

RISK MEASURE PRESERVING PIECEWISE LINEAR APPROXIMATION OF EMPIRICAL DISTRIBUTIONS

PHILIPP ARBENZ AND WILLIAM GUEVARA-ALARCÓN

April 4, 2016

ABSTRACT

Stochastic models used for pricing, reserving, or capital modelling in insurance companies are often very complex, which is why resulting distributions are typically approximated by Monte Carlo simulations. Both the market and regulators exert increasing pressure not to discard the resulting sample distributions, but rather to store them for future review, audit, or validation, as well as to transfer them between actuarial systems. The present work introduces a compression algorithm which approximates an empirical univariate distribution function through a piecewise linear distribution. In contrast to keeping the full sample, such an approximation facilitates the storage and data transfer of the results by drastically reducing memory requirements. The approximation algorithm preserves the mean and imposes a uniformly bounded relative error over a space of coherent risk measures (TVaR). An efficient, open source implementation is provided.

KEYWORDS

Monte Carlo simulation, empirical distribution, piecewise linear distribution, compression, coherent risk measures, TVaR.

1. INTRODUCTION

In the insurance industry, stochastic actuarial models can be found in areas such as pricing, reserving, natural catastrophe modelling, or capital modelling. Many of these models use distributions which possess non-trivial features such as heavy tails, jumps, or modifications arising from terms and conditions of (re)insurance contracts. Thereby, many of the random variables of interest cannot be represented in a simple manner through a parametric distribution. In these cases, empirical distributions, obtained by means of a simulated sample of realisations from a stochastic model, are often used as an approximation of the random variable of interest. For the purpose of reasonably capturing extreme values of the

risks, the size of the samples involved in this process is generally large - millions or even more.

There is an increased tendency among insurance companies to store, transfer and reuse simulated distributions mentioned above, due to several reasons such as the following. Firstly, proliferating **review requirements**: inputs and outputs from actuarial models are more frequently reviewed, from both technical and non-technical stakeholders, such as independent validation or audit. Actuarial peer reviews can form important steps in the sign-off governance. Furthermore, in some jurisdictions, reinsurance contracts need to be documented in a way that auditors can check *risk transfer test* requirements (see IASB (2004)). Secondly, tightening **regulatory standards**: in Europe, the adoption of Solvency 2 requires various actuarial models which need to be evaluated and documented; not only for calculating the SCR (Solvency Capital Requirement), but also to conduct ORSA (Own Risk and Solvency Assessment), and to write actuarial function holder reports. In addition, risk management and internal control systems (Pillar 2 of Solvency 2) mandate a consistent, controlled, and transparent use of these models and their results. As described in De Brauw Blackstone Westbroek (2014), a great amount of information needs to be managed and recorded to ensure that minimum record retention periods are adhered to, i.e., that data can be traceable and accessible for several years. Thirdly, increasingly **interconnected models**: information systems and applications inside insurance companies have become complex and are increasingly connected and dependent of each other. Interconnected systems imply that part of the data is accessible to a large number of stakeholders, so that it must be consistent and understandable to all of them. The *use test*, which is mandatory for internal models under Solvency 2 and SST (see EP-CEU (2009) and FINMA (2008)), fosters the precise and coherent evaluation of risks and consistent use of methods and parameters within insurance companies.

In this paper, we develop an algorithm to approximate a large univariate sample distribution which needs to be stored on a long term basis, or transferred between actuarial systems. The approximation is performed through a piecewise linear (PWL) distribution, allowing a massive reduction in memory and bandwidth requirements compared to storing the full sample. The approximation algorithm preserves the mean and imposes a uniformly bounded relative error over a space of coherent risk measures. At the same time, our approach preserves the shape of the distribution which is not possible when using another conventional solution: storing only key statistics, such as mean and standard deviation. From a

statistical point of view, the proposed method can also be interpreted as a non-parametric estimator for univariate sample distributions.

There is a considerable amount of literature on approximations through piecewise linear functions in areas such as engineering, computational fluid dynamics or computer graphics, see e.g. Nguyen and Oommen (1997), Hakimi and Schmeichel (1991), or Qiu (2003). However, the authors are not aware of any publication on using PWL distributions as a means to approximate sample distributions in an insurance or risk measure context. The present work was partially inspired by a so called *pruning algorithm* invented and extensively used in the reinsurance companies Converium and SCOR (see Hummel (2005)). The pruning and our algorithm share the same purpose: approximating a sample distribution through a PWL distribution. However, the probabilistic framework is entirely different, the pruning algorithm uses the Kolmogorov-Smirnov statistic to infer the approximation error, whereas in our approach we will use the space of spectral risk measures.

The paper is structured as follows. Section 2 introduces piecewise linear distributions while Section 3 focuses on the TVaR risk measure. Section 4 defines the approximation error. Section 5 specifies the approximation algorithm. In Section 6, we give an error bound for spectral risk measures. Section 7 investigates the numerical complexity. In Section 8 and Section 9, we give examples of results and implementation, respectively. We compare the algorithm to alternatives in Section 10 and conclude in Section 11.

2. PIECEWISE LINEAR DISTRIBUTIONS

In this section, we introduce the notion of a piecewise linear (PWL) distribution. As usual, for a cumulative distribution function (cdf) $F_X(x) = \mathbb{P}[X \leq x]$ of a random variable $X : \Omega \mapsto \mathbb{R}$ defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ we define the quantile function $F_X^- : (0, 1] \mapsto \mathbb{R}$ through the generalised inverse $F_X^-(t) = \inf\{x \in \mathbb{R} : F_X(x) \geq t\}$.

Definition 2.1. *Let $\mathbf{x} = (x_1, x_2, \dots, x_K)$ and $\mathbf{y} = (y_1, y_2, \dots, y_K)$ be a pair of vectors of equal length $K \in \mathbb{N}$, such that*

$$-\infty < x_1 \leq x_2 \leq \dots \leq x_K < \infty,$$

$$0 = y_1 \leq y_2 \leq \dots \leq y_K = 1.$$

A random variable $X : \Omega \rightarrow \mathbb{R}$ with cdf $G(t) = \mathbb{P}[X \leq t]$ ($t \in \mathbb{R}$) has a piecewise linear distribution $G = \text{PWL}(\mathbf{x}, \mathbf{y})$ if

$$G(t) = \begin{cases} 0, & \text{if } t < x_1, \\ y_k + (y_{k+1} - y_k) \frac{t - x_k}{x_{k+1} - x_k}, & \text{if } x_k < t < x_{k+1}, \\ \max_{\{k=1, \dots, K \mid x_k = t\}} \{y_k\}, & \text{if } t = x_k \text{ for some } k, \\ 1, & \text{if } t > x_K. \end{cases} \quad (2.1)$$

Hence, the cdf of a PWL distribution is given by the linear interpolation between the *interpolation points* $\{(x_1, y_1), (x_2, y_2), \dots, (x_K, y_K)\}$. The only special case occurs if there are several interpolation points with the same x -coordinate. In this case, the distribution has an atom and the cdf is defined such that it is right-continuous, thus its value will be the maximum ordinate y of all interpolation points with the same x -coordinate of interest.

Example 2.2. Let $K = 4$, $\mathbf{x} = (1, 4, 4, 9)$, $\mathbf{y} = (0, 0.6, 0.8, 1)$, and $G = \text{PWL}(\mathbf{x}, \mathbf{y})$. Then,

$$G(t) = \begin{cases} 0, & \text{if } t < 1, \\ 0.2(t - 1), & \text{if } 1 \leq t < 4, \\ 0.8 + 0.04(t - 4), & \text{if } 4 \leq t < 9, \\ 1, & \text{if } 9 \leq t. \end{cases}$$

Note that $G(1) = 0$, $G(9) = 1$. There is an atom at 4 such that $\mathbb{P}[X = 4] = 0.8 - 0.6 = 0.2$, $\lim_{t \uparrow 4} G(t) = 0.6$ and $G(4) = \max_{\{k=1, \dots, 4 \mid x_k = 4\}} \{y_k\} = \max\{0.6, 0.8\} = 0.8$ for $X \sim G$. The distribution $G(t)$ is illustrated in Figure 1.

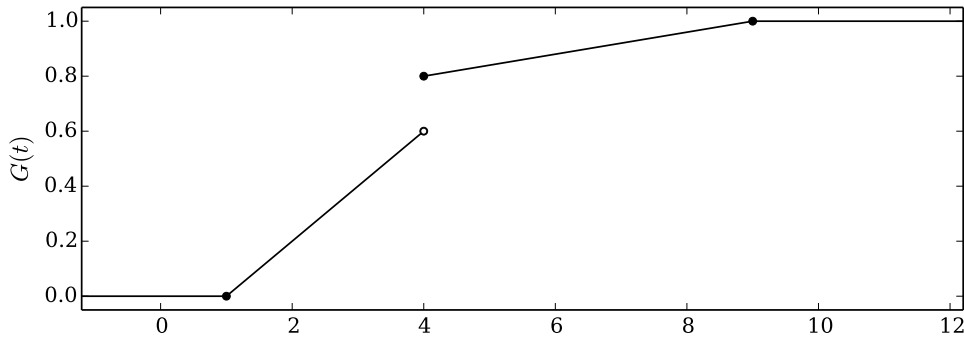


FIGURE 1. An illustration of the cdf $G(t)$ for $G = \text{PWL}((1, 4, 4, 9), (0, 0.6, 0.8, 1))$.

Remark 2.3. We can distinguish the following cases with respect to continuity of a PWL distribution $G = \text{PWL}(\mathbf{x}, \mathbf{y})$:

- **Continuous case:** In case the components of the vector (x_1, x_2, \dots, x_K) are strictly increasing, i.e., if $x_1 < x_2 < \dots < x_K$, then the PWL distribution is absolutely continuous and the density is piecewise constant with

$$\frac{d}{dt}G(t) = \frac{y_{k+1} - y_k}{x_{k+1} - x_k} \quad \text{for } x_k \leq t < x_{k+1}.$$

- **Discrete case:** In case every line segment of the PWL distribution is either horizontal or vertical (i.e., if $(y_{k+1} - y_k)(x_{k+1} - x_k) = 0$ for every $k = 1, \dots, K-1$), then the distribution is discrete.
- **General case:** In case none of the two special cases above apply, then the distribution has both discrete and absolutely continuous parts.

The following lemma shows that not only the cdf can easily be represented as a piecewise linear function, but also the quantile function (inverse cdf) is piecewise linear.

Lemma 2.4. For a PWL distribution $G = \text{PWL}(\mathbf{x}, \mathbf{y})$, the inverse $G^{-}(t)$ for $0 < t \leq 1$ is given by

$$G^{-}(t) = \begin{cases} x_k + \frac{x_{k+1} - x_k}{y_{k+1} - y_k}(t - y_k), & \text{if } y_k < t < y_{k+1}, \\ \min_{\{k=1, \dots, K \mid y_k = t\}} \{x_k\}, & \text{else.} \end{cases}$$

Proof. For $y_k < t < y_{k+1}$, the inversion is immediate. If $t = y_k$ for some k , the result follows from the definition $G^{-}(t) = \inf\{x \in \mathbb{R} : G(x) \geq t\}$. \square

Example 2.5. Let $G = \text{PWL}(\mathbf{x}, \mathbf{y})$ as in Example 2.2. Then the inverse cdf is given by

$$G^{-}(t) = \begin{cases} 1 + 5t, & \text{if } 0 < t < 0.6, \\ 4, & \text{if } 0.6 \leq t < 0.8, \\ 4 + 25(t - 0.8), & \text{if } 0.8 \leq t \leq 1. \end{cases}$$

One of the desirable properties of the PWL distribution is that many functionals of interest can be calculated analytically, such as the moments given in the following lemma.

Lemma 2.6. Let $G = \text{PWL}(\mathbf{x}, \mathbf{y})$ and $m > 0$. Then, the m -th moment $\mathbb{E}[X^m]$ of a random variable $X \sim G$ is given by

$$\mathbb{E}[X^m] = \sum_{k=1}^{K-1} \int_{y_k}^{y_{k+1}} G^{-}(t)^m dt,$$

where

$$\int_{y_k}^{y_{k+1}} G^{-}(t)^m dt = \begin{cases} \frac{x_{k+1}^{m+1} - x_k^{m+1}}{m+1} \frac{y_{k+1} - y_k}{x_{k+1} - x_k}, & \text{if } x_{k+1} > x_k, \\ x_k^m (y_{k+1} - y_k), & \text{if } x_{k+1} = x_k. \end{cases}$$

Proof. We have $\mathbb{E}[X^m] = \mathbb{E}[(G^-(U))^m]$, where U is uniformly distributed on $[0, 1]$. Hence,

$$\mathbb{E}[X^m] = \mathbb{E}[G^-(U)^m] = \int_0^1 G^-(t)^m dt = \sum_{k=1}^{K-1} \int_{y_k}^{y_{k+1}} G^-(t)^m dt,$$

which leads to the desired result by using Lemma 2.4. \square

Example 2.7. Let $G = \text{PWL}(\mathbf{x}, \mathbf{y})$ as in Example 2.2 and $X \sim G$. Then, the mean is given by

$$\mathbb{E}[X] = \frac{4^2 - 1^2}{2} \frac{0.6 - 0}{4 - 1} + 4(0.8 - 0.6) + \frac{9^2 - 4^2}{2} \frac{1 - 0.8}{9 - 4} = 3.6.$$

Analogously, the second moment is equal to

$$\mathbb{E}[X^2] = \frac{4^3 - 1^3}{3} \frac{0.6 - 0}{4 - 1} + 4^2(0.8 - 0.6) + \frac{9^3 - 4^3}{3} \frac{1 - 0.8}{9 - 4} = 16.2667,$$

which also leads to $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = 3.3067$.

In order to simplify the notation in the following sections, we define the basis of a PWL distribution.

Definition 2.8. For a PWL distribution $G = \text{PWL}(\mathbf{x}, \mathbf{y})$, the vector $\text{basis}(G) = \mathbf{z} = (z_1, z_2, \dots, z_S)$ is defined such that $0 = z_1 < z_2 < \dots < z_S = 1$ and for every z_s there is a k with $z_s = y_k$. Furthermore, for $\text{basis}(G) = (z_1, z_2, \dots, z_S)$, the semi-closed intervals $(z_s, z_{s+1}]$ between two consecutive points z_s and z_{s+1} in the basis are called **segments** of G .

