European Journal of Operational Research 269 (2018) 1180–1184

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Short Communication

Comment on “An algorithm for moment-matching scenario generation with application to financial portfolio optimization”

Juan Pablo Contreras, Paul Bosch*, Mauricio Herrera

Facultad de Ingeniería, Universidad del Desarrollo, Ave. Plaza 680, Las Condes, Santiago de Chile, Chile

ARTICLE INFO

Article history:
Received 29 August 2016
Accepted 13 February 2018
Available online 15 March 2018

Keywords:
Scenarios
Finance
Moment matching

ABSTRACT

A paper by Ponomareva, Roman, and Date proposed a new algorithm to generate scenarios and their probability weights matching exactly the given mean, the covariance matrix, the average of the marginal skewness, and the average of the marginal kurtosis of each individual component of a random vector. In this short communication, this algorithm is questioned by demonstrating that it could lead to spurious scenarios with negative probabilities. A necessary and sufficient condition for the appropriate choice of algorithm parameters is derived to correct this issue.

1. Introduction

Stochastic programming (*SP*) is a framework to model decision making under uncertainty. In *SP* some parameters are uncertain and described by statistical distributions. In cases when these distributions are unknown, it is usual to build a discrete approximation with a finite number of outcomes or *scenarios*, by taking advantage of the fact that probability distributions governing the data are known or can be estimated. An overview of scenario generation procedures for *SP* was given by Kaut and Wallace in Kaut and Wallace (2007).

An important class of scenario generation methods is based on matching a set of statistical properties, e.g. moments. Generally, the mean vector and covariance matrix computed from data are the usual statistics to match. However, higher order moments can also be used in order to get better approximation. The skewness and the kurtosis are some of the higher order moments that can be matched by using, for example, complex non-convex optimization problems (Høyland & Wallace, 2001).

In Date, Jalen, and Mamon (2008), a new sigma point filtering algorithm for state estimation in non-linear time series and non-Gaussian systems was developed. This algorithm combines numerical simplicity of the ensemble filter, along with exact moment matching properties. A modified version of this algorithm was proposed in Ponomareva and Date (2013) in which sigma points and the corresponding probability weights were modified at each step to match exactly the predicted values for the average marginal skewness and the average marginal kurtosis, besides matching the mean and covariance matrix. In Ponomareva, Roman, and Date (2015), this algorithm for generating scenarios and the corresponding probability weights was used in the context of financial optimization in order to decrease the computational complexity and th e time for solving *SP* problems. However, in the works of Ponomareva et al. (2015), some assumptions were inconsistent because several important considerations were simply overlooked. As a direct consequence of incorrect estimates for certain bounds, using the proposed algorithm, we obtained some scenarios with negative probabilities. In this short communication, we introduce a necessary and sufficient condition to be imposed on data and parameters in order to solve this issue. In addition, we suggest some changes in specific steps of the algorithm to enable it to deploy correctly.

This paper is organized as follows: Section 2 briefly reviews the algorithm given in Ponomareva et al. (2015). In Section 3, we present the main comment regarding the work by Ponomareva et al. (2015). Section 4 summarizes our proposition for a modified algorithm. In Section 5, we deploy and compare both algorithms on a data set from the Chilean stock exchange to illustrate all comments set out in Section 3. Conclusions are drawn in Section 6.

2. Algorithm for generating scenarios by matching: mean, covariance matrix, average marginal skewness and average marginal kurtosis

The method proposed in Ponomareva et al. (2015) generates $2ns+3$ scenarios and their corresponding probabilities, where n is the dimension of the random vector and s is a positive integer chosen by the user. The inputs for the algorithm are: the mean vector

* Corresponding author.
E-mail addresses: juan.contrerasff@usach.cl (J.P. Contreras), pbosch@udd.cl (P. Bosch), mherrera@ingenieros.udd.cl (M. Herrera).

https://doi.org/10.1016/j.ejor.2018.02.028
0377-2217/

μ, the covariance matrix Σ, and the average of third and fourth moments denoted by $\bar{\xi}$ and $\bar{\kappa}$, respectively.

