Professor Ramon ALEMANY, PhD[1]
E-mail: ralemany@ub.edu
Dept. of Econometrics, Riskcenter - IREA, University of Barcelona
Professor Catalina BOLANCÉ, PhD[1]
E-mail: bolance@ub.edu
Dept. of Econometrics, Riskcenter - IREA, University of Barcelona
Professor Montserrat GUILLÉN, PhD[1]
E-mail: mguillen@ub.edu
Dept. of Econometrics, Riskcenter - IREA, University of Barcelona
Alemar E. PADILLA-BARRETO, PhD Candidate
E-mail: alemarpadillabarreto@ub.edu
Dept. of Econometrics, Riskcenter - IREA, University of Barcelona

COMBINING PARAMETRIC AND NON-PARAMETRIC METHODS TO COMPUTE VALUE-AT-RISK

Abstract. *We design a system for calculating the quantile of a random variable that allows us combining parametric and non-parametric estimation methods. This approach is applicable to evaluate the severity of potential losses from existing data records; therefore, it is useful in many areas of economics and risk evaluation. The procedure is based on an initial parametric model assumption and then a nonparametric correction is introduced. In addition, a second correction is proposed so that the value at risk estimator is asymptotically optimal. Our procedure allows smoothing the tail behavior of the empirical distribution. Due to the lack of sample information for extreme values, smoothness in the tail cannot be achieved if classical nonparametric estimators are used. We apply this method to a real problem in the area of motor insurance.*
Keywords*: quantile, nonparametric, loss models, extremes, risk evaluation*

[1] Support received from the Spanish Ministry of Science / FEDER ECO2010-21787-C03-01 is acknowledged. Guillén thanks ICREA Academia.

JEL Classification: C14

1. Introduction and motivation

Risk quantification is often carried out in two steps: first, a tolerance level is fixed and, second, value at risk is calculated. Value at risk is the amount that is exceeded by one individual loss with probability equal to the corresponding tolerance level. So this notion corresponds to the concept of quantile of a statistical distribution (for a general notion of risk see [16]).

In practice, data records from observed losses conform the basis for a statistical analysis that leads to risk quantification. However, risk managers need to rely on assumptions about the statistical behavior of these existing data. Three classical statistical approaches to estimating value at risk can be followed: i) the empirical statistical distribution of the loss or a smoothed version can be used, ii) a Normal or Student's t distribution can be assumed, or iii) another parametric model can be fitted (see [10] and [14]). Sample size is a key factor in determining the eventual method. To use the empirical distribution function, a minimum sample size is required. The Normal approximation provides a straightforward expression for the value at risk, but unfortunately many losses or their transformations (for example, logarithmic transformation) are far from having a Normal shape or even a Student's t distribution. Alternatively, a suitable parametric density to which the loss data should fit could be found. Note that the methods proposed by [15] for estimating the quantile are not suitable for highly asymmetric distributions as has been shown in [1]. A nonparametric approach, such as classical kernel estimation (CKE), smooths the shape of the empirical distribution and "extrapolates" its behavior when dealing with extremes. However, when the losses variable is right skewness the number of sample observations in the right tail of the distribution is scarce, for this reason the CKE cannot smooth the shape of the empirical distribution and, therefore, it cannot extrapolate the shape of the distribution above the maximum observed value in the sample. For this reason, we propose a two-step estimation procedure. First, we fit a parametric model. Second, we use a transformed kernel estimation (TKE) method, thus ensuring that the final result is asymptotically optimal and it is guaranteed that the shape of the right tail is extrapolated. The TKE method is based on a transformation of the original data so that the transformed data follow a distribution that can be estimated optimally with the CKE. The procedure is based on the ideas of stochastic simulation, i.e. we generate values from a Uniform(0,1) distribution using a parametric cumulative distribution function (cdf) and second we use the inverse of another cdf that allows us to implement the CKE.

In this paper we present a system to quantify risk and show that it is suitable to estimate extreme quantiles of a skewed distribution with heavy right tail. There are lots of applications in finance and economics that require the analysis of extreme values with skewed data and for which it is questionable to assume a particular statistical distribution parametric hypothesis. Figures 1 and 2 present how a risk quantification basic system should be implemented, either directly using a parametric or a non-parametric method (Figure 1) or combining both (Figure 2).

