Can we use kernel smoothing to estimate Value at Risk and Tail Value at Risk?

Ramon Alemany, Catalina Bolancé and Montserrat Guillén
Department of Econometrics, Riskcenter-IREA, University of Barcelona,
Av. Diagonal, 690, 08034 Barcelona, Spain

February 27, 2013

Abstract

In this paper we analyse nonparametric methods to estimate risk measures in loss distributions. We study kernel estimation for Value-at-Risk and Tail Value-at-Risk based on transformation of the original data. The proposed method consists of a double transformation kernel estimation. We show that a suitable bandwidth selection criterion has a direct expression for the optimal smoothing parameter. The bandwidth can accommodate to the given extreme quantile level and expected shortfall. The procedure is useful for large data sets frequently available in finance and insurance.

1 Introduction

Risk measures and their mathematical properties have been widely studied in the literature (see, for instance, the books by McNeil et al. (2005) and Jorion (2007) or articles such as Dhaene et al. (2006) among many others). Most of those contributions and applications in risk management usually assume a parametric distribution for the loss random variable[1]. Deviations from parametric hypothesis can be critical in the extremes and produce inaccurate results (see, Kupiec, 1995). Krätschmer and Zähle (2011) investigated the error made even when the normal approximation is plugged in a general distribution-invariant risk measure.

Our approach is nonparametric as in Peng et al. (2012), Cai and Wang (2008), Jones and Zitikis (2007). We propose a method to estimate extreme quantiles and expected shortfall

[1]Standard industry models as CreditRisk$^+$ are parametric. See, Fan and Gu (2003), and references therein for semiparametric models.

that is based on a nonparametric estimate of the cumulative distribution function with an optimal bandwidth at the desired quantile level. Eling (2012) recently used a similar benchmark nonparametric fit to describe claims severity distributions in property-liability insurance (see, Bolancé et al., 2012, for details) but the choice of the smoothing parameter needs further analysis. Besides, Eling (2012) was interested in the fit for the density of claims severity, not on risk measurement or quantiles[2]. We present the nonparametric estimation approach and focus on the bandwidth choice. We also carry out a simulation exercise.

A risk measure widely used to quantify the risk is the value-at-risk (VaR) with level α. It is defined as follows,

$$VaR_{\alpha}\left(X\right)=\inf\left\{x,F_{X}\left(x\right)\geq\alpha\right\}=F_{X}^{-1}\left(\alpha\right), \tag{1}$$

where X is a random variable with probability distribution function (pdf) f_X, and cumulative distribution function (cdf) F_X. Artzner et al. (1999) discussed other risk measures, but they stated that expected shortfall is preferred in practice due to its better properties, although value-at-risk is widely used in applications.

Tail value-at-risk with level α is defined as follows[3],

$$TVaR_{\alpha}\left(X\right)=E\left(X|X\geq VaR_{\alpha}\left(X\right)\right). \tag{2}$$

In this paper we propose a method to estimate the VaR_{α} and $TVaR_{\alpha}$ in extreme quantiles, based on transformed kernel estimation (TKE) of the cdf of a loss distribution. The proposed method consists of a double transformation kernel estimation (DTKE), and it works well for very extreme levels and a large sample size. It also improves quantile estimation compared to existing methods. An additional contribution is that we propose a simple expression for an optimal bandwidth parameter.

Some previous research has already studied nonparametric estimation of quantiles. On the one hand Azzalini (1981) suggested to estimate the cdf and then to obtain the quantile from its inverse function. On the other hand Harrell and Davis (1982) proposed an alternative quantile estimator, based in a weighted sum of sample observations. Later, Sheather and Marron (1990) analysed the existing kernel methods for quantile estimation and proposed a smoothing parameter. None of those contributions, however, focused on highly skewed or heavy tailed distributions, which most often appear in financial and insurance risk management.

[2]Eling (2012) worked with two empirical data sets. The first dataset is US indemnity losses and the second is comprised of Danish fire losses. His work indicated that the transformation kernel (Bolancé et al., 2003) is the best and second best approach when compared with the parametric distributions in terms of the log likelihood value in his applications. The transformation kernel approach performed extremely well there and confirmed the results presented by Bolancé et al. (2008a) for auto insurance.

[3]Note that there is some confusion in the literature, but TVaR is also referred to as expected shortfall. For detailed definitions, see Denuit et al. (2005).

Only a few contributions have studied kernel estimation for TVaR, besides their results are not satisfactory, as mentioned by Chen (2007). This author concludes that a simple empirical approach is better than nonparametric smoothing. Here we introduce a new transformed kernel estimation to tail value-at-risk that is not worse than an empirical distribution approach in some cases.

Recently, Swanepoel and Van Graan (2005) presented kernel estimation of a cdf using nonparametric transformation, i.e. a simple form of transformed kernel estimation. Instead, Bolancé et al. (2008b) used a parametric transformation, which provides good results in the estimation of conditional tail expectation. Alemany et al. (2013) propose an improved non-parametric procedure to estimate the VaR in finance and insurance applications and derive an optimal expression for the bandwidth parameter. Here, we extend the analysis to TVaR estimation.