The method begins by choosing a vector $Z \in \mathbb{R}^n$ such that $\Sigma - ZZ^\top > 0$, and then a square matrix L is chosen such that $\Sigma = LL^\top + ZZ^\top$. Then, the s numbers $p_k \in (0,1)$, $k = 1, \dots, s$ are chosen such that

$$\sum_{k=1}^{s} p_k < \frac{1}{2n} \qquad \wedge \qquad \sum_{k=1}^{s} \frac{1}{p_k} < \gamma, \tag{1}$$

where the upper bound γ is defined as

$$\gamma = 2s^2 \frac{n\bar{\kappa} - \frac{3}{4}\left(\sum_{i=1}^n Z_i^4\right)\left(\frac{n\bar{\xi}}{\sum_{i=1}^n Z_i^3}\right)^2}{\sum_{i,j} L_{ij}^4}. \tag{2}$$

In addition, the following parameters are defined:

- An extra probability $p_{s+1} = 1 - 2n\sum_{k=1}^s p_k$
- Auxiliary parameters $\phi_1 = \frac{n\bar{\xi}\sqrt{p_{s+1}}}{\sum_{i=1}^n Z_i^3}$ and $\phi_2 = p_{s+1}\frac{n\bar{\kappa} - \frac{1}{2s^2}(\sum_{i,j} L_{ij}^4)(\sum_{k=1}^s \frac{1}{p_k})}{\sum_{i=1}^n Z_i^4}$
- The symmetric parameters $\alpha = \frac{1}{2}(\phi_1 + \sqrt{4\phi_2 - 3\phi_1^2})$ and $\beta = \frac{1}{2}(-\phi_1 + \sqrt{4\phi_2 - 3\phi_1^2})$
- The coefficients $\omega_0 = 1 - \frac{1}{\alpha\beta}$, $\omega_1 = \frac{1}{\alpha(\alpha+\beta)}$ and $\omega_2 = \frac{1}{\beta(\alpha+\beta)}$

The first $2\,ns$ scenarios of the distribution and their corresponding probabilities are fixed as:

$$\mathbb{P}\left\{X_{ik}^+ = \mu + \frac{1}{\sqrt{2sp_k}}L_i\right\} = \mathbb{P}\left\{X_{ik}^- = \mu - \frac{1}{\sqrt{2sp_k}}L_i\right\} = p_k, \quad i = 1, \dots, n;\, k = 1, \dots, s,$$

were L_i represents the i-column of matrix L. Then, three extra scenarios are added, $\mathbb{P}\{X_0 = \mu\} = p_{s+1}w_0$, and

$$\mathbb{P}\left\{X_\alpha = \mu + \frac{\alpha}{\sqrt{p_{s+1}}}Z\right\} = p_{s+1}w_1, \qquad \mathbb{P}\left\{X_\beta = \mu - \frac{\beta}{\sqrt{p_{s+1}}}Z\right\} = p_{s+1}w_2. \tag{3}$$

Finally, Proposition 1 in Ponomareva et al. (2015) summarizes the properties of the generated discrete asymmetric distribution denoted by $\mathcal{X}$. The authors showed that $\mathbb{E}[\mathcal{X}] = \mu$, $\mathbb{E}[(\mathcal{X} - \mu)(\mathcal{X} - \mu)^\top] = \Sigma$, and:

i. $2n\sum_{k=1}^s p_k + p_{s+1}\sum_{i=0}^3 w_i = 1$,
ii. $\sum_{i=1}^n \sum_{k=1}^{2ns+3} \mathbb{E}(\mathcal{X}_{ij} - \mu_i)^3 = n\bar{\xi}$,
iii. $\sum_{i=1}^n \sum_{k=1}^{2ns+3} \mathbb{E}(\mathcal{X}_{ij} - \mu_i)^4 = n\bar{\kappa}$.

Thus, the average of the marginal skewness and the average of the marginal kurtosis are matched.

3. Comment on the previous algorithm

Our main concerns regarding the work by Ponomareva et al. (2015) are summarized in three points, which will be discussed separately.

3.1. Constraints over data

The condition $\sum_{k=1}^s \frac{1}{p_k} < \gamma$, which restricts the definition of the probabilities, is a necessary condition to ensure $4\phi_2 - 3\phi_1^2 > 0$, and to guarantee that previously defined parameters α and β be real numbers.