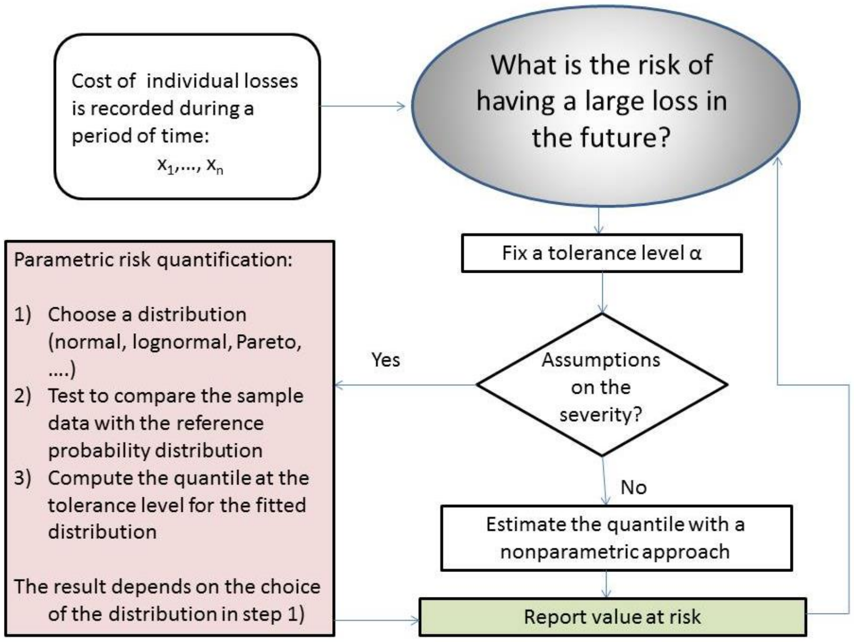

Figure 1: Description of a classical risk quantification procedure

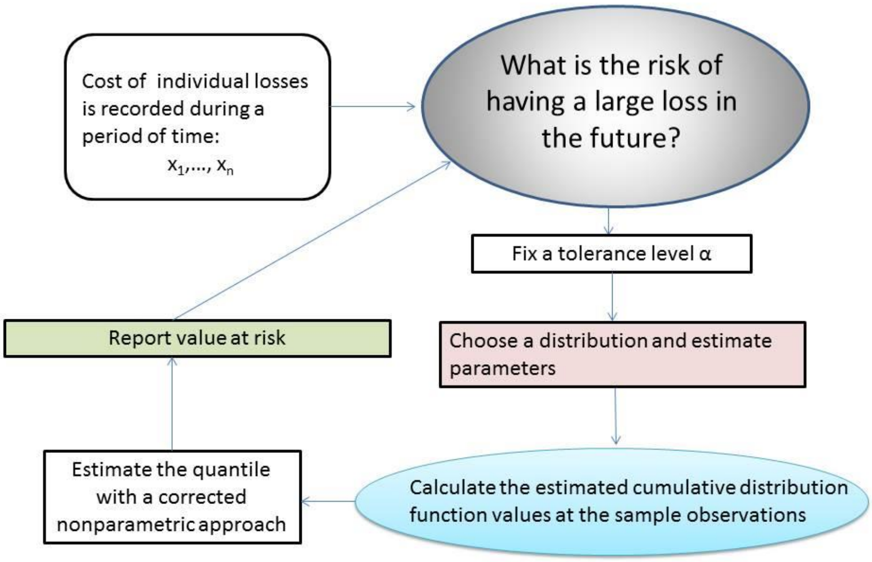

Figure 2: Proposed risk quantification system based on combining parametric and nonparametric methods

2. Notation

Let X be a random variable that represents a loss with cdf F_X. For instance, X may refer to the cost of an operational failure or an accident. The larger the cost is, the larger the severity of the loss event. The value at risk (VaR) is also known as the quantile of F_X and it is defined as:

$$VaR_\alpha(X) = inf\,\{x, F_X(x) \geq \alpha\} = F_X^{-1}(\alpha) \qquad (1)$$

where the confidence level α is a probability close to 1. So, we calculate a quantile in the right tail of the distribution. VaR_α is the cost level that an α proportion of losses does not exceed. So, a fraction of losses (1 − α) would exceed that level.