2 Motivation and outline

Our motivation is found on the statistical assumptions underlying the random behaviour of loss distributions. In practice, calculating VaR_α and $TVaR_\alpha$ requires to assume a particular stochastic behaviour of losses. Assumptions have generally been based on three possible statistical principles: i) the empirical statistical distribution of the loss or some smoothed version, ii) assuming that the loss follows a Normal or Student t distribution and iii) some other alternative parametric approximations. Sample size is a key factor to determine the method to estimate the quantile. In order to use the empirical distribution function, a minimum sample size is required. The Normal approximation provides a straightforward expression for the most popular risk measures, although the loss may be far from having a Normal shape or even a Student t distribution. Alternatively, one should find a suitable heavy tailed parametric distribution to which the loss data should fit (see, for example, McNeil et al., 2005; Jorion, 2007; Bolancé et al., 2012). Extreme value theory can be used to locate the tail of the distribution (see, Reiss and Thomas, 1997; Hill, 1975; Guillén et al., 2011).

A principal difference between our transformed kernel estimation and the fit of a heavy tailed parametric loss distribution is that we use sample information to estimate the parameters of a initial parametric model and, later, we also use the sample information to correct this initial fit. The proposed method works when losses have heavy tailed distributions, it is easy to implement and it provides consistent results. It is very flexible, so it is comparable to the empirical distribution approach. The method proposed in this work smooths the shape of the empirical distribution and extrapolates its behaviour when dealing with extremes, where data are very scarce.

The results of our simulation study show that our double transformed kernel estimation method can be applied to very extreme risk measurement and is specially suitable when the sample size is large. This is useful when basic parametric densities provide a poor fit in the tail. In the transformed kernel approach, no parametric form is imposed on the loss distribution, but, most importantly, this method avoids defining where the tail of the loss distribution starts in order to apply extreme value theory.

We introduce kernel estimation notation and basic nonparametric concepts in Section 3 in order to make the presentation self-contained. In Section 3 we present nonparametric estimation of a pdf and a cdf. We also describe nonparametric estimation of cdf in connection with estimation of value-at-risk and tail value-at-risk. Section 4 introduces transformation kernel estimation of a cdf and a new result on its asymptotic properties. Double transformation kernel estimation of a cdf and the selection of the smoothing parameter are studied in Section 5. Section 6 presents a simulation study where we can confirm the properties of the methods proposed in the previous sections. The most relevant conclusions and a discussion are given in the last section. Implementation tools in R are available from the authors and detailed hands-on examples of transformation kernel estimation can be found in Bolancé et al. (2012).

3 Nonparametric estimation of a cumulative distribution function

Let X be a random variable which represents a loss amount; its cdf is F_X. Let us assume that X_i $i = 1, ..., n$ denotes data observations from the loss random variable X. For instance, loss data may also arise from historical simulation or they may have been generated in a Monte Carlo analysis. A natural nonparametric method to estimate cdf is the empirical distribution,

$$\widehat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x), \tag{3}$$

where $I(\cdot) = 1$ if condition between parentheses is true. Then, the empirical estimate of value-at-risk is:

$$\widehat{VaR}_\alpha\left(X\right)_n = \inf\left\{x, \widehat{F}_n\left(x\right) \geq \alpha\right\}. \tag{4}$$

Estimation of the empirical distribution is very simple, but it cannot extrapolate beyond the maximum observed data point. This is especially troublesome if the sample is not too large, and one may suspect that the probability of a loss larger than the maximum observed loss in the data sample is not zero.

The empirical estimator of tail value-at-risk is:

$$\widehat{TVaR}_{\alpha}\left(X\right)_{n}=\frac{\sum_{i=1}^{n}X_{i}I\left(X_{i}>\widehat{VaR}_{\alpha}\left(X\right)_{n}\right)}{\sum_{i=1}^{n}I\left(X_{i}>\widehat{VaR}_{\alpha}\left(X\right)_{n}\right)},\tag{5}$$

where $I(\cdot)$ is a indicator function that takes value 1 if condition between parenthesis is true and $\widehat{VaR}_{\alpha}\left(X\right)_{n}$ is defined in 4.

