DOCUMENTS DE TREBALL

DE LA FACULTAT D'ECONOMIA I EMPRESA

Col.lecció d'Economia

E09/219

Transformation kernel density estimation of actuarial loss functions

Catalina Bolancé (1)

Montserrat Guillen (2)

Jens Perch Nielsen (3)

Adreça correspondència:

(1) Catalina Bolancé
Dept. Econometrics, RFA-IREA, University of Barcelona,
Diagonal, 690, 08034 Barcelona, Spain,

(2) Montserrat Guillen
Dept. Econometrics, RFA-IREA, University of Barcelona,
Diagonal, 690, 08034 Barcelona, Spain,

(3) Jens Perch Nielsen
Cass Business School, City University London
106 Bunhill Row, London EC1Y 8TZ, UK

Abstract:
A transformation kernel density estimator that is suitable for heavy-tailed distributions is discussed. Using a truncated Beta transformation, the choice of the bandwidth parameter becomes straightforward. An application to insurance data and the calculation of the value-at-risk are presented.

Keywords:
Heavy-tailed distributions, Value at Risk, Non-parametric methods

JEL Classification:
C14, G22

Resum:
Es presenta un estimador nucli transformat que és adequat per a distribucions de cua pesada. Utilitzant una transformació basada en la distribució de probabilitat Beta l'elecció del paràmetre de finestra és molt directa. Es presenta una aplicació a dades d'assegurances i es mostra com calcular el Valor en Risc.

1 Introduction

The severity of claims is measured in monetary units and is usually referred to as insurance loss or claim cost amount. The probability density function of claim amounts is usually right skewed, showing a big bulk of small claims and some relatively infrequent large claims. For an insurance company, density tails are therefore of special interest due to their economic magnitude and their influence on the re-insurance agreements.

It is widely known that large claims are highly unpredictable while they are responsible for financial instability and so, since solvency is a major concern for both insurance managers and insurance regulators, there is a need to estimate the density of claim cost amounts and to include the extremes in all the analyses.

This paper is about estimating the density function nonparametrically when data are heavy-tailed. Other approaches are based on extremes, a subject that has received much attention in the economics literature. Embrechts et al (1999), Coles (2001), Reiss and Thomas (2001) have treated extreme value theory (EVT) in general. Chavez-Demoulin and Embrechts (2004), based on Chavez-Demoulin and Davison (2005), have discussed smooth extremal models in insurance. Their focus is devoted to highlight the nonparametric trends, as a time-dependence is present in many catastrophic risk situations (such as storms or natural disasters) and in the finansial marked. A recent work by Cooray and Ananda (2005) combine the lognormal and the Pareto distribution and derive a distribution which has a suitable shape for small claims and can handle heavy tails. Others have addressed this subject with the g-and-h distribution, like Dutta and Perry (2006) for operation risk analysis.

In previous papers, we have analysed claim amounts in a one-dimensional setting and we have realized that a nonparametric approach that accounts for the asymmetric nature of the density is preferred for insurance loss distributions

(Bolance et al. 2003, Buch-Larsen et al, 2005). Moreover, we have applied the method on a liability data set and compared the nonparametric kernel density estimation procedure to classical methods (Buch-Larsen, 2006). Several authors (Clements et al., 2003) have devoted much interest to transformation kernel density estimation, which was initially proposed by Wand et al. (1991) for asymmetrical variables and based on the shifted power transformation family. The original method provides a good approximation for heavy-tailed distributions. The statistical properties of the density estimators are also valid when estimating the cumulative density function (cdf). Transformation kernel estimation turns out to be a suitable approach to estimate quantiles near 1 and therefore, it can be used to estimate value-at-risk (VaR) in financial and insurance related applications.

Buch-Larsen et al. (2005) proposed an alternative transformations based on a generalization of the Champernowne distribution, simulation studies have shown that it is preferable to other transformation density estimation approaches for distributions that are Pareto-like in the tail. In the existing contributions, the choice of the bandwidth parameter in transformation kernel density estimation is still a problem. One way of undergoing bandwidth choice is to implement the transformation approach so that transformation leads to a beta distribution, then use existing theory to optimize bandwidth parameter choice on beta distributed data and backtransform to the original scale. The main drawback is that the beta distribution may be very steep in the domain boundary, which causes numerical instability when the derivative of the inverse distribution function is needed for the backward transformation. In this work we propose to truncate the beta distribution and use the truncated version at transformation kernel density estimation. The results on the optimal choice of the bandwidth for kernel density estimation of beta density are used in the truncated version directly. In the simulation study we see that our approach produces very good results for heavy-tailed data. Our results are particularly relevant for applications in insurance, where the claims amounts are analyzed and usually small claims (low cost) coexist with only a few large claims (high cost).