Hence, the basis of $G = \text{PWL}(\mathbf{x}, \mathbf{y})$ can be obtained by removing all duplicate values of \mathbf{y} .

Example 2.9. Let $G = \text{PWL}(\mathbf{x}, \mathbf{y})$ as in Example 2.2. Then, $\text{basis}(G) = \mathbf{z} = (0, 0.6, 0.8, 1)$. We have $\text{basis}(G) = \mathbf{y}$ since there are no duplicate values in \mathbf{y} . The three segments of G are given by $(0, 0.6]$, $(0.6, 0.8]$, and $(0.8, 1]$.

The family of PWL distributions has been chosen for approximation purposes because of its simplicity, flexibility and further desirable characteristics. Section 10 gives additional justification for this selection, details on PWL distributions, and comparisons.

Note that in the area of application of this paper, we are mainly interested in quantities which can be written as functionals of the cdf (such as means, VaR's, TVaR's). The density, if it exists, is of lesser interest. Furthermore, PWL distributions in this paper have a PWL cdf, and not a PWL density, as it is sometimes used in other publications (see also Remark 2.3).

3. THE TAIL VALUE-AT-RISK (TVaR) AND TVaR EQUIVALENCE

In this section, we introduce the risk measure tail value-at-risk (TVaR) and its deviation measure TVaR^Δ . These will later be used to define PWL distributions which are admissible approximations of the sample distribution. The approximation employs TVaR, since it is the most frequently used coherent risk measure.

In an abuse of notation, we write $\mathbb{E}[F]$ and $\text{TVaR}^\Delta(F)$ instead of $\mathbb{E}[X]$ and $\text{TVaR}^\Delta(X)$ for $X \sim F$ whenever the meaning is clear, which is justified for law invariant functionals.

Definition 3.1. *For a distribution F with $\mathbb{E}[F] < \infty$ and a level $0 < \alpha \leq 1$, the risk measure $\text{TVaR}_\alpha(F)$ and its deviation $\text{TVaR}_\alpha^\Delta(F)$ (see Rockafellar et al. (2006)) are given as*

$$\begin{aligned}\text{TVaR}_\alpha(F) &= -\frac{1}{\alpha} \int_0^\alpha F^-(t) dt, \\ \text{TVaR}_\alpha^\Delta(F) &= \text{TVaR}_\alpha(F - \mathbb{E}[F]) = \mathbb{E}[F] - \frac{1}{\alpha} \int_0^\alpha F^-(t) dt.\end{aligned}$$

Different names such as *expected shortfall* (ES) or *conditional tail expectation* (CTE) are used by other authors to refer to the former defined TVaR or slightly modified risk measures. All those quantities coincide in the case that the distribution F is continuous.

Note that the sign convention in Definition 3.1 is chosen such that $\text{TVaR}_\alpha^\Delta(F) \geq 0$ for any α and distribution F . We have $\text{TVaR}_\alpha^\Delta(F) = 0$ for $\alpha < 1$ if and only if F represents a constant distribution. Therefore, TVaR^Δ can be seen as a volatility measure, which measures how strongly the lower tail expectation of the distribution deviates from the mean $\mathbb{E}[F]$.

In other literature, the upper tail notation $\left(\frac{1}{\alpha} \int_{1-\alpha}^1 F^-(t) dt\right)$ is sometimes used. We will later see that this is algorithmically equivalent. The algorithm is easier to derive with the current notation. The following example illustrates TVaR and TVaR^Δ .

Example 3.2. *Let $G = \text{PWL}((1, 4, 4, 9), (0, 0.6, 0.8, 1))$ as in Example 2.2. As shown in Example 2.7, we have $\mathbb{E}[G] = 3.6$. For a level $\alpha = 0.2$, we get $\text{VaR}_{0.2}(G) = G^-(0.2) = 2$, $\text{TVaR}_{0.2}(G) = -1.5$, and $\text{TVaR}_{0.2}^\Delta(G) = 3.6 - 1.5 = 2.1$, which is illustrated in Figure 2.*

In order to avoid confusion and to create a consistent notation throughout the rest of the paper, we will make the following notational convention.

Notation 3.3. *Throughout the remainder of the paper:*

- G denotes a PWL distribution;
- F denotes an empirical distribution $F(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{X_i \leq x\}$ of a sample $\{X_1, X_2, \dots, X_n\}$ with sample size $n \in \mathbb{N}$, where $\mathbb{1}\{\cdot\}$ denotes the indicator function.

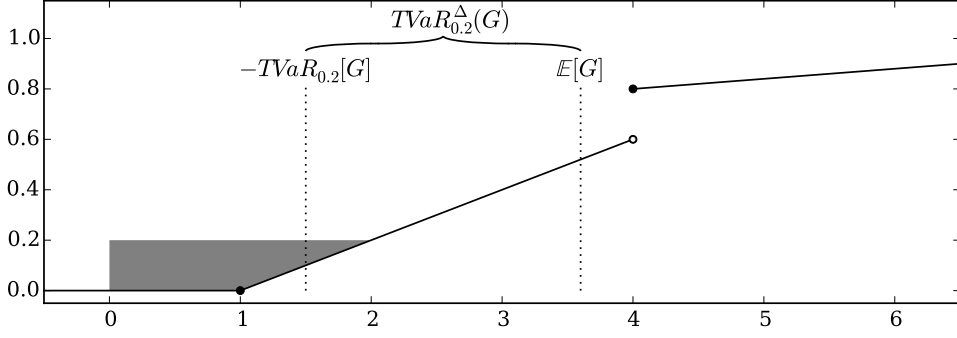


FIGURE 2. An illustration of $E[G]$, $\text{TVaR}_{0.2}(G)$, and $\text{TVaR}_{0.2}^{\Delta}(G)$. The shaded area illustrates $\text{TVaR}_{0.2}(G) = -1.5$, which is given by the negative average of G^- for quantiles between 0 and 0.2.

Note that in this paper, we are not concerned with the sampling algorithm leading to $\{X_1, X_2, \dots, X_n\}$, but only discuss the approximation step from F to G . Of course, the sampling itself may also present very challenging technical and mathematical aspects. We also do not make any assumption on the distribution from which $\{X_1, X_2, \dots, X_n\}$ was sampled.

As F is discrete, it is easy to calculate the quantile function $F^-(\cdot)$:

$$F^-(t) = X_{(i)} \quad \text{for} \quad \frac{i-1}{n} < t \leq \frac{i}{n}, \quad (3.1)$$

where $X_{(i)}$ is the i -th order statistic such that $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n-1)} \leq X_{(n)}$.

The following two lemmas provide analytical formulae for the TVaR for PWL and sample distributions. Of course, the corresponding formulae for TVaR^{Δ} are then trivial to deduce.

Lemma 3.4. *Suppose $G = \text{PWL}(\mathbf{x}, \mathbf{y})$ and $y_m < \alpha \leq y_{m+1}$ for some $m \in \{1, 2, \dots, K-1\}$. Then,*

$$\text{TVaR}_{\alpha}(X) = -\frac{1}{\alpha} \left(\sum_{k=1}^{m-1} \frac{x_k + x_{k+1}}{2} (y_{k+1} - y_k) + (\alpha - y_m) x_m + \frac{(\alpha - y_m)^2}{2} \frac{x_{m+1} - x_m}{y_{m+1} - y_m} \right).$$

Proof. Analogous to the proof of Lemma 2.6. □

For $\alpha = y_m$, Lemma 3.4 simplifies to $\text{TVaR}_{y_m}(X) = -1/y_m \sum_{k=1}^{m-1} (x_k + x_{k+1})(y_{k+1} - y_k)/2$.

Example 3.5. *Let $G = \text{PWL}(\mathbf{x}, \mathbf{y})$ as in Example 2.2. Then, we have*

$$\text{TVaR}_{0.8}(X) = -\frac{1}{0.8} \left(\frac{1+4}{2} (0.6 - 0) + \frac{4+4}{2} (0.8 - 0.6) \right) = -2.875$$

and $\text{TVaR}_{0.8}^{\Delta}(X) = 3.6 - 2.875 = 0.725$.

Lemma 3.6. *For the empirical sample distribution F , we have $E[F] = 1/n \sum_{i=1}^n X_i$. Furthermore, for $\alpha \in \{1/n, 2/n, \dots, n/n\}$ we have*

$$\text{TVaR}_{\alpha}(X) = -\frac{1}{\alpha n} \sum_{i=1}^{\alpha n} X_{(i)}.$$

Proof. Immediate consequence of (3.1). \square

Example 3.7. Suppose we have the following sample of size 10:

$$\{X_1, X_2, \dots, X_{10}\} = \{1, 1.6, 4.3, 4.6, 6, 7.1, 13, 13.4, 16, 18.8\}.$$

This leads to $F(x) = 1/10 \sum_{i=1}^{10} \mathbb{1}\{X_i \leq x\}$. Then,

$$\mathbb{E}[F] = \frac{1}{10} \sum_{i=1}^{10} X_i = 8.58, \quad \text{TVaR}_{0.6}(F) = -\frac{1}{6} \sum_{i=1}^6 X_{(i)} = -4.1, \quad \text{TVaR}_{0.6}^\Delta(F) = 4.48.$$

In the following, we will define the notion of TVaR equivalence. This concept will allow to deduce an efficient divide-and-conquer algorithm.

Definition 3.8. A PWL distribution $G = \text{PWL}(\mathbf{x}, \mathbf{y})$ and a sample distribution F are called TVaR equivalent if $\mathbb{E}[G] = \mathbb{E}[F]$ and

$$\text{TVaR}_{y_k}^\Delta(G) = \text{TVaR}_{y_k}^\Delta(F) \quad \text{for } k = 2, 3, \dots, K.$$

Definition 3.8 implies that, apart from having the same mean, G and F have the same $\text{TVaR}_\alpha^\Delta$ for all levels $\alpha = y_2, y_3, \dots, y_K$. The level $y_1 = 0$ is omitted, since TVaR_0^Δ is not defined. Due to $\mathbb{E}[G] = \mathbb{E}[F]$, TVaR equivalence also implies $\text{TVaR}_{y_k}(G) = \text{TVaR}_{y_k}(F)$ for $k = 2, 3, \dots, K$. Note that TVaR equivalence imposes no constraints for $\text{TVaR}_\alpha^\Delta(G)$ with $\alpha \neq y_1, \dots, y_K$.

Example 3.9. Given the sample distribution F as defined in Example 3.7. Furthermore, let $G = \text{PWL}((0.8, 7.4, 12.9, 17.7), (0, 0.6, 0.6, 1))$. Note that $\text{basis}(G) = (0, 0.6, 1)$. Figure 3 illustrates F and G .

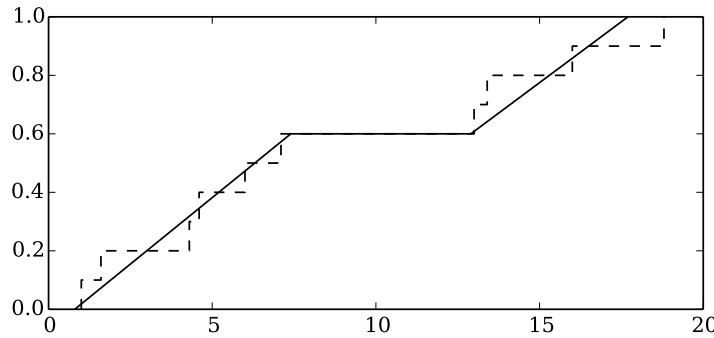


FIGURE 3. An illustration of F (dashed, empirical cdf as in Example 3.7) and G (solid line, PWL distribution) where G and F are TVaR equivalent.

By using Lemma 3.4 and Lemma 3.6, we see that G and F are TVaR equivalent since

$$\mathbb{E}[G] = \mathbb{E}[F] = 8.58, \quad \text{TVaR}_{0.6}^\Delta(G) = \text{TVaR}_{0.6}^\Delta(F) = 4.48, \quad \text{TVaR}_1^\Delta(G) = \text{TVaR}_1^\Delta(F) = 0.$$

Note that for $\alpha \neq 0.6, 1$, the equality $\text{TVaR}_\alpha^\Delta(F) = \text{TVaR}_\alpha^\Delta(G)$ does generally not hold.

The following lemma shows that if $G = \text{PWL}(\mathbf{x}, \mathbf{y})$ and F are TVaR equivalent, then the average over every segment $(y_k, y_{k+1}]$ is equal as well.

Lemma 3.10. *Suppose $G = \text{PWL}(\mathbf{x}, \mathbf{y})$ and F are TVaR equivalent. Then, we have*

$$\int_{y_k}^{y_{k+1}} G^{\leftarrow}(t) dt = \int_{y_k}^{y_{k+1}} F^{\leftarrow}(t) dt$$

for every $k = 1, \dots, K - 1$.

Proof. Due to TVaR equivalence, we have $y_k \text{TVaR}_{y_k}^{\Delta}(G) - y_k \text{TVaR}_{y_k}^{\Delta}(F) = 0$. Hence,

$$\begin{aligned} 0 &= \left(y_{k+1} \text{TVaR}_{y_{k+1}}^{\Delta}(G) - y_k \text{TVaR}_{y_k}^{\Delta}(G) \right) - \left(y_{k+1} \text{TVaR}_{y_{k+1}}^{\Delta}(F) - y_k \text{TVaR}_{y_k}^{\Delta}(F) \right) \\ &= \int_{y_k}^{y_{k+1}} F^{\leftarrow}(t) dt - \int_{y_k}^{y_{k+1}} G^{\leftarrow}(t) dt. \quad \square \end{aligned}$$

The following theorem shows that for TVaR equivalent distributions and $\alpha \in (y_k, y_{k+1}]$, the difference $\text{TVaR}_{\alpha}^{\Delta}(F) - \text{TVaR}_{\alpha}^{\Delta}(G)$ can be written as an expression which depends only on F^{\leftarrow} and G^{\leftarrow} evaluated on $(y_k, y_{k+1}]$.