As we are interested in calculating VaR_α, we need an assumption regarding the stochastic behavior of losses or, as we suggest, we need to estimate the cdf F_X with no distributional assumptions.

3. Nonparametric quantile estimation

3.1. Empirical distribution

Estimation of VaR_α is straightforward when F_X in (1) is replaced by the empirical distribution:

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x), \qquad (2)$$

where I (·) is an indicator function which takes values 1 or 0. I (·) = 1 if the condition between parentheses is true, then

$$VaR_\alpha(X) = inf\{x, \hat{F}_n(x) \geq \alpha\}. \qquad (3)$$

The bias of the empirical distribution is zero and its variance is:

$$(F_X(x)[1 - F_X(x)])/n.$$

The empirical distribution is straightforward and it is an unbiased estimator of the cdf, but it cannot be extrapolated beyond the maximum observed data point. This is particularly troublesome if the sample is not too large and it is suspected that a loss larger than the maximum observed loss in the data sample might occur.

3.2. Classical Kernel Methods

Classical kernel estimation (CKE) of cdf F_X is obtained by integration of the classical kernel estimation of its probability density function (pdf) f_X, which is defined as follows:

$$\hat{f}_X(x) = \frac{1}{nb}\sum_{i=1}^{n} k\left(\frac{x-X_i}{b}\right), \tag{4}$$

where k is a pdf, which is known as the kernel function. Some examples of very common kernel functions are the Epanechnikov and the Gaussian kernel (see [17]). Parameter b is known as the bandwidth or smoothing parameter. It controls the smoothness of the cdf estimate. The larger b is, the smoother the resulting cdf. Function K is the cdf of k.

The usual expression for the kernel estimator of a cdf is easily obtained:

$$\hat{F}_X(x) = \int_{-\infty}^{x} \hat{f}_X(u)du = \int_{-\infty}^{x} \frac{1}{nb}\sum_{i=1}^{n} k\left(\frac{u-X_i}{b}\right) du \tag{5}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\frac{x-X_i}{b}} k(t)dt = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x-X_i}{b}\right).$$

To estimate VaR_α, the Newton-Raphson method is applied:

$$\hat{F}_X\left(\widehat{VaR}_\alpha(X)\right) = \alpha.$$

The classical kernel estimation of a cdf as defined in (5) bears many similarities to the expression of the well-known empirical distribution in (2). In (5) $K\left(\frac{x-X_i}{b}\right)$ should be replaced by $I(X_i \leq x)$ in order to obtain (2). The main difference between (2) and (5) 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, but it gives more weight to the observations that are smaller than x than it does to the observations that are greater than x. It has already been noted by [12] and [2] that, when $n \to \infty$, the mean squared error (MSE) of $\hat{F}_X(x)$can be approximated by:

$$E\left\{\hat{F}_X(x) - F_X(x)\right\}^2 \sim \frac{F_X(x)[1 - F_X(x)]}{n} - f_X(x)\frac{b}{n}\left(1 - \int_{-1}^{1} K^2(t)dt\right)$$

$$+ b^4\left(\frac{1}{2}(x) f_X'(x) \int t^2 k(t)dt\right)^2. \tag{6}$$

The resulting first two terms in (6) correspond to the asymptotic variance and the third term is the squared asymptotic bias. The kernel cdf estimator has less variance than that of the empirical distribution estimator, but it has some bias which tends to zero if the sample size is large.

The value for the smoothing parameter b that minimizes (6) asymptotically is:

$$b_x^* = \left(\frac{f_X(x)\int K(t)[1-K(t)]dt}{(f_X'(x)\int t^2 k(t)dt)^2}\right)^{\frac{1}{3}} n^{-\frac{1}{3}}, \tag{7}$$

where subindex x indicates that the smoothing parameter is optimal at this point. Moreover, Azzalini(1981) in [2] showed that (7) is also optimal when calculating the quantiles (i.e. VaR_α). However, in practice, calculating b_x^* is not simple because it depends on the true value of $f_X(x)$ and the quantile x is also unknown.