Let $f_{\mathbf{x}}$ be a density function. Terrell and Scott (1985) and Terrell (1990) analyzed several density families that minimize functionals $\int \left\{ f_{\mathbf{x}}^{(p)}(x) \right\}^2 dx$, where superscript (p) refers to the p-th derivative of the density function. We will use these families in the context of transformed kernel density estimation. The results on those density families are very useful to improve the properties of the transformation kernel density estimator.

Given a sample $X_1, ..., X_n$ of independent and identically distributed (iid) observations with density function $f_{\mathbf{x}}$, the classical kernel density estimator is:

$$\hat{f}_{\mathbf{x}}(x) = \frac{1}{n} \sum_{i=1}^{n} K_b (x - X_i), \tag{1}$$

where b is the bandwidth or smoothing parameter and $K_b(t) = K(t/b)/b$ is the kernel. In Silverman (1986) or Wand and Jones (1995) one can find an extensive revision of classical kernel density estimation.

An error distance between the estimated density $\hat{f}_{\mathbf{x}}$ and the theoretical density $f_{\mathbf{x}}$ that has widely been used in the analysis of the optimal bandwidth b is the mean integrated squared error ($MISE$):

$$E \left\{ \int \left(f_{\mathbf{x}}(x) - \hat{f}_{\mathbf{x}}(x) \right)^2 dx \right\}. \tag{2}$$

It has been shown (see, for example, Silverman, 1986, chapter 3) that the $MISE$ is asymptotically equivalent to $A - MISE$:

$$\frac{1}{4} b^4 (k_2)^2 \int \left\{ f_{\mathbf{X}}''(x) \right\}^2 dx + \frac{1}{nb} \int K(t)^2 dt, \tag{3}$$

where $k_2 = \int t^2 K(t)\, dt$. If the second derivative of $f_{\mathbf{x}}$ exists (and we denote it by $f_{\mathbf{X}}''$), then $\int \{ f_{\mathbf{x}}''(x) \}^2 dx$ is a measure of the degree of smoothness because the smoother the density, the smaller this integral is. From the expression for $A - MISE$ it follows that the smoother $f_{\mathbf{x}}$, the smaller the value of $A - MISE$.

Terrell and Scott (1985, Lemma 1) showed that $Beta\,(3,3)$ defined on the domain $(-1/2, 1/2)$ minimizes the functional $\int \left\{f''_{\mathbf{x}}\,(x)\right\}^2 dx$ within the set of beta densities with same support. The $Beta\,(3,3)$ distribution will be used throughout our work. Its pdf and cdf are:

$$g\,(x) = \frac{15}{8}\left(1-4x^2\right)^2, -\frac{1}{2} \leq x \leq \frac{1}{2}, \quad (4)$$

$$G\,(x) = \frac{1}{8}\left(4-9x+6x^2\right)(1+2x)^3. \quad (5)$$

We assume that a transformation exists so that $T\,(X_i) = Z_i\ i = 1, ..., n$ is assumed from a $Uniform(0,1)$ distribution. We can again transform the data so that $G^{-1}\,(Z_i) = Y_i\ i = 1, ..., n$ is a random sample from a random variable $\mathbf{y}$ with a $Beta(3,3)$ distribution, whose pdf and cdf are defined in (5).