Theorem 3.11. *Suppose $G = \text{PWL}(\mathbf{x}, \mathbf{y})$ and F are TVaR equivalent. Let $\alpha \in (y_k, y_{k+1}]$ for some k . Then,*

$$\text{TVaR}_{\alpha}^{\Delta}(F) - \text{TVaR}_{\alpha}^{\Delta}(G) = \frac{1}{\alpha} \left(\int_{y_k}^{\alpha} G^{\leftarrow}(t) dt - \int_{y_k}^{\alpha} F^{\leftarrow}(t) dt \right).$$

Proof. Lemma 3.10 provides the identity $1/y_k \int_0^{y_k} G^{\leftarrow}(t) dt = 1/y_k \int_0^{y_k} F^{\leftarrow}(t) dt$. Subtracting these and $\mathbb{E}[F] - \mathbb{E}[G] = 0$ from $\text{TVaR}_{\alpha}^{\Delta}(F) - \text{TVaR}_{\alpha}^{\Delta}(G)$ leads to the desired result. \square

We will see that Theorem 3.11 constitutes a critical component of the approximation algorithm, since it will allow to decompose the approximation problem into independent problems, one for each quantile segment.

4. ERROR BOUNDS AND ADMISSIBILITY

In this section we provide a mathematical framework which defines a set of PWL distributions which are admissible approximations to the sample distribution. The approximation is done by limiting the relative error over a set of risk measures $\text{TVaR}_{\alpha}^{\Delta}$, with different levels α and a certain accuracy of choice.

Definition 4.1. *For a sample distribution F with sample size n , a PWL distribution G is called an **admissible approximation** of F with accuracy $\epsilon > 0$ if $\mathbb{E}[F] = \mathbb{E}[G]$ and*

$$\max_{\alpha \in \{\frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}\}} \frac{|\text{TVaR}_{\alpha}^{\Delta}(G) - \text{TVaR}_{\alpha}^{\Delta}(F)|}{\text{TVaR}_{\alpha}^{\Delta}(F)} \leq \epsilon. \quad (4.1)$$

Definition 4.1 implies that, apart from having the same mean, the $\text{TVaR}_\alpha^\Delta$ relative error of the approximation G does not exceed ϵ for levels implied through the sample points, i.e., $|\text{TVaR}_\alpha^\Delta(G) - \text{TVaR}_\alpha^\Delta(F)| \leq \epsilon \text{TVaR}_\alpha^\Delta(F)$ for all $\alpha \in \{1/n, 2/n, \dots, (n-1)/n\}$. For now, we restrict ourselves to $\alpha \in \{1/n, 2/n, \dots, (n-1)/n\}$ since this allows to derive a very efficient algorithm. The extension to the entire interval $\alpha \in (0, 1)$ will be done in Section 6.

We will later provide some guidance on selecting the ϵ parameter for practical purposes.

Example 4.2. *Let F be defined in Example 3.7 with sample size $n = 10$. Then, a PWL distribution G is an admissible approximation of F with accuracy ϵ if $\mathbb{E}[G] = \mathbb{E}[F] = 8.58$ and*

$$\begin{aligned} |\text{TVaR}_{0.1}^\Delta(G) - \text{TVaR}_{0.1}^\Delta(F)| &\leq \epsilon \text{TVaR}_{0.1}^\Delta(F), \\ |\text{TVaR}_{0.2}^\Delta(G) - \text{TVaR}_{0.2}^\Delta(F)| &\leq \epsilon \text{TVaR}_{0.2}^\Delta(F), \\ &\vdots \\ |\text{TVaR}_{0.9}^\Delta(G) - \text{TVaR}_{0.9}^\Delta(F)| &\leq \epsilon \text{TVaR}_{0.9}^\Delta(F). \end{aligned}$$

We will provide a further definition as well as a lemma that decomposes the set of inequalities (4.1) into a set of numerically tractable problems.

Definition 4.3. *A PWL distribution $G = \text{PWL}(\mathbf{x}, \mathbf{y})$ is called sample compatible with respect to an empirical distribution F with sample size n if $y_k \in \{0/n, 1/n, 2/n, \dots, n/n\}$ for all $k = 1, 2, \dots, K$.*

An example illustrating Definition 4.3 will be provided in the following section. Using sample compatibility, the following lemma decomposes the TVaR equivalence condition (4.1) into a set of independent inequalities, one for each segment.

Theorem 4.4. *Suppose $G = \text{PWL}(\mathbf{x}, \mathbf{y})$ and F are TVaR equivalent and sample compatible. Let $(z_1, z_2, \dots, z_S) = \text{basis}(G)$. Then, G is an admissible approximation of F if and only if the following inequality is satisfied for every $s = 1, 2, \dots, S-1$:*

$$\max_{\alpha \in \{z_s + \frac{1}{n}, z_s + \frac{2}{n}, \dots, z_{s+1} - \frac{1}{n}\}} \frac{1}{\alpha \cdot \text{TVaR}_\alpha^\Delta(F)} \left| \int_{z_s}^\alpha G^-(t) dt - \int_{z_s}^\alpha F^-(t) dt \right| \leq \epsilon. \quad (4.2)$$

Proof. Immediate consequence of Theorem 3.11 and (4.1). \square

Note that for a fixed s , (4.2) evaluates $G^-(t)$ only for $t \in (z_s, z_{s+1}]$. Thereby, under the assumption of TVaR equivalence and sample compatibility, the calculations necessary for the different segments can be done independently. We will later show how this leads to a significant simplification of the numerical complexity in the algorithm.

In the remainder of this section, we show that the algorithm is symmetric.

The other common way to define TVaR is by taking the average over the upper quantiles, i.e., $\text{TVaR}_\alpha^{up}(F) = \frac{1}{1-\alpha} \int_\alpha^1 F^\leftarrow(t) dt$ and $\text{TVaR}_\alpha^{\Delta, up}(F) = \text{TVaR}_\alpha^{up}(F) - \mathbb{E}[F]$ for $0 < \alpha < 1$.

Lemma 4.5. *For $0 < \alpha < 1$, we have $\alpha \text{TVaR}_\alpha^\Delta(F) = (1 - \alpha) \text{TVaR}_\alpha^{\Delta, up}(F)$*

Proof. Plugging in the definition yields

$$\alpha \text{TVaR}_\alpha^\Delta(F) - (1 - \alpha) \text{TVaR}_\alpha^{\Delta, up}(F) = \alpha \mathbb{E}[F] - \int_0^\alpha F^\leftarrow(t) dt - \int_\alpha^1 F^\leftarrow(t) dt + (1 - \alpha) \mathbb{E}[F] = 0 \quad \square$$

By replacing the result of Lemma 4.5 into the admissibility condition (4.1), we see that any PWL distribution satisfying the admissibility condition also satisfies the analogous inequality for $\text{TVaR}_\alpha^{\Delta, up}(F)$. Hence, the algorithm is symmetric with respect to sign changes.

5. ALGORITHM

The basic idea on how to implement an algorithm to find an admissible PWL approximation of F is very simple: Start with $\mathbf{z} = (0, 1)$ and iteratively insert values into \mathbf{z} until there exists an admissible PWL distribution G with $\text{basis}(G) = \mathbf{z}$. However, for a concrete implementation, several mathematical and numerical problems related to this basic idea need to be clarified. This will be done in the following sections. Readers who are mainly interested in applications and examples may proceed directly to Section 8.

5.1. Reparametrisation of the PWL distribution.

In the following, we define an alternative parametrisation $G = \text{PWL}(\mathbf{z}, \boldsymbol{\mu}, \boldsymbol{\delta})$ of PWL distributions. The parametrisation $G = \text{PWL}(\mathbf{z}, \boldsymbol{\mu}, \boldsymbol{\delta})$ is graphically less intuitive than the previously used $G = \text{PWL}(\mathbf{x}, \mathbf{y})$, but it is more suitable to represent the numerical intricacies of the algorithm. Describing the algorithm with the $G = \text{PWL}(\mathbf{x}, \mathbf{y})$ parametrisation would be too cumbersome.

Definition 5.1. *Let $S \in \mathbb{N}$, $\mathbf{z} \in [0, 1]^S$, $\boldsymbol{\mu} \in \mathbb{R}^{S-1}$, and $\boldsymbol{\delta} \in [0, \infty)^{S-1}$ such that*

$$\begin{aligned} 0 &= z_1 < z_2 < \dots < z_S = 1, \quad \text{and} \\ \mu_s + \delta_s &\leq \mu_{s+1} - \delta_{s+1} \quad \text{for } s = 1, 2, \dots, S-2. \end{aligned} \quad (5.1)$$

Then, $G = \text{PWL}(\mathbf{z}, \boldsymbol{\mu}, \boldsymbol{\delta})$ if simultaneously $G = \text{PWL}(\mathbf{x}, \mathbf{y})$ with $\mathbf{x} \in \mathbb{R}^{2S-2}$ and $\mathbf{y} \in [0, 1]^{2S-2}$, such that $x_{2s-1} = \mu_s - \delta_s$, $x_{2s} = \mu_s + \delta_s$, $y_{2s-1} = z_s$, $y_{2s} = z_{s+1}$ for $s = 1, \dots, S-1$. Hence,

$$\begin{aligned} \mathbf{x} &= (\mu_1 - \delta_1, \mu_1 + \delta_1, \mu_2 - \delta_2, \mu_2 + \delta_2, \dots, \mu_{S-1} - \delta_{S-1}, \mu_{S-1} + \delta_{S-1}), \\ \mathbf{y} &= (z_1, z_2, z_2, z_3, \dots, z_{S-1}, z_S). \end{aligned}$$

Note that $\text{basis}(G) = \mathbf{z}$ for $G = \text{PWL}(\mathbf{z}, \boldsymbol{\mu}, \boldsymbol{\delta})$. Some points (x_k, y_k) may be duplicated - these can be discarded. Furthermore, using Definition 5.1 $G^-(t)$ can be rewritten for $t \in (z_s, z_{s+1}]$ through a simple formula:

$$G^-(t) = \mu_s + \delta_s \left(2 \frac{t - z_s}{z_{s+1} - z_s} - 1 \right) \quad \text{for } t \in (z_s, z_{s+1}]. \quad (5.2)$$

There is a natural interpretation of the parameters: μ_s is the average of G^- over the quantiles in the segment $(z_s, z_{s+1}]$ and δ_s represents a slope parameter. Note that for $\delta_s = 0$, G^- is constant on $(z_s, z_{s+1}]$. We have that $\lim_{t \downarrow z_s} G^-(t) = \mu_s - \delta_s$ and $G^-(z_{s+1}) = \mu_s + \delta_s$, which is illustrated in the following example.

Example 5.2. Let $G = \text{PWL}(\mathbf{x}, \mathbf{y})$ with $\mathbf{x} = (0.8, 7.4, 12.9, 17.7)$ and $\mathbf{y} = (0, 0.6, 0.6, 1)$, as in Example 3.9. Equivalently, we can reparametrise $G = \text{PWL}(\mathbf{z}, \boldsymbol{\mu}, \boldsymbol{\delta})$ with $S = 3$,

$$\mathbf{z} = (z_1, z_2, z_3) = (0, 0.6, 1), \quad \boldsymbol{\mu} = (\mu_1, \mu_2) = (4.1, 15.3), \quad \boldsymbol{\delta} = (\delta_1, \delta_2) = (3.3, 2.4).$$

Note that $\text{basis}(G) = \mathbf{z} \neq \mathbf{y}$. Equation (5.2) translates into

$$G^-(t) = 4.1 + 3.3(2(t - 0.0)/(0.6 - 0) - 1) \quad \text{for } t \in (0, 0.6],$$

$$G^-(t) = 15.3 + 2.4(2(t - 0.6)/(1 - 0.6) - 1) \quad \text{for } t \in (0.6, 1].$$

The former equation is illustrated in Figure 4. Note that

$$\lim_{t \downarrow 0} G^-(t) = \mu_1 - \delta_1 = 4.1 - 3.3 = 0.8 = x_1, \quad G^-(0.6) = \mu_1 + \delta_1 = 4.1 + 3.3 = 7.4 = x_2.$$

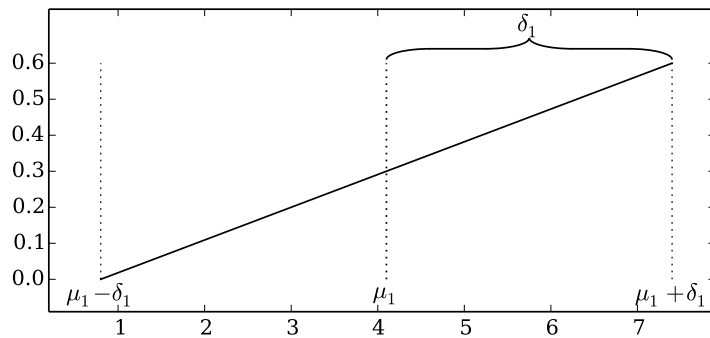


FIGURE 4. An illustration of the parametrisation of $G^-(t)$ (solid line) for $t \in (0, 0.6]$ through $\mu_1 = 4.1$ and $\delta_1 = 3.3$ as given in (5.2) and Example 5.2.

The following lemma shows that sample compatibility and TVaR equivalence allow to obtain a simple analytical expression for μ_s .

Lemma 5.3. *Suppose $G = \text{PWL}(\mathbf{z}, \boldsymbol{\mu}, \boldsymbol{\delta})$ and the empirical distribution F with sample size n , are TVaR equivalent and sample compatible. Then, for $s = 1, 2, \dots, S-1$ we have that*

$$\mu_s = \frac{1}{n(z_{s+1} - z_s)} \sum_{i=nz_s+1}^{nz_{s+1}} X_{(i)}.$$

Proof. From (5.2), we get $\frac{1}{z_{s+1} - z_s} \int_{z_s}^{z_{s+1}} G^-(t) dt = \mu_s$. Combining the former with Lemma 3.10 and (3.1) yields the desired result. \square

As Lemma 5.3 shows, TVaR equivalence uniquely determines μ_s . Namely, μ_s is equal to the average of all order statistics corresponding to quantiles between z_s and z_{s+1} .