Classical kernel estimation (CKE) of cdf F_X is obtained by integration of the classical kernel estimation of its pdf f_X. By means of a change of variable, the usual expression for the kernel estimator of a cdf is obtained:

$$\begin{aligned}\widehat{F}_X(x)&=\int_{-\infty}^{x}\widehat{f}_X(u)du=\int_{-\infty}^{x}\frac{1}{nb}\sum_{i=1}^{n}k\left(\frac{u-X_i}{b}\right)du\\&=\frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\frac{x-X_i}{b}}k\left(t\right)dt=\frac{1}{n}\sum_{i=1}^{n}K\left(\frac{x-X_i}{b}\right),\end{aligned}\tag{6}$$

where $k(\cdot)$ is a pdf, which is known as the kernel function. It is usually a symmetric pdf, but this does not imply that the final estimate of F_X is symmetric. Some examples of very common kernel functions are the Epanechnikov and the Gaussian kernel (see, Silverman, 1986). Parameter b is the *bandwidth* or the smoothing parameter. It controls the smoothness of the cdf estimate. The larger b is, the smoother the resulting cdf. Function $K(\cdot)$ is the cdf of $k(\cdot)$.

The classical kernel estimation of a cdf as defined in (6) is not much different to the expression of the well-known empirical distribution in (3). Indeed, in (6) one should replace $K\left(\frac{x-X_i}{b}\right)$ by $I(X_i \leq x)$ in order to obtain (3). The main difference between (3) and (6) is that the empirical cdf only uses data below x to obtain the point estimate of $F_X(x)$, while the classical kernel cdf estimator uses all the data above and below x. In other words, the empirical cdf gives more weight to the observations that are smaller than x than it does to the observations that are larger than x. Properties of kernel cdf estimator were analyzed by Reiss (1981) and Azzalini (1981).

In practice, to estimate VaR_α from $\widehat{F}_X\left(\cdot\right)$, we use the Newton-Raphson method to solve the equation:

$$\widehat{F}_X\left(x\right)=\alpha.\tag{7}$$

As an alternative, the kernel quantile estimator (KQE) is a classical method to estimate the VaR_α directly. For instance, Sheather and Marron (1990) proposed the following kernel quantile estimator:

$$KQ\left(\alpha\right)=\frac{\frac{1}{nb}\sum_{i=1}^{n}K\left(\frac{\frac{i-\frac{1}{2}}{n}-\alpha}{b}\right)X_{(i)}}{\frac{1}{nb}\sum_{i=1}^{n}K\left(\frac{\frac{i-\frac{1}{2}}{n}-\alpha}{b}\right)},\tag{8}$$

where $X_{(i)}$ is the i-th ordered observation.

As propose by Chen (2007), the kernel estimator of $TVaR_\alpha$ is based on the kernel estimation of the survival function, $\widehat{S}_X(x) = 1 - \widehat{F}_X(x)$, so:

$$\widehat{TVaR}_\alpha(X) = \frac{1}{n}\sum_{i=1}^{n} X_i \left(1 - K\left(\frac{\widehat{VaR}_\alpha(X) - X_i}{b}\right)\right), \tag{9}$$

There are ample evidence that kernel estimations of VaR and KQE improve the empirical approach (see, for example, Alemany et al. (2013)). However, kernel estimation of tail value-at risk does not improve the empirical approach (see, Chen (2007)).

4 Transformed Kernel Estimation

Let $T(\cdot)$ be a concave transformation where $Y = T(X)$ and $Y_i = T(X_i)$, $i = 1 \ldots n$ are the transformed observed losses. Then the kernel estimator of the transformed cumulative distribution function is:

$$\widehat{F}_Y(y) = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{y - Y_i}{b}\right) = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{T(x) - T(X_i)}{b}\right). \tag{10}$$

The transformated kernel estimation (TKE) of $F_X(x)$ is:

$$\widehat{F}_X(x) = \widehat{F}_{T(X)}(T(x)).$$

In expression (10), b and $K(\cdot)$ refer to the bandwidth and the cdf of the kernel function as in 6. In order to estimate VaR_α, the Newton-Raphson method solves the equation $\widehat{F}_{T(X)}(T(x)) = \alpha$ and once the result is obtained, the inverse of the transformation is applied.

To obtain the transformed kernel estimate, it is necessary to determine which is the transformation to use. Several authors have analyzed the transformation kernel estimate of the density function (see, Wand et al., 1991; Bolancé et al., 2003; Buch-Larsen et al., 2005; Pitt et al., 2012; Ruppert and Cline, 1994). However, few studies analyzed the transformed kernel estimate of the distribution function and the quantile (see, Swanepoel and Van Graan, 2005). In general, transformations are classified into parametric and nonparametric and, in turn, they may or may not correspond to a distribution function. The core objective of transformation $T(\cdot)$ is that the distribution of the transformed variable can be easily estimated using the classical kernel.

Alemany et al. (2013) have proved that the MSE of the transformed kernel estimation of a cdf based on (10) is asymptotically equal to:

$$
\begin{aligned}
& E\left\{\hat{F}_{T(X)}\left(T\left(x\right)\right)-F_{T(X)}\left(T\left(x\right)\right)\right\}^{2} \\
= & \tfrac{F_X(x)\left[1-F_X(x)\right]}{n}-\tfrac{1}{T'(x)}u\left(x\right)+\tfrac{1}{T'(x)}\left(1-\frac{\frac{f_X(x)}{f'_X(x)}}{\frac{T'(x)}{T''(x)}}\right)^{2}v\left(x\right)b^{4},
\end{aligned} \tag{11}
$$

where u and v are the same as before (see, Azzalini, 1981).

