

# Kernel density estimation of actuarial loss functions.\*

Catalina Bolancé<sup>†</sup>  
University of Barcelona

Montserrat Guillen<sup>‡</sup>  
University of Barcelona

Jens Perch Nielsen<sup>§</sup>  
Codan

April 4, 2000

## Abstract

In this paper we estimate actuarial loss functions based on a symmetrized version of the semiparametric transformation approach to kernel smoothing. We apply this method to an actuarial study of automobile claims. The method gives a good overall impression while estimating actuarial loss functions, since it is capable of estimating both the initial mode and the heavy tail that is so typical for actuarial and other economic loss distributions. We study the properties of the transformation kernel density estimation and show the differences with the multiplicative bias corrected estimator with variable bandwidth. We add insight into the kernel smoothing transformation method through an extensive simulation study with a particular view to the performance of the estimation at the tail.

KEY WORDS: Loss models; Transformation; Skewness; Weighted integrated squared error.

## 1 Introduction

One of the main concerns in actuarial science is to study a group of risks. There are two important reasons for this study, firstly the solvency question regarding the companies ability to keep the business in a good shape after bad year with many expensive claims, secondly the question of profitability. It is by now a well known fact from the financial literature that nobody wishes to take on unnecessary risk, this can also be stated as risk has a price. It is important for a given company that it does not take over a risk at the wrong price. Exactly for the same reasons as it is important for a financial dealer that he does not buy an asset for more than the

---

\*We are grateful to Steven Marron and Ricardo Cao for their help; Paul Kofman and the Centre for Insurance Statistics for assistance of getting the data used in this study. The first two authors have received support from CICYT grant SEC1999-0693.

<sup>†</sup>Dept. of Econometrics. University of Barcelona. Diagonal, 690, 08034 Barcelona, Spain. E-mail address: bolance@eco.ub.es.

<sup>‡</sup>Dept. of Econometrics. University of Barcelona. Diagonal, 690, 08034 Barcelona, Spain. E-mail address: guillen@eco.ub.es.

<sup>§</sup>Head of Actuarial Product Development at Codan, Codanhus, 60 Gammel Kongevej, DK-1790 Copenhagen V, Denmark. Email: npj@codan.dk.

arbitrage free price. The insurance market is however in many ways much more diverse than the financial market since all kinds of risks are insured at all kinds of conditions.

An insurance policy is the agreement between the insured and the insurer that fixes the conditions for receiving a financial compensation in the random future time if the loss occurs. The insurer has to make sure that the consumers believe this promise by transmitting confidence on the grounds of past experience by showing that claims have been paid and that the insurer will stay solvent. Therefore the solvency of the insurer is important also when it comes to the marketing possibilities of the company. The actuaries have to come up with solvency considerations that give confidence to all the people with interest in the company, the management, the stock-holders and the customers.

Several unknown sources of randomness are present in insurance contracts: whether the loss will occur or not and if it occurs, when it will occur. Additionally, severity is also random and so is the financial compensation to be paid. All the policyholders having a similar kind of insurance contract are usually referred to as a portfolio. It is often the case that a portfolio is reinsured in such a way that the loss on one particular portfolio of the company can not exceed a certain level.

The compensation received by the insured person in the event of a loss covered by the insurance contract is called the payment or the single claim amount. Loss distributions describe the payments to the insured, which may or may not exclude payments of zero (Hogg and Klugman 1984; Klugman, Panjer and Wilmot 1998).

The actuarial literature is abundant in studies of how to estimate the amount of a single claim (the loss distribution, or, equivalently, loss function), because it is the basis for the development of the insurance system that in turn gives the information of the solvency situation and of the price of risk. The loss distribution is the probability distribution of the amount to be paid to the insured for the damage. In the day-to-day business insurers need to estimate loss distributions for the obvious reasons, namely to be able set the right price of risk, to make provisions and furthermore to write reinsurance to reduce unwanted risk. The usual practice in the insurance industry is to face risk up to a certain level using the company's own financial resources and to reinsure the rest (reinsurance may be written using the excess of loss method or the aggregate stop-loss method, among others). In order to make decisions on the kind of reinsurance, the insurer has to evaluate the probability that a certain claim amount exceeds a fixed limit. Upper quantiles are important in this context because the occurrence of a large claim may cause insolvency or at least threaten the financial equilibrium of a firm. It is also interesting to calculate the probability of losses within a given interval for risk assessment (Daykin, Pentikäinen and Pesonen 1994, pp. 127-130). In practise the study of loss distributions is combined with the study of the random number of claims generated by a policy or a portfolio over a fixed period of time. In this paper we will however restrict ourselves to the question of loss distributions.

