UB Riskcenter Working Paper Series

University of Barcelona

Research Group on Risk in Insurance and Finance www.ub.edu/riskcenter

Working paper 2014/05 \\ Number of pages 25

Accounting for severity of risk when pricing insurance products

Ramon Alemany, Catalina Bolancé and Montserrat Guillén

Riskcenter

Accounting for severity of risk when pricing insurance products

Ramon Alemany[a,1,*], Catalina Bolancé[a,1], Montserrat Guillén[a,1]

[a] *Dept. of Econometrics, Riskcenter-IREA, University of Barcelona, Av. Diagonal, 690 08041 Barcelona, Spain*

Abstract

We design a system for improving the calculation of the price to be charged for an insurance product. Standard pricing techniques generally take into account the expected severity of potential losses. However, the severity of a loss can be extremely high and the risk of a severe loss is not homogeneous for all policy holders. We argue that risk loadings should be based on risk evaluations that avoid too many model assumptions. We apply a nonparametric method and illustrate our contribution with a real problem in the area of motor insurance.

Keywords: quantile, value-at-risk, loss models, extremes

1. Introduction

A central problem faced by the insurance industry is to calculating the price at which to underwrite an insurance contract, that is how much a policyholder should be required to pay an insurer in order to obtain coverage. In principle, the price is proportional to the insured risk and, as such, the insurer needs to estimate the possibility of loss and its potential magnitude. Nevertheless, it is not easy to evaluate the risk and, therefore, to evaluate the price that have to pay the policyholder in exchange for a coverage for the insured risk. In general,

*Corresponding author. Phone: +349340370340. Fax: +34934021821.

Email addresses: ralemany@ub.edu (Ramon Alemany), bolance@ub.edu (Catalina Bolancé), mguillen@ub.edu (Montserrat Guillén)

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

there are a lot of features that affect the value of the insured risk that are difficult to measure, such as attitudes, responsibility, commitment, ... (see Baublyte et al., 2012). Also, the perception of risk by the insured differs depending on many factors: the customs, culture, risk aversion, ... (see Hayakawa et al., 2011). Therefore, it is essential to establish criteria to help us quantify the real risk, both from the point of view of analysis and risk assessment and from the perspective of their coverage. Then, it is important that insurers have historical information about the accident of their insureds and, simultaneously, the researchers in statistics should provide new methods that can help to improve the quantification of risk based on data available to insurance companies.

In non-life insurance (including motor and house insurance), policyholders are typically placed in risk categories or risk groups. Thus, in the case of motor insurance, all policyholders classified as belonging to a particular risk group pay the same price - the *a priori* premium - during the first contract period, which usually has a duration of one year (see, for example, [4] and [29]). At the end of this year, insurers will have gathered information on the losses suffered by members of a given group and as a result it will become evident that the profiles within each group are not necessarily homogeneous.

In this article we propose a system for calculating surcharges and price rebates on -*a priori* premiums. We develop a method for approximating the price surcharge based on the risk margin. It is widely accepted that within a given risk category, policyholders are distinct and present a heterogeneous claims behavior. This in part is purely randomness, but it may also be attributed to certain unobservable factors. In motor insurance, a driver's aggressiveness or reflexes will determine his or her propensity to have an accident, as they may impact driving habits. Many of these unobservable factors are, however, quite often difficult to measure.

The classical risk surcharge is associated with a measure of dispersion of the claims history presented by a risk group - the greater the degree of heterogeneity of policyholders within a given category, the greater the uncertainty about the expected total claims for that group. Thus, even though policyholders in the same risk category have to pay the same *pure* premium, the surcharge is proportional to the uncertainty of the claims outcome for that group.

Thus, we propose a novel system for calculating the risk premium that takes into account the cost of claims and which captures the particular shape of the distribution of the severity of a claim. It is well known that small claims are much more frequent than large claims, i.e. the statistical distribution of claim severities is highly skewed to the right, but usually only the mean and the variance are of importance in the calculation of prices.

Several studies have proposed nonparametric methods as an alternative to parametric models to estimate the shape of the severity distribution (see [20], [10], [12], [13]), when there is no evidence that parametric assumptions are suitable for fitting a loss distribution. Such nonparametric methods are suited to analyze the characteristics of severities beyond those of their mean and variance.