There is a probability vector with properties (1) *if and only if* the interval $(\frac{s}{\gamma}, \frac{1}{2ns})$ is not empty. In this case, the probabilities can be obtained using a random generation of $p_k \sim U(\frac{s}{\gamma}, \frac{1}{2ns})$.

The existence of this interval requires some restrictions over the data set in order that inequality $\frac{s}{\gamma} < \frac{1}{2ns}$ be fulfilled. After some algebraic manipulation, we get the following relationship that the data set should satisfy to be used for generating scenarios:

$$n < \frac{n\bar{\kappa}}{\sum_{i,j} L_{ij}^4} - \frac{3\sum_{i=1}^n Z_i^4}{4\sum_{i,j} L_{ij}^4}\left(\frac{n\bar{\xi}}{\sum_{i=1}^n Z_i^3}\right)^2. \tag{4}$$

As in Ponomareva et al. (2015), the choice of Z and L depends only on data. Similar constraints over data appear in Date et al. (2008), but are less restrictive than condition (4).

It is easy to see that constraints over data are not sufficient to guarantee that w_1, w_2, w_0 be positives. If one of these parameters becomes negative, then it could result in ill-defined or negative probabilities for some scenarios. To avoid this situation more restrictive constraints over data should be imposed.

First, to guarantee $w_1 > 0$ and $w_2 > 0$, it is sufficient to ensure $\alpha > 0$ and $\beta > 0$, that we get by putting $\sqrt{4\phi_2 - 3\phi_1^2} > |\phi_1|$, or equivalently $\phi_2 > \phi_1^2$. Second, in order to get $w_0 > 0$, the condition $\phi_2 > \phi_1^2 + 1$ is required. It is easy to see that this last condition is also sufficient to get all the previous conditions; thus, we arrive at a *necessary and sufficient condition* to correctly apply the method for generation of scenarios with well-defined probabilities:

$$\phi_2 - \phi_1^2 > 1 \Leftrightarrow p_{s+1}\left[A - B\sum_{k=1}^s \frac{1}{p_k}\right] > 1, \tag{5}$$

where parameters A and B are defined by:

$$A = \frac{n\bar{\kappa}}{\sum_{i=1}^n Z_i^4} - \left(\frac{n\bar{\xi}}{\sum_{i=1}^n Z_i^3}\right)^2, \qquad B = \frac{\sum_{i,j} L_{ij}^4}{2s^2\sum_{i=1}^n Z_i^4}. \tag{6}$$

3.2. The choice of parameters

Condition (5) is sensitive to the choice of parameters Z, L and probabilities p_k. For choosing parameters we start with Z, then L and finally the probabilities p_k. We present some alternatives to choose these parameters.

3.2.1. The choice of Z

Vector $Z \in \mathbb{R}^n$ should be taken such that $\Sigma - ZZ^\top > 0$. Authors in Ponomareva et al. (2015) proposed to take $Z = \rho\sqrt{\text{diag}(\Sigma)}$, where $\text{diag}(\Sigma)$ is the diagonal of the covariance matrix and $\rho \in (0, 1)$. Under the assumption that Σ is a positive definite matrix, this choice works well for some sufficiently small values of ρ.

Another alternative is to choose Z using eigenvalues and eigenvectors of the covariance matrix. Let $0 < \lambda_1 \le \cdots \le \lambda_n$ be the eigenvalues of Σ and $\mathrm{v}^1, \dots, \mathrm{v}^n$ be the respective orthonormal eigenvectors, and set $Z = \rho\sqrt{\lambda_l}\,\mathrm{v}^l$ with $\rho \in (0, 1)$ and v^l any eigenvector, then:

$$\left(\Sigma - ZZ^\top\right)\mathrm{v}^j = \begin{cases} \lambda_j \mathrm{v}^j & \text{if } j \ne l \\ \lambda_l(1-\rho^2)\mathrm{v}^l & \text{if } j = l \end{cases}. \tag{7}$$

The eigenvectors and eigenvalues of $\Sigma - ZZ^\top$ are the same as the Σ, except for lth eigenvalue, which is $\lambda_l(1-\rho^2)$. This eigenvalue is also positive and therefore $\Sigma - ZZ^\top$ is a positive definite matrix.