An alternative to the smoothing parameter in (7) is to use the rule-of-thumb proposed in [17], but since the objective in this paper is to estimate a quantile in the right tail of a distribution, [1] recommended calculating the bandwidth using a smoothing parameter that minimizes the weighted integrated squared error (WISE) asymptotically, i.e.:

$$WISE\{\hat{F}_X\} = E\left\{\int \left[F_X(x) - \hat{F}_X(x)\right]^2 x^2\, dx\right\}.$$

The value of b that minimizes WISE asymptotically is:

$$b^{**} = \left(\frac{\int f_X(x)x^2dx \int K(t)[1-K(t)]dt}{\int\left[f_X'(x)\right]^2 x^2 d_x (\int t^2 k(t)dt)^2}\right)^{\frac{1}{3}} n^{-\frac{1}{3}}. \tag{8}$$

and when replacing the theoretical true density f_X by the Normal pdf, the estimated smoothing parameter is:

$$\hat{b}^{**} = \sigma_X^{\frac{5}{3}} \left(\frac{8}{3}\right)^{\frac{1}{3}} n^{-\frac{1}{3}}. \tag{9}$$

Various methods to calculate b exist. For instance, cross-validation and plug-in methods (see, for example, [8]) are very usual. However, these methods require considerable computational effort in large data sets.

3.3. Transformed Kernel Estimation

Transformed kernel estimation (TKE) is better than classical kernel density estimation when estimating distributions with right skewness (see [4], [9], [5], [7] and [3]). Even if a large sample is available, the number of observations in the right tail are scarce and standard nonparametric estimates are inefficient to estimate an extreme quantile, such as when $\alpha = 0.995$.

Transformed kernel estimation is based on applying a transformation to the original variable so that the transformed variable has a symmetric distribution. Once classical kernel estimation is implemented on the transformed data, the inverse transformation returns to the original scale.

Let $T(\cdot)$ be a concave transformation, $Y = T(X)$ and $Y_i = T(X_i)$, $i = 1 \dots n$ are the transformed data, the transformed kernel estimation of the original cdf is:

$$\hat{F}_X(x) = \hat{F}_{T(X)}(T(x)) = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{T(x)-T(X_i)}{b}\right) \tag{10}$$

where b and K are as defined in Section 3.2.

When estimating VaR_α, the following equation needs to be solved to find $T(X)$:

$$\hat{F}_{T(X)}\left(T\left(\widehat{VaR}_\alpha(X)\right)\right) = \alpha$$

and then $\widehat{VaR}_\alpha$ is estimated using the inverse of the transformation on $T(X)$.

The smoothing parameter in the transformed kernel estimation of a cdf or quantile is the same as the smoothing parameter in the classical kernel estimation of cdf associated to the transformed variable. We can calculate the bandwidth in (9) if σ_X is replaced by σ_Y.

Many studies have proposed transformations in the context of the transformed kernel estimation of the pdf (see [19], [4], [9], [11], [13] and [3]). However only a few studies analyze the transformed kernel estimation of the cdf and quantile (see [1]). These transformations can be classified into those that are a cdf and those that do not correspond to a specific cdf. Moreover, non- parametric cdf transformations can also be considered. If $T(x)$ is a parametric cdf, then transformed kernel estimation in (10) can be interpreted as a parametric estimation with a nonparametric correction. In [1] it has been shown that the MSE of TKE is:

$$= E\left\{\hat{F}_X(x) - F_X(x)\right\}^2 \sim \frac{F_X(x)[1-F_X(x)]}{n} - \frac{1}{T'(x)} f_X(x)\frac{b}{n}\left(1 - \int_{-1}^{1} K^2(t)dt\right) + \frac{1}{T'(x)}\left(1 - \frac{\frac{f_X(x)}{f'_X(x)}}{\frac{T'(x)}{T''(x)}}\right)^2 \left[\frac{1}{2} f'_X(x) \int_{-1}^{1} t^2 k(t)dt\right]^2 b^4 \tag{11}$$

In expression (11) it can be seen that when $T(x)$ is the same as the theoretical cdf $F_X(x)$, then the bias of the transformed kernel estimation is zero. Moreover, if $T(x)$ is not the same as the theoretical cdf $F_X(x)$, then the TKE is a consistent estimator, given that $nb \to 0$ when $n \to \infty$. However, if we use a parametric model and this is not the same as the $F_X(x)$, then the parametric estimator is not consistent. The problem with TKE when T (x) is a cdf is that TKE is not useful to obtain $\hat{F}_{T(X)}(T(x))$.