In this work, we use a parametric transformation $T(\cdot)$, namely the modified Champernowne cdf as proposed by Buch-Larsen et al. (2005)

Let us define the kernel estimator of the density function for the transformed variable:

$$\hat{g}\,(y) = \frac{1}{n}\sum_{i=1}^{n} K_b\,(y - Y_i)\,, \quad (6)$$

which should be as close as possible to a $Beta(3,3)$. We can obtain an exact value for the bandwidth parameter that minimizes $A - MISE$ of $\hat{g}$. If $K\,(t) = (3/4)\,(1-t^2)\,1\,(|t| \leq 1)$ is the Epanechnikov kernel, where $1\,(\cdot)$ equals one when the condition is true and zero otherwise, then we show that the optimal smoothing parameter for $\hat{g}$ if $\mathbf{y}$ follows a $Beta(3,3)$ is:

$$b = \left(\frac{1}{5}\right)^{-\frac{2}{5}}\left(\frac{3}{5}\right)^{\frac{1}{5}}(720)^{-\frac{1}{5}}\,n^{-\frac{1}{5}}, \quad (7)$$

Finally, in order to estimate the density function of the original variable, since

$y = G^{-1}(z) = G^{-1}\{T(x)\}$, the transformation kernel density estimator is:

$$\hat{f}_{\mathbf{x}}(x) = \hat{g}(y)\left[G^{-1}\{T(x)\}\right]' T'(x) = \tag{8}$$

$$= \frac{1}{n}\sum_{i=1}^{n} K_b\left(G^{-1}\{T(x)\} - G^{-1}\{T(X_i)\}\right)\left[G^{-1}\{T(x)\}\right]' T'(x). \tag{9}$$

The estimator in (8) asymptotically minimizes $MISE$ and the properties of the transformation kernel density estimation (8) are studied in Bolancé et al. (2008). Since we want to avoid the difficulties of the estimator defined in (8), we will construct the transformation so that we avoid the extreme values of the beta distribution domain.

2 Estimation procedure

Let $\mathbf{z} = \mathbf{T}(\mathbf{x})$ be a $Uniform(0,1)$, we define a new random variable in the interval $[1-l, l]$, where $1/2 < l < 1$. The values for l should be close to 1. The new random variable is $z^* = T^*(x) = (1-l) + (2l-1)T(x)$. We will discuss later the value of l.

The pdf of the new variable $y^* = G^{-1}(z^*)$ is proportional to the $Beta(3,3)$ pdf, but it is in $[-a, a]$ interval, where $a = G^{-1}(l)$. Finally, our proposed transformation kernel density estimation is:

$$\hat{f}_{\mathbf{x}}(x) = \frac{\hat{g}(y^*)\left[G^{-1}\{T^*(x)\}\right]' T^{*\prime}(x)}{(2l-1)} = \hat{g}(y^*)\left[G^{-1}\{T^*(x)\}\right]' T'(x) \tag{10}$$

$$= \frac{1}{n}\sum_{i=1}^{n} K_b\left(G^{-1}\{T^*(x)\} - G^{-1}\{T(X_i)\}\right)\left[G^{-1}\{T^*(x)\}\right]' T'(x) \tag{11}$$

The value of $A - MISE$ associated to the kernel estimation $\hat{g}(y^*)$, where the random variable y^* is defined on an interval that is smaller than $Beta(3,3)$ domain is:

$$A - MISE_a = \frac{1}{4} b^4 (k_2)^2 \int_{-a}^{a} \{g''(y)\}^2 dy + \frac{1}{nb}\int_{-a}^{a} g(y)\, dy \int K(t)^2 dt. \tag{12}$$

And finally, the optimal bandwidth parameter based on the asymptotic mean integrated squared error measure for $\hat{g}(y^*)$ is:

$$
\begin{aligned}
b_g^{opt} &= k_2^{-\frac{2}{5}} \left(\int_{-1}^{1} K(t)^2 \, dt \int_{-a}^{a} g(y) \, dy \right)^{\frac{1}{5}} \left(\int_{-a}^{a} \{g''(y)\}^2 \, dy \right)^{-\frac{1}{5}} n^{-\frac{1}{5}} && (13) \\
&= \left(\frac{1}{5} \right)^{-\frac{2}{5}} \left(\frac{3}{5} \left(\frac{1}{4} a \left(-40a^2 + 48a^4 + 15 \right) \right) \right)^{\frac{1}{5}} && (14) \\
&\quad \times \left(360a \left(-40a^2 + 144a^4 + 5 \right) \right)^{-\frac{1}{5}} n^{-\frac{1}{5}}, && (15)
\end{aligned}
$$

The difficulty that arises when implementing the transformation kernel estimation expressed in (10) is the selection of the value of l. This value can be chosen subjectively as discussed in the simulation results by Bolancé et al. (2008). Let X_i, $i = 1, ..., n,$ be iid observations from a random variable with an unknown density $f_{\mathbf{x}}$. The transformation kernel density estimator of $f_{\mathbf{x}}$ is called KIBMCE (kernel inverse beta modified Champernowne estimator).