3.2.2. The choice of L

The matrix L must be chosen such that $LL^\top = \Sigma - ZZ^\top$. The authors in Ponomareva et al. (2015) propose to use the square root of matrix $\Sigma - ZZ^\top$. Because $\Sigma - ZZ^\top$ is a positive definite matrix, there is a unique symmetric positive definite matrix L that satisfies $LL^\top = \Sigma - ZZ^\top$.

The above election is unique, however there are other non-symmetric matrices that can be used. For instance, L can be calculated as Cholesky decomposition of $\Sigma - ZZ^\top$, obtaining in this case, a triangular matrix. Different choices of L produce different values for the term $\sum_{i,j} L_{ij}^4$, affecting the values of parameters A and B in condition (5).

3.2.3. The choice of p_k

Once Z and L are fixed, we look for values of probabilities p_k and p_{s+1} such that condition (5) be satisfied. Consider the following optimization problem that maximize the value of $\phi_2 - \phi_1^2$:

$$\begin{aligned} \max_{p_1,\dots,p_s,p_{s+1}} \quad & p_{s+1}\left[A - B\sum_{k=1}^{s}\frac{1}{p_k}\right] \\ \text{s.t.} \quad & \\ & p_{s+1} = 1 - 2n\sum_{k=1}^{s} p_k \\ & p_{s+1}, p_k \geq 0, \; k = 1,\dots,s \end{aligned} \tag{8}$$

We have consider $A > 0$, otherwise condition (5) is impossible to meet. The point $P^* = (p^* \; p^* \; \cdots \; p^*)^\top$, where $p^* = \sqrt{\frac{B}{2nA}}$ and $p^*_{s+1} = 1 - 2nsp^*$, is a stationary point for problem (8) with an optimal value $(\sqrt{A} - \sqrt{2nBs})^2$. Moreover, when expression $\sqrt{A} - \sqrt{2nBs}$ is positive and using the second order sufficient optimality conditions, it is possible to prove that (P^*, p^*_{s+1}) is the unique maximum for the optimization problem.

Finally, the necessary and sufficient condition (5) can be rewritten as follows:

$$\frac{n\overline{\kappa}}{\sum_{i=1}^n Z_i^4} - \left(\frac{n\overline{\xi}}{\sum_{i=1}^n Z_i^3}\right)^2 > 1 + 2\sqrt{n\frac{\sum_{i,j} L_{ij}^4}{\sum_{i=1}^n Z_i^4}} + n\frac{\sum_{i,j} L_{ij}^4}{\sum_{i=1}^n Z_i^4} \tag{9}$$

Note that above condition is independence of probabilities.

3.3. Changing the optimal probability vector P^*

When $p_k = p^*$ for all $k = 1, 2, \dots, s$ the scenarios

$$X_{ik}^{\pm} = \mu \pm \frac{1}{\sqrt{2sp^*}} L_i,$$

are the same for each $i = 1, \dots, n$.

Most of the approaches for solving stochastic programming problems numerically are based on replacing the probability distribution by a discrete distribution with finite support, where each scenario appears with its probability. A good approximation of probability distribution requires a large number of scenarios and repeated scenarios mean a reduction of elements in the support of the probability distribution.

For the purposes of the method, probabilities P^* are not better than any other vector of probabilities that satisfies the necessary and sufficient condition (5), it is just a one precise way to choose them. To avoid repeated scenarios in the generation procedures, we propose some minor changes in the optimal probability vector P^* without losing the fulfillment of condition (5).

Suppose Z, L were already chosen, and the condition (9) is met, then the optimal value of the problem (8) is greater than 1. In this case there exists a neighborhood of P^* where the condition (5) is still fulfilled, and the probabilities p_k are different from each other.