The double transformed kernel estimation (DTKE) method for estimating the quantile was proposed by [1]. First, we estimate the parameters of a parametric distribution. In our case, we propose to choose among the following set of parametric models: exponential, Weibull, Lognormal, Champernowne distribution and its generalizations. We use the generalized Champernowne cdf[1] given that it provides the best fit to our real data set. Then, the data are transformed with a cdf function. Second, the transformed data are again transformed using the inverse of the cdf of a $Beta(3,3)$ distribution defined on the domain $[-1,1]$, this cdf and their pdf are:

$$M(x) = \frac{3}{16}x^5 - \frac{5}{8}x^3 + \frac{15}{16}x + \frac{1}{2}$$

$$m(x) = \frac{15}{16}(1-x^2)^2, -1 \leq x \leq 1$$

(see [1] for further details and [6] for computer codes in SAS and R). The double transformation approach is based on the fact that the cdf of a $Beta(3,3)$ can be estimated optimally using classical kernel estimation (see [18]). Given that double transformed data have a distribution that is close to the $Beta(3,3)$ distribution, an accurate optimal bandwidth for estimating VaR_α can be used. For example, based on expression (8) and replacing f_X by the pdf of the Beta(3,3), i.e. $f_X(x) = m(x)$, we obtain:

$$\hat{b}^{**} = \left(\frac{9}{7}\right)^{\frac{1}{3}} n^{-\frac{1}{3}}. \quad (12)$$

4. Data Study

We analyze a data set obtained from a Spanish insurance company that contains a sample of 5,122 automobile claim costs. This is a standard insurance loss data set with observations on the cost of accident losses, i.e. a large, heavy-

[1]A generalized Champernowne distribution has the following cdf:

$$T_x(x) = \frac{((x+c)^\gamma - c^\gamma)}{((x+c)^\gamma + (M+c)^\gamma - 2c^\gamma)},$$

where $c, \gamma, M > 0, x > -c$.

tailed sample containing many small values and a few large extremes. The sample represents 10% of all insured losses reported to the company's motor insurance section.

The original data are divided into two groups: claims from policyholders who were under 30 years of age (younger policyholders) when the accident took place and claims from policyholders who were 30 years old or over (older policyholders) when they had the accident that gave rise to the claim for compensation. The first group consists of 1,061 observations in the cost interval of 1 to 126,000 and the second group comprises 4,061 observations in the interval ranging from 1 to 17,000. In Table 1 we present some descriptive statistics. The loss distributions of both the younger and older policyholders present right skewness and, furthermore, the distribution of claim severity for younger drivers presents a heavier tail than that associated with the older drivers (see [4]).

For each data set of younger and older drivers, respectively, we seek to estimate the VaR_α with $\alpha = 0.95$ and $\alpha = 0.995$. The value at risk is needed to determine which of the two groups is more risky in terms of accident severity, so that a larger premium loading can be imposed on that group. The following nonparametric methods are implemented: i) The empirical distribution (Emp) as in expression (2), ii) the classical kernel estimation of a cdf (CKE), as described in section 3.2 with a bandwidth based on the minimization of WISE and iii) the double transformed kernel estimation of cdf (DTKE), as described in section 3.3 with a bandwidth based on the minimization of MSE at $x = VaR_\alpha$. Epanechnikov kernel functions are used for CKE and DTKE.

In Table 2 we show the values of estimates $VaR_{0.95}$ and $VaR_{0.995}$ using the original samples. For $\alpha = 0.95$, all methods produce similar estimated values. However, with $\alpha = 0.995$, the results differ from one method to another. We observe that for the older drivers, the classical kernel estimation produces a $VaR_{0.995}$ estimate similar to the empirical quantile, while for the younger drivers, who they are the group with most risk, our proposed approach, which we call DTKE, provides estimates that lie considerably above the empirical quantile.

Table 1: Summary of the younger and older drivers' accident cost data

	Younger	Older	All
Number of observations	1,061	4,061	5,122
Mean	243.1	402.7	276.1
Median	66	68	67
Std Deviation	3,952.3	704.6	1,905.5
Maximum	126,000.0	17,000.0	126,000.0

Source: Own's data. Cost of claims in monetary units.