We will continue this study with a short introduction to insurance pricing and, in Section 3, we describe suitable nonparametric methods. In Section 4 we present the application, and we compare different nonparametric approaches to risk estimation in the motor insurance sector. Our two data sets contain automobile claim cost values from policyholders under 30 years and from those that are 30 years or older, respectively. Earlier studies similarly compare the distribution of accident severity for these two groups of policyholder (see [12], [20], [13] and [10]), but no attempt has previously been made to calculate the premium price correction based on a risk adjustment of severity information. Our conclusions and a discussion of the implementation of our proposal in insurance companies are provided in the last section.

2. Insurance pricing

Insurance companies define risk groups in accordance with information that is available at policy issuance. For instance, in the motor insurance sector, both the insured person and the insured vehicle matter. Typical variables, or risk factors, for this type of insurance include: age, zone of residence, car value, power of the car, etc.

A generalized linear model (GLM) is estimated to predict the expected number of claims given the risk factors and the expected cost per claim[2]. Additionally, cluster analysis (see [22]) can also be useful to partition the portfolio of insurance contracts into clusters of policyholders presenting a similar risk. The so-called *pure* premium is calculated as the product of the expected number of claims times the expected cost (see, for example, [4], [50]). A price surcharge is always added to the *pure* premium in order to obtain the -*a priori* premium. This loading covers managerial expenses, marketing, and security margins and solvency requirements. In subsequent periods, the -*a priori* premium may be corrected on the basis of the observed claims experience of the policyholders in a procedure known as *experience* rating. Having been corrected, the price is referred to as the -*a posteriori* premium.

Price correction is conducted in line with the information obtained from the accumulated claims experience. Thus, policyholders that have no claims in the previous year typically receive a bonus in the form of a reduction in their premium on renewing their contract. By contrast, customers that claimed compensation for an accident are usually charged a higher price. Such bonus-malus systems seek to correct the -*a priori* risk assessment, but they also discourage claims and play a central role in customer retention ([63], [39], [33]). As such, *experience rating* serves as a deterrent to policyholders who prefer not to claim for small accidents and so lose their price rebate. Moreover, as insurance companies receive fewer claims than they would if they did not operate a bonus-malus system, they are able to improve their claims handling and customer service. Many studies have focused on the estimation of -*a posteriori* premiums (see, for example, [44], [50], [14], [15],[8], [9] and [4]), but procedures to guarantee fair price loadings remain controversial. This is, in part, explained by the fact that the initial price setting is based primarily on expectations about the new

[2]Alternative modeling approaches, including gradient boosting trees, have also been proposed (see [30])

customer. In accident insurance in the construction industry, Imriyas [36] argues that experience rating is not efficient and calls for further research in the analysis of risk.

Here, we focus on the distribution of the claim costs in a given risk category, i.e. the severity distribution. This is reported to present right skewness with a long right tail (see [4] and [29]). This means that certain risk measures, including variance, standard deviation and the coefficient of variation, which are useful for identifying heterogeneous groups when a distribution is symmetric, cannot be used for asymmetric severity distribution. Instead, risk measures based on the right tail of the distribution, such as the Value-at-Risk (VaR) or the Tail-Value-at-Risk (TVaR), are more useful in accounting for the heterogeneity of the claims behavior within a given risk category. Moreover, assessing the risk of occurrence of a large loss in a given risk category can provide -*a priori* information to the insurers about their policyholders and the price structure they are marketing.

Bali and Theodossiou ([7]) analyzed different families of extreme value parametric distributions to estimate the VaR, and showed that the choice of a particular parametric distribution had a significant impact on VaR estimates and, as a consequence, on the risk premium and the final -*a priori* premium. Dowd and Blake ([25]) pointed out that nonparametric methods can serve as a good alternative to parametric methods because they avoid "the danger of mis-specifying the distribution", although they can be imprecise in tail regions where data are especially sparse. However, recent contributions avoid the potential imprecision of nonparametric methods in the tail. In [2] a nonparametric method based on the transformed kernel estimation was proposed to estimate the VaR with no parametric assumptions.

