defined as follows (Rockafellar & Uryasev, 2000).

$$VaR_\alpha[Y] := \min\{t|F(t) \geq \alpha\} = \min\{t|P(Y \leq t) \geq \alpha\} \tag{3}$$

$$CVaR_\alpha[Y] := \frac{1}{1-\alpha}\int_\alpha^1 VaR_\gamma[Y]\, d\gamma, \tag{4}$$

where $F(.)$ is the cdf of random variable Y, and $\alpha \in [0,1]$ represents the risk preference.

Based on (Rockafellar & Uryasev, 2000), for the following minimization problem, one optimal solution and the optimal value are $VaR_\alpha[Y]$ and $CVaR_\alpha[Y]$, respectively.

$$\min_{\eta \in R}\left\{\eta + \frac{1}{1-\alpha}E[(Y-\eta)_+]\right\} \tag{5}$$

where $(Y-\eta)_+$ is $\max\big((Y-\eta), 0\big)$.

Considering CVaR instead of expectation in equation (2), we obtain equation (1).

If we substitute $G(x, \xi)$ for Y, the following optimization problem is obtained:

$$\min_{x\in X,\ \eta \in R}\left\{\eta + \frac{1}{1-\alpha}E[(G(x,\xi)-\eta)_+]\right\} \tag{6}$$

The quantity $(G(x,\xi)-\eta)_+$ is zero for any fixed x and scenario ξ such that $G(x,\xi) \leq \eta$. So, if $G(x,\xi)$ is a cost function, $CVaR_\alpha[G(x,\xi)]$ is the expected cost of scenarios whose costs fall in the upper $(1-\alpha)$-probability tail of the cost distribution. As a result, the scenarios ξ for which $G(x,\xi)$ does not fall in the tail d0 not contribute to the optimal solution.

In this paper, we present a method to assess scenario reliability with respect to the upper $(1-\alpha)$-probability tail of the cost distribution. The procedure to test our method is to simulate a set of synthetic instances and a set of scenarios for each instance. By systematically varying the extent to which the distribution of scenarios matches that of the synthetic observations, we can test the ability of our tool to recognize when observations are indistinguishable from scenarios or diagnose the type and extent of mismatch between them.

3. Methodology

We develop a method to assess the reliability of scenarios based on the cumulative distribution function of their costs. To achieve this goal, we need a way to represent that the cost of scenarios and observations are interchangeable. We do this by constructing the empirical cdf of scenario costs and evaluate the observation cost under it. This value represents the probability that the scenario costs do not exceed the cost of the observation. The histogram of those probabilities over all instances provides a visual depiction of scenarios' reliability or unreliability.

Suppose that we have n instances, for each of which we have m scenarios. Let $\Xi_i = \{\xi_{i1}, \xi_{i2}, \dots, \xi_{in}\}$ be a set of scenarios for instance i, $i = 1, \dots, m$, where ξ_{ij} has probability P_{ij} and $\sum_{j=1}^{n} P_{ij} = 1$, and let ξ_{io} be the observed value of ξ in instance i. For each instance, i, we use Ξ_i and its distribution P_i to solve the problem and obtain an optimal solution, x_i^*. If $G(x_i^*, \xi_{io}) < VaR_\alpha[G(x_i^*, .)]$ then discard instance i. Otherwise, identify the corresponding set of *effective* scenarios:

$$\overline{D}_i^* = \{\xi \in \Xi_i: G(x_i^*, \xi) \geq VaR_\alpha[G(x_i^*, .)]\} \tag{7}$$

The effective scenarios are the subset that determine the CVaR at optimality. Also define the partition of scenarios:

$$\begin{aligned} D_i^* &= \{\xi \in \Xi_i: G(x_i^*, \xi) < VaR_\alpha[G(x_i^*, .)]\} \\ E_i^* &= \{\xi \in \Xi_i: G(x_i^*, \xi) = VaR_\alpha[G(x_i^*, .)]\} \\ F_i^* &= \{\xi \in \Xi_i: G(x_i^*, \xi) > VaR_\alpha[G(x_i^*, .)]\} \end{aligned} \tag{8}$$

Motivated by (Arpón et al., 2018), we revise the scenario probabilities to:

$$\begin{aligned} &\hat{P}_{ij} = 0, && j \in D_i^* \\ &\hat{P}_{ij} = P_{ij}/(1-\alpha), && j \in F_i^* \\ &\hat{P}_{ij} = \left(1 - \alpha - \sum\nolimits_{j \in F_i^*} P_{ij}\right) / \big(|E_i^*|(1-\alpha)\big), && j \in E_i^* \end{aligned} \tag{9}$$

The scenario set for instance i is replaced by the corresponding set of values of $G(x_i^*, \xi_{ij}), j = 1, \dots, n$, and the observation for instance i is replaced by $G(x_i^*, \xi_{io})$. In each non-discarded instance, i, we use the revised probabilities to construct the cumulative distribution function of $G(x_i^*, \xi_{ij}), j = 1, \dots, n$.

To construct the histogram of the probabilities of the objective function of the observations, suppose that k_i is the number of scenarios of instance i whose revised probabilities are greater than zero, and they are sorted in order of increasing cost. The observation probability is computed as:

$$q_i = \begin{cases} 0 & G(x_i^*, \xi_{io}) < G(x_i^*, \xi_{i1}) \\ \hat{P}_{i1} & G(x_i^*, \xi_{i1}) \leq G(x_i^*, \xi_{io}) < G(x_i^*, \xi_{i2}) \\ \vdots & \vdots \\ \hat{P}_{i1} + \hat{P}_{i2} + \dots + \hat{P}_{ik_i} = 1 & G(x_i^*, \xi_{ik_i}) < G(x_i^*, \xi_{io}) \end{cases} \tag{10}$$

Thus, if we have n_1 non-discarded instances, we will have n_1 probabilities, $\left\{\hat{P}_{ij} | i \in \{1, \dots, n_1\}, j \in \{1, \dots, k_i\}\right\}$, and we can construct their histogram to depict their pattern.

4. Investment Problem

To set the optimal value of allocation of wealth to risk-free and risky assets which not only can maximize the expected excess return but also minimize the investment risk, we can use the mean-risk stochastic optimization model that allows short-selling of the risky asset. Suppose f_t is the return of a risk-free asset; R_t is the uncertain return of a risky asset; w_t is the weight assigned to the risky asset; and $1 - w_t$ as the weight assigned to the risk-free asset. A generic mean-risk optimization model using ρ as a risk measure and $\lambda \in [0, 1]$ as a risk-aversion parameter at time *t* can be formulated as:

$$\min_{w_t} (1-\lambda)E[-(R_t - f_t)w_t] + \lambda\rho[-(R_t - f_t)w_t] \tag{11a}$$

$$-1 \le w_t \le 1 \tag{11b}$$

(Guo & Ryan, 2021a) showed that for a coherent risk measure, like $CVaR_\alpha$, and according to equations (5) and (6), the risk-averse formulation is:

$$\min_{w_t,\ \eta_{t+1}} (1-\lambda)E[-(R_t - f_t)w_t] + \lambda\left\{\eta_{t+1} + \frac{1}{1-\alpha}E[(-(R_t - f_t)w_t - \eta_{t+1})_+]\right\} \tag{12a}$$

$$-1 \le w_t \le 1 \tag{12b}$$

where η_{t+1} is the α-quantile of the negative excess return distribution at time *t* + 1.