Example 5.4. *Let G and F be defined as in Example 3.9. Then, G is sample compatible with respect to F since $\{0, 0.6, 1\} \subseteq \{0/10, 1/10, 2/10, \dots, 9/10, 10/10\}$, where 10 is the sample size underlying the empirical cdf F . Consequently, Lemma 5.3 yields*

$$\mu_1 = \frac{1}{6} \sum_{i=1}^6 X_{(i)} = 4.1, \quad \text{and} \quad \mu_2 = \frac{1}{4} \sum_{i=7}^{10} X_{(i)} = 15.3.$$

Note that $\int_{z_s}^{\alpha} G^-(t) dt$ for $\alpha \in (z_s, z_{s+1}]$ and $G = \text{PWL}(\mathbf{z}, \boldsymbol{\mu}, \boldsymbol{\delta})$ is a function of α , μ_s , and δ_s . Since the parameter $\boldsymbol{\mu}$ is given through Lemma 5.3 if we assume the PWL distribution to be TVaR equivalent and sample compatible, the only remaining free parameter is $\boldsymbol{\delta}$. The following theorem provides a link between the $\boldsymbol{\delta}$ and the admissibility condition (4.2). In fact, the theorem shows that a $\text{PWL}(\mathbf{z}, \boldsymbol{\mu}, \boldsymbol{\delta})$ distribution satisfies the inequality (4.2) if and only if δ_s lies in a certain interval, which is non-empty in case a solution to (4.2) exists. The proof of Theorem 5.5 and formulae for δ_s^{\min} and δ_s^{\max} are provided in Appendix A since they are rather technical.

Theorem 5.5. *Suppose $G = \text{PWL}(\mathbf{z}, \boldsymbol{\mu}, \boldsymbol{\delta})$ and F are TVaR equivalent and sample compatible. Then, for a fixed $s = 1, 2, \dots, S-1$, inequality (4.2) is satisfied if and only if $\delta_s \in \Delta_s$, where*

- $\Delta_s = \emptyset$ (in case there is no solution), or
- $\Delta_s = [\delta_s^{\min}, \delta_s^{\max}] \subset \mathbb{R}$ (in case there is a solution).

Furthermore, Δ_s can be calculated with numerical complexity $O(n(z_{s+1} - z_s))$.

Proof. See Appendix A. \square

In case there is a solution to (4.2), we obtain an interval $[\delta_s^{\min}, \delta_s^{\max}]$ such that (4.2) is satisfied for every $\delta_s \in [\delta_s^{\min}, \delta_s^{\max}]$, which is illustrated in the following example. In general, the interval Δ_s becomes larger when the required accuracy (ϵ) of the approximation increases, or when the range covered by the segment $(z_s, z_{s+1}]$ decreases.

Example 5.6. Given the sample distribution F as defined in Example 3.7, $\mathbf{z} = (0, 0.6, 1)$, and $\epsilon = 0.25$. Then, Theorem 5.5 yields that $\Delta_1 \neq \emptyset$ with $\delta_1^{min} = 1.470$, $\delta_1^{max} = 5.994$, and $\Delta_1 = [1.470, 5.994]$, as it is illustrated in Figure 5. With the same F and \mathbf{z} , but reduced $\epsilon = 0.1$, the interval Δ_1 reduces to $\Delta_1 = [3.108, 4.630]$.

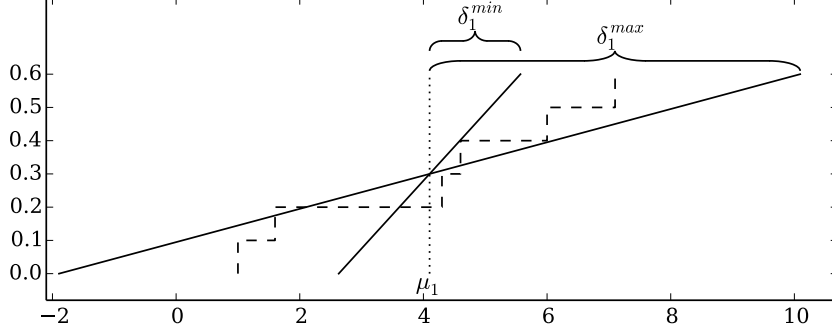


FIGURE 5. An illustration of $\delta_1^{min} = 1.470$ and $\delta_1^{max} = 5.994$ resulting out of Theorem 5.5 using $\epsilon = 0.25$. For $\delta_1 = \delta_1^{min}$, G represents an interpolation between the points $(2.63, 0)$ and $(5.57, 0.6)$ on the segment $(0, 0.6]$. Similarly, for $\delta_1 = \delta_1^{max}$, G interpolates between $(-1.894, 0)$ and $(10.094, 0.6)$.

5.2. Default Slope Parameter.

In this section, we provide a default value for δ_s to be used in case $\Delta_s \neq \emptyset$. We propose to take the δ_s which minimises the L_2 distance between F^- and G^- measured on the segment $(z_s, z_{s+1}]$. This can be seen as a mean-square regression parameter.

Theorem 5.7. Suppose $G = \text{PWL}(\mathbf{z}, \boldsymbol{\mu}, \boldsymbol{\delta})$ and F are TVaR equivalent and sample compatible. For some fixed s , we denote with δ_s^{reg} the δ_s parameter which minimises the L_2 distance between G^- and F^- on $(z_s, z_{s+1}]$ when varying δ_s :

$$\delta_s^{reg} = \operatorname{argmin}_{\delta_s} \int_{z_s}^{z_{s+1}} (F^-(t) - G^-(t))^2 dt$$

Then, δ_s^{reg} is given by

$$\delta_s^{reg} = -3\mu_s \frac{z_{s+1} + z_s}{z_{s+1} - z_s} + \frac{6}{(z_{s+1} - z_s)^2} \frac{1}{n^2} \sum_{i=nz_s+1}^{nz_{s+1}} X_{(i)}(i - 1/2).$$

Proof. Let $D(\delta_s) = \int_{z_s}^{z_{s+1}} (F^\leftarrow(t) - G^\leftarrow(t))^2 dt$. By plugging (5.2) into $D(\delta_s)$, we get

$$\begin{aligned} D(\delta_s) &= \int_{z_s}^{z_{s+1}} \left(F^\leftarrow(t) - \left(\mu_s + \delta_s \left(2 \frac{t - z_s}{z_{s+1} - z_s} - 1 \right) \right) \right)^2 dt \\ &= \int_{z_s}^{z_{s+1}} (F^\leftarrow(t) - \mu_s)^2 dt - \delta_s 2 \int_{z_s}^{z_{s+1}} (F^\leftarrow(t) - \mu_s) \left(2 \frac{t - z_s}{z_{s+1} - z_s} - 1 \right) dt \\ &\quad + \delta_s^2 \int_{z_s}^{z_{s+1}} \left(2 \frac{t - z_s}{z_{s+1} - z_s} - 1 \right)^2 dt. \end{aligned}$$

The last component can be calculated as $\int_{z_s}^{z_{s+1}} \left(2 \frac{t - z_s}{z_{s+1} - z_s} - 1 \right)^2 dt = \frac{z_{s+1} - z_s}{3}$. Furthermore, due to $\int_{z_s}^{z_{s+1}} (F^\leftarrow(t) - \mu_s) dt = 0$, we get

$$\begin{aligned} \int_{z_s}^{z_{s+1}} (F^\leftarrow(t) - \mu_s) \left(2 \frac{t - z_s}{z_{s+1} - z_s} - 1 \right) dt &= \frac{2}{z_{s+1} - z_s} \int_{z_s}^{z_{s+1}} (F^\leftarrow(t) - \mu_s) t dt \\ &= \left(\frac{2}{z_{s+1} - z_s} \int_{z_s}^{z_{s+1}} F^\leftarrow(t) t dt \right) - \mu_s (z_{s+1} + z_s). \end{aligned}$$

Since we assumed G to be sample compatible, we can write

$$\int_{z_s}^{z_{s+1}} F^\leftarrow(t) t dt = \sum_{i=nz_s+1}^{nz_{s+1}} \int_{(i-1)/n}^{i/n} F^\leftarrow(t) t dt = \sum_{i=nz_s+1}^{nz_{s+1}} \int_{(i-1)/n}^{i/n} X_{(i)} t dt = \frac{1}{n^2} \sum_{i=nz_s+1}^{nz_{s+1}} X_{(i)} (i - 1/2).$$

Using the previous calculations, we can deduce the derivative

$$\frac{d}{d\delta_s} D(\delta_s) = \frac{-4}{z_{s+1} - z_s} \frac{1}{n^2} \sum_{i=nz_s+1}^{nz_{s+1}} X_{(i)} (i - 1/2) + 2\mu_s (z_{s+1} + z_s) + 2\delta_s \frac{z_{s+1} - z_s}{3}.$$

We obtain δ_s^{reg} by setting the above equation to zero. \square

Example 5.8. Given the sample distribution F as defined in Example 3.7 and $\mathbf{z} = (0, 0.6, 1)$ as in Example 5.6. Then, Theorem 5.7 yields $\delta_1^{reg} = 3.667$ and $\delta_2^{reg} = 3.750$.

Further illustrations of Theorem 5.7 and δ_s^{reg} will be provided in Section 5.4.

5.3. Best Bisection Point for a Segment.

Using the results of Section 5.1, we can reformulate the basic idea of the algorithm: it starts with $\mathbf{z} = (0, 1)$ and inserts points into \mathbf{z} until $\Delta_s \neq \emptyset$ for every segment $(z_s, z_{s+1}]$. In this section, we describe where to bisect $(z_s, z_{s+1}]$ in case $\Delta_s = \emptyset$.

Algorithm 5.9. In case $\Delta_s = \emptyset$, insert \tilde{z} into \mathbf{z} , where

$$\tilde{z} = \operatorname{argmax}_{\alpha \in \{z_s + \frac{1}{n}, z_s + \frac{2}{n}, \dots, z_{s+1} - \frac{1}{n}\}} \left| \int_{z_s}^{\alpha} G^\leftarrow(t) dt - \int_{z_s}^{\alpha} F^\leftarrow(t) dt \right|,$$

with μ_s and δ_s chosen according to Lemma 5.3 and Theorem 5.7, respectively, i.e., bisect the segment $(z_s, z_{s+1}]$ into two new segments $(z_s, \tilde{z}]$ and $(\tilde{z}, z_{s+1}]$.

The value \tilde{z} represents a bisection of the segment at the level where the difference of the scaled TVaR $^\Delta$ between F and the corresponding linear approximation is largest. The performance of the given choice of \tilde{z} is very good in terms of convergence and run time. Further mathematical justification will be given in Section 7.

Example 5.10. *Given the sample distribution F as defined in Example 3.7 and $\mathbf{z} = (0, 1)$. Applying Algorithm 5.9 on the segment $(z_1, z_2] = (0, 1]$ yields $\tilde{z} = 0.6$ since the function $\alpha \mapsto \left| \int_{z_1}^\alpha G^-(t)dt - \int_{z_1}^\alpha F^-(t)dt \right|$ evaluates to $(0.14, 0.14, 0.21, 0.11, 0.04, 0.29, 0.15, 0.17, 0.12)$ for $\alpha = (0.1, 0.2, \dots, 0.9)$.*

5.4. Ensuring Segment Compatibility.

Recall that the approximation algorithm will start with finding a basis \mathbf{z} such that $\Delta_s \neq \emptyset$ for all segments. However, such a basis together with the choice $\delta_s = \delta_s^{reg}$ does not necessarily imply a well defined PWL approximation. Namely, it needs to be ensured that the resulting PWL distribution satisfies the compatibility condition (5.1). This section describes how to ensure that a parameter triple $(\mathbf{z}, \boldsymbol{\mu}, \boldsymbol{\delta})$ satisfies (5.1).

In Section 5.2, we propose to use a default value $\delta_s = \delta_s^{reg}$. In most cases, this choice will lead to compatible segments. Suppose now that

$$\mu_s + \delta_s^{reg} > \mu_{s+1} - \delta_{s+1}^{reg},$$

for some $s \in \{1, 2, \dots, S-2\}$, i.e., (5.1) is not satisfied with $\delta_s = \delta_s^{reg}$ and $\delta_{s+1} = \delta_{s+1}^{reg}$. We can now distinguish two cases: either the segments can be compatible using $\delta_s = \delta_s^{min}$ and $\delta_{s+1} = \delta_{s+1}^{min}$ (i.e., $\mu_s + \delta_s^{min} \leq \mu_{s+1} - \delta_{s+1}^{min}$), or the opposite is true (i.e., $\mu_s + \delta_s^{min} > \mu_{s+1} - \delta_{s+1}^{min}$). Algorithm 5.11 describes a solution to the former problem and Algorithm 5.13 to the latter.

Algorithm 5.11. *Suppose that for some fixed s , we have*

$$\mu_s + \delta_s^{reg} > \mu_{s+1} - \delta_{s+1}^{reg}, \quad \text{and} \quad \mu_s + \delta_s^{min} \leq \mu_{s+1} - \delta_{s+1}^{min}.$$

Consequently, the intervals $\{\mu_s + \delta_s : \delta_s \in \Delta_s\} \subset \mathbb{R}$ and $\{\mu_{s+1} - \delta_{s+1} : \delta_{s+1} \in \Delta_{s+1}\} \subset \mathbb{R}$ overlap. In this case, we propose to proceed by resetting δ_s and δ_{s+1} such that $\mu_s + \delta_s$ and $\mu_{s+1} - \delta_{s+1}$ are equal to the midpoint of the interval

$$\left[\mu_s + \delta_s^{min}, \mu_s + \delta_s^{reg} \right] \cap \left[\mu_{s+1} - \delta_{s+1}^{reg}, \mu_{s+1} - \delta_{s+1}^{min} \right]. \quad (5.3)$$

The following example illustrates Algorithm 5.11.

Example 5.12. Suppose F is defined as in Example 3.7. Furthermore, let $\mathbf{z} = (0, 0.3, 1)$ and $\epsilon = 0.25$. We have

$$\begin{aligned} \mu_1 &= 2.3, & \delta_1^{min} &= 0, & \delta_1^{reg} &= 2.2, & \delta_1^{max} &= 4.224, \\ \mu_2 &= 11.271, & \delta_2^{min} &= 6.398, & \delta_2^{reg} &= 8.437, & \delta_2^{max} &= 11.01, \end{aligned}$$

which leads to

$$4.500 = \mu_1 + \delta_1^{reg} > \mu_2 - \delta_2^{reg} = 2.835, \quad \text{and} \quad 2.300 = \mu_1 + \delta_1^{min} \leq \mu_2 - \delta_2^{min} = 4.873.$$

The interval (5.3) is given by $[2.300, 4.500] \cap [2.835, 4.873] = [2.835, 4.500]$ with a midpoint $(2.835 + 4.500)/2 = 3.667$. By using Algorithm 5.11, we set $\delta_1 = 1.367$ and $\delta_2 = 7.604$. After this adjustment, the two segments connect at the midpoint. Figure 6 provides an illustration.