Let s an even number and $\varepsilon_1, \dots, \varepsilon_{s/2}$ be positive constants. Consider a variation in the probability vector as:

$$\hat{P} = \left(p^* + \varepsilon_1 \;\; p^* + \varepsilon_2 \; \cdots \; p^* + \varepsilon_{s/2} \;\; p^* - \varepsilon_{s/2} \; \cdots \; p^* - \varepsilon_2 \;\; p^* - \varepsilon_1\right)^\top, \tag{10}$$

with this variation, a complementary probability $\hat{p}_{s+1}$ is maintained, and the value of $\phi_2 - \phi_1^2$ decreases according to the following equation:

Table 1
Stocks considered in numeric analysis.

AESGENER	BSANTANDER	CONCHATORO	FALABELLA	SALFACORP
AGUAS-A	CAP	COPEC	FORUS	SECURITY
ANDINA-B	CCU	CORPBANCA	IAM	SK
ANTARCHILE	CENCOSUD	ECL	LAN	SM-CHILE-B
BANMEDICA	CHILE	ENDESA	PARAUCO	SONDA
BCI	CMPC	ENERSIS	QUINECO	SQM-B
BELASCO	COLBUN	ENTEL	RIPLEY	VAPORES

Table 2
Values of probabilities.

$p_1 = 0.001402$	$p_6 = 0.001262$
$p_2 = 0.001281$	$p_7 = 0.001419$
$p_3 = 0.001368$	$p_8 = 0.001293$
$p_4 = 0.001310$	$p_9 = 0.001207$
$p_5 = 0.001355$	$p_{10} = 0.001228$

$$\phi_2 - \phi_1^2 = p^*_{s+1}\left[A - B\frac{s}{p^*}\right] - 2\frac{p^*_{s+1}}{p^*}B\sum_{k=1}^{s/2}\frac{\varepsilon_k^2}{p^{*2} - \varepsilon_k^2} \tag{11}$$

In order to make $\hat{P}$ feasible in condition $\phi_2 - \phi_1^2 > 1$ is needed the following:

$$2\frac{p^*_{s+1}}{p^*}B\sum_{k=1}^{s/2}\frac{\varepsilon_k^2}{p^{*2} - \varepsilon_k^2} < p^*_{s+1}\left[A - B\frac{s}{p^*}\right] - 1. \tag{12}$$

One way to select the values ε_k, for example, is picking $s/2$ numbers $u_k \sim U(0, 1)$ and taking:

$$\varepsilon_k = u_k p^* \sqrt{\frac{\nu^* - 1}{p_{s+1}A - 1}} \quad k = 1, \dots, s/2. \tag{13}$$

where ν^* is the optimal value of problem (8).

Eq. (12) is satisfied, and we obtain a new probability vector satisfying condition (5). Proposition 1 of Ponomareva et al. (2015) shows that the marginal central moments are matched independent of the value of the probabilities, then this perturbed probabilities conserve the desired properties in the generated scenarios.

4. Reformulating the algorithm for generating scenarios

Considering the above observations we propose the following reformulated algorithm:

INPUT: μ mean vector, Σ covariance matrix, $\overline{\xi}$ third moment average and $\overline{\kappa}$ fourth moment average, s an even number of scenarios.

OUTPUT: Discrete probability distribution that matches mean vector, covariance matrix, and third and fourth average moments.

Step 1: Find $Z \in \mathbb{R}^n$ and $L \in M_n(\mathbb{R})$ such as $\Sigma - ZZ^\top > 0$ and $\Sigma = LL^\top + ZZ^\top$.

Step 2: If

$$\frac{n\overline{\kappa}}{\sum_{i=1}^n Z_i^4} - \left(\frac{n\overline{\xi}}{\sum_{i=1}^n Z_i^3}\right)^2 > 1 + 2\sqrt{n\frac{\sum_{i,j} L_{ij}^4}{\sum_{i=1}^n Z_i^4}} + n\frac{\sum_{i,j} L_{ij}^4}{\sum_{i=1}^n Z_i^4}$$

go to *Step 3*, else go to *Step 1* and change Z or L.

Step 3: Set

$$A = \frac{n\overline{k}}{\sum_{i=1}^n Z_i^4} - \left(\frac{n\overline{\xi}}{\sum_{i=1}^n Z_i^3}\right)^2, \qquad B = \frac{\sum_{i,j} L_{ij}^4}{2s^2\sum_{i=1}^n Z_i^4}$$

and define $p^* = \sqrt{\frac{B}{2nA}}$, $p^*_{s+1} = 1 - 2nsp^*$ and $\nu^* = (\sqrt{A} - \sqrt{2nBs})^2$.