In the following sections we characterize the severity distribution using a nonparametric estimation of the VaR. We then compare different risk groups and present a method for approximating the risk premium. Our aim is to show that nonparametric statistical methods do not need to rely on assumptions regarding the severity of claims and provide a flexible tool to charge policyholders according to their risk profile. We focus on the VaR because it can be readily obtained using nonparametric methods ([59] provides an extensive review of risk valuation). Moreover, the properties of some nonparametric methods for estimating the VaR when the distribution is right skewed have been previously established in [2].

3. Nonparametric quantile estimation

Let X be a random variable that represents a loss, i.e. a claim cost, with cumulative distribution function (cdf) F_X. The VaR is also known as the quantile of F_X, i.e. it is defined as:

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

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 claims does not exceed. So, a fraction of claims $(1-\alpha)$ would exceed that level.

As we are interested in calculating VaR_α, we need an assumption regarding the stochastic behavior of losses and/or we need to estimate the cdf F_X. In practice, three classical statistical approaches to estimating F_X 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 approximation can be assumed (see [47]). 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 VaR_α, but unfortunately insurance claim losses 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 (see [41]). Note that the methods proposed by [34] and [58] for estimating VaR_α are not suitable for highly asymmetric distributions as has been shown in [2]. A nonparametric approach, such as classical kernel estimation (CKE), smooths the shape of the empirical distribution and "extrapolates" its behavior when dealing with extremes. In this study we use transformed kernel estimation and consider it suitable to estimate extreme quantiles of a skewed distribution.

3.1. Empirical distribution

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

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

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

$$VaR_\alpha\,(X) = \inf\left\{x, \widehat{F}_n\,(x) \geq \alpha\right\}. \tag{3}$$

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

$$\left(F_X\,(x)\,[1 - F_X\,(x)]\right)/n.$$

The empirical distribution is very 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 very 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 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:

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

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 [60]). 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:

$$\begin{aligned}\widehat{F}_X(x) &= \int_{-\infty}^{x} \widehat{f}_X(u)du = \int_{-\infty}^{x} \frac{1}{nb}\sum_{i=1}^{n} k\left(\frac{u-X_i}{b}\right) du \\ &= \frac{1}{n}\sum_{i=1}^{n} \int_{-\infty}^{\frac{x-X_i}{b}} k\left(t\right) dt = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x-X_i}{b}\right).\end{aligned} \tag{5}$$

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

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

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 [52] and [6] that, when $n \to \infty$, the mean squared error (MSE) of $\widehat{F}_X\left(x\right)$ can be approximated by:

$$\begin{aligned}E\left\{\widehat{F}_X\left(x\right) - F_X\left(x\right)\right\}^2 \sim \frac{F_X(x)[1-F_X(x)]}{n} &- f_X\left(x\right)\frac{b}{n}\left(1 - \int_{-1}^{1} K^2\left(t\right)dt\right) \\ &+ b^4\left(\frac{1}{2}f_X'\left(x\right)\int t^2 k\left(t\right)dt\right)^2.\end{aligned} \tag{6}$$

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:

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

where subindex x indicates that the smoothing parameter is optimal at this point. Moreover, Azzalini in [6] 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 [60], but since the objective in this paper is to estimate a quantile in the right tail of a distribution, [2] recommended calculating the bandwidth using a smoothing parameter that minimizes the weighted integrated squared error (WISE) asymptotically, i.e.:

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

The value of b that minimizes WISE asymptotically is:

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

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

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

Various methods to calculate b exist. For instance, cross-validation and plug-in methods (see, for example, [18] and [3]) are very usual. However, these methods require considerable computational effort in large data sets.

3.3. Transformed Kernel Estimation

Transformed kernel estimation is better than classical kernel density estimation when estimating distributions with right skewness (see [12], [20], [13] and [10]). 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:

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

where b and K are as defined in Section 3.2.

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

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

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 [64], [12], [20], [54] and [10]). However, only a few studies analyze the transformed kernel estimation of the cdf and quantile (see [2], [61] and [1]). These transformations can be classified into those that are a cdf and those that do not correspond to a specific cdf. Moreover, nonparametric cdf transformations can also be considered.