3 VaR estimation

In finance and insurance, the VaR represents the magnitude of extreme events and therefore it is used as a risk measure, but VaR is a quantile. Let $\mathbf{x}$ a loss random variable with distribution function $F_{\mathbf{x}}$, given a probability level p, the VaR of $\mathbf{x}$ is $VaR(\mathbf{x}, \mathbf{p}) = \inf \{x, F_{\mathbf{x}}(x) \geq p\}$. Since $F_{\mathbf{x}}$ is a continuous and nondecreasing function, then $VaR(\mathbf{x}, \mathbf{p}) = F_{\mathbf{x}}^{-1}(p)$, where p is a probability near 1 (0.95, 0.99,...). One way of approximating $VaR(\mathbf{x}, \mathbf{p})$ is based on the empirical distribution function, but this has often been criticized because the empirical estimation is based only on a limited number of observations, and even np may not be an integer number. As an alternative to the empirical distribution approach, classical kernel estimation of the distribution function can be useful, but this method will be very imprecise for asymmetrical or heavy-tailed variables.

Swanepoel and Van Graan (2005) propose to use a nonparametric transformation, that is equal to a classical kernel estimation of the distribution function. We propose to use a parametric transformation based on a distribution function.

Given a transformation function $Tr\left(\mathbf{x}\right)$ it follows that $F_{\mathbf{x}}\left(x\right)=F_{Tr(\mathbf{x})}\left(Tr\left(x\right)\right)$. So, the transformation kernel estimation of $VaR\left(\mathbf{x},\mathbf{p}\right)$ is based in the kernel estimation of the distribution function of the transformed variable.

Kernel estimation of the distribution function is (Azzalini, 1981 and Reiss, 1981):

$$\hat{F}_{Tr(\mathbf{x})}\left(Tr\left(x\right)\right)=\frac{1}{n}\sum_{i=1}^{n}\int_{-1}^{\frac{Tr(x)-Tr(X_i)}{b}}K\left(t\right)dt, \tag{16}$$

Therefore, the $VaR\left(\mathbf{x},\mathbf{p}\right)$ can be found as:

$$VaR\left(\mathbf{x},\mathbf{p}\right)=Tr^{-1}\left[VaR\left(Tr\left(\mathbf{x}\right),p\right)\right]=Tr^{-1}\left[\hat{F}^{-1}_{Tr(\mathbf{x})}\left(p\right)\right] \tag{17}$$

4 Simulation study

This section presents a comparison of our inverse beta transformation method with the results presented by Buch-Larsen, et al. (2005) based only on the modified Champernowne distribution. Our objective is to show that the second transformation, that is based on the inverse of a Beta distribution, improves density estimation.

In this work we analyze the same simulated samples as in Buch-Larsen, et al. (2005), which were drawn from four distributions with different tails and different shapes near 0: *lognormal, lognormal-Pareto, Weibull* and *truncated logistic.* The distributions and the chosen parameters are listed in Table 1.

Table 1: Distributions in simulation study.

Distribution	Density	Parameters
Mixture of pLognormal(μ,σ) and $(1-p)$Pareto(λ,ρ,c)	$f(x) = p\frac{1}{\sqrt{2\pi\sigma^2}x}e^{-\frac{(\log x-\mu)^2}{2\sigma^2}}$ $+(1-p)(x-c)^{-(\rho+1)}\rho\lambda^{\rho}$	$(p,\mu,\sigma,\lambda,\rho,c)$ $= (0.7,0,1,1,1,-1)$ $= (0.3,0,1,1,1,-1)$
Lognormal(μ,σ)	$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}x}e^{-\frac{(\log x-\mu)^2}{2\sigma^2}}$	$(\mu,\sigma)=(0,0.5)$
Weibull(γ)	$f(x)=\gamma x^{(\gamma-1)}e^{-x^{\gamma}}$	$\gamma=1.5$
Normal(μ,σ)	$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$(\mu,\sigma)=(5,1)$

Buch-Larsen, et al. (2005) evaluate the performance of the KMCE estimators compared to the estimator described by Clements, et al. (2003), the estimator described by Wand, et al. (1991) and the estimator described by Bolancé, et al. (2003). The Champernowne transformation substantially improve the results from previous authors. Here we see that if the second transformation based on the inverse beta transformation improves the results presented in Buch-Larsen, et al. (2005), this means that the double-transformation method presented here is a substantial gain with respect to existing methods.