They also proved that the optimal solution for the above problem is as follows:

$$w_t^* = \begin{cases} -1 & if \quad E[R_t] - f_t < -\lambda d_{\alpha,t}^+ \\ 0 & if \quad -\lambda d_{\alpha,t}^+ \le E[R_t] - f_t \le \lambda d_{\alpha,t}^- \\ 1 & if \quad \lambda d_{\alpha,t}^- < E[R_t] - f_t \end{cases} \tag{13}$$

where $d_{\alpha,t}^+ \equiv CVaR_\alpha[R_t] - E[R_t] \ge 0$ and $d_{\alpha,t}^- \equiv CVaR_\alpha[-R_t] + E[R_t] \ge 0$.

We consider a special case of this problem for our numerical experiments in which we set the risk aversion parameter, λ, to one. Then the optimal solution has the following form:

$$w_t^* = \begin{cases} -1 & if \quad f_t > CVaR_\alpha[R_t] \\ 0 & if \quad -CVaR_\alpha[-R_t] \le f_t \le CVaR_\alpha[R_t] \\ 1 & if \quad f_t < -CVaR_\alpha[-R_t] \end{cases} \tag{14}$$

As a result, the optimal value of the objective function is:

$$OB^* = \begin{cases} R_t - f_t & if \quad f_t > CVaR_\alpha[R_t] \\ 0 & if \quad -CVaR_\alpha[-R_t] \le f_t \le CVaR_\alpha[R_t] \\ -R_t + f_t & if \quad f_t < -CVaR_\alpha[-R_t] \end{cases} \tag{15}$$

5. Numerical Results

In this study, the monthly rate of return of the risk-free investment is considered to be $f_t = 0.0005$, selected based on returns of the US Treasury bill (T-bill) according to the CRSP database (www.crsp.com) between January 1, 2001, and December 31, 2019. We consider two probability models for the risky rate of return.

5.1. Normally Distributed Rate of Return

In first experiment, all scenarios and observations are generated from Normal distributions. We simulated *m* = 5000 instances, with *n* = 5000 scenarios randomly generated for each. For the risky asset, we randomly sampled 5000 rates of return from $N(\mu = 0.005,\ \sigma = 0.033)$ for the simulated observations. For scenarios, we systematically varied the parameters of the simulated scenario distribution with the 18 combinations of $\mu \in \mathrm{A} = \{-0.01, 0.005, 0.04\}$ and $\sigma \in \mathrm{B} = \{0.002, 0.003, 0.01, 0.033, 0.05, 0.06\}$. For each value of $\alpha \in \{0, 0.5, 0.75, 0.9\}$, Figures 1-3 show the related panels of the histograms. The rows represent the three mean values of the scenario distributions, which also quantify their bias, and the columns correspond to the various values of the scenario standard deviation. In each figure, the panel with bars shaded white shows results for when the scenario distribution matches that of the observations.

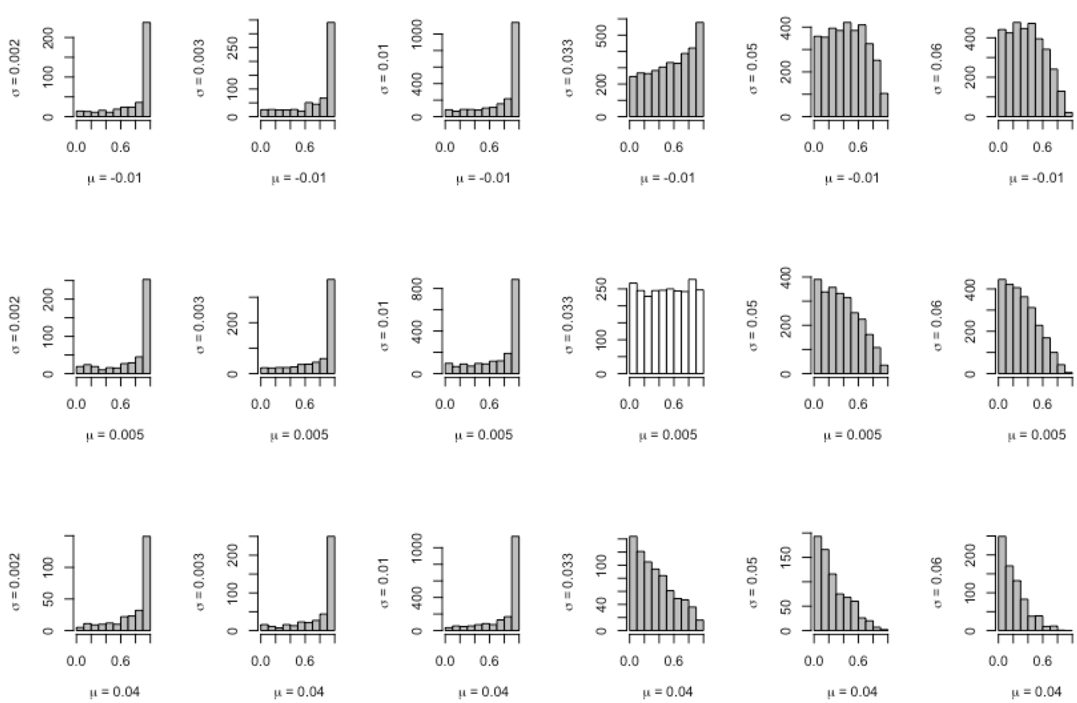

Figure 1: *q-histogram of normal distribution simulation experiment for α=0.5. Rows are labeled with μ, and columns are labeled with σ.*

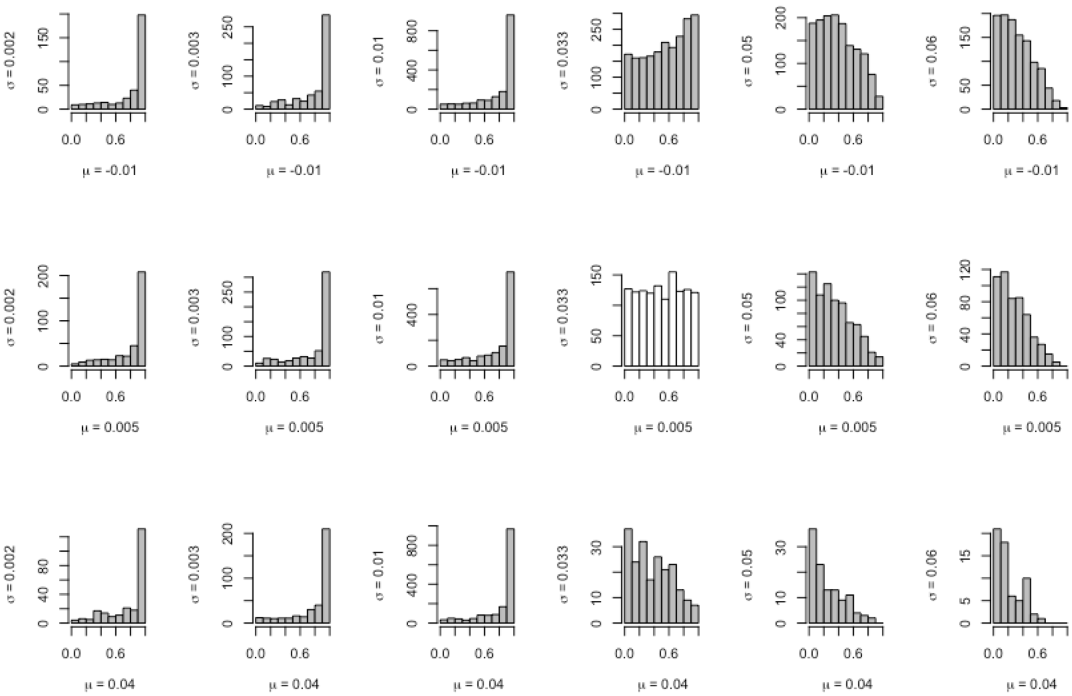

Figure 2: *q-histogram of normal distribution simulation experiment for α=0.75. Rows are labeled with μ, and columns are labeled with σ.*