Traditional methods for loss distributions use parametric models. Two of the most popular shapes are based on the lognormal distribution and the pareto distribution. The impression seems to be that while the pareto distribution is best to estimate in the tail, then the lognormal has the best overall shape. It has been suggested that a useful approach to model loss data is nonparametric smoothing, since this ideally should be able to get the advantages of both the lognormal distribution and the pareto distribution. This approach is typically most suitable when the number of observations is large, as in the actuarial practice where the size of insurance

portfolios sometimes is very large. The standard actuarial curriculum considers kernel density estimation as a fundamental topic (Klugman et al. 1998). Actuaries, however, have not made an extensive use of nonparametric methods in this field. Our opinion is that actuaries will only use a method that is particularly good at estimating the density at the tails, a question that classical kernel smoothing fails to do. On the other hand, if nonparametric methods are to be widely accepted by practitioners, both actuaries and economists, there should be a rather simple way to produce the estimates. In this paper we show that a slightly adjusted version of the semiparametric transformed method of Wand, Marron and Ruppert (1991) is indeed suitable to study loss distributions. We show that the estimation is easy to implement (including the often complicated question of choosing the amount of smoothing) and that this method provides good and smooth estimates of the upper tail. Parametric models have often been justified due to their simplicity and their ability to solve the problem of lack of smoothness of the empirical approach. We now feel that we provide actuaries with a simple nonparametric alternative to their traditional parametric estimation techniques.

Actuaries are interested in having good estimates at all the values in the domain range: small losses because they are very frequent, medium losses causing a dramatic increase of expenses (demanding liquidity) and large losses that may mean that reinsurance contracts should be reconsidered. In this paper we estimate several actuarial loss function of automobile claims using nonparametric methods. We focus on our version of the semiparametric transformation approach to kernel smoothing introduced by Wand et al. (1991) and compare it with the standard kernel estimator and the multiplicative bias correction method (Hjort and Glad 1995; Jones, Linton and Nielsen 1995). We advocate in this paper that the semiparametric transformation method behaves excellently when it comes to estimating actuarial loss functions. It is a very suitable approach when estimating functions with a lognormal type of shape and a pareto type of tail. Loss functions have typically one mode for the low loss values and then a long heavy tail. We show by a simulation study that the method is able to estimate all three possible kind of tails, as defined in Embrechts, Klüppelberg and Mikosch (1999, p. 152), namely the Fréchet type, the Weibull type and the Gumbel type. This makes this method extremely powerful for actuaries at all levels, i.e. the non-life actuary calculating the risk of his auto claims, when a life actuary consider the risk of a group life insurance, or when these actuaries consider the relevant price for a reinsurance contract, where the reinsurer takes over the risk corresponding to the tail of the distribution. Besides, when the actuary calculates the total risk of a portfolio, perhaps using the famous recursion formula of Panjer (1981) or some more recent generalization of this formula, then the loss function of the individual claims is needed.

We show that our version of the semiparametric transformation principle of Wand et al. (1991) also is able to estimate the risk of a heavy-tailed distribution beyond the data. A heavy-tailed distribution is here of pareto shape. When we use our non-parametric estimator to fit the closest possible Pareto distribution in the tail we see, that at least for the estimator we tried out in our simulations, then our method is much better than the Hill estimator (Hill 1975), that is widely used in actuarial science and finance, see Embrechts et al. (1999, p. 330) and Danielson and Vries (1997). On top of this our estimation technique does not have to bother about where the tail begins such as the Hill estimator has to. This is often a difficult question and fatal errors leading to massive losses of insurance or reinsurance companies may happen. The beginning of the tail can be estimated by a methodology that can be compared to the trade-off between bias and variance in kernel density estimation, but the only (to our

knowledge) published method on this (Hall 1990) involves sub-bootstrapping, that is hard to understand and implement for the practitioner. For an expert implementation of Hill's method in the important field of estimating extreme values in finance, see Danielson and Vries (1997). Recently another method has appeared that so far has not reached the attention of an actuarial audience, namely the semiparametric robust estimator of Feuerverger and Hall (1999). Compared to this estimator, our estimator has the advantage of being directly linked to the non-parametrically estimated loss function. The connection between the loss function and the tail index is therefore immediate in our estimator of the tail index. We consider this to be a considerable competitive advantage of our method when it comes to the application in the fields of actuarial and financial loss models. We believe that our method is a big jump forward for the practitioner, whether it is an actuary or a financial analyst, to get a quick and easily understandable estimate of the loss distribution.