We measure the performance of the estimators by the error measures based in L_1 norm, L_2 norm and $WISE$. This last weighs the distance between the estimated and the true distribution with the squared value of x. This results in an error measure that emphasizes the tail of the distribution, which is very relevant in practice when dealing with income or cost data:

$$\left(\int_0^{\infty}\left(\widehat{f}(x)-f(x)\right)^2 x^2\,dx\right)^{1/2}. \tag{18}$$

The simulation results can be found in Table 2. For every simulated density and for sample sizes $N=100$ and $N=1000$, the results presented here correspond to the following error measures L_1, L_2 and $WISE$ for different values of the trimming parameter $l = 0.99, 0.98$. The benchmark results are labelled KMCE and they correspond to those presented in Buch-Larsen, et al. (2005).

In general, we can conclude that after a second transformation based on the inverse of a certain Beta distribution cdf the error measures diminish with respect to the KMCE method. In some situations the errors diminish quite substantially with respect to the existing approaches.

We can see that the error measure that shows improvements when using the KIBMCE estimator is the $WISE$, which means that this new approach is fitting the tail of positive distributions better than existing alternatives. The $WISE$ error measure is always smaller for the KIBMCE than for the KMCE, at least for one of the two possible value of l that have been used in this simulation study. This would make the KIBMCE estimator specially suitable for positive heavy-tailed distributions. When looking more closely at the results for the mixture of a log-normal distribution and a Pareto tail, we see that larger values of l are needed to improve the error measures that were encountered with the KMCE method.

We can see that for the Truncated logistic distribution, the lognormal distribution and the Weibull distribution, the method presented here is clearly better than the existing KMCE. We can see in Table 2 that for $N = 1000$, the KIBMCE $WISE$ is about 20% lower than the KMCE $WISE$ for these distributions. A similar behavior is shown by the other error measures, L_1 and L_2, for $N = 1000$, are about 15% lower for the KIBMCE.

Note that the KMCE method was studied in Buch-Larsen, et al. (2005) and the simulation study showed that it improved on the error measures for the existing methodological approaches (Clements, et al., 2003 and Wand, et al., 1991).

Table 2: The estimated error measures for KMCE and KIBMCE
with $l = 0.99$ and $l = 0.98$ for sample size 100 and 1000 based on 2000 repetitions

			Log-Normal	Log-Pareto		Weibull	Tr. Logist.
				p =.7	p =.3		
N=100	L1	KMCE	0.1363	0.1287	0.1236	0.1393	0.1294
		l=0.99	0.1335	0.1266	0.1240	0.1374	0.1241
		l=0.98	0.1289	0.1215	0.1191	0.1326	0.1202
	L2	KMCE	0.1047	0.0837	0.0837	0.1084	0.0786
		l=0.99	0.0981	0.0875	0.0902	0.1085	0.0746
		l=0.98	0.0956	0.0828	0.0844	0.1033	0.0712
	WISE	KMCE	0.1047	0.0859	0.0958	0.0886	0.0977
		l=0.99	0.0972	0.0843	0.0929	0.0853	0.0955
		l=0.98	0.0948	0.0811	0.0909	0.0832	0.0923
N =1000	L1	KMCE	0.0659	0.0530	0.0507	0.0700	0.0598
		l=0.99	0.0544	0.0509	0.0491	0.0568	0.0497
		l=0.98	0.0550	0.0509	0.0522	0.0574	0.0524
	L2	KMCE	0.0481	0.0389	0.0393	0.0582	0.0339
		l=0.99	0.0394	0.0382	0.0393	0.0466	0.0298
		l=0.98	0.0408	0.0385	0.0432	0.0463	0.0335
	WISE	KMCE	0.0481	0.0384	0.0417	0.0450	0.0501
		l=0.99	0.0393	0.0380	0.0407	0.0358	0.0393
		l=0.98	0.0407	0.0384	0.0459	0.0369	0.0394