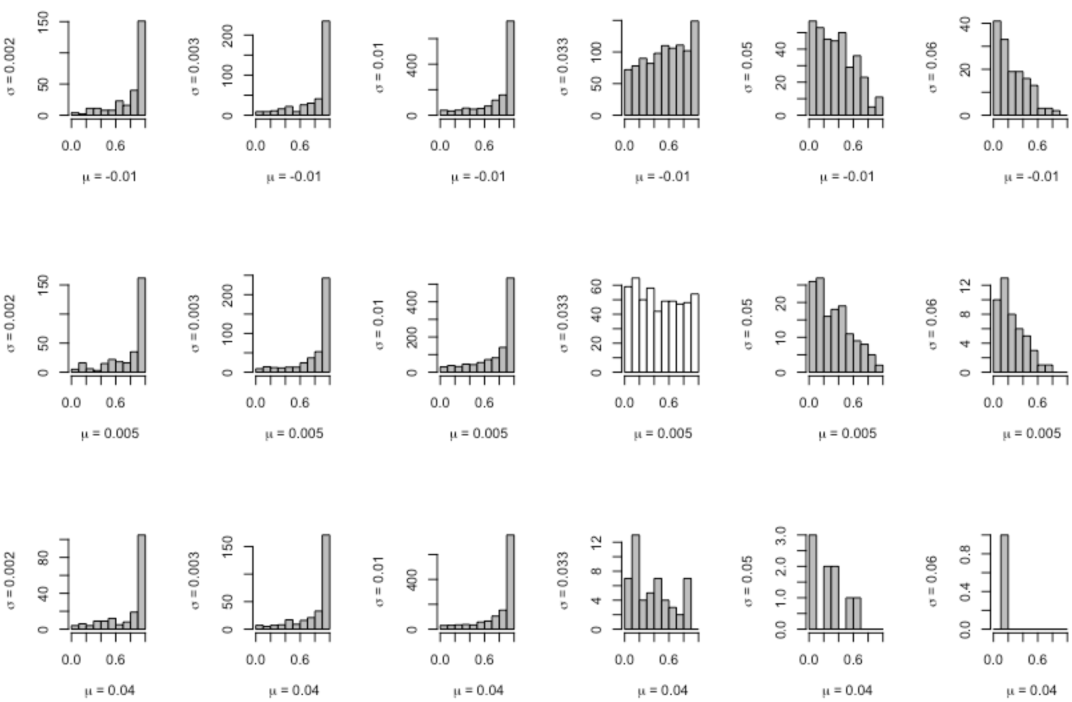

Figure 3: *q-histogram of normal distribution simulation experiment for α=0.9. Rows are labeled with μ, and columns are labeled with σ.*

The results for different values of $\alpha \in \{0.5, 0.75, 0.9\}$ show that when the standard deviation of scenarios is smaller than that of observations; i.e., $\sigma_{scen}^2/\sigma_{obs}^2 < 1$, the resulting histograms tend to have increasing trends. In other words, most of the observation cost probabilities are near one since the cost of an observation that falls in the upper tail exceeds the costs of the scenarios. On the contrary, bigger standard deviations for scenarios compared to observations, $\sigma_{scen}^2/\sigma_{obs}^2 > 1$, yield decreasing trends in the resulting histograms. In these cases, the costs of observations in the retained instances tend to be smaller than the scenario costs in the upper tail. When both scenarios and observations are generated from the same distributions, the histogram has no increasing or decreasing trend. When $\sigma_{scen}^2 = \sigma_{obs}^2$ and $\mu_{scen} < \mu_{obs}$, the trend is increasing since most observation fall in the more extreme upper tail of the cost distribution. On the other hand, when $\sigma_{scen}^2 = \sigma_{obs}^2$ and $\mu_{scen} > \mu_{obs}$, the trend is decreasing because those observations that fall in the upper tail are not as extreme. These patterns are more regular for smaller values of α because a larger number of instances is retained for inclusion in the histograms.

5.2. Heavy-Tailed Rate of Return Distribution

For second experiment, we consider a modified version of the heavy-tailed distribution introduced by (Lim et al., 2011), which is a mixture of a normal distribution and a negative exponential tail. Let $I(\epsilon)$ and $I(0.5)$ be Bernoulli random variables with respective parameters ϵ and 0.5, W be normally distributed with parameters μ and σ, and Y be an exponential random variable with rate parameter $\lambda = 1/\sigma$.

Assume all of these random variables are mutually independent, and define:

$$X = \big(1 - I(\epsilon)\big)W + I(\epsilon)[I(0.5)(Y + \mu - \sigma) + \big(1 - I(0.5)\big)(-Y + \mu + \sigma)] \qquad (18)$$

Then $E[X] = \mu$, $Var[X] = \sigma^2$, and the skewness equals zero. The kurtosis depends on ϵ and σ, as illustrated empirically in Figure 4, which shows histograms of samples of size 10,000 for various combinations of these parameters. As for the normal distribution, we simulated $m = 5000$ instances, with $n = 5000$ scenarios randomly generated for each. The mean of normal distribution is set to 0.005. If we consider three different values for $\sigma \in \{0.033,\ 0.05,\ 0.07\}$, the corresponding negative exponential distribution parameter values would be $\lambda \in \{1/0.033,\ 1/0.05,\ 1/0.07\}$. The Bernoulli distribution parameter is selected from $\epsilon \in \{0.05,\ 0.3,\ 0.5,\ 0.7,\ 0.95\}$. Figure 4 shows how when λ increases for a fixed value of ϵ, the kurtosis is not affected much. On the other hand, when λ is fixed, the kurtosis increases with ϵ since the distribution of X is dominated by the combination of the exponential distributions.