When it comes to the actual implementation of the method of Wand et al. (1991), we do deviate slightly since we only consider transformations that give a symmetric distribution. Firstly, we asked ourselves the question: "Why is it possible to estimate beyond the data in density estimation, when this is impossible in related estimation areas, such as regression and hazard estimation?" (Embrechts et al. 1999). The answer to this question seems to be that the density integrates to one. So, even if we do not have information on the tail, the restrictions on the density estimator provide us with valuable information. Acknowledging this logic, we decided to estimate the entire density at once using the method by Wand et al. (1991) but forcing the transformation to result in a symmetric density. Our simulation study shows that the approach has many advantages. We emphasize that our purpose in this paper has been to construct a good overall method for actuaries and financial analysts to use both for the entire domain of the loss function and for its tail. Therefore, we present a general method for practitioners to be used in a wide variety of situations, at least for a preliminary analysis of data. In our simulations we have used the Integrated Squared Error (ISE) and another Weighted Integrated Squared Error (WISE) that is more fit to evaluate the performance of actuarial loss functions in the tail. The weighted version penalizes the deviation at the tail, which corresponds to large claims (large amounts of money). In this regard, we do not follow the classical point of view that studies tail behavior using only the observations in the tail. We recommend the applied book of Reiss and Thomas (1997) for a solid introduction to the practical aspects of loss function estimation in insurance and finance. Embrechts et al. (1999) gives a more theoretical introduction to the area.

In sections 2 and 3, we study the properties of the transformation kernel density estimation and show the differences with the multiplicative bias corrected estimator with variable bandwidth. Section 4 describes the use of the shifted power transformation family. We apply these techniques to the automobile claims data, that are described in section 5. Section 6 presents the results, showing how to estimate actuarial loss distributions in practice. Finally, in section 7 we compare the transformation method to the standard kernel estimator (with boundary correction) and to the multiplicative bias correction estimator in a simulation study.

## 2 Use of a fixed transformation in kernel density estimation

We assume that we observe  $n$  independent identically distributed stochastic variables  $X_1, \dots, X_n$  with density  $f(\cdot)$  on the positive real line. The traditional nonparametric kernel estimator is:

$$\hat{f}(x) = n^{-1} \sum_{i=1}^n b^{-1} K \{b^{-1}(x - X_i)\} = n^{-1} \sum_{i=1}^n K_b(x - X_i),$$

where  $b$  is the bandwidth and the kernel  $K$  is a density. Now let  $F$  be some transformation function that is at least two times continuously differentiable.  $F$  does not have to be a cumulated density function. Then the estimator based on estimating the standard kernel density estimator of the transformed data and then transforming back is

$$\hat{f}(x, F) = F'(x) n^{-1} \sum_{i=1}^n K_b \{F(x) - F(X_i)\}.$$

We show below that this estimator has a close approximation to the multiplicatively bias corrected estimators with variable bandwidth  $b \{F'(x)\}^{-1}$  (for the fixed bandwidth case, see Hjort and Glad 1995). This estimator is

$$\hat{f}_M(x, F) = F'(x) n^{-1} \sum_{i=1}^n K_b \{F'(x)(x - X_i)\} \frac{F'(x)}{F'(X_i)}.$$

Following Hjort and Glad (1995) we get that

$$nb \left[ \hat{f}_M(x, F) - E \left\{ \hat{f}_M(x, F) \right\} \right] \Rightarrow N \{0, F'(x) f(x) \alpha(K)\}$$

and

$$E \left\{ \hat{f}_M(x, F) \right\} - f(x) = \frac{1}{2} k_2 b^2 \{F'(x)\}^{-1} \{f(x)/F'(x)\}'' + O_P(b^4),$$

where  $\alpha(K) = \int_{-\infty}^{+\infty} [K(t)]^2 dt$  and  $k_2 = \int_{-\infty}^{+\infty} t^2 K(t) dt$ . Hjort and Glad (1995) did assume that  $F$  was a probability distribution, but this condition is not necessary. We only need that  $F'(x) > 0$ . The variable bandwidth multiplicative estimator has the important weakness that it does not integrate to one. This can be expected to be of considerable importance when estimating the tail. We show below that the transformation method makes a clever local adjustment to the variable bandwidth multiplicative correction method such that it integrates to one.