5 Data study

In this section, we apply our estimation method to a data set that contains automobile claim costs from a Spanish insurance company for accidents occurred in 1997. This data set was analyzed in detail by Bolancé et al. (2003). It is a typical insurance claims amount data set, i.e. a large sample that looks heavy-tailed. The data are divided into two age groups: claims from policyholders who are less than 30 years old, and claims from policyholders who are 30 years old or older. The first group consists of 1,061 observations in the interval [1;126,000] with mean value 402.70. The second group contains 4,061 observations in the interval [1;17,000] with mean value 243.09. Estimation of the parameters in the modified

Champernowne distribution function for the two samples of is, for young drivers $\widehat{\alpha}_1 = 1.116$, $\widehat{M_1} = 66$, $\widehat{c}_1 = 0.000$ and for older drivers $\widehat{\alpha}_2 = 1.145$, $\widehat{M_2} = 68$, $\widehat{c}_2 = 0.000$, respectively. We notice that $\alpha_1 < \alpha_2$, which indicates that the data set for young drivers has a heavier tail than the data set for older drivers.

Figure 1 and 2 plot the estimated densities. For small costs, we see that the KIBMCE density in the mode is greater than for the KMCE approach proposed by Buch-Larsen et al. (2005) both for young and older drivers. For both methods, the tail in the estimated density of young policyholders is heavier than the tail of the estimated density of older policyholders. This can be taken as evidence that young drivers are more likely to claim a large amount than older drivers. The KIBMCE method produces lighter tails than the KMCE methods. Based on the results in the simulation study presented in Bolancé et al (2008), we believe that the KIBMCE method improves the estimation of the density in the extreme claims class.

Figure 1: KIBMCE with $l = 0.99$ and $l = 0.98$ (solid line) and KMCE (dashed line) estimators of automobile claims for younger policyholders.

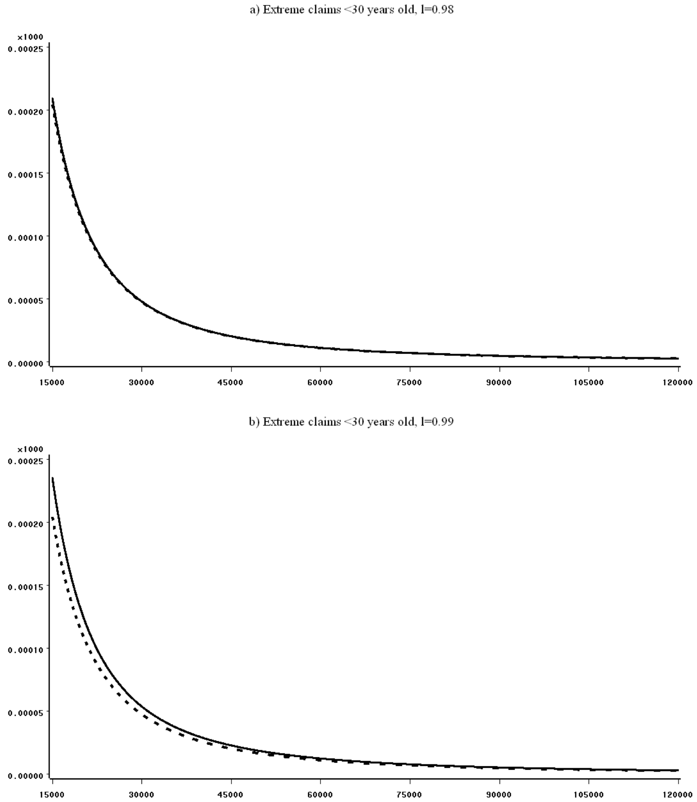
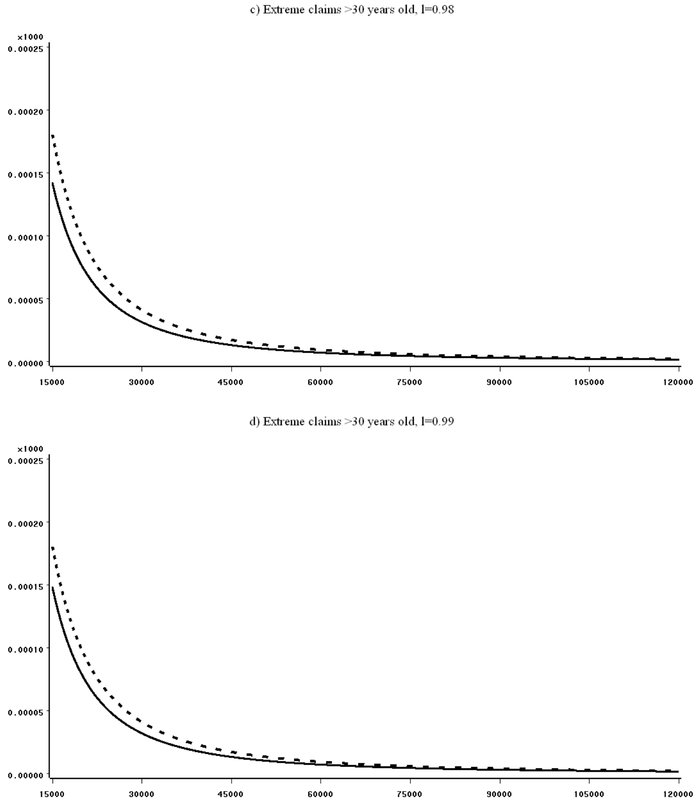