The results in Table 2 show that the double transformation kernel estimation does not underestimate the risk. As expected, it is a suitable method "to extrapolate the extreme quantile" in the zones of the distribution where almost no sample information is available. The estimated $VaR_{0.995}$ with this method is higher than the alternative nonparametric methods.

Table 2: Value at risk with tolerance level α (VaR_{α}) estimated for automobile claim cost data

		Younger	Older	All
$\alpha = 0.95$				
	Emp	1104.00	1000.00	1013.00
	CKE	1293.00	1055.33	1083.26
	DTKE	1257.33	1005.98	1048.51
$\alpha = 0.995$				
	Emp	5430.00	3000.00	4678.00
	CKE	5465.03	4040.40	4695.80
	DTKE	7586.27	4411.11	4864.08

Source: Own's data. Cost of claims in monetary units.

In Figure 3, we plot the estimated VaR_{α} for a grid of α between 0.99 and 0.999 for younger and older drivers, using the empirical distribution (Emp), the classical kernel estimation (CKE) and the double transformed kernel estimation (DTKE). Plots in Figure 3 show that Emp and CKE are very similar, i.e. in the zone where the data are scarce CKE does not smooth Emp. In both plots we observe that DTKE is a smoother version than Emp and CKE and, therefore, it allows the extrapolation of the VaR_{α} beyond the maximum observed in the sample with a smoothed curve.

It is immediately apparent that the risk of a severe accident among the group of younger policyholders is higher than that recorded among the older policyholders. As a consequence, the risk loading should be proportionally higher for this younger age group. In other words, younger drivers should pay proportionally higher insurance premiums because they are more likely to be involved in severe accidents.