In practice, the use of a suitable transformation that reduces the variance at the expense of increasing the bias of the estimation (see Alemany et al. (2013)).

5 Double Transformed Kernel Estimation

The estimation that we describe in this section is based on the method proposed by Bolancé et al. (2008a) in the context of density functions. Here the objective is to estimate the cdf, from which we can obtain the transformed kernel estimator of VaR_α and $TVaR_\alpha$.

To obtain a smoothing parameter that is asymptotically optimal, it is sufficient to minimize A-MISE (Asymptotic mean square error):

$$
\frac{1}{4}b^4\int\left[f'_Y\left(y\right)\right]^2dy\left(\int t^2k\left(t\right)dt\right)^2-\frac{1}{n}b\int K\left(t\right)\left[1-K\left(t\right)\right]dt,
$$

where, given b and $k(\cdot)$, the value is minimum when functional $\int\left[f'_Y\left(y\right)\right]^2dy$ is minimum. Therefore, the proposed method is based on the transformation of the variable in order to achieve a distribution that minimizes the previous expression.

Terrell (1990) showed that the density of a $Beta\left(3,3\right)$ distribution defined on the domain $[-1,1]$ minimizes $\int\left[f'_Y\left(y\right)\right]^2dy$, in the set of all densities with known variance. Its pdf and cdf are, respectively:

$$
\begin{aligned}
m\left(x\right) &= \frac{15}{16}\left(1-x^2\right)^2, -1\leq x\leq 1, \\
M\left(x\right) &= \frac{3}{16}x^5-\frac{5}{8}x^3+\frac{15}{16}x+\frac{1}{2}.
\end{aligned} \tag{12}
$$

The double transformation kernel estimation method requires an initial transformation of the data $T\left(X_i\right)=Z_i$, where a transformed variable distribution that is close to a Uniform$(0,1)$ is obtained. Afterwards, the data are transformed again using the inverse of the distribution function of a $Beta\left(3,3\right)$, $M^{-1}\left(Z_i\right)=Y_i$. The resulting variable after the double transformation has a distribution that is close to a $Beta\,(3,3)$ (see, Bolancé, 2010). The double transformation kernel estimator (DTKE) of a cdf is:

$$\begin{aligned}\widehat{F}_X\,(x) &= \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{M^{-1}\,(T\,(x)) - M^{-1}\,(T\,(X_i))}{b}\right) \\ &= \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{y - Y_i}{b}\right). \end{aligned} \tag{13}$$

The distribution of the double transformed variable is similar to a $Beta\,(3,3)$. So, its pdf and the cdf are defined in (12). It is possible to find the optimal value of the smoothing parameter precisely at the point where VaR_α is located , i.e. the optimal bandwidth is found at the quantile of the $Beta\,(3,3)$ distribution, from the following expression:

$$b_{T(x)}^{Clas} = \left(\frac{u\,(T\,(x))}{4v\,(T\,(x))}\right)^{\frac{1}{3}} n^{-\frac{1}{3}}, \tag{14}$$

where

$$u\,(T\,(x)) = m\,(y)\left(1 - \int_{-1}^{1} K^2\,(t)\,dt\right),$$

and

$$v\,(T\,(x)) = \left[\frac{1}{2}m'\,(y)\int_{-1}^{1} t^2 k\,(t)\,dt\right]^2.$$

For example, the optimal bandwidth for the $VaR_{0.99}(X)$ is:

$$b_{VaR_{0.99}}^{Clas} = \left(\frac{0.133\,90\frac{9}{35}}{\left[\frac{1}{5}\right]^2 1.249\,4}\right)^{\frac{1}{3}} n^{-\frac{1}{3}} = 0.883\,21 n^{-\frac{1}{3}}.$$

Finally, $VaR_\alpha(X)$ is estimated from:

$$\hat{F}_X\left(M^{-1}\left(T\left(\widehat{VaR}_\alpha(X)\right)\right)\right) = \alpha \tag{15}$$

The double transformed kernel estimation of $\widehat{TVaR}_\alpha$ is based on the double transformed kernel estimator of survival function, i.e.

$$\widehat{S}_X\,(x) = \widehat{S}_X\left(M^{-1}\,(T\,(x))\right) = 1 - \widehat{F}_X\left(M^{-1}\,(T\,(x))\right),$$