Consider the bias term of the transformed estimator using that:

$$E \left\{ \hat{f}(x, F) \right\} = F'(x) \int K_b \{F(x) - F(y)\} f(y) dy.$$

We define  $F(x) - F(y) = F'(x)(x - y_0)$  and substitute variable  $y$ , then,

$$\begin{aligned} E \left\{ \hat{f}(x, F) \right\} &= F'(x) \int K_b \{F'(x)(x - y_0)\} \frac{F'(x)}{F'(y)} f(y) dy_0 \\ &= E \left\{ \hat{f}_M(x, F) \right\} + F'(x) \int K_b \{F'(x)(x - y_0)\} \left\{ \frac{f(y)}{F'(y)} - \frac{f(y_0)}{F'(y_0)} \right\} F'(x) dy_0. \end{aligned}$$

The difference between the bias of the variable bandwidth multiplicative bias correction and the transformation method is therefore a subtle local shift of the key fraction  $\frac{f}{F'}$  assuring that

the estimator integrates to one. This shift does also imply a slightly changed asymptotic bias. We analyze this by considering the second term above. Note that

$$\left\{ \frac{f(y)}{F'(y)} - \frac{f(y_0)}{F'(y_0)} \right\} = \left( \frac{f}{F'} \right)' (y^*)(y - y_0),$$

where  $y^*$  is between  $y_0$  and  $y$  and that

$$(y - y_0) = \frac{F''(x^*)}{2F'(x^{**})}(x - y_0)^2,$$

where  $x^*$  is between  $x$  and  $y_0$  and  $x^{**}$  is between  $y_0$  and  $y$ . This gives us the result that

$$\left\{ \frac{f(y)}{F'(y)} - \frac{f(y_0)}{F'(y_0)} \right\} = \left( \frac{f}{F'} \right)' (y^*) \frac{F''(x^*)}{2F'(x^{**})} (x - y_0)^2.$$

Now let

$$h(x) = \left( \frac{f}{F'} \right)' (x) \frac{F''(x)}{2\{F'(x)\}^2}.$$

Finally,

$$\begin{aligned} E \{ \hat{f}(x, F) \} &= E \{ \hat{f}_M(x, F) \} - k_2 b^2 h(x) + o_P(b^2) \\ &= \frac{1}{2} k_2 b^2 \left[ \left( \frac{f}{F'} \right)' (x) \frac{1}{\{F'(x)\}^2} \right]' + o_P(b^2). \end{aligned}$$

It is in addition easy to see that the asymptotic variances of  $E \{ \hat{f}(x, F) \}$  and  $E \{ \hat{f}_M(x, F) \}$  are asymptotically equivalent, such that,

$$nb \left[ \hat{f}(x, F) - E \{ \hat{f}(x, F) \} \right] \Rightarrow N \{ 0, F'(x) f(x) \alpha(K) \}.$$

### 3 Transformations in kernel density estimation

In this section we consider the question of estimating a suitable transformation and then use the same type of approach as above. Let

$$\mathbb{F} = \{ F_\theta \mid \theta \in \Theta \}$$

be a set of twice continuously differentiable transformations.  $\mathbb{F}$  could be a set of cumulated distribution functions or some other suitable set of functions.  $\Theta$  can correspond to a parametric family of transformation functions as in Wand et al.(1991) and also in Yang and Marron (1999) or it can correspond to a bigger nonparametric class of functions as in Ruppert and Cline (1994) and Hössjer and Ruppert (1995). For a given  $\theta$  the transformed density can be written as, see Wand et al. (1991),

$$\hat{f}(x, \theta) = F'_\theta(x) n^{-1} \sum_{i=1}^n K_b \{ F_\theta(x) - F_\theta(X_i) \},$$

where we have assumed that the transformations are differentiable.

Now let  $\hat{\theta}$  be some suitable estimator for  $\theta$ , then the resulting transformed density can be stated as

$$\hat{f}(x, \hat{\theta}) = F'_\theta(x) n^{-1} \sum_{i=1}^n K_b \{F_{\hat{\theta}}(x) - F_{\hat{\theta}}(X_i)\}.$$

Now assume that  $\hat{\theta}$  is a square-root-n consistent estimator of  $\theta$ . Then it is straightforward to notice that the asymptotic theory of  $\hat{f}(x, \hat{\theta})$  equals the asymptotic theory of  $\hat{f}(x, \theta)$ , so we get

$$E \{ \hat{f}(x, \hat{\theta}) \} = \frac{1}{2} k_2 b^2 \left[ \left( \frac{f}{F'_\theta} \right)' (x) \frac{1}{\{F'_\theta(x)\}} \right]' + o_P(b^2).$$