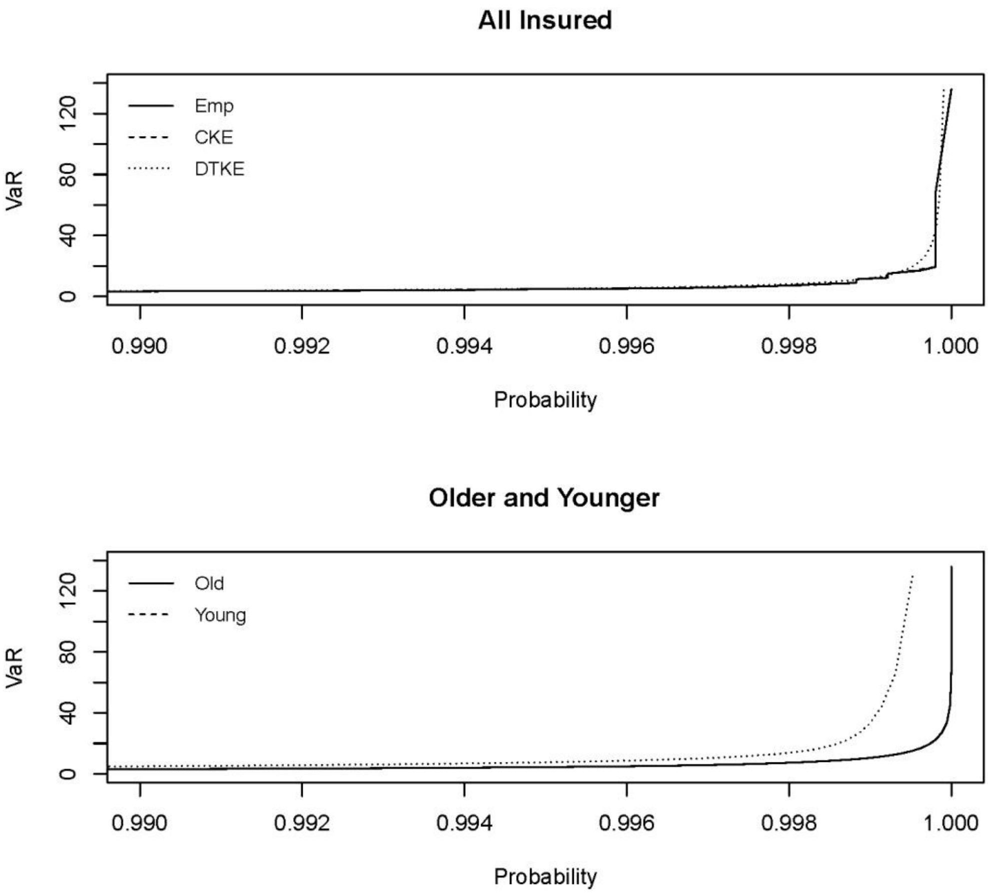

Figure 3: Estimated Value-at-Risk for tolerance levels (x-axis) above 99%. Above: Comparison of three methods for all policyholders. Solid, dashed and dotted lines correspond to the empirical, the classical kernel and the transformed kernel estimation method, respectively. Below: Value-at-Risk estimated with double transformed kernel estimation given the tolerance level. Solid line and dotted line correspond to older and younger policyholders, respectively.

To analyze the accuracy of the different methods we generate 1,000 bootstrap random samples of the costs of the younger and older policyholders. Each random sample has the same size as the original sample, but observations are chosen with a replacement so that some can be repeated and some can be excluded. We estimate the VaR_α for each bootstrap sample. In Table 3 we show the mean and the coefficient of variation (CV). The coefficient of variation is used to compare accuracy given that the nonparametric estimates, except for the empirical estimation, have some bias in finite sample size. The mean and the CV of the estimated VaR_α for the bootstrap samples, with α = 0.95 and α = 0.995, is shown for the claim costs of younger drivers, for the claim cost of older drivers and for all drivers together. The empirical distribution supposes that the maximum possible loss is the maximum observed in the sample. However, as the sample is finite and the extreme values are scarce, these extreme values may not provide a precise estimate of VaR_α. So, we need "to extrapolate the quantile", i.e. we need to estimate the VaR_α in a zone of the distribution where we have almost no sample information. In Table 3 we observe that the bootstrap means are similar for all methods at $\alpha = 0.95$, but differ when $\alpha = 0.995$. Moreover, if we analyze the coefficients of variation we observe that, for the younger policyholders, the two kernel-based methods are more accurate than the empirical estimation.

Given that the means of the VaR_α estimates for younger driver are larger than the means for the older drivers, we conclude that the younger drivers have a distribution with a heavier tail than that presented by the older policyholders.

Table 3: Results of bootstrap simulation for value at risk estimation with tolerance level α in the automobile claim cost data

		Younger		Older		All	
		Mean	CV	Mean	CV	Mean	CV
$\alpha = 0.95$	Emp	1145.02	0.124	1001.57	0.040	1021.92	0.034
	CKE	1302.19	0.104	1060.24	0.051	1086.88	0.045
	DTKE	1262.58	0.105	1008.28	0.054	1049.64	0.045
$\alpha = 0.995$	Emp	5580.67	0.297	4077.89	0.134	4642.61	0.093
	CKE	5706.69	0.282	4134.66	0.123	4643.42	0.087
	DTKE	7794.70	0.217	4444.75	0.095	4883.85	0.080

Source: Own's data. Cost of claims in monetary units. Emp refers to the empirical distribution method, CKE is the method based on classical kernel estimation and DTKE is the double transformed method.

For older drivers, and similarly for all the policyholders, empirical estimation seems the best approach at α = 0.95, but not at α = 0.995.

When $\alpha = 0.995$, underestimation of the Empirical distribution method (Emp) is evident compared to the lower quantile level at α = 0.95. The DTKE method has the lowest coefficient of variation compared to the other methods.

The double transformation kernel estimation is, in this case, the most accurate method for estimating extreme quantiles, as is shown in the bootstrap approach described above. Therefore, the DTKE is a method that can be recommended to produce risk estimates at large tolerance levels such $99{,}5\%$.

5. Conclusions

When analyzing the distribution of losses in a given risk class, we are aware that right skewness is frequent. As a result, certain risk measures, including variance and standard deviation, which are useful for identifying groups when the distribution is symmetric, are unable to discriminate distributions that contain a number of infrequent extreme values. By way of alternative, risk measures that focus on the right tail, such as quantiles, can be useful to quantify risk and for comparing risk classes.