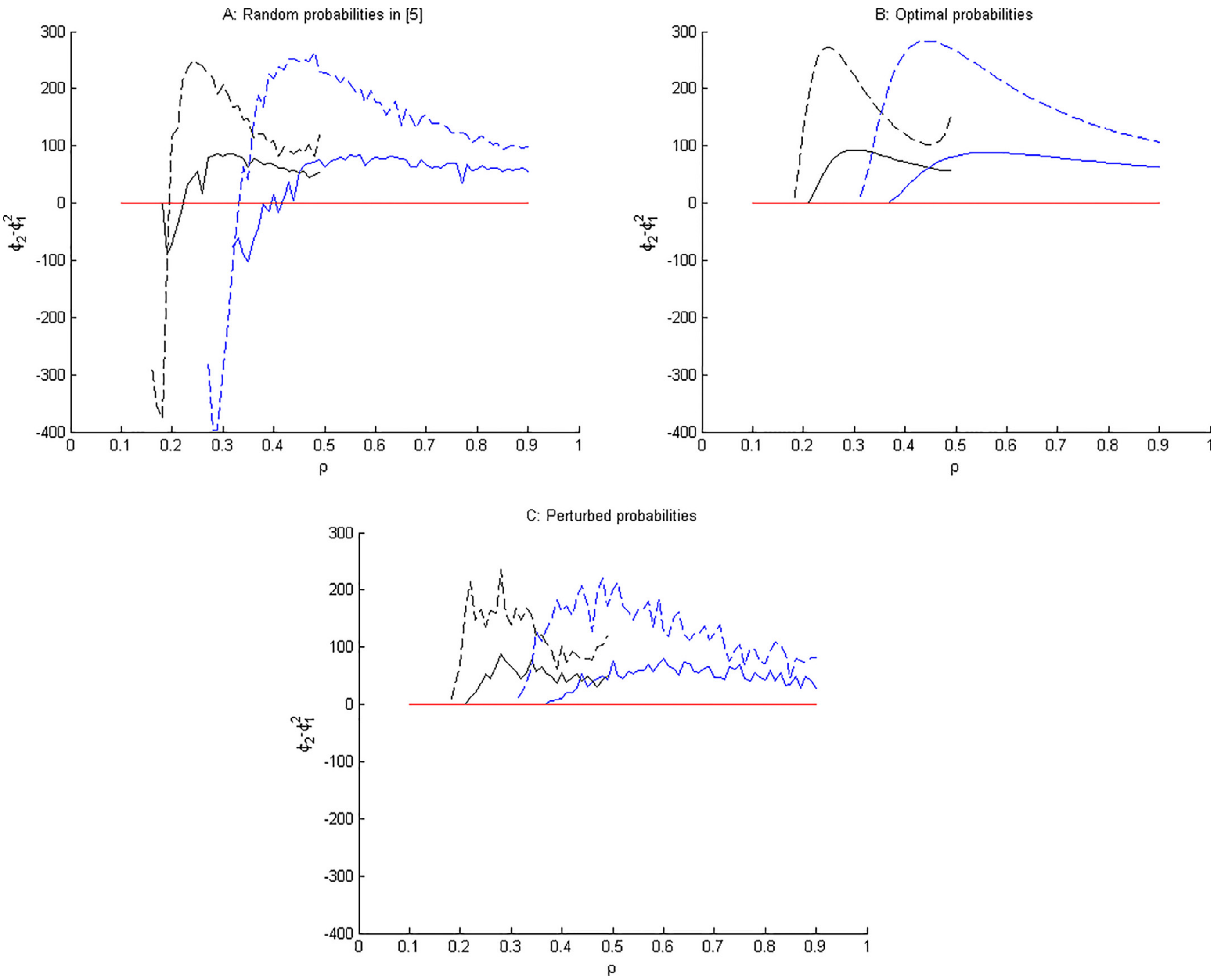

Fig. 1. Values for the expression $\phi_2 - \phi_1^2$ for different Z, L and probabilities selection.