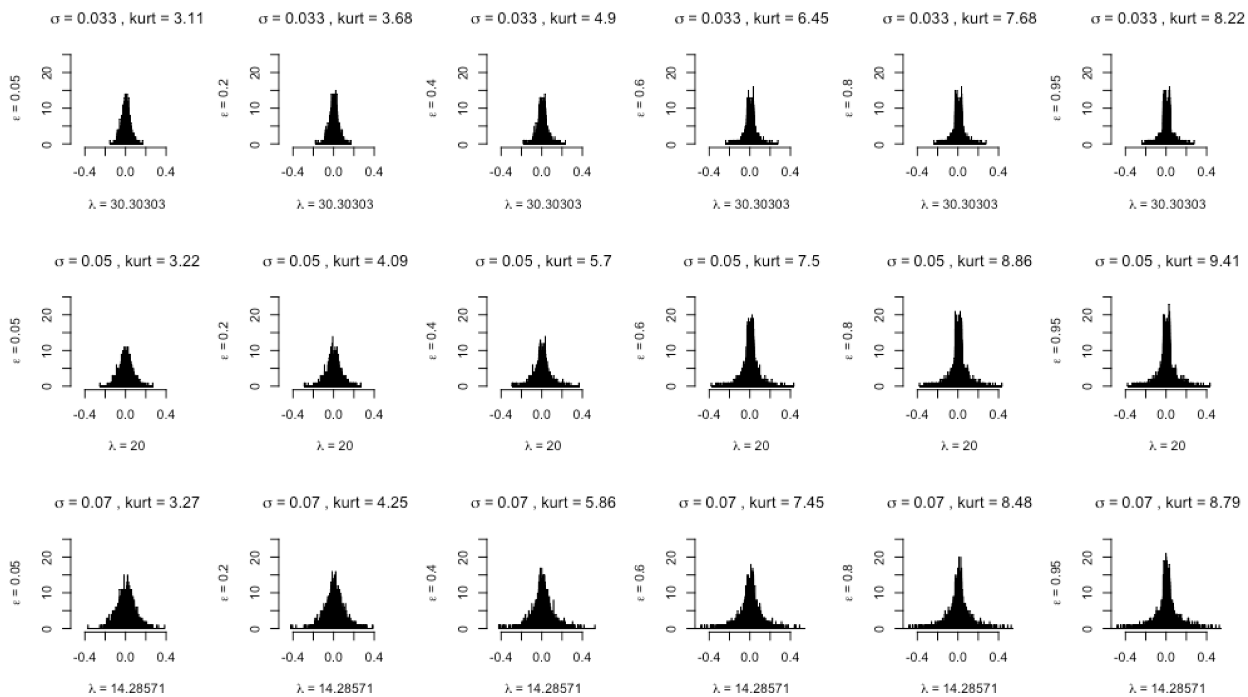

Figure 4: The effect of Bernoulli and negative exponential parameters on the shape of heavy-tail distribution.

If we plot the q-histograms for this distribution considering $\mu = 0.005$, and 18 combinations of λ and ϵ, Figures 5-7 will be obtained. The observations are generated using $\lambda = 1/0.033$ and $\epsilon = 0.05$.

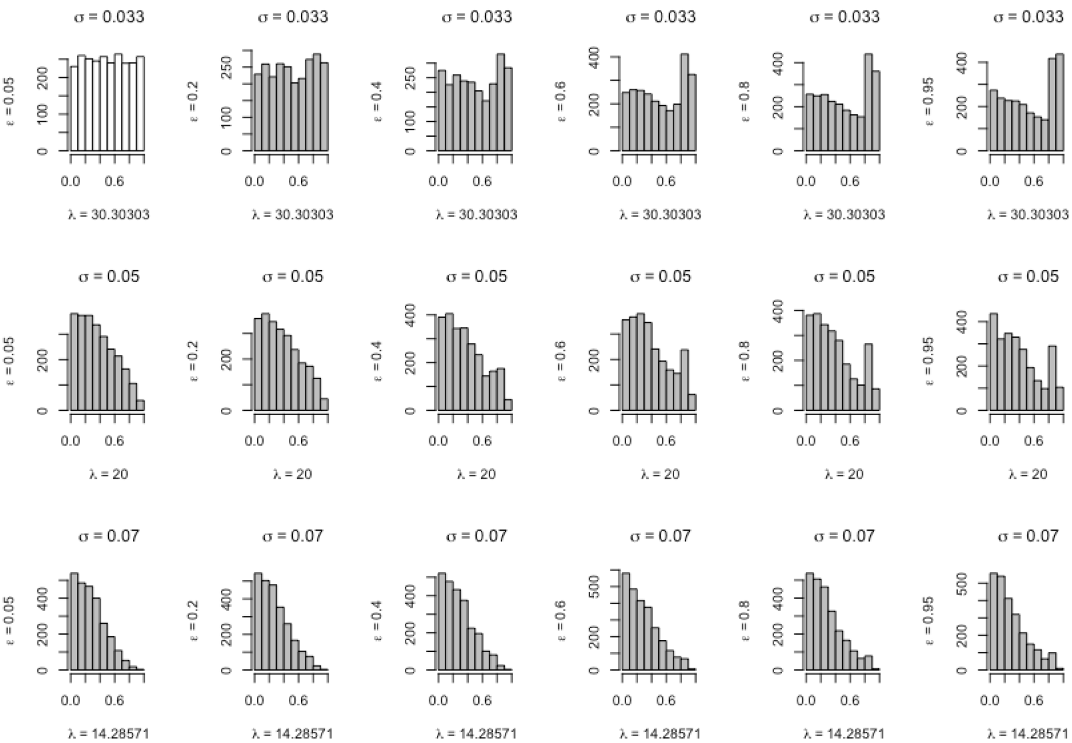

Figure 5: *q-histograms of heavy-tailed distribution simulation experiment for α=0.5. Rows are labeled with λ, and columns are labeled with ϵ.*

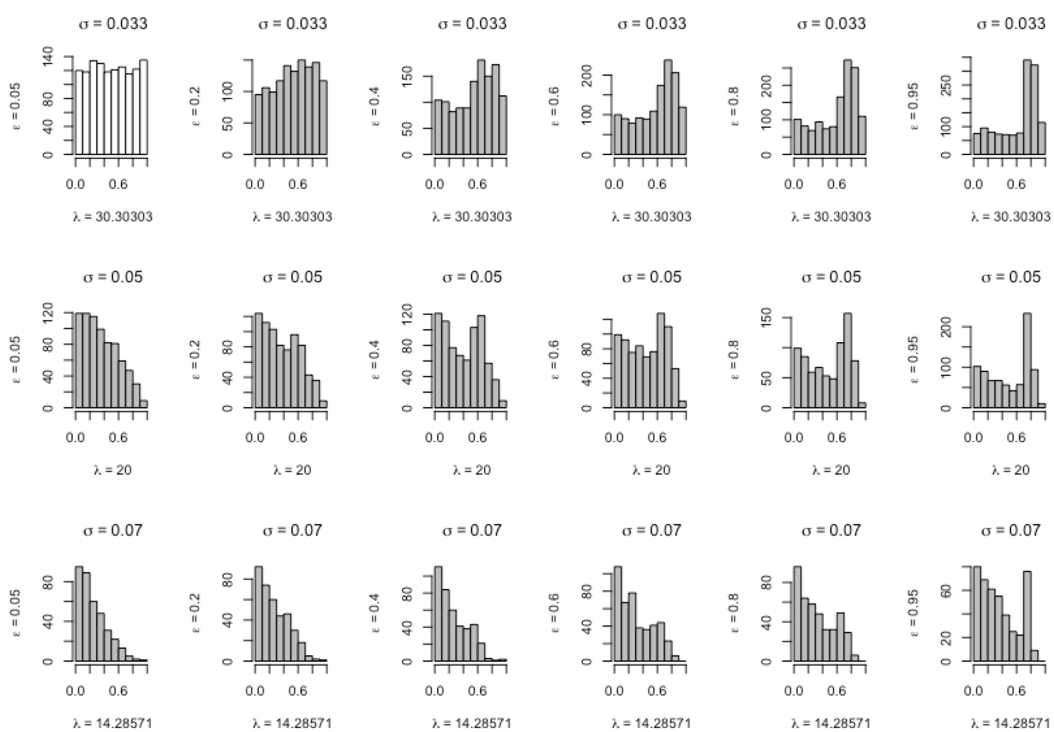

Figure 6: *q-histograms of heavy-tailed distribution simulation experiment for α=0.75. Rows are labeled with λ, and columns are labeled with ϵ.*