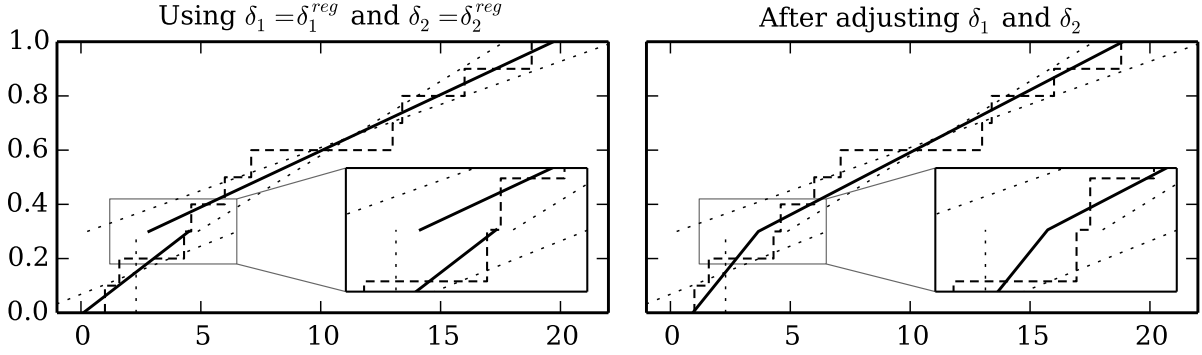


FIGURE 6. Illustration of Algorithm 5.11. Dotted: δ_s^{min} and δ_s^{max} . Left: With $\delta_1 = \delta_1^{reg}$ and $\delta_2 = \delta_2^{reg}$, the two segments (solid) are incompatible. Right: Setting $\delta_1 = 1.367$ and $\delta_2 = 7.604$ yields compatible and connecting segments.

The following algorithm proposes an approach to resolve the situation where two segments are also incompatible when using $\delta_s = \delta_s^{min}$ and $\delta_{s+1} = \delta_{s+1}^{min}$.

Algorithm 5.13. Suppose that for some fixed s , we have

$$\mu_s + \delta_s^{min} > \mu_{s+1} - \delta_{s+1}^{min}.$$

Consequently, the intervals $\{\mu_s + \delta_s : \delta_s \in \Delta_s\}$ and $\{\mu_{s+1} - \delta_{s+1} : \delta_{s+1} \in \Delta_{s+1}\}$ are disjoint. In this case, we propose to proceed by bisecting the bigger of the two segments using the bisection point defined in Algorithm 5.9.

The following example illustrates Algorithm 5.13

Example 5.14. Suppose F is defined as in Example 3.7. Furthermore, let $\mathbf{z} = (0, 0.3, 1)$ and $\epsilon = 0.08$. We get

$$\begin{aligned} \mu_1 &= 2.3, & \delta_1^{min} &= 1.253, & \delta_1^{reg} &= 2.2, & \delta_1^{max} &= 2.860, \\ \mu_2 &= 11.271, & \delta_2^{min} &= 8.146, & \delta_2^{reg} &= 8.437, & \delta_2^{max} &= 9.440, \end{aligned}$$

which leads to

$$3.553 = \mu_1 + \delta_1^{min} > \mu_2 - \delta_2^{min} = 3.126.$$

We cannot resolve the incompatibility through resetting δ_1 and δ_2 for the purpose of fulfilling condition (5.1). Therefore, applying Algorithm 5.13, we bisect the upper segment $(0.3, 1]$ into the two segments $(0.3, 0.6]$ and $(0.6, 1]$ by using $\tilde{z} = 0.6$. Figure 7 provides an illustration.

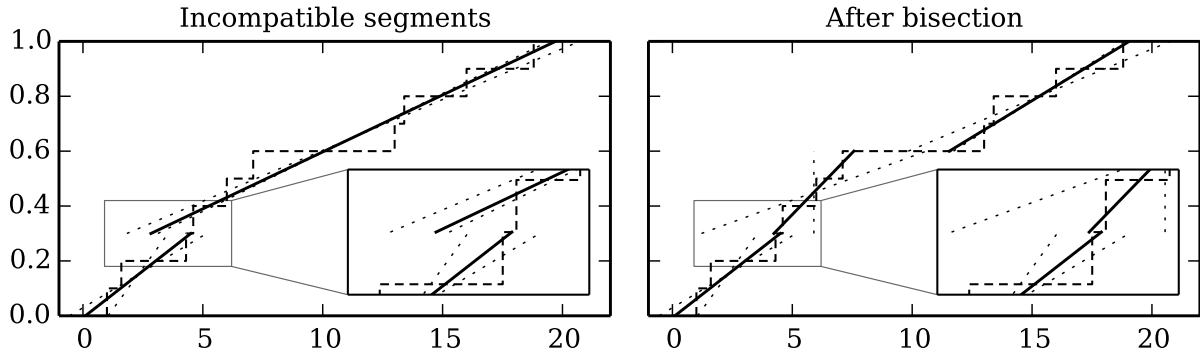


FIGURE 7. For $\mathbf{z} = (0, 0.3, 1)$ and $\epsilon = 0.08$, the parameters δ_1 and δ_2 cannot be chosen such that the resulting PWL is admissible. The problem is resolved by bisecting the upper segment at $\tilde{z} = 0.6$.

Note that after this operation, the first and second segment are still not compatible, but they can be connected through Algorithm 5.11. Note that F and \mathbf{z} are equal in Example 5.12 and Example 5.14. However, choosing a smaller ϵ ($\epsilon = 0.08$ instead of $\epsilon = 0.2$, respectively) leads to narrower intervals Δ_1 and Δ_2 .

5.5. Smoothing.

Suppose we have found an admissible solution, i.e., a parameter triple $(\mathbf{z}, \boldsymbol{\mu}, \boldsymbol{\delta})$ which satisfies (4.1) and (5.1). For certain applications of the PWL approximation algorithm, we may know that the underlying distribution from which F was sampled is smooth. In this case, it would be favourable if G^- were continuous. However, in many cases G^- , will have jumps, i.e., $\mu_s + \delta_s < \mu_{s+1} - \delta_{s+1}$ for some s . It is possible to reset δ_s and δ_{s+1} to larger values in Δ_s and Δ_{s+1} , respectively, such that G^- becomes continuous.

Algorithm 5.15. For $s = \{1, \dots, S-2\}$:

(1) Check whether the following interval is nonempty:

$$[\mu_s + \delta_s^{reg}, \mu_s + \delta_s^{max}] \cap [\mu_{s+1} - \delta_{s+1}^{max}, \mu_{s+1} - \delta_{s+1}^{reg}]. \quad (5.4)$$

(2) In case interval (5.4) is nonempty, and δ_s as well as δ_{s+1} can be adjusted without violating (5.1), then reset δ_s and δ_{s+1} such that $\mu_s + \delta_s$ and $\mu_{s+1} - \delta_{s+1}$ are equal to the midpoint of (5.4).

The following example illustrates Algorithm 5.15.

Example 5.16. Suppose F is defined as in Example 3.7. Furthermore, let $\mathbf{z} = (0, 0.6, 1)$ and $\epsilon = 0.25$. We have

$$\begin{array}{llll} \mu_1 = 4.1, & \delta_1^{min} = 1.47, & \delta_1^{reg} = 3.667, & \delta_1^{max} = 5.994, \\ \mu_2 = 15.3, & \delta_2^{min} = 1.26, & \delta_2^{reg} = 3.75, & \delta_2^{max} = 8.073, \end{array}$$

Choosing $\delta_1 = \delta_1^{reg}$ and $\delta_2 = \delta_2^{reg}$ leads to compatible segments since

$$7.767 = \mu_1 + \delta_1^{reg} < \mu_2 - \delta_2^{reg} = 11.55.$$

Note that the admissible ranges overlap, since $10.094 = \mu_1 + \delta_1^{max} > \mu_2 - \delta_2^{max} = 7.227$. The interval (5.4) is given by $[7.767, 10.094]$. Therefore, Algorithm 5.15 suggests resetting $\delta_1 = 4.830$ and $\delta_2 = 6.370$ such that the two segments connect at the midpoint 8.930. Figure 8 provides an illustration.

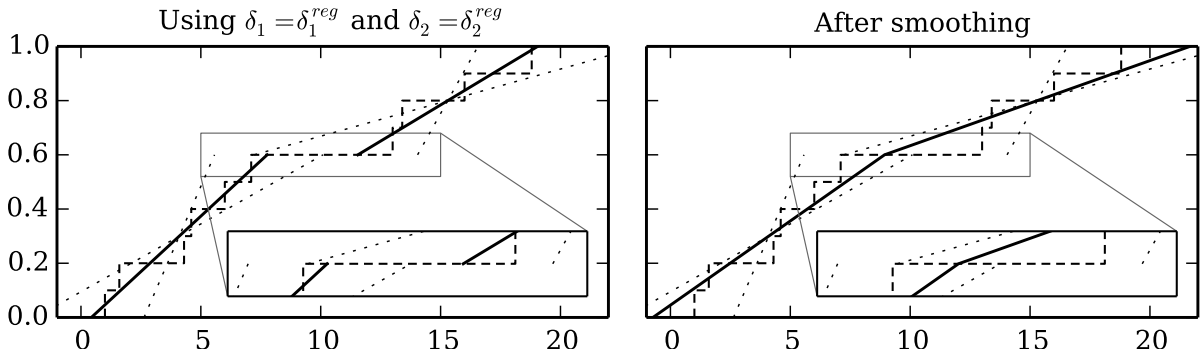


FIGURE 8. Illustration of smoothing as described in Algorithm 5.15. After smoothing, $G^-(t)$ is continuous at $t = z_2 = 0.6$.

5.6. Full Algorithm.

Using the results of the previous sections, we now have all components to describe the full approximation algorithm.

Algorithm 5.17. Fix $\epsilon > 0$ and initialise $\mathbf{z} = (0, 1)$. It is also possible to use a predefined set of quantiles $\mathbf{z} = (0, z_1, \dots, z_R, 1)$, (with $0 < z_1 < \dots < z_R < 1$), if the user requires that the approximation preserves $\text{TVaR}_\alpha(F) = \text{TVaR}_\alpha(G)$ for $\alpha = z_1, \dots, z_R$.

- (1) Determine Δ_s for every segment $(z_s, z_{s+1}]$ in \mathbf{z} through Theorem 5.5.
- (2) If there are segments with $\Delta_s = \emptyset$: Bisect all of them as described in Section 5.3 and return to Point (1).
- (3) Set $\delta_s = \delta_s^{reg}$ for all segments (see Section 5.2).
- (4) In case there are segments which do not satisfy (5.1): If possible, use Algorithm 5.11. Otherwise, use Algorithm 5.13 and go back to Point (3).
- (5) Optional: Apply smoothing as described in Section 5.5.

Note that the algorithm is shift and scale invariant since the relative error in terms of TVaR^Δ also possesses these properties. Step (2) requires that the segments are bisected until $\Delta_s \neq \emptyset$ for all s , which implies that the resulting intervals are usually small.

The algorithm can be run with and without smoothing. Smoothing will make G have less flat parts, which is desirable for underlying smooth distributions. Furthermore, since some pairs of segments are connected at the same point when smoothing is applied, a smoothed PWL approximation generally has less interpolation points and hence requires less memory space. On the other hand, G may be further apart from the mean square regression (δ_s^{reg}) than without smoothing. The numerical complexity added by the application of smoothing is negligible as well as the additional run time. The effect of smoothing in the performance of the algorithm is illustrated in Section 8. Overall, the authors generally recommend to use smoothing.

In order to select the ϵ parameter, the user should assess the required accuracy as well as the error of F in terms of TVaR^Δ compared to the underlying distribution from which F was sampled. In actuarial practice, the authors found ϵ parameters between 0.001 and 0.01 provide a good compromise between goodness-of-fit and number of interpolation points.

6. STRICT ADMISSIBILITY AND SPECTRAL RISK MEASURES

In this section we will show that the error bounds used on TVaR^Δ also extend to larger spaces of risk measures. Additionally, we will strengthen the admissibility criterion leading to a more general error bound for all spectral risk measures.

Definition 6.1. Let $\phi \in L^1(0, 1)$ be an admissible risk spectrum as defined in Acerbi (2002) (i.e., $\phi : [0, 1] \rightarrow [0, \infty)$ is non-negative, decreasing, and satisfies $\int_0^1 \phi(t) dt = 1$). Then, the **spectral risk measure** M_ϕ maps a distribution F to

$$M_\phi(F) = - \int_0^1 F^-(t) \phi(t) dt.$$

Analogous to the TVaR^Δ , we define the deviation $M_\phi^\Delta(F) = \mathbb{E}[F] - \int_0^1 F^-(t) \phi(t) dt$.

We refer to Gzyl and Mayoral (2006) for an overview on spectral risk measures.

Example 6.2. By setting $\phi(t) = 1/\alpha \mathbb{1}\{t \leq \alpha\}$ for $t \in [0, 1]$, we see that TVaR_α is a spectral risk measure.

The following lemma shows that for a large set of spectral risk measures with piecewise constant spectrum, an error bound analogous to the bound for TVaR^Δ applies.