Step 4: Choose $s/2$ numbers $u_k \sim U(0, 1)$ and define

$$\varepsilon_k = u_k p^* \sqrt{\frac{\nu^* - 1}{p_{s+1}A - 1}};$$

set probabilities $p_k = p^* + \varepsilon_k$ for $k = 1, \ldots, s/2$ and $p_k = p^* - \varepsilon_k$ for $k = s/2 + 1, \ldots, s$. Finally set $p_{s+1} = p^*_{s+1}$, and define:

$$\phi_1 = \frac{n\overline{\xi}\sqrt{p_{s+1}}}{\sum_{i=1}^n Z_i^3}, \qquad \phi_2 = p_{s+1}\frac{n\overline{\kappa} - \frac{1}{2s^2}\left(\sum_{i,j} L_{ij}^4\right)\left(\sum_{k=1}^s \frac{1}{p_k}\right)}{\sum_{i=1}^n Z_i^4},$$

$$\alpha = \frac{1}{2}\left(\phi_1 + \sqrt{4\phi_2 - 3\phi_1^2}\right), \qquad \beta = \frac{1}{2}\left(-\phi_1 + \sqrt{4\phi_2 - 3\phi_1^2}\right),$$

$$w_0 = 1 - \frac{1}{\alpha\beta}, \qquad w_1 = \frac{1}{\alpha(\alpha+\beta)}, \qquad w_2 = \frac{1}{\beta(\alpha+\beta)}.$$

Step 5: Return the scenarios

$$X_{ik}^{\pm} = \mu \pm \frac{1}{\sqrt{2sp_k}}L_i, \qquad X_0 = \mu, \qquad X_\alpha = \mu + \frac{\alpha}{\sqrt{p_{s+1}}}Z,$$

$$X_\beta = \mu - \frac{\beta}{\sqrt{p_{s+1}}}Z.$$

and their respective probabilities $\mathbb{P}(X_{ik}^+) = \mathbb{P}(X_{ik}^-) = p_k$, $\mathbb{P}(X_0) = p_{s+1}w_0$, $\mathbb{P}(X_\alpha) = p_{s+1}w_1$ and $\mathbb{P}(X_\beta) = p_{s+1}w_2$.

5. A numeric example

IPSA is the main stock index of the Chilean stock market, consisting of the forty most traded stocks in the local market. In our numerical example we consider a data set with thirty-five stocks from *IPSA* (see Table 1), that consists of daily returns or closing prices for each day. To construct the input data for the algorithm on a specific day, a time window of three years (about 753 trading days) was used. The data set can be downloaded from http://www.bolsadesantiago.com/mercado/Paginas/detalleindicesbursatiles.aspx?indice=IPSA.

To give an example for which the proposed algorithm in Ponomareva et al. (2015) fails, we considered what happened on the date 08-22-2013. Scenarios were generated for estimating the behavior of returns on this day. Following the procedure outlined in Ponomareva et al. (2015), the vector Z was taken as $Z_i = \rho\sqrt{\Sigma_{ii}}$ with $\rho = 0.7$, and the matrix L was taken as the only symmetric positive definite matrix, such that $LL^\top = \Sigma - ZZ^\top$. We consider $s = 10$, obtaining a value of $9.1709 \cdot 10^{-3}$ for the parameter γ. Following recommended steps given in Ponomareva et al. (2015), the probabilities p_k were randomly generated in the interval $(\frac{s}{\gamma}, \frac{1}{2ns}) = (0.001136, 0.001428)$. The values of these probabilities and the algorithm parameters are shown in Table 2 and Table 3, respectively.

Table 3
Values used for the numerical example.

$p_{s+1} = 0.0809$	$\phi_1 = -20.5801$	$\phi_2 = 422.89$
$4\phi_2 - 3\phi_1^2 = 420.93$		
$\alpha = -0.0317$ and $\beta = 205.484$		
$w_1 = -1.5374$	$w_2 = 0.0023$	$w_0 = 25.350$
$\phi_2 - \phi_1^2 = -0.6505$		

In this case, we obtained a negative value for the parameter α (see Table 3), and consequently some negative probabilities appeared in the last three scenarios. This error happened because condition (5) is not met ($\phi_2 - \phi_1^2 = -0.6505 \not> 1$) for the recommended random choice of probabilities in Ponomareva et al. (2015). Nevertheless, it is possible to operate the method by changing the choice of parameters Z, L and p_k so that condition (5) is met, as we explained in the former section.