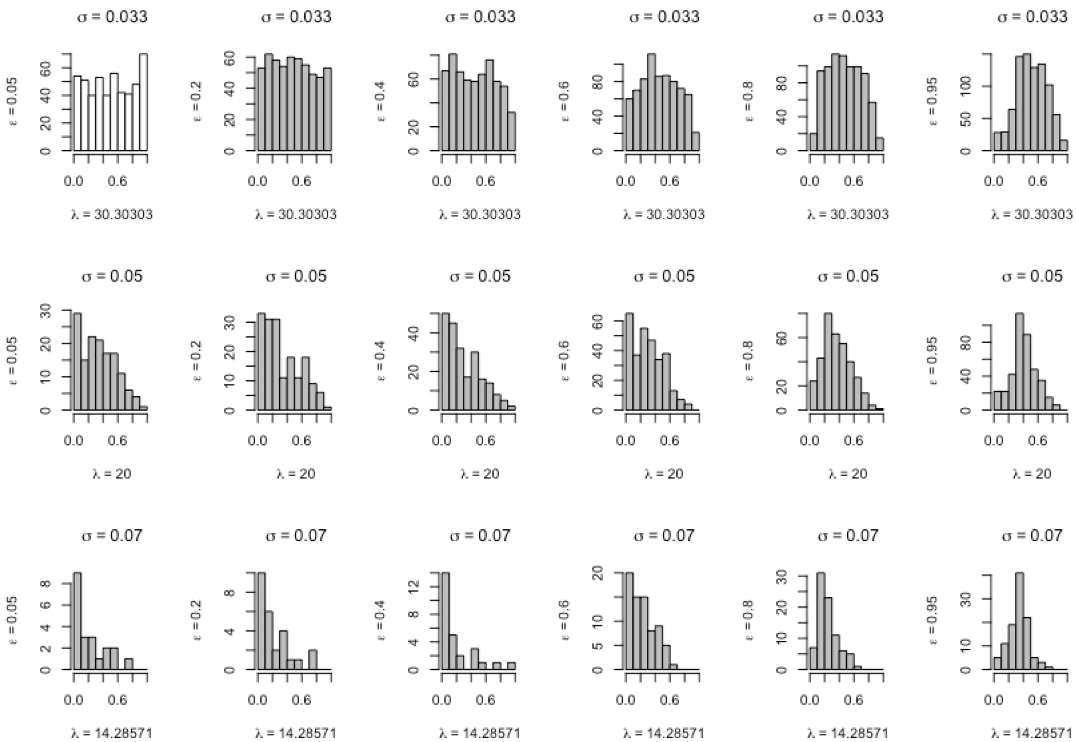

Figure 7: *q-histograms of heavy-tailed distribution simulation experiment for α=0.9. Rows are labeled with λ, and columns are labeled with ϵ.*

As can be understood from Figures 5-7, decreasing values of λ (i.e., increasing variance) direct the observation probabilities to smaller values when ϵ is fixed. On the other hand, as the values of ϵ approach 1, downward-sloping histograms are replaced by hill-shaped ones. Unlike in the first experiment, these patterns are more evident for larger values of α that represent a higher level of risk aversion.

5.3. Comparison between the results of q-histogram and MTD rank histogram

In this section, we compare the resulting histograms obtained by our tool for the cost of scenarios and observations and MTDRh package (Sari & Ryan, 2016) for the costs of scenarios and observations considering $\alpha = 0$ and also for the scenarios and observations directly. All parameter values are as described in section 5.2. and 5.3. We simulated m = 1000 instances, with n = 1000 scenarios randomly generated for each instance. Comparing Figure 8 with Figures 9 and 10, and Figure 11 with Figures 12 and 13, we conclude that q-histograms are better able to diagnose unreliability with respect to heavy tails. In MTDRhs implemented for the cost of scenarios and observations as well as for the scenarios and observations directly under the normal distribution, when the mean and standard deviation of scenario and observations are similar, the resulting histograms are flat. However, there are other histograms which are also flat, which could cause false confidence in unreliable scenario sets. In MTDRh for the cost of heavy-tailed distribution, the results are more confusing compared to the normal distribution. In Figure 12, most histograms appear flat, and it is very difficult to distinguish reliable scenario sets. However, when we consider heavy-tailed distribution, the MTDRh implemented for the scenarios and observations directly can diagnose the reliable scenario set. Also, the trends for $\sigma^2_{scen}/\sigma^2_{obs} < 1$ and $\sigma^2_{scen}/\sigma^2_{obs} > 1$ are very different between the MTDRhs and q-histograms. Considering data generated from a normal distribution, for $\alpha = 0$ (where CVaR represents ordinary expectation) when $\sigma^2_{scen}/\sigma^2_{obs} < 1$, the q-histograms are U-shaped. On the other hand, when $\sigma^2_{scen}/\sigma^2_{obs} > 1$, the resulting q-histograms are hill-shaped. Still, we observe an almost uniform q-histogram when the scenario distribution matches that of the observations. As can be understood from Figures 9 and 10, MTDRh trends for the cost of scenarios and observations, and those for the scenarios and observations directly, are very similar to each other, and different from the trends of the q-histogram. An increasing trend for $\sigma^2_{scen}/\sigma^2_{obs} < 1$ and a decreasing trend for $\sigma^2_{scen}/\sigma^2_{obs} > 1$ is understandable. In heavy-tailed distribution, the trends in all panels of histograms are different while for all of them the reliable scenarios are flat. Overall, it appears that the MTDRh is not able to diagnose unreliability with respect to the heaviness of tails, but the new q-histogram works well for this purpose.