And for the asymptotic variance we get

$$nb \left[ \hat{f}(x, \hat{\theta}) - E \{ \hat{f}(x, \hat{\theta}) \} \right] \Rightarrow N \{0, F'_\theta(x) f(x) \alpha(K)\}.$$

**Remark 1** *It has been pointed out in Jones (1990) and Wand and Jones (1995, p. 44) that an argument based on Taylor expansion shows that the transformation method is related to both the known types of variable bandwidth methods, see Jones (1990), and that it can be considered as somewhere in between these two methods. While this is certainly correct, this approach gives no insight into the ability of the purely nonparametric transformation method to improve the rate of convergence of the resulting bias of the estimator, see Hössjer and Ruppert (1995), since the only variable window method that can reduce the rate of convergence of the bias is based on a variable bandwidth that is equal to the square root of the inverse density. Our interpretation says that the bias reducing property comes from the relation between the transformation method and multiplicative bias correction, see Hjort and Glad (1995) and Jones et al. (1995). Moreover the transformation method turns out to approximate to a variable bandwidth version of multiplicative bias correction.*

## 4 The shifted power transformation family

Transformed kernel estimation used here is based on a transformation of the observations of  $X_1, \dots, X_n$ . The purpose of this transformation is to obtain a new set of variables whose pdf's are approximately symmetric as in the normal distribution. The reason for this approach is three-fold. First, a transformation that results in a symmetric distribution is bound to have had a significant influence on a possible heavy tail of the original distribution. Secondly, it makes it quite reasonable to use a simple rule of thumb bandwidth selection rule while estimating the density of the transformed distribution. Thirdly, the boundary problem will more or less disappear since the transformed distribution can be expected to level off slowly. Therefore we do not bother about boundary kernels or other forms of correcting at the boundary (Jones and Signorini 1997; Zhang, Karunamuni and Jones 1999).

The reason for restricting ourselves to transformations resulting in symmetric functions is therefore not based on asymptotic properties, but rather that it seems practical from the point of view of applying the method to real data. The strategy of non-parametric statistics and model selection is to find natural and useful restrictions that do not bias the estimator too much, but help removing variability. We think the symmetry restriction lives up to the point.

Our approximation to this symmetry consideration is based on the estimated skewness that we force to be zero.

The shifted power transformation family used by Wand et al. (1991) is

$$y = g_\lambda(x) = \begin{cases} (x + \lambda_1)^{\lambda_2} & \text{if } \lambda_2 \neq 0 \\ \ln(x + \lambda_1) & \text{if } \lambda_2 = 0 \end{cases},$$

where  $\lambda = (\lambda_1, \lambda_2)$  with  $\lambda_1 > -\min(X_1, \dots, X_n)$  and  $\lambda_2 < 1$  (for right-skewed distributions). The parameters  $(\lambda_1, \lambda_2)$  will be called the transformation parameters. The density for the transformed variable is therefore:

$$f_y(y, \lambda) = f \left\{ g_\lambda^{-1}(y) \right\} \left( g_\lambda^{-1} \right)'(y).$$

For a given pair  $(\lambda_1, \lambda_2)$  the transformed density,  $f_y$ , can be estimated by the standard kernel density estimator and the following estimator of the original density  $f$  is:

$$\hat{f}(x, \lambda) = g_\lambda'(x) \hat{f}_y(y, \lambda) = g_\lambda'(x) n^{-1} \sum_{i=1}^n K_b(g_\lambda(x) - g_\lambda(X_i)).$$

As mentioned in the introduction, there are three classes of density tail, the Weibull (with light tail), the lognormal (with medium sized tail) and the Pareto (with a heavy tail) belong to each of the three classes, see Embrechts et al. (1999, p. 330). Note that the Pareto tail shape actually belongs to our class of transformations. Combining this fact with the asymptotic theoretical properties of our transformed estimator makes us expect that our estimator behaves extremely well in the tail of distributions of heavy tails. Our simulation study in section 7 confirms our beliefs.

To implement the transformation approach, a method to select the transformation parameters and the bandwidth is necessary. Our purpose is the following: firstly, we restrict ourselves to the set of  $\lambda$  parameters that approximately give zero skewness for the transformed data  $Y_1, \dots, Y_n$  (which have also been scaled to have the same variance as the original sample, see Wand et al. 1991). We define skewness as  $\hat{\gamma}_y = \left\{ n^{-1} \sum_{i=1}^n (Y_i - \bar{Y})^3 \right\} / \left\{ n^{-1} \sum_{i=1}^n (Y_i - \bar{Y})^2 \right\}^{\frac{3}{2}}$ , where  $\bar{Y}$  is the sample mean.