Table 1: Distributions in the simulation study

Distribution	$F_X(x)$	Parameters
Weibull	$1-e^{-x^{\gamma}}$	$\gamma=1.5$
LogNormal	$\int_{-\infty}^{\log x}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(t-\mu)^2}{2\sigma^2}}dt$	$(\mu,\sigma)=(0,0.5)$
Mixture Lognormal	$p\int_{-\infty}^{\log x}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(t-\mu)^2}{2\sigma^2}}dt$	$(p,\mu,\sigma,\lambda,\rho,c)=(0.7,0,1,1,1,-1)$
-Pareto	$+(1-p)\left(1-\left(\frac{x-c}{\lambda}\right)^{-\rho}\right)$	$(p,\mu,\sigma,\lambda,\rho,c)=(0.3,0,1,1,1,-1)$

so,

$$\widehat{TVaR}_{\alpha}(X)=\frac{1}{n}\sum_{i=1}^{n}X_i\left(1-K\left(\frac{M^{-1}\left(T\left(\widehat{VaR}_{\alpha}(X)\right)\right)-M^{-1}\left(T\left(X_i\right)\right)}{b}\right)\right), \tag{16}$$

where the bandwidth is recalculated using 14and x is replaced by $\widehat{VaR}_{\alpha}(X)$ from (15).

6 Simulation Study

We summarize the results of a simulation study, where compare the mean squared error (MSE) when estimating VaR of our proposed double transformed kernel estimation (DTKE) with the empirical estimation (Emp). Alemany et al. (2013) show that DTKE of VaR_{α} outperforms some classical methods. As shown by Chen (2007), the classical kernel estimation does not improve the empirical distribution method. We present a simulation study, including a wide range of distributions comparing DTKE and Emp.

We generated $2,000$ samples of size $n=500$ and $2,000$ samples of size $n=5,000$ from each distribution in Table 1 and in 2. We selected four distributions with positive skewness and different tail shapes: Lognormal, Weibull and two mixtures of Lognormal-Pareto.

For each sample of size $n=500$ we estimated the VaR_{α} and $TVaR_{\alpha}$, with $\alpha=0.95$ and $\alpha=0.995$. When the sample size is $n=5,000$, in addition, we estimated VaR_{α} with $\alpha=0.999$.

Using the $2,000$ replication estimates we estimated MSE for each method. To calculate MSE we used the theoretical value of VaR_{α} in Tables 3, 4 and 5 and the theoretical value of $TVaR_{\alpha}$ in Tables 6, 7 and 8.

Results shown in Tables 9 to 14 refer to the ratio between MSE of the DTKE and the MSE of the empirical method (Emp). Moreover, sub-index x indicates that we used a smoothing

Table 2: Distributions in the simulation study for sensitivity analysis

Distribution	$F_X(x)$	Parameters	
		Lower	Upper
Weibull	$1 - e^{-x^{\gamma}}$	$\gamma = 0.75$	$\gamma = 3$
LogNormal	$\int_{-\infty}^{\log x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$	$(\mu, \sigma) = (0, 0.25)$	$(\mu, \sigma) = (0, 1)$
Mixture Lognormal	$p \int_{-\infty}^{\log x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$	$(p, \rho) = (0.7, 0.9)$	$(p, \rho) = (0.7, 1.1)$
-Pareto*	$+(1-p)\left(1 - \left(\frac{x-c}{\lambda}\right)^{-\rho}\right)$	$(p, \rho) = (0.3, 0.9)$	$(p, \rho) = (0.3, 1.1)$

The remaining parameters are the same that those in Table 1.

Table 3: True VaR_α in the simulated distributions in Table 1

Distribution	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.999$
Weibull	2.0781	3.0392	3.6271
Lognormal	2.2760	3.6252	4.6885
Mixture Lognormal-Pareto$_{p=0.7}$	7.5744	59.1892	299.0013
Mixture Lognormal-Pareto$_{p=0.3}$	13.4079	139.0034	699.0001

parameter based on minimization of MSE for the corresponding α level. A ratio value smaller than one indicates that tye DTKE is preferred to the Emp.

Table 4: True VaR_α in the simulated distributions (lower parameters in Table 2)

Distribution	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.999$
Weibull	4.3185	9.2367	13.1558
Lognormal	1.5086	1.9040	2.1653
Mixture Lognormal-Pareto$_{p=0.7}$	8.6258	93.6051	564.4016
Mixture Lognormal-Pareto$_{p=0.3}$	18.0137	241.4306	1448.5061

Table 5: True VaR_α in the simulated distributions (upper parameters in Table 2)

Distribution	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.999$
Weibull	1.4416	1.7433	1.9045
Lognormal	5.1802	13.1422	21.9821
Mixture Lognormal-Pareto$_{p=0.7}$	6.8606	40.9029	177.6320
Mixture Lognormal-Pareto$_{p=0.3}$	10.5928	88.3539	384.8806

Table 6: True $TVaR_\alpha$ in the simulated distributions in Table 1

Distribution	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.999$
Weibull	13.814276	177.1484	900.99958
Lognormal	2.8586	4.2957	5.4341
Mixture Lognormal-Pareto$_{p=0.7}$	∞	∞	∞
Mixture Lognormal-Pareto$_{p=0.3}$	∞	∞	∞