Lemma 6.3. Suppose G is an admissible approximation of F with accuracy ϵ . Furthermore, let $\phi : [0, 1] \rightarrow [0, \infty)$ be a piecewise constant spectrum, such that $\phi(t) = \tilde{\phi}_i$ for $t \in ((i-1)/n, i/n]$, where $\tilde{\phi}_1, \tilde{\phi}_2, \dots, \tilde{\phi}_n \in \mathbb{R}$ is a non-negative decreasing sequence satisfying $1/n \sum_{i=1}^n \tilde{\phi}_i = 1$. Then,

$$\left| M_\phi^\Delta(G) - M_\phi^\Delta(F) \right| \leq \epsilon M_\phi^\Delta(F).$$

Proof. Let $\psi_1, \psi_2, \dots, \psi_n$ be defined through $\psi_n = \tilde{\phi}_n$ and $\psi_i = (\tilde{\phi}_i - \tilde{\phi}_{i+1}) \frac{i}{n}$ for $i = 1, 2, \dots, n-1$. We have $\phi(t) = \sum_{i=1}^n \psi_i \frac{1}{i/n} \mathbb{1}\{t \leq i/n\}$, which leads to $M_\phi^\Delta(F) = \sum_{i=1}^n \psi_i \text{TVaR}_{i/n}^\Delta(F)$ and

$$\begin{aligned} \left| M_\phi^\Delta(G) - M_\phi^\Delta(F) \right| &\leq \left| \sum_{i=1}^n \psi_i (\text{TVaR}_{i/n}^\Delta(G) - \text{TVaR}_{i/n}^\Delta(F)) \right| \\ &\leq \sum_{i=1}^n \psi_i |\text{TVaR}_{i/n}^\Delta(G) - \text{TVaR}_{i/n}^\Delta(F)| \leq \sum_{i=1}^n \psi_i \epsilon \text{TVaR}_{i/n}^\Delta(F) = \epsilon M_\phi^\Delta(F). \quad \square \end{aligned}$$

Example 6.4. Given F as in Example 3.7, i.e., $n = 10$. Let $\phi(t) = 6 \cdot \mathbb{1}\{t \leq 0.1\} + \mathbb{1}\{0.1 < t \leq 0.5\}$ be the risk spectrum. Then, Lemma 6.3 yields that for an admissible PWL approximation $G = \text{PWL}(\mathbf{x}, \mathbf{y})$ with accuracy ϵ , we have $|M_\phi^\Delta(G) - M_\phi^\Delta(F)| \leq \epsilon M_\phi^\Delta(F)$. Furthermore, note that $M_\phi(F) = 0.5 \text{TVaR}_{0.1}(F) + 0.5 \text{TVaR}_{0.5}(F)$.

In the following definition, we will tighten the admissibility condition, which will then allow to deduce an even stronger bound for the set of all spectral risk measures.

Definition 6.5. A PWL distribution G is called a **strictly admissible approximation** of F with accuracy $\epsilon > 0$ if $\mathbb{E}[F] = \mathbb{E}[G]$ and

$$\sup_{0 < \alpha < 1} \frac{|\text{TVaR}_\alpha^\Delta(G) - \text{TVaR}_\alpha^\Delta(F)|}{\text{TVaR}_\alpha^\Delta(F)} \leq \epsilon. \quad (6.1)$$

Definition 6.5 implies that, apart from having the same mean, the relative error of the approximation G measured through the $\text{TVaR}_\alpha^\Delta$ does not exceed ϵ for any $0 < \alpha < 1$, i.e.,

$$|\text{TVaR}_\alpha^\Delta(G) - \text{TVaR}_\alpha^\Delta(F)| \leq \epsilon \text{TVaR}_\alpha^\Delta(F) \quad (6.2)$$

for all $0 < \alpha < 1$. The difference between the definition of admissibility and strict admissibility is that in Definition 4.1, (6.2) needs to be satisfied only for $\alpha \in \{1/n, 2/n, \dots, (n-1)/n\}$, whereas for full admissibility, (6.2) holds for every $0 < \alpha < 1$. Strict admissibility allows to extend the result of Lemma 6.3 to the set of all spectral risk measures.

Theorem 6.6. Suppose $G = \text{PWL}(\mathbf{x}, \mathbf{y})$ is a strictly admissible approximation of F with accuracy ϵ . Furthermore, let M_ϕ^Δ be a spectral risk measure. Then,

$$|M_\phi^\Delta(G) - M_\phi^\Delta(F)| \leq \epsilon M_\phi^\Delta(F). \quad (6.3)$$

Proof. It is possible to express $M_\phi(F) = \int_0^1 \text{TVaR}_\alpha(F) d\Psi(\alpha)$ where the measure $d\Psi(\alpha)$ is such that $d\Psi(\alpha) = -\alpha d\phi(\alpha)$ and $\int_0^1 d\Psi(\alpha) = 1$ (see Acerbi (2002)). Plugging the above into $M_\phi^\Delta(G) - M_\phi^\Delta(F)$ yields

$$\begin{aligned} |M_\phi^\Delta(G) - M_\phi^\Delta(F)| &= \left| \int_0^1 (\text{TVaR}_\alpha(G) - \text{TVaR}_\alpha(F)) d\Psi(\alpha) \right| \\ &\leq \int_0^1 |\text{TVaR}_\alpha(G) - \text{TVaR}_\alpha(F)| d\Psi(\alpha) \leq \int_0^1 \epsilon \text{TVaR}_\alpha^\Delta(F) d\Psi(\alpha) = \epsilon M_\phi^\Delta(F). \quad \square \end{aligned}$$

The following theorem provides a result on the numerical complexity of checking whether a certain PWL distribution is strictly admissible. Due to the technicality, the proof is given in Appendix B.

Theorem 6.7. Suppose $G = \text{PWL}(\mathbf{x}, \mathbf{y})$ is an admissible and sample compatible approximation of F with accuracy ϵ . Then, the segments where G violates the strict admissibility condition (6.1) can be found with numerical complexity $O(n)$.

Proof. See Appendix B. □

The following algorithm describes how to obtain a strictly admissible PWL approximation.

Algorithm 6.8. Set $\epsilon > 0$.

- (1) Execute Algorithm 5.17 in order to obtain an admissible approximation.
- (2) Test through Theorem 6.7 whether the resulting PWL distribution is strictly admissible. If not, bisect all segments which violate the strict admissibility condition and go back to Point (2) in Algorithm 5.17.

Executing Algorithm 5.17 without testing for strict admissibility may potentially result in a PWL approximation which does not satisfy the condition $|\text{TVaR}_\alpha^\Delta(G) - \text{TVaR}_\alpha^\Delta(F)| \leq \epsilon \text{TVaR}_\alpha^\Delta(F)$ for some $\alpha \notin \{1/n, 2/n, \dots, n/n\}$. In practice, this is not a problem. On the one hand, admissible PWL approximations are often also strictly admissible due to continuity of $\alpha \mapsto \text{TVaR}_\alpha$. On the other hand, for a large sample size n for which the algorithm is intended, the grid $\{1/n, 2/n, \dots, n/n\}$ is so dense that thresholds between the grid points are not of practical relevance. Strict admissibility is mainly interesting from a mathematical point of view, since Theorem 6.6 shows that it is numerically feasible to find a PWL approximation which satisfies (6.3) for *any* spectral risk measure. The effect of strict admissibility in the performance of the algorithm is analysed in Section 8. Overall, the authors do recommend to enforce strict admissibility only if n is very small or if spectral risk measures, different of TVaR, are applied to the PWL approximation.

7. CONVERGENCE RATE AND NUMERICAL COMPLEXITY

In this section, we show abstract and empirical results allowing us to conclude that the algorithm is numerically efficient. To do so, we will first define a distance measure between a general distribution and its linear approximation over a segment. It will later allow to deduce an error reduction rate.

Definition 7.1. For a distribution F , the distance $\mathcal{D}(F, (z_s, z_{s+1}], \alpha)$ on the segment $(z_s, z_{s+1}]$ for $\alpha \in (z_s, z_{s+1}]$ is defined as:

$$\mathcal{D}(F, (z_s, z_{s+1}], \alpha) = \left| \int_{z_s}^{\alpha} L_{(z_s, z_{s+1}]}^{\leftarrow}(t) dt - \int_{z_s}^{\alpha} F^{\leftarrow}(t) dt \right|,$$

where $L_{(z_s, z_{s+1}]}^{\leftarrow}(t)$ is a least-squares linear approximation given through (5.2) with parameters μ_s and δ_s analogous to Lemma 5.3 and Theorem 5.7. (i.e., $\mu_s = 1/(z_{s+1} - z_s) \int_{z_s}^{z_{s+1}} F^{\leftarrow}(t) dt$ as well as $\delta_s^{reg} = \operatorname{argmin}_{\delta_s} \int_{z_s}^{z_{s+1}} (F^{\leftarrow}(t) - L_{(z_s, z_{s+1}]}^{\leftarrow}(t))^2 dt$). The overall distance $\mathcal{D}(F, [z_s, z_{s+1}])$ is defined as

$$\mathcal{D}(F, [z_s, z_{s+1}]) = \max_{z_s < \alpha \leq z_{s+1}} \mathcal{D}(F, (z_s, z_{s+1}], \alpha).$$

We see that Definition 7.1 provides a distance measure between a distribution F and its linear (mean square regression) approximation, over the segment $(z_s, z_{s+1}]$. The distance is measured through the maximum difference between the integrals over F and its linear approximation. The foregoing is analogous to the maximum (4.2), when dropping the normalisation term $\alpha \text{TVaR}_\alpha^\Delta$, which provides a reasonable approximation, since for small segments, this factor will not vary much over the segment.

Theorem 7.2. *Let F_{pol} be a distribution such that $F_{pol}^-(t) = a + bt + ct^2$ for $t \in (z_s, z_{s+1}]$. Analogous to Algorithm 5.9, define $\tilde{z} = \operatorname{argmax}_\alpha \{\mathcal{D}(F_{pol}, [z_s, z_{s+1}], \alpha)\}$. Then, bisection at the optimal point \tilde{z} multiplicatively reduces the approximation error:*

$$\max\{\mathcal{D}(F_{pol}, [z_s, \tilde{z}]), \mathcal{D}(F_{pol}, [\tilde{z}, z_{s+1}])\} = \left(\frac{\max\{\tilde{z} - z_s, z_{s+1} - \tilde{z}\}}{z_{s+1} - z_s}\right)^3 \mathcal{D}(F_{pol}, [z_s, z_{s+1}]).$$

Proof. See Appendix C. □

Theorem 7.2 establishes that for a distribution with quadratic inverse, bisection of a segment multiplicatively reduces the distance measured analogous to (4.2). The result of Theorem 7.2 cannot directly be applied to empirical sample distributions, which are discontinuous by definition. However, if we assume that F is sampled from a smooth distribution, the sample size is large, and the segment is sufficiently small (i.e., $1/n \ll z_{s+1} - z_s \ll 1$) then F^- on $(z_s, z_{s+1}]$ can be well approximated through a quadratic function. Therefore, we believe that also for empirical sample distributions an error reduction behaviour as shown in Theorem 7.2 holds. We propose to investigate a proof of this conjecture in future research.

The following example corroborates that the multiplicative error reduction behaviour is also observed in the PWL approximation algorithm.

Example 7.3. *We consider an empirical distribution F sampled from a lognormal random variable with mean 10, standard deviation 1, and sample size $n = 10^6$. We start with $\mathbf{z}^{(0)} = (0, 1)$ and iteratively define $\mathbf{z}^{(k+1)}$ by bisecting every segment in $\mathbf{z}^{(k)}$ using Algorithm 5.9. The error of the k -th iteration is then defined as*

$$E^{(k)} = \max_{(z_s, z_{s+1}] \in \mathbf{z}^{(k)}} \mathcal{D}(F, (z_s, z_{s+1}]).$$

Figure 9 shows how the approximation error $E^{(k)}$ decreases exponentially in the number of bisections applied.

Theorem 7.2 and Example 7.3 indicate that the approximation error reduces exponentially with the number of bisections. Due to Theorem 5.5, calculating Δ_s for a segment requires $O(n(z_{s+1} - z_s))$ operations. In combination, we obtain that applying Algorithm 5.17

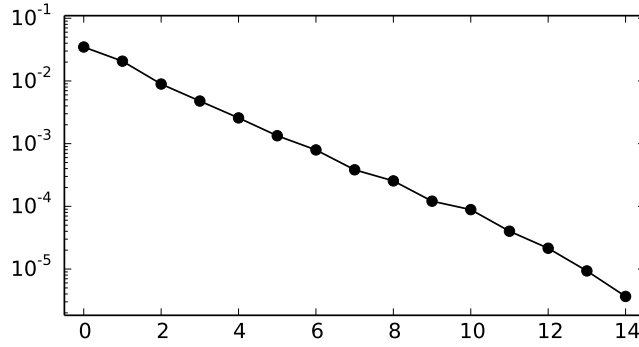


FIGURE 9. Errors $E^{(k)}$ for iterations $k = 0, \dots, 14$.

on an ordered sample requires $O(n \log(1/\epsilon))$ operations. Sorting a sample of size n requires $O(n \log(n))$ operations with usual sorting algorithms. Therefore, the combined numerical complexity of the PWL approximation algorithm is $O(n \log(n/\epsilon))$. With the same argument, we can also conclude that the number of interpolation points is of order $O(\log(1/\epsilon))$. Both statements will be empirically confirmed in Section 8.

8. ILLUSTRATIONS OF FULL ALGORITHM

In this section, we provide some examples illustrating results and performance of the approximation algorithm.

In Figure 10, we provide an empirical distribution F sampled from a loss distribution of an excess-of-loss reinsurance treaty. In order to make visible the discrete nature of the sample distribution, we use a small sample size $n = 1000$. The treaty is assumed to have a limit of 10, a deductible of 12, and an aggregate limit of 30. Furthermore, we assume the loss frequency to be Poisson($\lambda = 2$) distributed and single losses to have a Pareto($x_0 = 10, \alpha = 2.5$) distribution. An admissible and smoothed PWL approximation with accuracy $\epsilon = 0.001$ is given by the dashed distribution. We see that the sample distribution has both discrete parts with atoms at 0, 10, 20, and 30, as well as continuous parts between the atoms.

As a further illustration, we execute the approximation algorithm with smoothing, for different sample sizes and accuracies. Namely, we use $n = 10^4, 10^5, 10^6$ and $\epsilon = 0.1, 0.01, 0.001, 0.0001$, which represent situations that would be encountered in practice. Table 1 shows the run time in milliseconds, as well as the number of interpolation points, both averaged over 100 repetitions on an ordinary computer (Intel Core i7, 2.93 GHz, 8 GB RAM).

We see that for a sample size of 1 million, the run time remains well below a second. The resulting number of interpolation points of the PWL distribution remains below 100, which

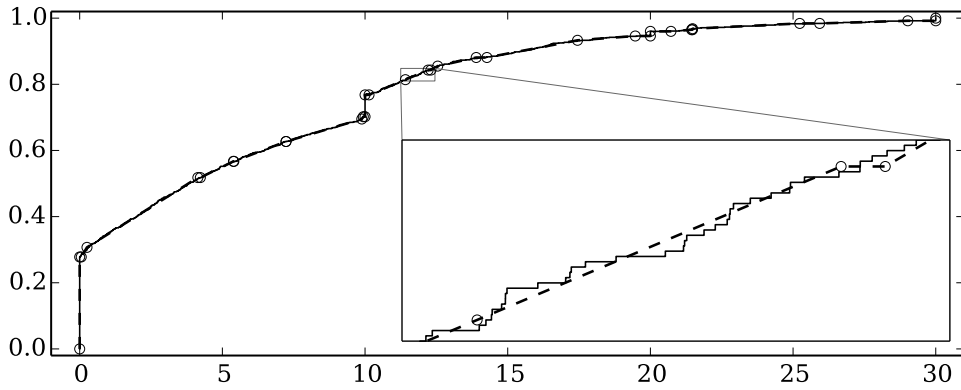


FIGURE 10. Sample distribution F (solid line) with $n = 1000$ and an admissible PWL approximation thereof (dashed, circles at grid points) with accuracy $\epsilon = 0.001$. The lower right sub-plot provides a zoom.