Fig. 1 shows the values of the expression $\phi_2 - \phi_1^2$ for different choices of parameters. The four curves in each graph was constructed by combining the two forms proposed to choose Z (eigenvectors associated to the maximal eigenvalue and diagonal of covariance matrix Σ), with two ways to choose L (symmetric matrix and Cholesky decomposition). In each case, the values of the parameter ρ for Z choice were taken between 0.1 and 0.9. In the graph A, the probabilities p_k were randomly selected according to Ponomareva et al. (2015), i.e. $p_k \sim U(\frac{s}{\gamma}, \frac{1}{2ns})$. Note the condition $\phi_2 - \phi_1^2 > 1$ is violated repeatedly, in these cases we cannot generate the scenarios properly.

In graph B, the probabilities were chosen from the optimal solution of problem (8), such that the value $\phi_2 - \phi_1^2$ is as large as possible. Now, the values of the expression stay always over the horizontal $\phi_2 - \phi_1^2 = 1$ for the same data set.

Finally, graph C is constructed by perturbing the optimal probabilities obtained from problem (8). The condition is still met and the algorithm can be used. This discussion shows that careless choice of probabilities could cause errors in generating scenarios method.

6. Conclusions

The algorithm described in Ponomareva et al. (2015) aims to generate scenarios by matching the first two moments and the averages of the third and the fourth moments of a distribution. In this short communication, we have shown that the algorithm as proposed in Ponomareva et al. (2015) cannot always be applied. Application should depend on the fulfillment of a condition imposed on data, along with the correct choice of algorithm parameters. We have shown that even this condition over data, by itself, is insufficient, and a more restrictive condition should be imposed. This newfound condition also depends on the choice of the probabilities for the scenarios.

In this short communication, new alternatives for choosing algorithm parameters were introduced, along with an appropriate method for determining the probabilities in order to meet the more restrictive necessary and sufficient condition, but at the same time to obtain enough scenarios that are different from each other.

We have introduced a reformulated algorithm, which is able to indicate when the scenario generation can be performed and when not, depending on the data and parameters used. The numerical example from the Chilean stock market, clearly illustrate how a careless choice of the algorithm parameters can lead to negative probabilities for some generated scenarios.

The autors of Ponomareva et al. (2015) suggest that freedom in the determination of the probabilities could be used in future works to match higher moments. However, in this short communication was introduced a stronger condition sensitive to the choice of the probabilities. This fact could make the matching of higher moments by using the proposed algorithm more difficult. Despite this difficult, the method proposed in Ponomareva et al. (2015) represents an efficient, fast and easy implementation way to generate scenarios.

We suggest as future research work to find more and better alternatives for choosing parameters Z and L, in order to met condition (9). For example, in the election of Z we propose to use eigenvalues and eigenvectors of covariance matrix, the immediate question that arises, which of these eigenvectors produced a good election of Z?.

Acknowledgement

We are grateful to an anonymous referee for their detailed review and constructive inputs.

References

Date, P., Jalen, L., & Mamon, R. (2008). A new algorithm for latent state estimation in non-linear time series models. *Applied Mathematics and Computation, 203*, 224–232.

Høyland, K., & Wallace, S. W. (2001). Generating scenario trees for multistage decision problems. *Management Science, 47*(2), 295–307.

Kaut, M., & Wallace, S. W. (2007). Evaluation of scenario-generation methods for stochastic programming. *Pacific Journal of Optimization, 3*(2), 257–271.

Ponomareva, K., & Date, P. (2013). Higher order sigma point filter: A new heuristic for nonlinear time series filtering. *Applied Mathematics and Computation, 221*, 662–671.

Ponomareva, K., Roman, D., & Date, P. (2015). An algorithm for moment-matching scenario generation with application to financial portfolio optimisation. *European Journal of Operational Research, 240*(3), 678–687.