The double transformed kernel estimation (DTKE) method for estimating the quantile was proposed by [2]. First, the data are transformed with a cdf function (for instance, the generalized Champernowne cdf[3] and, second, the transformed data are again transformed using the inverse cdf of a $Beta(3,3)$ distribution defined on the domain $[-1,1]$ (see [2] for further details and [16] for computer codes in SAS and R). The double transformation approach is

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

$$T_X(x) = ((x+c)^\gamma - c^\gamma) \,/\, ((x+c)^\gamma + (M+c)^\gamma - 2c^\gamma)\,.\ \ c, \gamma, M > 0 \quad -c < x$$

based on the fact that the cdf of a $Beta(3,3)$ can be estimated optimally using classical kernel estimation (see [62]). Given that double transformed data have a distribution that is close to the $Beta(3,3)$ distribution, an optimal bandwidth for estimating VaR_α can be used. Details as to how this optimal bandwidth can be calculated are to be found in [2].

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 data set with observations on the cost of accident claims, i.e. a large, heavy-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 in 1997.

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 claim cost interval of 1 to 126,000 and the second group comprises 4,061 observations in the interval ranging from 1 to 17,000. Costs are expressed in monetary units. 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 policyholders presents a heavier tail than that associated with the older policyholders (see [12]).

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 heterogeneous in terms of accident severity, so that a larger premium loading can be imposed on that group. We also compare the relative size of risk between the group of younger and older policyholders. 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 were used for CKE and DTKE.

Table 1: Summary of the younger and older policyholders' claims cost data

Data	n	Mean	Median	Std. Deviation	Coeff. Variation
All	5,122	276.1497	67	1,905.5420	690.0394
Younger	1,061	402.7012	66	3,952.2661	981.4388
Older	4,061	243.0862	68	704.6205	289.86445

Cost of claims are expressed in monetary units.

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 younger drivers, the classical kernel estimation produces a $VaR_{0.995}$ estimate similar to the empirical quantile, while for the older drivers this nonparametric method provide estimates above the empirical quantile.

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 empirical quantile.

In Figure 1, 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 1 show that Emp and CKE are very similar, i.e. in the

Table 2: VaR_α results for automobile claim cost data.

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

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.

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 in the appendix. Therefore, we can conclude that DTKE is a nonparametric method that can be used to produce risk estimates at large tolerance levels such as 99,5%.

It is immediately apparent that the risk among the group of younger policyholders is higher than that recorded among the older policyholders. Thus, a young claimant is more likely to make a large claim and this risk is higher for the younger policyholders than for their older counterparts. As a consequence, the risk loading should be proportionally higher for this younger age group. In other words, younger drivers should pay higher insurance premiums because they are more likely to be involved in severe accidents. Moreover, once involved in an accident, young drivers present a higher risk than older drivers of presenting a higher claim. The frequency of claims has not been specifically examined here; yet, it is also the case that younger drivers with similar characteristics (other than age) to those of older drivers usually present a higher expected number of