	Run time in milliseconds			Number of points (K)		
	$n = 10^4$	$n = 10^5$	$n = 10^6$	$n = 10^4$	$n = 10^5$	$n = 10^6$
$\epsilon = 0.1$	2	14	155	11	11	11
$\epsilon = 0.01$	4	19	171	19	17	17
$\epsilon = 0.001$	5	25	192	31	27	28
$\epsilon = 0.0001$	7	35	219	62	48	45

TABLE 1. Average run time (in milliseconds) and number of interpolation points of the algorithm (over 100 repetitions) for an admissible and smoothed approximation

for $n = 10^6$ implies a compression ratio above 10000. Due to the fact that the sample distribution gets smoother with increasing sample points, the number of interpolation points decreases from $n = 10^4$ to $n = 10^6$. However, the opposite behaviour is usually observed when the sample is drawn from an unbounded distribution.

In order to show the effect of applying smoothing and enforcing strict admissibility on the algorithm, we analyse its performance when $n = 10^6$ and $\epsilon = 0.001$ for the excess-of-loss and a lognormal distribution with mean 10 and standard deviation 1. Table 2 shows the results over 100 repetitions. The additional time required to smooth the approximation is negligible, while the number of interpolation points of a smoothed approximation is lower than without smoothing. When also enforcing strict admissibility, the run time of the algorithm almost doubles for the excess-of-loss and more than triples for the lognormal distribution. The extra points obtained to ensure strict admissibility are very few for large sample sizes. Therefore,

enforcing strict admissibility does not seem to be worthwhile given the required additional computational effort.

	Admissible non smoothed		Admissible smoothed		Strictly admissible smoothed	
	Run time	Points (K)	Run time	Points (K)	Run time	Points (K)
Excess-of-loss	190	30	191	28	340	28
Lognormal	311	83	311	69	1112	70

TABLE 2. Effect of non-smoothing and strict admissibility on the average run time (in milliseconds) and number of interpolation points of the algorithm (over 100 repetitions with $n = 10^6$ and $\epsilon = 0.001$).

Overall, the authors recommend to generally use smoothing; and to enforce strict admissibility only if n is very small or if spectral risk measures, different of TVaR, are applied to the PWL approximation.

9. IMPLEMENTATION

We provide implementations of the algorithm in three programming languages: Python, C++ and R. The code is provided under the permissive and free *MIT license*; it can be obtained through the authors or at the following internet address:

<https://sites.google.com/site/philipparbenz/home/pwl-approximation>

Example 9.1. *The Python implementation of the PWL approximation algorithm takes two arguments: the sample (as a list or numpy array) as well as an accuracy parameter.*

LISTING 1. Minimal code example in Python

```
from compressor import PWLcompressor
Sample = [1, 1.6, 4.3, 4.6, 6, 7.1, 13, 13.4, 16, 18.8]
PWLapprox = PWLcompressor(Sample, Accuracy=0.01)
```

The C++ code provides a statically typed and very fast implementation. Dependencies are limited to the C++ standard template library (STL).

Example 9.2. *The C++ implementation of the PWL approximation algorithm takes two arguments: the sample (as a vector of doubles) as well as an accuracy parameter.*

LISTING 2. Minimal code example in C++

```
#include <vector>
#include "compressor.h"
int main(){
    std::vector<double> Sample = {1,1.6,4.3,4.6,6,7.1,13,13.4,16,18.8};
    PWLCompressor Comp = PWLCompressor(Sample, 0.01); }

```

The R code is implemented in the form of a simple to use package.

Example 9.3. *The R implementation of the PWL approximation algorithm takes two arguments: the sample (as a vector) as well as an accuracy parameter.*

LISTING 3. Minimal code example in R

```
library (PWLCompressor)
Sample <- c(1, 1.6, 4.3, 4.6, 6, 7.1, 13, 13.4, 16, 18.8)
PWLapprox <- PWLCompressor (Sample, 0.01)

```

All three implementations have optional parameters allowing to enforce smoothing and strict admissibility.

10. COMPARISON TO ALTERNATIVE APPROACHES

In this section, we justify the selection of the class of PWL distributions as approximating distributions. Furthermore, we compare the PWL approximation algorithm to other approaches commonly applied when information on univariate sample distributions should be preserved or transferred between actuarial systems.

The family of PWL distributions possesses several desirable properties:

- PWL distributions allow to efficiently calculate quantities such as cdf, quantiles (inverse cdf), moments (e.g., expectation) and risk measures (e.g., tail value-at-risk).
- PWL distributions can approximate all kind of sample distribution shapes and characteristics, such as continuous and discrete distributions. Heavy tailed distributions can be approximated with arbitrary precision.
- PWL distributions are straightforward to visualise, simple to export to spreadsheet applications, and are easily explained to non-technical stakeholders.
- PWL distributions are parsimonious, in the sense that the number of parameters (i.e., the size of the parameter vectors \mathbf{x} and \mathbf{y}) scales with the complexity of the shape of the sample distribution. Generally, the number of PWL parameters is massively lower

than the size of the sample to be approximated. For a degenerate constant sample, only two interpolation points are needed. Hence, PWL distributions are efficient with respect to memory and network bandwidth resources.

Other often used methods for non-parametric approximations also employ piecewise defined functions, based on higher degree polynomials. We restrict ourselves to linear segments since polynomials would make calculations of statistics much more difficult, provide only minor additional accuracy, and would render the algorithm significantly less efficient.

We now compare the PWL approximation algorithm with other approaches. This comparison is made with respect to the following desirable qualitative characteristics:

- A method is *shape preserving* if it allows to preserve important quantitative information on the empirical distribution shape, e.g., on smoothness (continuous vs discrete vs mixed), or tail decay (heavy tail or exponential decay).
- A method is *IT system independent* if it is not tied to a certain state of the IT system, such as the version of the operating system, software or hardware.
- A method is *memory and bandwidth efficient* if storing and transmitting the retained information on the sample distribution is efficient.
- A method is *mean and risk preserving* if the approximation has the same mean as the sample and an explicit error bound can be given for a set of risk measures.

As an illustration for the importance of the shape preserving property, consider that a normal, binomial, and Pareto distribution may have the same mean and standard deviation, but represent distributions with entirely different characteristics. With the introduction of the Solvency II, SST, and Basel 3 regimes, it has become crucial that risks are not only understood in terms of key statistics, but also in terms of the entire distribution.

The easiest and historically most frequently used solution is to **store only key statistics**, such as mean and standard deviation of the sample, and discard the sample itself. This approach loses most information about the shape of the distribution, but it is of course memory efficient and IT system independent. Any information on risk measures not explicitly stored is lost.

Storing the full sample allows to retain full information on the sample distribution. However, this approach is not memory and bandwidth efficient at all. The computing power of ordinary computers is already so large that for many stochastic models, millions of simulations can be generated in few seconds. Storing a sample of size 10 million requires around

640 Megabytes of memory. Hence, repeatedly storing all simulation results quickly exhausts memory and bandwidth resources, even in modern computing environments.

A third alternative is to **store software and random number generator** (RNG) underlying the simulation. However, even if this is simple in the short term, it is hardly practical in the long term: software or IT systems might be changed, upgraded or switched off; random number generators might be replaced; licenses for commercial closed-source software may expire or become unavailable. Furthermore, it leads to the necessity to run and maintain multiple versions of old systems and code in parallel.

Another alternative approach is to **store a predetermined set of quantiles**. It is possible to preserve the shape of the empirical distribution under this approach but unlike the algorithm presented, the mean will only be close, but not exactly the same. The magnitude of the difference between the empirical distribution and the approximation in terms of some risk measure of interest will be unknown for the quantiles that are not predetermined. Additionally, if this approach is implemented in a production environment dealing with thousands of distributions coming from different underlying models, the amount of stored data will be the same in all cases. On the contrary, the PWL approximations have a number of interpolation points which is adapted to the complexity of the shape of the empirical distribution. PWL approximation stores more points for the parts where the empirical distribution is convex/concave and only few points in the parts where the distribution is constant or linear.

Table 3 provides a comparative overview on the qualitative properties of the different presented approaches. Only PWL approximation possesses all four desirable properties.

	Shape preserving	IT system independent	Memory and bandwidth efficient	Mean and risk preserving
Store key statistics	No	Yes	Yes	No
Store full sample	Yes	Yes	No	Yes
Store RNG seed	Yes	No	No	Yes
Store fixed quantiles	Yes	Yes	(Yes)	No
PWL approximation	Yes	Yes	Yes	Yes

TABLE 3. Qualitative properties of the different presented approaches which can be used to retain information on a univariate sample distribution.

11. CONCLUSION

This paper introduces an algorithm that computes a piecewise linear approximation for a given empirical univariate sample distribution. The algorithm is designed in such a way that the empirical distribution and its approximation have the same mean, and that the relative error for a set of spectral risk measures is uniformly bounded. The algorithm can be applied to distributions with any kind of shape: continuous, discrete or mixed. It compresses the information included in samples of size of millions to piecewise linear distributions consisting of only a couple of hundred interpolation points. Therefore, this compression allows to significantly reduce the amount of memory required to store large empirical sample distributions, which may have to be kept for several years. An efficient, free and open source software implementation including a full set of test cases is provided.

In future research, we propose to investigate whether the error bound imposed on TVaR^Δ also implies error bounds for cdf, moments, quantile function, or other risk measures. It could be interesting to extend the algorithm to another distance measure that is not insurance inspired, for instance to one of the distance measures described in Gibbs and Su (2002).

ACKNOWLEDGEMENTS

As SCOR Fellow, William Guevara-Alarcón thanks SCOR for financial support. The authors thank Hansjörg Albrecher, Mathieu Cambou, Christoph Hummel, George Kottanattu, Eric Dal Moro, Andres Rosero, and Johanna Ziegel for their helpful comments. The authors would like to thank the anonymous referee, a co-editor and the editor for their valuable comments which improved the paper.

REFERENCES

- Acerbi, C. (2002). Spectral measures of risk: A coherent representation of subjective risk aversion. *Journal of Banking & Finance*, 26(7):1505–1518.
- De Brauw Blackstone Westbroek (2014). *European document retention guide: a comparative view across 16 countries to help you better understand legal requirements and records management best practice*.
- Eidgenössische Finanzmarktaufsicht (FINMA) (2008). *Rundschreiben 2008/44 Schweizer Solvenztest (SST)*.

- European Parliament and Council of the European Union (EP-CEU) (2009). *Directive 2009/138/EC of the European Parliament and of the Council of 25 November 2009 on the taking-up and pursuit of the business of Insurance and Reinsurance (Solvency II)*.
- Gibbs, A. L. and Su, F. E. (2002). On Choosing and Bounding Probability Metrics. *International Statistical Review*, 70(3):419–435.
- Gzyl, H. and Mayoral, S. (2006). On a relationship between distorted and spectral risk measures. *MPRA paper Munich University Library, Germany*, (916).
- Hakimi, S. L. and Schmeichel, E. F. (1991). Fitting polygonal functions to a set of points in the plane. *CVGIP: Graphical Models and Image Processing*, 53(2):132–136.
- Hummel, C. (2005). Specification for Tunnel Compressing. *Converium Internal Document*.
- International Accounting Standards Board (IASB) (2004). *International Financial Reporting Standard (IFRS) 4 Insurance Contracts*.
- Nguyen, T. B. and Oommen, B. J. (1997). Moment-preserving piecewise linear approximations of signals and images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(1):84–91.
- Qiu, P. (2003). A jump-preserving curve fitting procedure based on local piecewise-linear kernel estimation. *Journal of Nonparametric Statistics*, 15(4-5):437–453.
- Rockafellar, R. T., Uryasev, S., and Zabarankin, M. (2006). Generalized deviations in risk analysis. *Finance and Stochastics*, 10(1):51–74.