Table 3 presents the $VaR\left(\mathbf{x}, \mathbf{0.95}\right)$ which are obtained from the empirical distribution estimation and those obtained with the KMCE and KIBMCE. We believe that the KIBMCE provides an adequate estimation of the VaR and it seems a recommendable approach to be used in practice.

Table 3: Estimation of $VaR\left(\mathbf{x}, \mathbf{0.95}\right)$, in thousands.

			KIBMCE	
	Empirical	KMCE	$l = 0.99$	$l = 0.98$
Young	1104	2912	1601	1716
Older	1000	1827	1119	1146

Figure 2: KIBMCE with $l = 0.99$ and $l = 0.98$ (solid line) and KMCE (dashed line) estimators of automobile claims for older policyholders.



References

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

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

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

[4] Dutta, K. and Perry, J. (2006): A Tale of Tails: An Empirical Analysis of Loss Distribution Models for Estimating Operational Risk Capital. *FRB of Boston Working Paper No. 06-13* Available at SSRN: http://ssrn.com/abstract=918880

[5] Buch-Larsen, T., Guillen, M., Nielsen, J.P. and Bolancé, C. (2005): Kernel density estimation for heavy-tailed distributions using the Champernowne transformation. *Statistics*, *39*, 503-518.

[6] Clements, A.E., Hurn, A.S. and Lindsay, K.A. (2003): Möbius-like mappings and their use in kernel density estimation. *Journal of the American Statistical Association*, *98*, 993-1000.

[7] Chavez-Denoulin, V. and Davison, P. (2005): Generalized additive modelling of sample extremes. *Applied Statistics*, 54, 1, 207-222.

[8] Chavez-Denoulin, V. and Embrechts, P. ((2004): Smooth Extremal Models in Finance and Insurance. *The Journal of Risk and Insurance*, 71, 2, 183-199.

[9] Coles, S. (2001): *An introduction to statistical modeling of extreme values*, Springer, London.

[10] Cooray, K. and Ananda, M.M.A. (2005): Modeling actuarial data with a composite lognormal-Pareto model. *Scandinavian Actuarial Journal*, 5, 321-334.

[11] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1999): *Modelling Extremal Events*, Springer, London.

[12] Jones, M.C. (1990): Variable kernel density estimation and variable kernel density estimation. *Australian Journal of Statistics, 32*, 361-371.

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

[14] Reis, R.D. and Thomas M.(2001): *Statistical Analysis of Extreme Values: with Applications to Insurance, Finance, Hydrology and Other Fields,* 2nd Edition. Birkhauser.

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

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

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

[18] Terrell, G.R. and Scott, D.W. (1990): Oversmoothed nonparametric density estimates. *Journal of the American Statistical Association, 80,* 209-214.

[19] Wand, M.P. and Jones, M.C. (1995): *Kernel Smoothing.* London: Chapman & Hall.

[20] Wand, P., Marron, J.S. and Ruppert, D. (1991): Transformations in density estimation. *Journal of the American Statistical Association*, *86,* 414, 343-361.

[21] Wu, T.-J., Chen, C.-F. and Chen, H.-Y. (2007): A variable bandwidth selector in multivariate kernel density estimation. *Statistics & Probability Letters, 77,* 462-467.

Acknowledgement 1 *The Spanish Ministry of Education and Science /FEDER support SEJ2007-63298 is acknowledged.*