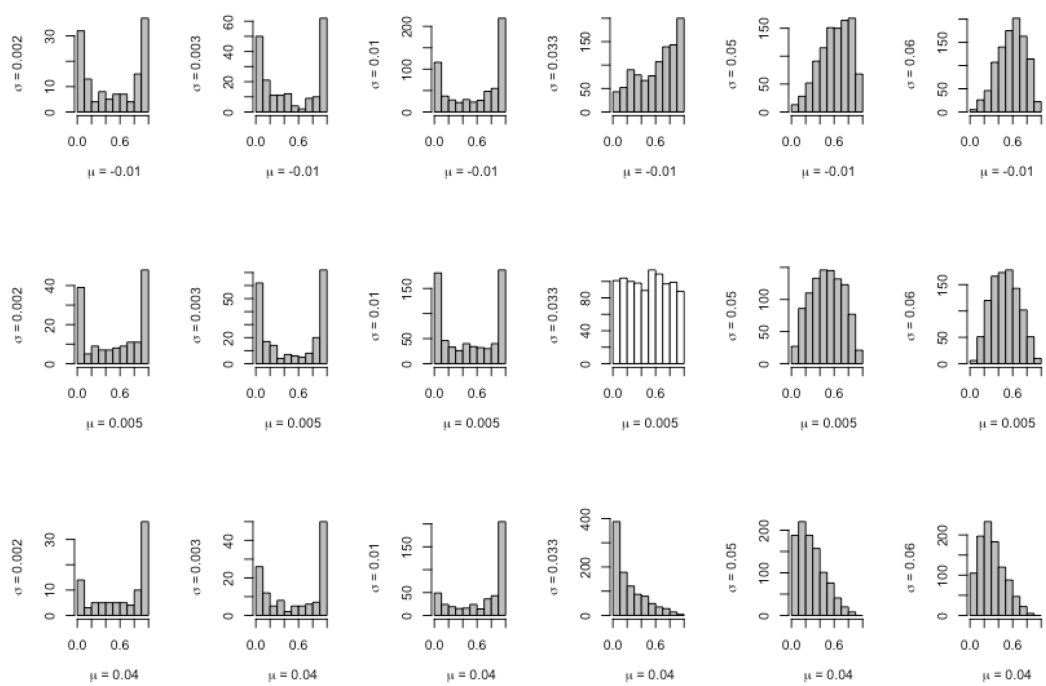

Figure 8: q-histograms of normal distribution simulation experiment for α=0. Rows are labeled with λ, and columns are labeled with ϵ.

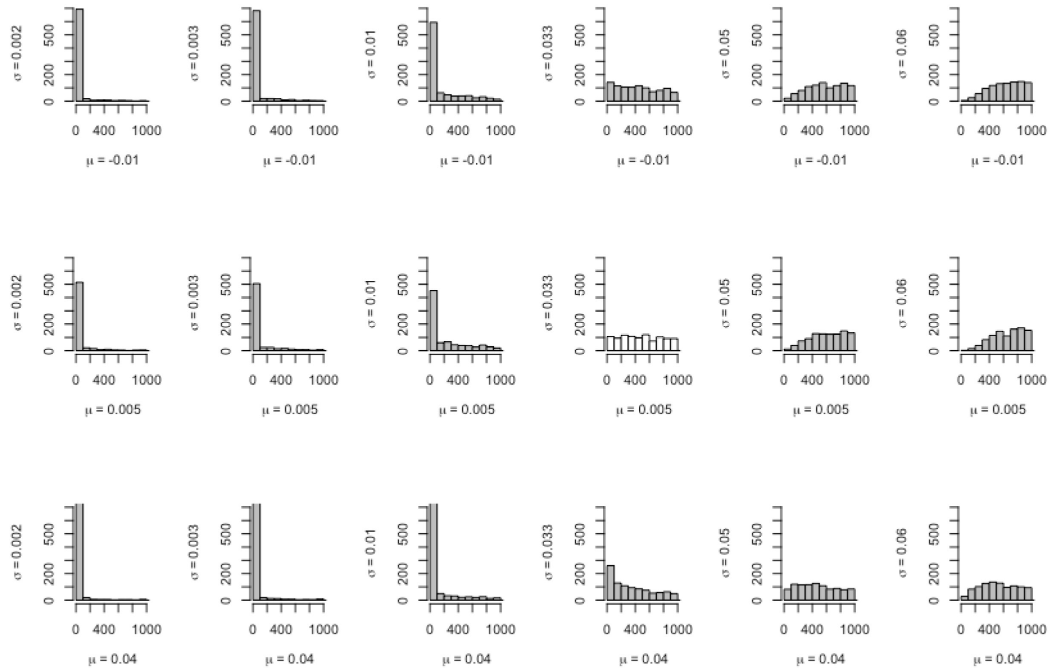

Figure 9: *MTDRh of cost of normal distribution simulation experiment for α=0. Rows are labeled with λ, and columns are labeled with ϵ.*

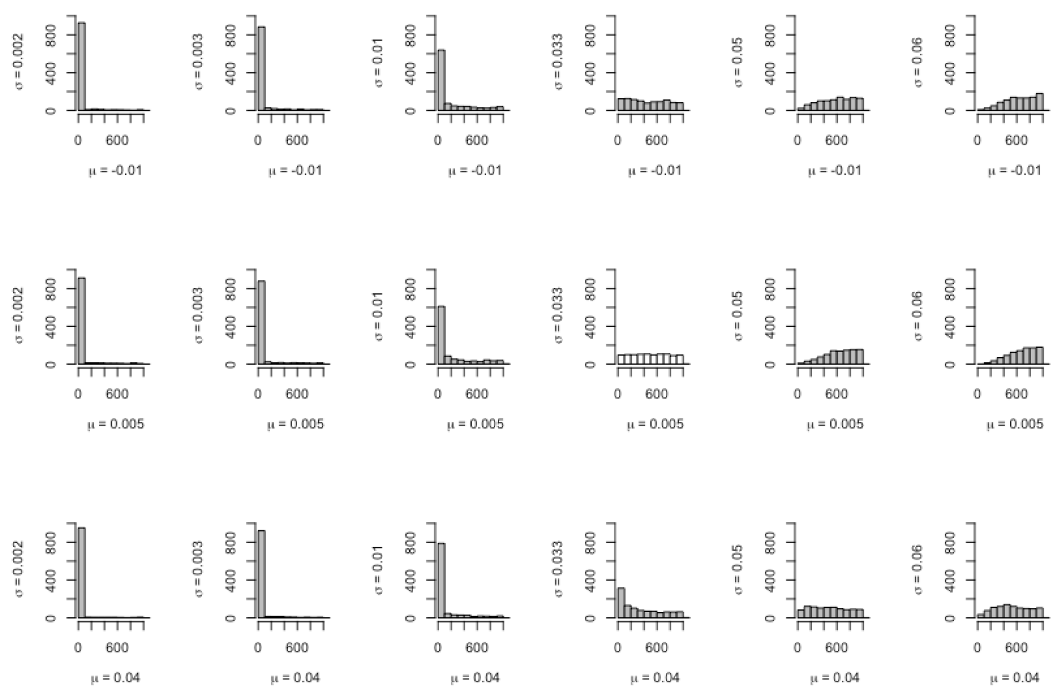

Figure 10: *MTDRh of normal distribution simulation experiment for α=0. Rows are labeled with λ, and columns are labeled with ϵ.*