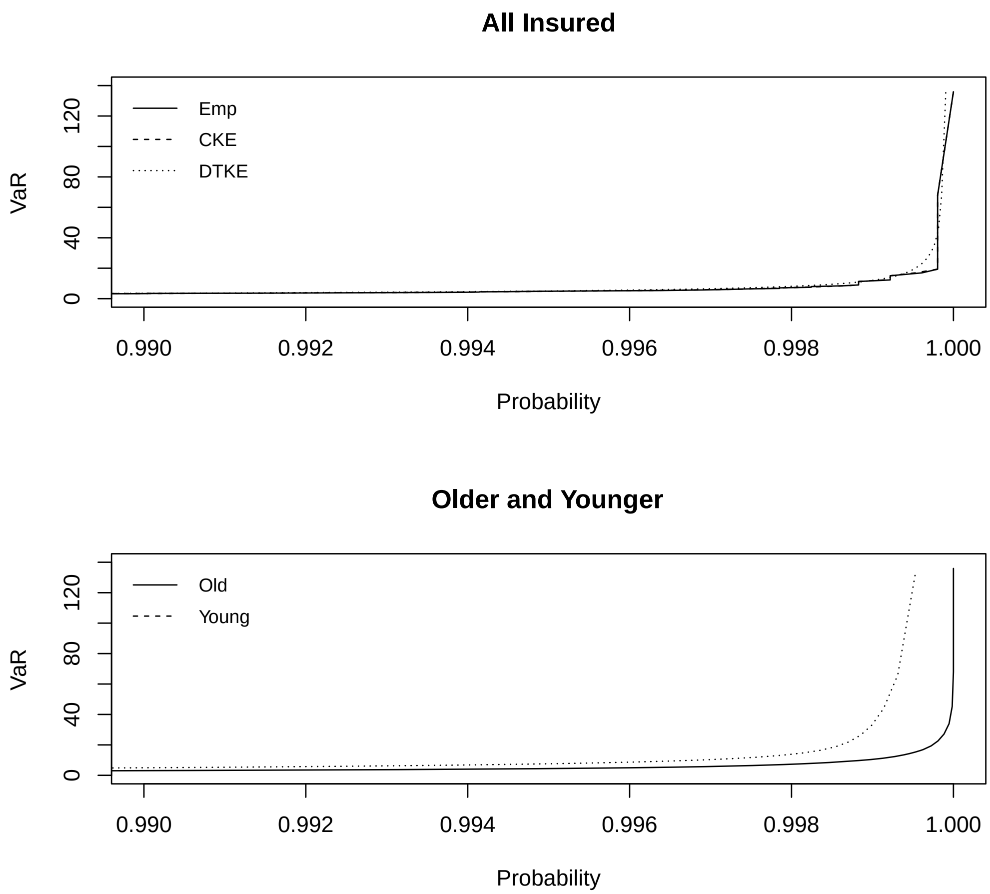

Figure 1: Estimated Value-at-Risk for tolerance levels (x-axis) above 99%. Above: Comparison of three methods for all policyholders. Solid, dashed and dotted lines correspond to the empirical, the classical kernel and the transformed kernel estimation method, respectively. Below: Value-at-Risk estimated with double transformed kernel estimation given the tolerance level. Solid line and dotted line correspond to older and younger policyholders, respectively.

claims[4].

In order to calculate the risk premium, when the loss severity distribution presents right skewness, we can compute $VaR_{0.995}$ for each group and then compare the risk groups. Here, for instance, the younger versus older policyholders presents a risk ratio equal to $7586/4411 = 1.72$ (see the last row in Table 2).

In Table 1 we can see that the mean cost of a claim for younger drivers is 402.7, while it is only 243.1 for older drivers. So, the *pure* premium, which serves as the basis for the price of an insurance contract, takes into account the fact that younger drivers should pay more than older drivers based on the average cost per claim[5].

In Table 1 the standard deviation for the younger group (3952) is more than five times greater than the standard deviation of the older group (705). thus, many insurers would charge younger drivers a risk premium loading that is five times higher. This increases the price of motor insurance for younger drivers significantly because, in practice, the price of the loading is proportional to the standard deviation. For instance the risk loading might be 5% times the standard deviation. In this case, older drivers would pay $243.1 + 0.05 \cdot 705 = 278.4$, but younger drivers would pay $402.7 + 0.05 \cdot 3952 = 600.3$. As a results, the premium paid by younger drivers would exceed that paid by older drivers by $600.3/278.4 = 115\%$.

We propose that the loading should, in fact, be proportional to a risk measure that takes into account the probability that a loss will be well above the average. For instance, VaR_α can be used with $\alpha = 99.5\%$. Given that the risk ratio for the younger versus the older driver at the 99.5% tolerance level equals 1.72, the risk premium loading for younger drivers $(0.005 \cdot 7586)$ should not be 72% higher than the risk premium loading for older drivers $((0.005 \cdot 4411)$ - note that

[4] We do not consider models for claim counts, limiting ourselves to claims severity only.

[5] A young driver with the same expected number of claims as an older driver should pay a premium that is 66% higher than that paid by an older driver $(402.7/243.1 = 1.66)$ due to this difference in the average claim cost.

$0.005 = 0.5\%$ is the risk level that corresponds to a tolerance of 99.5%. Thus, the price for older drivers is $243.1 + 0.005 \cdot 4411 = 265.2$ while the price for younger drivers should be equal to $402.7 + 0.005 \cdot 7586 = 440.63$. In this way, although the price of motor insurance is higher for younger drivers, it is only 66% higher than the price charged to older drivers.

Finally, we should stress that to determine the final product price, the expected number of claims needs to be taken into account. Thereafter, general management expenses and other safety loadings, such as expenses related to reinsurance, should be added to obtain the final commercial price.