APPENDIX A. PARAMETERS FOR ADMISSIBLE APPROXIMATIONS

Theorem A.1. *Suppose $G = \text{PWL}(\mathbf{z}, \boldsymbol{\mu}, \boldsymbol{\delta})$ and F are TVaR equivalent and sample compatible. Then, for some fixed s ,*

$$\max_{\alpha \in \{z_s + \frac{1}{n}, z_s + \frac{2}{n}, \dots, z_{s+1}\}} \frac{|\text{TVaR}_\alpha^\Delta(G) - \text{TVaR}_\alpha^\Delta(F)|}{\text{TVaR}_\alpha^\Delta(F)} \leq \epsilon$$

is satisfied if and only if $a_k \leq \delta_s b_k \leq c_k$, for $k = nz_s + 1, nz_s + 2, \dots, nz_{s+1}$; where:

$$\begin{aligned} a_k &= -\epsilon \left(k\mu - \sum_{i=1}^k X_{(i)} \right) + \left(\sum_{i=nz_s+1}^k X_{(i)} \right) - (k - nz_s)\mu_s, \\ b_k &= \frac{(k - nz_s)(k - nz_{s+1})}{n(z_{s+1} - z_s)}, \\ c_k &= +\epsilon \left(k\mu - \sum_{i=1}^k X_{(i)} \right) + \left(\sum_{i=nz_s+1}^k X_{(i)} \right) - (k - nz_s)\mu_s. \end{aligned}$$

Proof. By using Theorem 3.11 and multiplying with $n\alpha$, we get

$$\frac{|\text{TVaR}_\alpha^\Delta(G) - \text{TVaR}_\alpha^\Delta(F)|}{\text{TVaR}_\alpha^\Delta(F)} \leq \epsilon \iff n \left| \int_{z_s}^\alpha G^-(t) dt - \int_{z_s}^\alpha F^-(t) dt \right| \leq n\alpha\epsilon \text{TVaR}_\alpha^\Delta(F).$$

Using (5.2) and (3.1), we can calculate for $\alpha \in \{z_s + 1/n, z_s + 2/n, \dots, z_{s+1}\}$ that

$$\int_{z_s}^\alpha G^-(t) dt = (\alpha - z_s)\mu_s + \delta_s \frac{(\alpha - z_s)(\alpha - z_{s+1})}{z_{s+1} - z_s}, \quad \int_{z_s}^\alpha F^-(t) dt = \frac{1}{n} \sum_{i=nz_s+1}^{\alpha n} X_{(i)}.$$

Combining the above, using that $|A| \leq B$ is equivalent to $-B \leq A \leq B$, and reshuffling, we get

$$\begin{aligned} & -n\alpha\epsilon \text{TVaR}_\alpha^\Delta(F) + n \int_{z_s}^\alpha F^-(t) dt - n(\alpha - z_s)\mu_s \\ & \leq \delta_s n \frac{(\alpha - z_s)(\alpha - z_{s+1})}{z_{s+1} - z_s} \leq \\ & + n\alpha\epsilon \text{TVaR}_\alpha^\Delta(F) + n \int_{z_s}^\alpha F^-(t) dt - n(\alpha - z_s)\mu_s. \end{aligned} \tag{A.1}$$

We obtain the desired result by plugging in $\alpha = k/n$, where $\alpha \in \{z_s + \frac{1}{n}, z_s + \frac{2}{n}, \dots, z_{s+1}\}$. \square

Theorem A.2. *Given three sequences $a_i, b_i, c_i \in \mathbb{R}$ for $i \in I$, where I is a finite subset of \mathbb{N} . Suppose at least one of the b_i is non-zero. Then there exists values $\delta \in \mathbb{R}$ such that the inequalities*

$$a_i \leq \delta b_i \leq c_i, \quad i \in I, \tag{A.2}$$

are all satisfied if and only if the following conditions are met:

$$\min_{i:b_i=0} a_i \leq 0, \quad \max_{i:b_i=0} c_i \geq 0, \quad \delta^{\min} \leq \delta^{\max},$$

where

$$\delta^{\min} = \max \left\{ \max_{i:b_i>0} \frac{a_i}{b_i}, \max_{i:b_i<0} \frac{c_i}{b_i} \right\}, \quad \delta^{\max} = \min \left\{ \min_{i:b_i>0} \frac{c_i}{b_i}, \min_{i:b_i<0} \frac{a_i}{b_i} \right\}.$$

In case these conditions hold, the admissible set of δ for (A.2) is given by the interval $\delta \in [\delta^{\min}, \delta^{\max}] \subset \mathbb{R}$.

Proof. For i with $b_i = 0$, the inequality (A.2) is equivalent to $a_i \leq 0 \leq c_i$. For $b_i > 0$, (A.2) is equivalent to $a_i/b_i \leq \delta \leq c_i/b_i$. For $b_i < 0$, (A.2) is equivalent to $c_i/b_i \leq \delta \leq a_i/b_i$. Combining and taking the maximum (minimum, respectively) yields the desired result. \square

Proof of Theorem 5.5: Theorem A.1 shows that (4.2) is equivalent to a set of inequalities in the form of (A.2). Using the result of Theorem A.2, we set $\Delta_s = \emptyset$ if $\delta^{\min} > \delta^{\max}$ and $\Delta_s = [\delta^{\min}, \delta^{\max}]$ if $\delta^{\min} \leq \delta^{\max}$. \square

APPENDIX B. TESTING STRICT ADMISSIBILITY

In order to obtain a proof of Theorem 6.7, we first rewrite the inequality (6.1) as a set of n functional inequalities $L_k(x) \leq M_k(x) \leq R_k(x)$ such that L_k , M_k , and R_k are simple functions which are numerically tractable. Namely, L_k and R_k are linear functions whereas M_k is quadratic and convex. Note that the deduction of (A.1) from (6.1) did not make any assumption on α . Hence, (6.1) holds if the inequality (A.1) holds for all $\alpha \in (0, 1)$.

Theorem B.1. *Suppose $G = \text{PWL}(\mathbf{z}, \boldsymbol{\mu}, \boldsymbol{\delta})$ and F are TVaR equivalent and sample compatible. Then, G is a strictly admissible approximation of F if and only if*

$$|X_{(1)} - x_1| \leq \epsilon (\mathbb{E}(F) - X_{(1)}), \quad (\text{B.1})$$

$$|X_{(n)} - x_K| \leq \epsilon (\mathbb{E}(F) - X_{(n)}), \quad (\text{B.2})$$

$$L_k(x) \leq M_k(x) \leq R_k(x), \quad (\text{B.3})$$

for all $k = 1, \dots, n$ and $x \in [0, 1]$. The values $X_{(1)}$ and $X_{(n)}$ are the minimum and maximum of the sample from which F is derived. The functions $L_k, M_k, R_k : [0, 1] \mapsto \mathbb{R}$ are given by

$$L_k(x) = L_k^0 + x(L_k^1 - L_k^0),$$

$$M_k(x) = M_k^0 + x(M_k^1 - M_k^0) + M_k^2 x(x-1),$$

$$R_k(x) = R_k^0 + x(R_k^1 - R_k^0),$$

with $L_k^0 = a_{k-1}$, $L_k^1 = a_k$, $M_k^0 = \delta_s b_{k-1}$, $M_k^1 = \delta_s b_k$, $M_k^2 = \delta_s / (nz_{s+1} - nz_s)$, $R_k^0 = c_{k-1}$, $R_k^1 = c_k$, where a_k, b_k , and c_k are defined as in Theorem A.1.

Proof. The conditions (B.1) and (B.2) appear from the following limit cases:

$$\lim_{\alpha \rightarrow 0} \frac{|\text{TVaR}_\alpha^\Delta(G) - \text{TVaR}_\alpha^\Delta(F)|}{\text{TVaR}_\alpha^\Delta(F)} = \frac{|X_{(1)} - x_1|}{\mathbb{E}(F) - X_{(1)}} \leq \epsilon$$

$$\lim_{\alpha \rightarrow 1} \frac{|\text{TVaR}_\alpha^{\Delta up}(G) - \text{TVaR}_\alpha^{\Delta up}(F)|}{\text{TVaR}_\alpha^{\Delta up}(F)} = \frac{|X_{(n)} - x_K|}{\mathbb{E}(F) - X_{(n)}} \leq \epsilon$$

The proof of (B.3) is based on the deduction of (A.1) from (6.1). With the analogous arguments, we can see that (A.1) for some fixed α is equivalent to $\tilde{L}(\alpha) \leq \tilde{M}(\alpha) \leq \tilde{R}(\alpha)$, where

$$\tilde{L}(\alpha) = -n\alpha\epsilon \text{TVaR}_\alpha^\Delta(F) + n \int_{z_s}^\alpha F^-(t) dt - n \int_{z_s}^\alpha \mu_s dt,$$

$$\tilde{M}(\alpha) = n \int_{z_s}^\alpha G^-(t) - \mu_s dt,$$

$$\tilde{R}(\alpha) = +n\alpha\epsilon \text{TVaR}_\alpha^\Delta(F) + n \int_{z_s}^\alpha F^-(t) dt - n \int_{z_s}^\alpha \mu_s dt,$$

with s chosen such that $z_s \leq \alpha \leq z_{s+1}$. Note that $\int_{z_s}^{\alpha} \mu_s dt$ and $\int_{z_s}^{\alpha} (G^-(t) - \mu_s) dt$ can be calculated analytically, see (A.1). Finally, we define

$$L_k(x) = \tilde{L}((k-1+x)/n), \quad M_k(x) = \tilde{M}((k-1+x)/n), \quad R_k(x) = \tilde{R}((k-1+x)/n),$$

for $k = 1, \dots, n$ and $x \in [0, 1]$. Recall that F is piecewise constant and $G^-(t)$ is linear on $(z_s, z_{s+1}]$. Therefore, $L_k(x)$ and $R_k(x)$ are linear, whereas $M_k(x)$ is quadratic. A close inspection of the integrals reveals that the boundary points $L_k(0)$, $M_k(0)$, $R_k(0)$ and $L_k(1)$, $M_k(1)$, $R_k(1)$ can easily be represented through the a_k , b_k , and c_k as defined Theorem A.1. \square

The set of inequalities in (B.3) can efficiently be solved through the following Theorem.

Theorem B.2. *Given three functions $L, M, R : [0, 1] \rightarrow \mathbb{R}$*

$$L(x) = L^0 + x(L^1 - L^0),$$

$$M(x) = M^0 + x(M^1 - M^0) + M^2 x(x-1),$$

$$R(x) = R^0 + x(R^1 - R^0),$$

where $L^0, L^1, M^0, M^1, M^2, R^0, R^1 \in \mathbb{R}$. Suppose the parameters satisfy the following inequalities: $L^0 \leq M^0 \leq R^0$, $L^1 \leq M^1 \leq R^1$, and $M^2 \geq 0$. Then there exists $x \in [0, 1]$ such that the inequality $L(x) \leq M(x) \leq R(x)$ is **not** satisfied if and only if the following three conditions are all satisfied:

$$\min\{M^0 - L^0, M^1 - L^1\} < \frac{1}{4}M^2, \tag{B.4}$$

$$|(M^1 - M^0) - (L^1 - L^0)| < M^2, \tag{B.5}$$

$$M(x^*) - L(x^*) < 0,$$

where $x^* = 1/2 - ((M^1 - M^0) - (L^1 - L^0)) / (2M^2)$.

Proof. The inequality $M(x) \leq R(x)$ holds since we assumed $M^2 \geq 0$ (implying that $M(x)$ is convex) and $M^0 \leq R^0$ as well as $M^1 \leq R^1$.

Proving the second part of the inequality, $L(x) \leq M(x)$, is less straight forward. Define $f : [0, 1] \rightarrow \mathbb{R}$ as $f(x) = M(x) - L(x)$. The inequality (B.4) is easy to deduce: We have

$$\begin{aligned} \min_{0 \leq x \leq 1} f(x) &\geq \min_{0 \leq x \leq 1} (M^0 - L^0 + x(M^1 - M^0 - L^1 + L^0)) + \min_{0 \leq x \leq 1} M^2 x(x-1) \\ &= \min\{M^0 - L^0, M^1 - L^1\} - 1/4M^2. \end{aligned}$$

Note that f is a convex quadratic function with $f(0) = M^0 - L^0$, $f(1) = M^1 - L^1$, and $d^2/dx^2 f(x) = 2M^2$. This implies that the minimum of f is located at 0 or 1 if $x^* \notin [0, 1]$. In case $x^* \in [0, 1]$, which is equivalent to (B.5), the minimum of f is at x^* , and $f(x^*) = M(x^*) - L(x^*)$. \square

Proof of Theorem 6.7: Theorem B.1 and Theorem B.2 allow to check whether an admissible and sample compatible approximation is also strictly admissible. Doing so requires $O(n)$ operations, since there are n functions L_k , M_k and R_k . Checking one triple of these functions through Theorem B.2 requires $O(1)$ operations, which leads to a total complexity of $O(n)$. \square

APPENDIX C. PROOF OF ERROR REDUCTION RATE

Lemma C.1. Let F_{pol} be a distribution such that $F_{pol}^-(t) = a + bt + ct^2$ for $t \in (z_s, z_{s+1}]$. Then,

$$\mathcal{D}(F_{pol}, [z_s, z_{s+1}], \alpha) = \frac{|c|}{6} \left| -2\alpha^3 + 3\alpha^2(z_{s+1} + z_s) - \alpha((z_{s+1} + z_s)^2 + 2z_s z_{s+1}) + z_s z_{s+1}(z_{s+1} + z_s) \right|$$

Proof. In this case μ_s is given by

$$\mu_s = \frac{1}{z_{s+1} - z_s} \int_{z_s}^{z_{s+1}} F_{pol}^-(t) dt = \frac{1}{z_{s+1} - z_s} \left(a(z_{s+1} - z_s) + \frac{b(z_{s+1}^2 - z_s^2)}{2} + \frac{c(z_{s+1}^3 - z_s^3)}{3} \right).$$

Analogous to Theorem 5.7, we obtain

$$\frac{d}{d\delta_s} D(\delta_s) = -2 \left(\frac{2}{z_{s+1} - z_s} \int_{z_s}^{z_{s+1}} F_{pol}^-(t) t dt - \mu_s(z_{s+1} + z_s) \right) + 2\delta_s \left(\frac{z_{s+1} - z_s}{3} \right).$$

Setting the above to zero yields

$$\delta_s^{reg} = \frac{6}{(z_{s+1} - z_s)^2} \left(\frac{(a - \mu_s)(z_{s+1}^2 - z_s^2)}{2} + \frac{b(z_{s+1}^3 - z_s^3)}{3} + \frac{c(z_{s+1}^4 - z_s^4)}{4} \right).$$

Plugging the above equations for μ_s and δ_s^{reg} into

$$\mathcal{D}(F_{pol}, [z_s, z_{s+1}], \alpha) = \left| \int_{z_s}^{\alpha} \left(\mu_s + \delta_s \left(2 \frac{t - z_s}{z_{s+1} - z_s} - 1 \right) \right) dt - \int_{z_s}^{\alpha} (a + bt + ct^2) dt \right|$$

leads to the desired result. \square

Lemma C.2. Let F_{pol} be a distribution such that $F_{pol}^-(t) = a + bt + ct^2$ for $t \in (z_s, z_{s+1}]$. Then,

$$\mathcal{D}(F_{pol}, [z_s, z_{s+1}]) = \frac{|c|\sqrt{3}(z_{s+1} - z_s)^3}{108}$$

Proof. Using the result of Lemma C.1, we can deduce the derivative

$$\frac{6}{|c|} \frac{d}{d\alpha} \mathcal{D}(F_{pol}, [z_s, z_{s+1}], \alpha) = -6\alpha^2 + 6\alpha(z_{s+1} + z_s) - ((z_{s+1} + z_s)^2 + 2z_s z_{s+1}).$$

Setting the above expression to zero yields

$$\alpha^* = \frac{z_{s+1} + z_s}{2} \pm \frac{\sqrt{3}(z_{s+1} - z_s)}{6}.$$

We obtain the desired result by plugging in $\mathcal{D}(F_{pol}, [z_s, z_{s+1}]) = \mathcal{D}(F_{pol}, [z_s, z_{s+1}], \alpha^*)$. \square

Proof of Theorem 7.2: Immediate consequence of Lemma C.2. \square