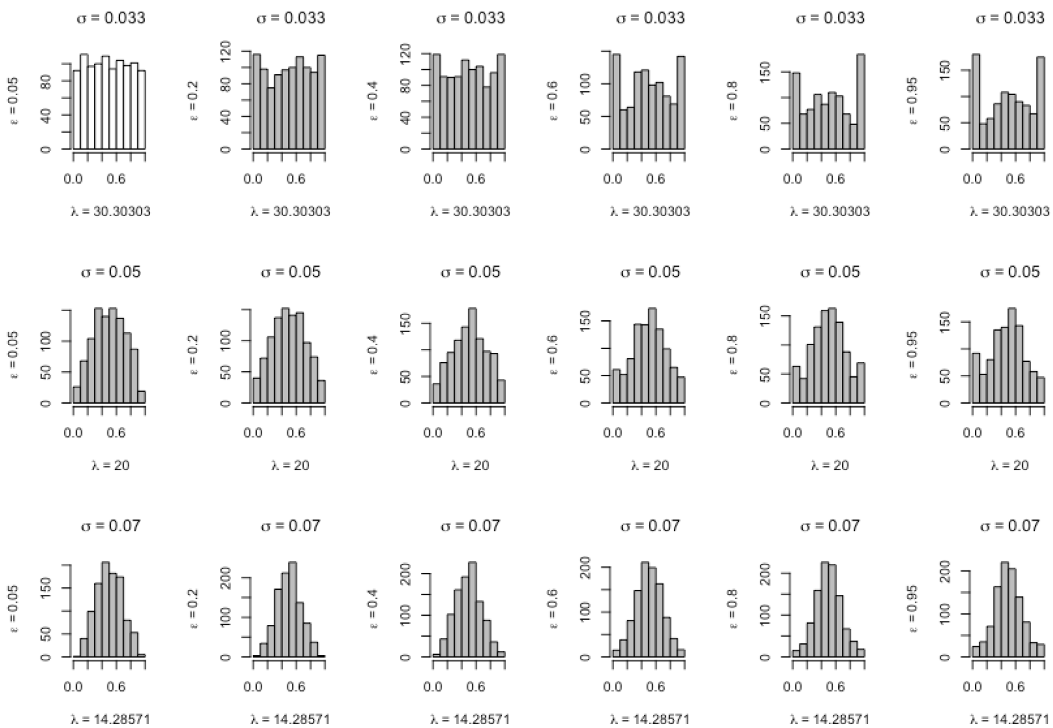

Figure 11: *q-histograms of heavy-tailed distribution simulation experiment for α=0. Rows are labeled with λ, and columns are labeled with ϵ.*

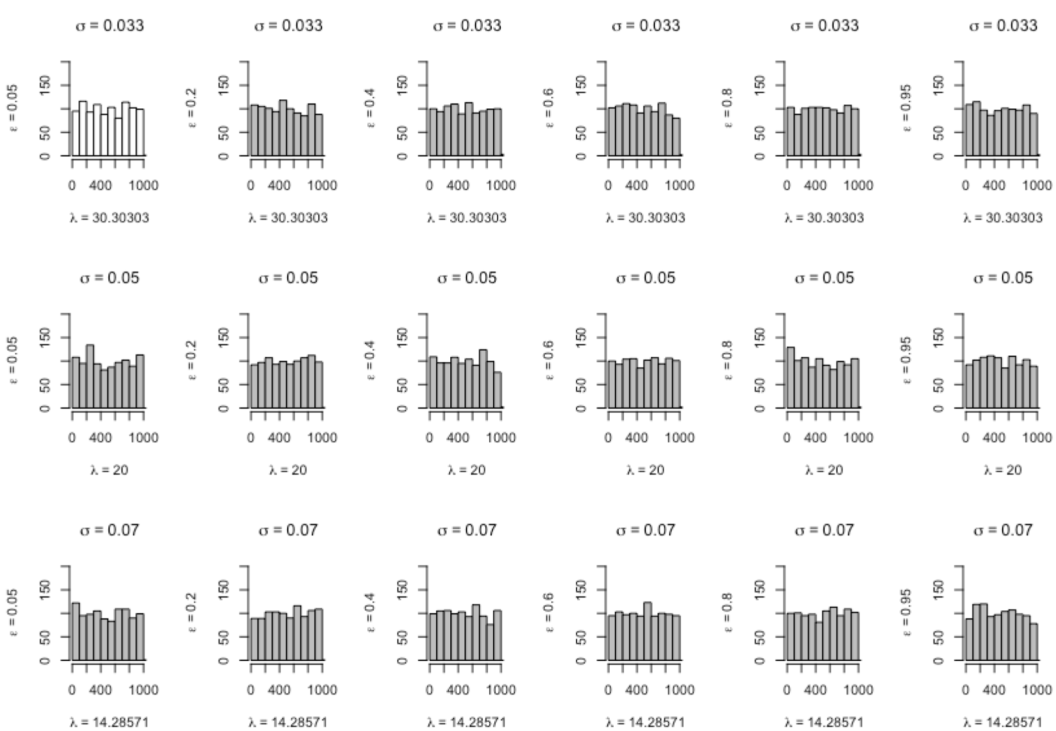

Figure 12: *MTDRh of cost of heavy-tailed distribution simulation experiment for α=0. Rows are labeled with λ, and columns are labeled with ϵ.*

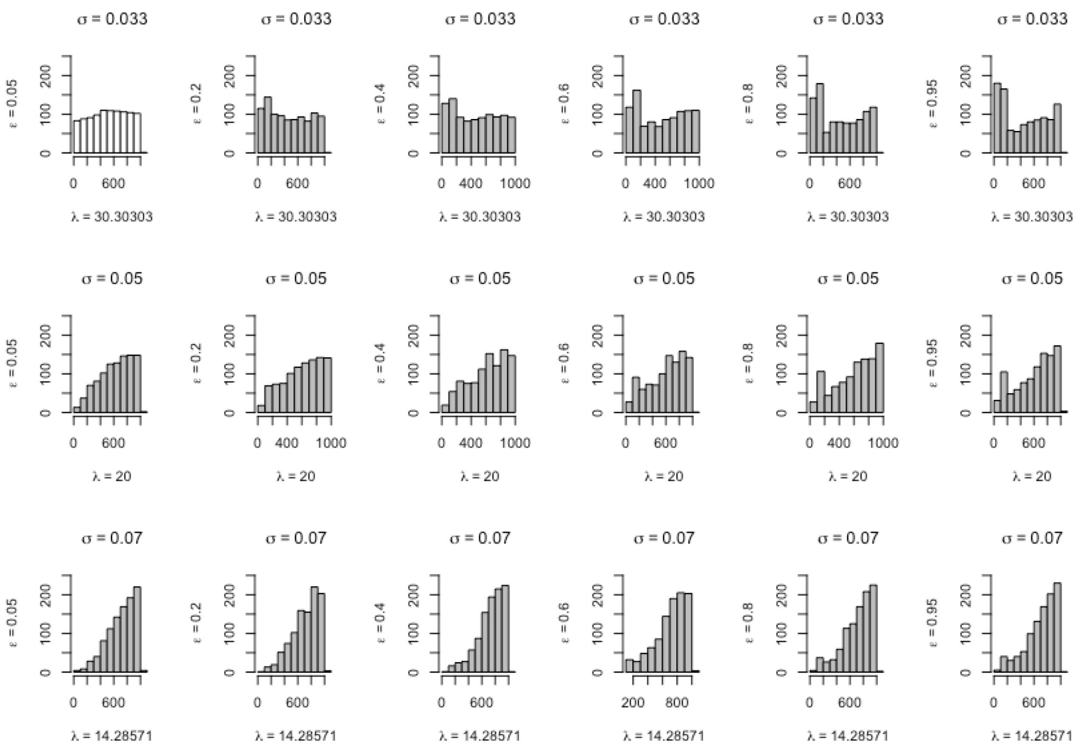

Figure 13: *MTDRh of heavy-tailed distribution simulation experiment for α=0. Rows are labeled with λ, and columns are labeled with ϵ.*

6. Conclusion

Assessing the reliability of scenario sets generated by alternative scenario generation methods without solving full-scale stochastic problems is important for finding high-quality scenarios in a reasonable amount of time. To assess the reliability of scenarios considering conditional value-at-risk (CVaR) of cost, we developed a visual tool based on a comparison between the costs of observations and scenarios in the upper tail. This tool can be used to diagnose bias or a mismatch in either dispersion or kurtosis in the scenario sets. We test our tool by applying it to synthetic instances of an investment problem considering two distributions of the risky rate of return. Our numerical example led us to conclude that our proposed tool can diagnose under- and over-dispersion as well as unreliability in terms of fatter tails. For the purpose of our simulation experiments, we used the optimal decision found with the whole set of scenarios to generate the empirical distribution in the tail and evaluate the cost of the observation.