In this paper we have proposed a system for measuring risk from loss data records that requires few statistical hypothesis on the distribution. We have also shown that certain modifications of the classical kernel estimation of cdf, such as transformations, give a risk measure estimate above the maximum observed in the sample without assuming a functional form that is strictly linked to a parametric distribution. Given the small number of values that are typically observed in the tail of a distribution, we believe our approach to be a practical method for risk analysts.

Our method can establish a distance between risk classes in terms of differences in the risk of extreme severities.

REFERENCES

[1] **Alemany, R., Bolancé, C. and Guillen, M. (2013),** ***A Nonparametric Approach to Calculating Value-at-Risk***. *Insurance: Mathematics and Economics*, 52(2), 255-262;

[2] **Azzalini, A. (1981),** ***A Note on the Estimation of a Distribution Function and Quantiles by a Kernel Method .****Biometrika*, 68(1), 326-328;

[3] **Bolancé, C. (2010), Optimal Inverse Beta(3,3)** ***Transformation in Kernel Density Estimation.*** *SORT-Statistics and Operations Research Transactions*, 34, 223-237;

[4] **Bolancé, C., Guillen, M. and Nielsen, J. P. (2003),** ***Kernel Density Estimation of Actuarial Loss Functions***. *Insurance: Mathematics and Economics*, 32(1), 19-36;

[5] **Bolancé, C., Guillen, M. and Nielsen, J. P. (2008),** ***Inverse Beta Transformation in Kernel Density Estimation***. *Statistics & Probability Letters*, 78(13), 1757-1764;

[6] **Bolancé, C., Guillen, M., Gustafsson, J. and Nielsen, J.P. (2012),** ***Quantitative Operational Risk Models***. *Chapman & Hall/CRC Finance Series;*

[7] **Bolancé, C., Guillen, M., Pelican, E. and Vernic, R. (2008), *Skewed Bivariate Models and Nonparametric Estimation for the CTE Risk Measure***. *Insurance: Mathematics and Economics*, 43(3), 386-393;
[8] **Bowman, A., Hall, P. and Prvan, T. (1998), *Bandwidth Selection for the Smoothing of Distribution Functions***. *Biometrika*, 85(4), 799-808;
[9] **Buch-Larsen, T., Guillén, M., Nielsen, J.P. and Bolancé, C. (2005), *Kernel Density Estimation for Heavy-Tailed Distributions Using the Champernowne Transformation***. *Statistics*, 39(6), 503-518;
[10] **Guillen, M., Prieto, F. and Sarabia, J. M. (2011), *Modelling Losses and Locating the Tail with the Pareto Positive Stable Distribution*.** *Insurance: Mathematics and Economics*, 49(3), 454-461;
[11] **Pitt, D., Guillen, M. and Bolancé, C. (2012), *An Introduction to Parametric and Non-Parametric Models for Bivariate Positive Insurance Claim Severity Distributions*** . Xarxa de Referencia en Economía Aplicada (XREAP). Working Papers XREAP2010-03;
[12] **Reiss, R.-D. (1981), *Nonparametric Estimation of Smooth Distribution Functions***. *Scandinavian Journal of Statistics*, 8, 116-119;
[13] **Ruppert, D. and Cline, D. B. (1994), *Bias Reduction in Kernel Density Estimation by Smoothed Empirical Transformations***. *The Annals of Statistics*, 22,185-210;
[14] **Sarabia, J. M. and Guillén, M. (2008), *Joint Modelling of the Total Amount and the Number of Claims by Conditionals***. *Insurance: Mathematics and Economics*, 43(3), 466-473;
[15] **Sheather, S. J. and Marron, J. S. (1990), *Kernel Quantile Estimators.*** *Journal of the American Statistical Association*, 85(410), 410-416;
[16] **Shen, L. and Elliott, R.J. (2012), *How to Value Risk***. Expert Systems with Applications, 39(5), 6111-6115;
[17] **Silverman, B.W. (1986), *Density Estimation for Statistics and Data Analysis***. *Chapman & Hall*;
[18] **Terrell, G. R. (1990), *The Maximal Smoothing Principle in Density Estimation***. *Journal of the American Statistical Association*, 85(410), 470-477;
[19] **Wand, M. P., Marron, J. S. and Ruppert, D. (1991), *Transformations in Density Estimation***. *Journal of the American Statistical Association*, 86(414), 343-353.