5. Conclusions

When analyzing the distribution of claim costs in a given risk class, we are aware that right skewness is frequent. As a result, certain risk measures, including variance, standard deviation and the coefficient of variation, 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 VaR_α, can be useful for comparing risk classes and, thus, calculating risk premium loadings.

Introducing a severity risk estimate in the calculation of risk premiums is of obvious interest. A direct interpretation of the quantile results in a straightforward implementation. The larger the distance between the average loss and the Value-at-Risk, the greater the risk for the insurer of deviating from the expected equilibrium between the total collected premium and the sum of all compensations.

In this paper we have proposed a system for comparing different insurance risk profiles using a nonparametric estimation. 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 and pricing departments. We show that the double transformation kernel estimation is a suitable method in this context, because no statistical hypothesis regarding the random distribution of severities is imposed.

Our method can establish a distance between risk classes in terms of differences in the risk of extreme severities. An additional feature of our system is that a surcharge to the a priori premium can be linked to the loss distribution of severities. The loadings for each risk class have traditionally been the same for all groups, i.e. insensitive to the risk measures, or proportional to the standard deviation of their respective severity distributions. We suggest that risk loadings should be proportional to the risk measured within the severity distribution of each group. Our approach has the advantage of needing no distributional assumptions and of being easy to implement.

References

[1] Alemany, R., Bolancé, C. and Guillén, M. (2012) Nonparametric estimation of Value-at-Risk, Working Paper XREAP2012-19. University of Barcelona.

[2] Alemany, R., Bolancé, C. and Guillén, M. (2013) A nonparametric approach to calculating value-at-risk, *Insurance: Mathematics and Economics*, 52(2), 255-262.

[3] Altman, N. and Léger, C. (1995) Bandwidth selection for kernel distribution function estimation, *Journal of Statistical Planning and Inference*, 46, 195-214.

[4] Antonio, A. and Valdez, E.A. (2012) Statistical concepts of *a priori* and *a posteriori* risk classification in insurance, *Advances in Statistical Analysis*, 96, 187-224.

[5] Artzner, P., Delbaen, F., Eber, J.M. and Heath, D. (1999) Coherent measures of risk, *Mathematical Finance*, 9, 203-228.

[6] Azzalini, A. (1981) A note on the estimation of a distribution function and quantiles by a kernel method, *Biometrika*, 68, 326-328.

[7] Bali, T.G. and Theodossiou, P. (2008) Risk measurement performance of alternative distribution function, *The Journal of Risk and Insurance*, 75, 411-437.

[8] Bermúdez, L. (2009) A priori ratemaking using bivariate Poisson regression models, *Insurance: Mathematics and Economics*, 44, 135-141.

[9] Bermúdez, L. and Karlis, D. (2011) Bayesian multivariate Poisson models for insurance ratemaking, *Insurance: Mathematics and Economics*, 48, 226-236.

[10] Bolancé, C. (2010) Optimal inverse Beta(3,3) transformation in kernel density estimation, *SORT-Statistics and Operations Research Transactions*, 34, 223-237.

[11] Bolancé, C., Guillén, M. and Ayuso, M. (2012) A nonparametric approach to analysing operational risk with an application to insurance fraud, *The Journal of Operational Risk*, 7, 57-75.

[12] Bolancé, C., Guillén, M. and Nielsen, J.P. (2003) Kernel density estimation of actuarial loss functions, *Insurance: Mathematics and Economics*, 32, 19-36.

[13] Bolancé, C., Guillén, M. and Nielsen, J.P. (2008) Inverse Beta transformation in kernel density estimation, *Statistics & Probability Letters*, 78, 1757-1764.

[14] Bolancé, C., Guillén, M. and Pinquet, J. (2003) Time-varying credibility for frequency risk models, *Insurance: Mathematics and Economics*, 33, 273-282.

[15] Bolancé, C., Guillén, M. and Pinquet, J. (2008) On the link between credibility and frequency premium, *Insurance: Mathematics and Economics*, 43, 209-213.

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

[17] Bolancé, C., Guillén, M., Pelican, E. and Vernic, R. (2008) Skewed bivariate models and nonparametric estimation for CTE risk measure, *Insurance: Mathematics and Economics*, 43, 386-393.