Table 7: True $TVaR_\alpha$ in the simulated distributions (lower parameters in Table 2)

Distribution	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.999$
Weibull	13.9584	222.46119	1177.2399
Lognormal	1.6824	2.0663	2.3256
Mixture Lognormal-Pareto$_{p=0.7}$	∞	∞	∞
Mixture Lognormal-Pareto$_{p=0.3}$	∞	∞	∞

Table 8: True $TVaR_\alpha$ in the simulated distributions (upper parameters in Table 2)

Distribution	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.999$
Weibull	15.157129	177.24489	892.48832
Lognormal	8.5572	18.9710	30.1691
Mixture Lognormal-Pareto$_{p=0.7}$	52.6554	417.3973	1778.6291
Mixture Lognormal-Pareto$_{p=0.3}$	111.4529	948.6322	4109.9091

Table 9: Results of MSE of VaR_α for Weibull and Lognormal

	n=500		n=5000		
	Weibull				
Method	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.999$
Emp	1	1	1	1	1
DTKEx	0.92	1.25	0.97	0.98	0.89
	Lognormal				
	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.999$
Emp	1	1	1	1	1
DTKEx	0.92	1.24	0.95	0.97	0.83

Table 10: Results of MSE of VaR_α for Weibull and Lognormal (parameters lower than in Table 2)

	n=500		n=5000		
	Weibull				
Method	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.999$
Emp	1	1	1	1	1
DTKEx	0.98	1.55	0.99	1.05	0.95
	Lognormal				
	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.999$
Emp	1	1	1	1	1
DTKEx	0.94	1.17	0.95	0.96	0.82

Table 11: Results of MSE of VaR_α for Weibull and Lognormal (parameters upper than in Table 2)

	n=500		n=5000		
	Weibull				
Method	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.999$
Emp	1	1	1	1	1
DTKEx	0.95	1.26	0.98	1.00	0.88
	Lognormal				
	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.999$
Emp	1	1	1	1	1
DTKEx	0.97	1.37	0.96	0.99	0.84

Table 12: Results of MSE of VaR_α for a mixture of Lognormal-Pareto

	n=500		**n=5000**		
	70% Lognormal-30% Pareto				
Method	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.999$
Emp	1	1	1	1	1
DTKEx	0.95	0.79	0.96	0.83	0.54
	30% Lognormal-70% Pareto				
	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.999$
Emp	1	1	1	1	1
DTKEx	0.87	1.12	0.93	0.88	0.66

Table 13: Results of MSE of VaR_α for a mixture of Lognormal-Pareto (parameters lower than in Table 2)

	n=500		**n=5000**		
	70% Lognormal-30% Pareto				
Method	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.999$
Emp	1	1	1	1	1
DTKEx	0.92	0.74	0.96	0.83	0.78
	30% Lognormal-70% Pareto				
	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.999$
Emp	1	1	1	1	1
DTKEx	0.89	0.57	0.94	0.85	0.72

The MSE for the two mixture Lognormal-Pareto distributions shows the advantages of the transformed kernel estimation ofVaR_α when the distribution has heavy tail. The results in Tables 12, 13 and 14 show that the DTKE method improves the empirical cdf method, specially for a large sample size and an extreme quantile. For example in Table 12, for a 70% Lognormal - 30% Pareto, and sample size $n = 5,000$, DTKE reduces the MSE of Emp in the estimation of the $VaR_{0.999}$ by 46%, i.e. the ratio between the MSE of DTKE and the MSE of Emp method is 0.54. We also see that DTKE reduces the MSE for the estimation of $VaR_{0.995}$ by 17% compared to the empirical method. For a 70% Lognormal - 30%, the reduction is 44% and 12%, respectively for $\alpha = 0.999$ and $\alpha = 0.995$. It is important to note that our proposal allows to calculate the asymptotically optimal bandwidth $b_{T(x)}^{Clas}$ in expression (14) without assuming a value for $T(x)$, given that we can calculate this exactly from the $Beta\,(3,3)$ distribution in expression (12).

Table 14: Results of MSE of VaR_α for a mixture of Lognormal-Pareto (parameters upper than in Table 2)

	n=500		**n=5000**		
	70% Lognormal-30% Pareto				
Method	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.999$
Emp	1	1	1	1	1
DTKEx	0.97	0.70	0.95	0.84	0.63
	30% Lognormal-70% Pareto				
	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.999$
Emp	1	1	1	1	1
DTKEx	0.91	0.70	0.95	0.85	0.67

Table 15: Results of MSE of $TVaR_\alpha$ for Weibull and Lognormal

	n=500		**n=5000**		
	Weibull				
Method	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.999$
Emp	1.00	1.00	1.00	1.00	1.00
DTKEx	1.27	0.92	1.12	1.00	0.79
	Lognormal				
	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.999$
Emp	1.00	1.00	1.00	1.00	1.00
DTKEx	1.16	1.11	1.06	1.06	0.99

Results for $TVaR_\alpha$ for Weibull and Lognormal distribution are shown in Tables 15, 16, 17. For the mixtures of Lognormal-Pareto we need can only calculate the MSE for $TVaR_\alpha$ when the Pareto parameter is smaller than one. Therefore, in Table 18 we only show the ratio between MSE of the DTKE and the MSE of the empirical method (Emp) for a mixture Lognomal-Pareto with the Pareto parameter smaller than one. In Tables 19 and 20 we show the mean of the results for the estimation of $TVaR_\alpha$, so that in this case the best outcome is to have a very large value.