The dependence of this method on the optimal solution approximated according to an arbitrary sample of scenarios is a limitation. In addition, a more precise definition of reliability, in the vein of approximation error or problem-dependent divergence, could help establish the theoretical performance of a reliability assessment method in general. In future work, one goal is to use a moderate-sized subset of scenarios to more efficiently approximate the optimal solution. Further testing is needed in other applications, with multi-dimensional scenarios and scenario generation methods other than sampling.

Acknowledgment: This material is based upon work supported by the National Science Foundation under Grant No. 2327671.

References

Arpón, S., Homem-de-Mello, T., & Pagnoncelli, B. (2018). Scenario reduction for stochastic programs with Conditional Value-at-Risk. *Mathematical Programming*, *170*(1), 327-356. https://doi.org/10.1007/s10107-018-1298-9

Bertsimas, D., & Mundru, N. (2022). Optimization-Based Scenario Reduction for Data-Driven Two-Stage Stochastic Optimization. *Operations Research*. https://doi.org/10.1287/opre.2022.2265

Emirhüseyinoğlu, G., Shahhosseini, M., Hu, G., & Ryan, S. M. (2023). Validation of scenario generation for decision-making using machine learning prediction models. *Optimization Letters*. https://doi.org/10.1007/s11590-023-02023-7

Fairbrother, J., Turner, A., & Wallace, S. W. (2022). Problem-driven scenario generation: an analytical approach for stochastic programs with tail risk measure. *Mathematical Programming*, *191*, 141-182. https://doi.org/10.1007/s10107-019-01451-7

Gneiting, T., & Raftery, A. E. (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, *102*(477), 359-378. https://doi.org/10.1198/016214506000001437

Guo, X., & Ryan, S. M. (2021a). Portfolio rebalancing based on time series momentum and downside risk. *IMA Journal of Management Mathematics*. https://doi.org/10.1093/imaman/dpab037

Guo, X., & Ryan, S. M. (2021b). Reliability assessment of scenarios generated for stock index returns incorporating momentum. *International Journal of Finance & Economics*, *26*, 4013-4031. https://doi.org/10.1002/ijfe.2002

Guo, X., & Ryan, S. M. (2023). Avoiding momentum crashes using stochastic mean-CVaR optimization with time-varying risk aversion. *The Engineering Economist*, 1-28. https://doi.org/10.1080/0013791x.2023.2229620

Henrion, R., & Römisch, W. (2022). Problem-based optimal scenario generation and reduction in stochastic programming. *Mathematical Programming*, *191*(1), 183-205. https://doi.org/10.1007/s10107-018-1337-6

Kaut, M. (2021). Scenario generation by selection from historical data. *Computational Management Science*, *18*(3), 411-429. https://doi.org/10.1007/s10287-021-00399-4

Kaut, M., & Wallace, S. W. (2007). Evaluation of scenario-generation methods for stochastic programming. *Pacific Journal of Optimization*, *3*(2), 257-271. http://work.michalkaut.net/papers_etc/SG_evaluation.pdf

Lim, A. E. B., Shanthikumar, J. G., & Vahn, G.-Y. (2011). Conditional value-at-risk in portfolio optimization: Coherent but fragile. *Operations Research Letters*, *39*(3), 163-171. https://doi.org/10.1016/j.orl.2011.03.004

Pinson, P., & Girard, R. (2012). Evaluating the quality of scenarios of short-term wind power generation. *Applied Energy*, *96*, 12-20. https://doi.org/10.1016/j.apenergy.2011.11.004

Prochazka, V., & Wallace, S. W. (2020). Scenario tree construction driven by heuristic solutions of the optimization problem. *Computational Management Science*, *17*(2), 277-307. https://doi.org/10.1007/s10287-020-00369-2

Rahimian, H., Bayraksan, G., & Homem-de-Mello, T. (2018). Identifying effective scenarios in distributionally robust stochastic programs with total variation distance. *Mathematical Programming*, *173*(1-2), 393-430. https://doi.org/10.1007/s10107-017-1224-6

Rockafellar, R. T., & Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk*, *2*(3), 21-41.

Rujeerapaiboon, N., Schindler, K., Kuhn, D., & Wiesemann, W. (2022). Scenario reduction revisited: fundamental limits and guarantees. *Mathematical Programming*, *191*(1), 207-242. https://doi.org/10.1007/s10107-018-1269-1

Ryan, S. M., & Shah Abadi, G. (2022). *Reliability Assessment of Scenarios for CVaR Minimization in Two-Stage Stochastic Programs* IISE Annual Conference & Expo, Seattle, WA.

Sari Ay, D., & Ryan, S. M. (2019). Observational data-based quality assessment of scenario generation for stochastic programs. *Computational Management Science*, *16*, 521-540. https://doi.org/10.1007/s10287-019-00349-1

Sari, D., Lee, Y., Ryan, S., & Woodruff, D. (2016). Statistical metrics for assessing the quality of wind power scenarios for stochastic unit commitment. *Wind Energy*, *19*, 873–893. https://doi.org/10.1002/we.1872

Sari, D., & Ryan, S. M. (2016). *Package MTDrh*. In The Comprehensive R Archive Network. https://cran.r-project.org/package=MTDrh

Sari, D., & Ryan, S. M. (2018). Statistical Reliability of Wind Power Scenarios and Stochastic Unit Commitment Cost. *Energy Systems*, *9*(4), 873-898. https://doi.org/10.1007/s12667-017-0255-7

Wang, J., Botterud, A., Bessa, R., Keko, H., Carvalho, L., Issicaba, D., Sumaili, J., & Miranda, V. (2011). Wind power forecasting uncertainty and unit commitment. *Applied Energy*, *88*(11), 4014-4023. https://doi.org/10.1016/j.apenergy.2011.04.011

Wilks, D. S. (2004). The Minimum Spanning Tree Histogram as a Verification Tool for Multidimensional Ensemble Forecasts. *Monthly Weather Review*, *132*, 1329-1340.

Zhang, W., Wang, K., Jacquillat, A., & Wang, S. (2023). Optimized Scenario Reduction: Solving Large-Scale Stochastic Programs with Quality Guarantees. *INFORMS Journal on Computing*. https://doi.org/10.1287/ijoc.2023.1295