[18] Bowman, A., Hall, P. and Prvan, T. (1998) Bandwidth selection for smoothing of distribution function, *Biometrika*, 85, 799-808.

[19] Brockett, P.L., Golden, L.L., Guillén, M., Nielsen, J.P., Parner, J. and Perez-Marin, A.M. (2008) Survival analysis of a household portfolio of insurance policies: how much time do you have to stop total customer defection?, *Journal of Risk and Insurance*, 75(3), 713-737.

[20] 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, 503-518.

[21] Cai, Z. and Wang, X. (2008) Nonparametric estimation of conditional VaR and expected shortfall, *Journal of Econometrics*, 147(1), 120-130.

[22] Campbell, M. (1986) An integrated system for estimating the risk premium of individual car models in motor insurance, *Astin Bulletin*, 16, 165-183.

[23] Chen, S.X. (2007) Nonparametric Estimation of Expected Shortfall, *Journal of Financial Econometrics*, 6(1), 87-107.

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

[25] Dowd, K. and Blake, D. (2006) After VaR: The theory, estimation, and insurance applications of quantile-based risk measures, *The Journal of Risk and Insurance*, 73, 193-229.

[26] Dhaene, J., Vanduffel, S., Tang, Q., Goovaerts, M.J., Kaas, R. and Vyncke, D. (2006) Risk measures and comonotonicity: A review, *Stochastic Models*, 22(4), 573-606.

[27] 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.

[28] Fan, J. and Gu, J. (2003) Semiparametric estimation of value-at-risk, *Econometrics Journal*, 6, 261-290.

[29] Gourieroux C. (1999) The Econometrics of Risk Classification in Insurance, *The Geneva Papers on Risk and Insurance Theory*, 24, 119-137.

[30] Guelman, L. (2012) Gradient boosting trees for auto insurance loss cost modeling and prediction, *Expert Systems with Applications*, 39(3), 3659-3667.

[31] Guillén, M., Nielsen, J.P. and Pérez-Marín, A.M. (2008) The need to monitor customer loyalty and business risk in the European insurance industry, *The Geneva Papers on Risk and Insurance-Issues and Practice*, 33(2), 207-218.

[32] Guillén, 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

[33] Guillén, M., Nielsen, J.P., Scheike, T.H. and Pérez-Marín, A.M. (2012) Time-varying effects in the analysis of customer loyalty: A case study in insurance, *Expert Systems with Applications*, 39(3), 3551-3558

[34] Harrell, F.E. and Davis, C.E. (1982) A new distribution-free quantile estimator, *Biometrika*, 69, 635-640.

[35] Hill, B.M. (1975) A simple general approach to inference about tail of a distribution, *Annals of Statistics*, 3, 1163-1174.

[36] Imriyas, K. (2009) An expert system for strategic control of accidents and insurers' risks in building construction projects, *Expert Systems with Applications*, 36(2), 4021-4034.

[37] Jones, B.L. and Zitikis, R. (2007) Risk measures, distortion parameters, and their empirical estimation, *Insurance: Mathematics and Economics*, 41(2), 279-297.

[38] Jorion, Ph. (2007) *Value at Risk*, McGraw-Hill, New York.

[39] Kaishev, V., Nielsen, J.P. and Thuring, F. (2013) Optimal customer selection for cross-selling of financial services products, *Expert Systems with Applications*, 40(5), 1748-1757.

[40] Kim, J.H.T. (2010) Bias correction for estimated distortion risk measure using the bootstrap, *Insurance: Mathematics and Economics*, 47(2), 198-205.

[41] Klugman, S.A., Panjer, H.H., Willmot, G.E. and Venter, G. (1998) *Loss models: from data to decisions*, Wiley New York.

[42] Krätschmer, V. and 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.

[43] Kupiec, P. (1995) Techniques for verifying the accuracy of risk measurement models, *Journal of Derivatives*, 3(2), 73-84.

[44] Lemaire, J. (1995) *Bonus-Malus System in Automobile Insurance*. Kluwer Academic Publisher.

[45] Lopez, O. (2012) A generalization of the Kaplan-Meier estimator for analyzing bivariate mortality under right-censoring and left-truncation with applications in model-checking for survival copula models, *Insurance: Mathematics and Economics*, 51(3), 505-516.