According to the simulation results and our theoretical approximations, we recommend to use a double transformation kernel estimation approximation with an optimal bandwidth to estimate VaR_α and $TVaR_\alpha$, for large databases and loss distributions that are heavy tailed.

Table 16: Results of MSE of $TVaR_\alpha$ for Weibull and Lognormal (parameters lower than in Table 2)

	n=500		**n=5000**		
	Weibull				
Method	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.999$
Emp	1.00	1.00	1.00	1.00	1.00
DTKEx	1.21	1.09	1.08	1.07	0.94
	Lognormal				
	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.999$
Emp	1.00	1.00	1.00	1.00	1.00
DTKEx	1.17	0.83	1.10	0.95	0.81

Table 17: Results of MSE of $TVaR_\alpha$ for Weibull and Lognormal (parameters upper than in Table 2)

	n=500		**n=5000**		
	Weibull				
Method	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.999$
Emp	1.00	1.00	1.00	1.00	1.00
DTKEx	1.17	0.61	1.09	1.06	1.67
	Lognormal				
	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.999$
Emp	1.00	1.00	1.00	1.00	1.00
DTKEx	1.05	1.21	1.02	1.05	1.04

Table 18: Results of MSE of $TVaR_\alpha$ for a mixture of Lognormal-Pareto (parameters upper than in Table 2)

	n=500		**n=5000**		
	70% Lognormal-30% Pareto				
Method	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.999$
Emp	1.00	1.00	1.00	1.00	1.00
DTKEx	1.00	1.35	1.00	1.00	1.00
	30% Lognormal - 70% Pareto				
	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.999$
Emp	1.00	1.00	1.00	1.00	1.00
DTKEx	1.00	1.42	1.00	1.00	1.00

Table 19: Mean of $TVaR_\alpha$ estimates for a mixture of Lognormal-Pareto

	n=500		**n=5000**		
	70% Lognormal-30% Pareto				
Method	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.999$
Emp	71.167185	472.400153	113.65721	989.078672	4436.4
DTKEx	70.721181	518.010988	113.56083	985.715346	4295.33
	30% Lognormal - 70% Pareto				
	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.999$
Emp	133.10523	850.241301	231.78149	2000.57	8822.78
DTKEx	132.11801	902.404185	231.57234	1989.97	8593.31

Table 20: Mean of $TVaR_\alpha$ estimates for a mixture of Lognormal-Pareto (parameters lower than in Table 2)

	n=500		**n=5000**		
	70% Lognormal-30% Pareto				
Method	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.999$
Emp	19298904	44491274.5	126963358	395458417	1216850728
DTKEx	37905378	379014093	259683249	2596832074	1.286E+10
	30% Lognormal - 70% Pareto				
	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.95$	$\alpha = 0.995$	$\alpha = 0.999$
Emp	3683615.6	11890814.7	359617512	1355925159	3100554859
DTKEx	10705656	106944643	1.179E+09	1.1791E+10	5.8954E+10

7 Conclusions

We have presented a method to estimate quantiles that is suitable when the loss is a random variable that is heavy tailed. The proposed double transformation kernel estimation does not depend on a parametric assumption for the random variable. Asymptotic properties have been proved, showing that when estimating extreme quantiles, the sample size needs to be large.

The proposed method is easily implemented and fast because the optimal smoothing parameter calculation is direct. Moreover, the proposed method is especially useful in many risk measurement settings because it does not require statistical distribution assumptions and can handle heavy tailed random variables. This is the case when analyzing some operational risk situations or in the analysis of severity distributions.

Our research provides a tractable nonparametric method that can be useful to avoid restrictive statistical hypothesis. We are working towards an iproved bandwidth selection method for nonparametric TVaR estimation.

Acknowledgements

We thank the Spanish Ministry of Science / FEDER support ECO2010-21787-C03-01 and the members and affiliates of Riskcenter at the University of Barcelona. Guillén acknowledges ICREA Academia.

References

Alemany, R., Bolancé, C., Guillén, M., 2013. A nonparametric approach to calculating value-at-risk. Insurance: Mathematics and Economics 52, 255–262.

Artzner, P., Delbaen, F., Eber, J., Heath, D., 1999. Coherent measures of risk. Mathematical Finance 9, 203–228.

Azzalini, A., 1981. A note on the estimation of a distribution function and quantiles by a kernel method. Biometrika 68, 326–328.