[46] McNeil, A. (1997) Estimating the tails of loss severity distributions using extreme value theory, *ASTIN Bulletin*, 27, 117-137.

[47] McNeil, A.J., Frey, R. and Embrechts, P. (2005) *Quantitative Risk Management: Concepts, Techniques, and Tools*. Princeton, University Press.

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

[49] Pinquet, J., Guillén, M. and Bolancé, C. (2000) Long-range contagion in automobile insurance data: estimation and implications for experience rating, *Working Paper 2000-43*, http://thema.u-paris10.fr.

[50] Pinquet, J., Guillén, M. and Bolancé, C. (2001) Allowance for the age of claims in bonus-malus systems, *Astin Bulletin*, 31, 337-348.

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

[52] Reiss, R.-D. (1981) Nonparametric estimation of smooth distribution functions, *Scandinavian Journal of Statistics*, 8, 116-119.

[53] Reiss R.-D. and Thomas, M. (1997) *Statistical Analysis of Extreme Values from Insurance, Finance, Hydrology and Others Fields*. Birkhäuser Verlag.

[54] Ruppert, D. R. and Cline, D. B. H. (1994) Bias reduction in kernel density estimation by smoothed empirical transformation, *Annals of Statistics*, 22, 185-210.

[55] 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.

[56] Sarda, P. (1993) Smoothing parameter selection for smooth distribution functions, *Journal of Statistical Planning and Inference*, 35, 65-75.

[57] Scaillet, O. (2004) Nonparametric estimation and sensitivity analysis of expected shortfall, *Mathematical Finance*, 14, 115-129.

[58] Sheather, S.J. and Marron, J.S. (1990) Kernel quantile estimators, *Journal of the American Statistical Association*, 85, 410-416.

[59] Shen, L. and Elliott, R.J. (2012) How to value risk, *Expert Systems with Applications*, 39(5), 6111-6115.

[60] Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman & Hall.

[61] Swanepoel, J.W.H. and Van Graan, F.C. (2005) A new kernel distribution function estimator based on a nonparametric transformation of the data, *Scandinavian Journal of Statistics*, 32, 551-562.

[62] Terrell, G.R. (1990) The maximal smoothing principle in density estimation, *Journal of the American Statistical Association*, 85, 270-277.

[63] Thuring, F., Nielsen, J.P., Guillén, M. and Bolancé, C. (2012) Selecting prospects for cross-selling financial products using multivariate credibility, *Expert Systems with Applications*, 39(10), 8809-8816.

[64] Wand, P., Marron, J.S. and Ruppert, D. (1991) Transformations in density estimation. *Journal of the American Statistical Association*, 86, 343-361.

Appendix

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

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

Table 3: Results of bootstrap simulation for Value-at-risk (VaR_α) estimation in the claim cost data sets.

	α=0.95					
Method	Younger		Older		All	
	Mean	CV	Mean	CV	Mean	CV
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
	α=0.995					
Method	Younger		Older		All	
	Mean	CV	Mean	CV	Mean	CV
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

UB·Riskcenter Working Paper Series

List of Published Working Papers

[WP 2014/01]. Bolancé, C., Guillén, M. and Pitt, D. (2014) "Non-parametric models for univariate claim severity distributions – an approach using R", UB Riskcenter Working Papers Series 2014-01.

[WP 2014/02]. Mari del Cristo, L. and Gómez-Puig, M. (2014) "Dollarization and the relationship between EMBI and fundamentals in Latin American countries", UB Riskcenter Working Papers Series 2014-02.

[WP 2014/03]. Gómez-Puig, M. and Sosvilla-Rivero, S. (2014) "Causality and contagion in EMU sovereign debt markets", UB Riskcenter Working Papers Series 2014-03.

[WP 2014/04]. Gómez-Puig, M., Sosvilla-Rivero, S. and Ramos-Herrera M.C. "An update on EMU sovereign yield spread drivers in time of crisis: A panel data analysis", UB Riskcenter Working Papers Series 2014-04.

[WP 2014/05]. Alemany, R., Bolancé, C. and Guillén, M. (2014) "Accounting for severity of risk when pricing insurance products", UB Riskcenter Working Papers Series 2014-05.