Bolancé, C., 2010. Optimal inverse beta(3,3) transformation in kernel density estimation. SORT-Statistics and Operations Research Transactions 34, 223–237.

Bolancé, C., Guillén, M., Gustafsson, J., Nielsen, J., 2012. Quantitative Operational Risk Models. Chapman & Hall/CRC Finance Series, London.

Bolancé, C., Guillén, M., Nielsen, J., 2003. Kernel density estimation of actuarial loss functions. Insurance: Mathematics and Economics 32, 19–36.

Bolancé, C., Guillén, M., Nielsen, J., 2008a. Inverse beta transformation in kernel density estimation. Statistics& Probability Letters 78, 1757–1764.

Bolancé, C., Guillén, M., Pelican, E., Vernic, R., 2008b. Skewed bivariate models and nonparametric estimation for cte risk measure. Insurance:Mathematics and Economics 43, 386–393.

Buch-Larsen, T., Guillén, M., Nielsen, J., Bolancé, C., 2005. Kernel density estimation for heavy-tailed distributions using the Champernowne transformation. Statistics 39, 503–518.

Cai, Z., Wang, X., 2008. Nonparametric estimation of conditional VaR and expected shortfall. Journal of Econometrics 147 (1), 120 – 130.

Chen, S., 2007. Nonparametric estimation of expected shortfall. Journal of Financial Econometrics 6 (1), 87 – 107.

Denuit, M., Dhaene, J., Goovaerts, M., Kass, R., 2005. Actuarial Theory for Dependent Risks: Measures, Orders and Models. John Wiley & Sons Ltd., New York.

Dhaene, J., Vanduffel, S., Tang, Q., Goovaerts, M., Kaas, R., Vyncke, D., 2006. Risk measures and comonotonicity: A review. Stochastic Models 22, 573–606.

Eling, M., 2012. Fitting insurance claims to skewed distributions: Are the skew-normal and skew-student good models? Insurance: Mathematics and Economics 51 (2), 239 – 248.

Fan, J., Gu, J., 2003. Semiparametric estimation of value-at-risk. Econometrics Journal 6, 261–290.

Guillén, M., Prieto, F., Sarabia, J., 2011. Modelling losses and locating the tail with the pareto positive stable distribution. Insurance: Mathematicsand Economics 49, 454–461.

Harrell, F., Davis, C., 1982. A new distribution-free quantile estimator. Biometrika 69, 635–640.

Hill, B., 1975. A simple general approach to inference about tail of a distribution. Annals of Statistics 3, 1163–1174.

Jones, B., Zitikis, R., 2007. Risk measures, distortion parameters, and their empirical estimation. Insurance: Mathematics and Economics 41 (2), 279 – 297.

Jorion, P., 2007. Value at Risk. McGraw-Hill, New York.

Krätschmer, V., Zähle, H., 2011. Sensitivity of risk measures with respect to the normal approximation of total claim distributions. Insurance: Mathematics and Economics 49 (3), 335 – 344.

Kupiec, P., 1995. Techniques for verifying the accuracy of risk measurement models. Journal of Derivatives 3 (2), 73–84.

McNeil, A., Frey, R., Embrechts, P., 2005. Quantitative Risk Management: Concepts, Techniques, and Tools. Princeton University Press, Princeton.

Peng, L., Qi, Y., Wang, R., Yang, J., 2012. Jackknife empirical likelihood method for some risk measures and related quantities. Insurance: Mathematics and Economics 51 (1), 142 – 150.

Pitt, D., Guillén, M., Bolancé, C., 2012. An introduction to parametric and non-parametric models for bivariate positive insurance claim severity distributions. Xarxa de Referncia en Economia Aplicada (XREAP). Working Papers XREAP2010-03.

Reiss, R.-D., 1981. Nonparametric estimation of smooth distribution functions. Scandinavian Journal of Statistics 8, 116–119.

Reiss, R.-D., Thomas, M., 1997. Statistical Analysis of Extreme Values from Insurance, Finance, Hydrology and Others Fields. Birkhäuser Verlag, Berlin.

Ruppert, D. R., Cline, D. B. H., 1994. Bias reduction in kernel density estimation by smoothed empirical transformation. Annals of Statistics 22, 185–210.

Sheather, S., Marron, J., 1990. Kernel quantile estimators. Journal of the American Statistical Association 85, 410–416.

Silverman, B., 1986. Density Estimation for Statistics and Data Analysis. Chapman & Hall/CRC Finance Series, London.

Swanepoel, J., Van Graan, F., 2005. A new kernel distribution function estimator based on a nonparametric transformation of the data. Scandinavian Journal of Statistics 32, 551–562.

Terrell, G., 1990. The maximal smoothing principle indensity estimation. Journal of the American Statistical Association 85, 270–277.

Wand, P., Marron, J., Ruppert, D., 1991. Transformations in density estimation. Journal of the American Statistical Association 86, 343–361.