To select the  $\lambda$  parameter vector, we aim at minimizing the mean integrated square error (MISE) of  $\hat{f}(x, \lambda)$ , which can be approximated by:

$$\frac{5}{4} \left[ k_2 \alpha(K)^2 \right]^{\frac{2}{5}} \beta_y(f_y'')^{\frac{1}{5}} n^{-\frac{4}{5}}, \quad (1)$$

where  $\beta_y(f_y'') = \int_{-\infty}^{+\infty} [f_y''(y, \lambda)]^2 dy$  (Wand et al. 1991). Minimizing (1) with respect to the transformation parameters is equivalent to minimizing  $\beta_y(f_y'')$ . Hall and Marron (1987) suggested the following estimator:

$$\hat{\beta}_y(f_y'') = n^{-1}(n-1)^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n c^{-5} K * K \left\{ c^{-1} (Y_i - Y_j) \right\}, \quad (2)$$

where  $c$  is the bandwidth used for this estimation and can be estimated by minimizing the mean square error (MSE) of  $\hat{\beta}_y(f_y'')$ . When it is assumed that  $f_y$  is a normal distribution  $c$



can be estimated by  $\hat{c} = \hat{\sigma}_x \left( \frac{21}{40\sqrt{2}n^2} \right)^{\frac{1}{13}}$ , where  $\hat{\sigma}_x = \sqrt{n^{-1} \sum_{i=1}^n (Y_i - \bar{Y})^2}$  (Park and Marron 1990; Wand et al. 1991).

Once the transformation parameters have been estimated, we have to make the selection of the bandwidth that is going to be used for the transformation. Here we simply use the rule-of-thumb developed by Silverman (1986, p. 45) for a standard normal density. Since our transformation aims at a transformed density with zero skewness, this approach seems very plausible. The final estimator of the bandwidth  $b$  is therefore,  $\hat{b} = 1.059\hat{\sigma}_x n^{-\frac{1}{5}}$  and the corresponding transformation estimator will be called  $\hat{f}(x, \hat{\lambda}; \hat{b})$ .

## 5 Automobile claims data

In this section we present two data sets that will be used for the estimation of loss functions. In the first one, we use data pooled from several insurance companies in The Netherlands. We use these data to show the performance of the different kernel density estimation methods. In the second one, we use data on claims that were provided by a Spanish insurance company. We will use this application to compare the shape of the loss function in subpopulations.

### 5.1 Automobile claims in The Netherlands

The data used below were collected by the Centrum voor Verzekeringsstatistiek (CVS) in The Netherlands in 1992. Dutch insurance companies provide their claims information to the Centre for comparative purposes. This practice is typical in most European countries, where companies are member of a nationwide insurers union. A relatively large number of companies participated in this study, such that close to 50% of all car collision claims that occurred during 1992 in The Netherlands are included in the sample.

The data considered only the so-called the WA-claims, which means that they correspond to payments for legal liability and bodily injury expenses involved in the car accident and covered by the policy contract. The expenses include personal injury damage, property damage, third party bodily injury damage and any other liability damage. Kalb, Kofman and Vorst (1996) used a parametric model to analyze these data.

The data set contains 164,183 observation of the cost of single claims in Dutch guilders. Table 1 shows some descriptive measures of the sample.

The data set has the typical characteristics of claims data in insurance. It is a large sample of strictly positive values with a heavy tail.

### 5.2 Automobile claims within an insurance company

In this second example we use information on claims from accidents occurred in Spain in 1997 that were provided by one of the major insurance companies in this country. The costs refer only to bodily injury payments in thousands of pesetas. Damage payments were not considered here. The claims included in the sample come from insured policyholders that have the most basic type of insurance contract, namely, third party liability coverage. We only included those policyholders that make a private use of the car and that accepted their fault in the accident. All bodily injury costs caused by the accident were added together, so that costs resulted from the

addition of medical bills and fracture, disfigurement, dismemberment or death compensation. Only accidents with a non-zero payment were examined.

Table 1. Descriptive statistics for automobile claims in the Netherlands, 1992.

Statistic	
Sample mean	4,658.85
Standard deviation	20,673.03
Skewness	37.16
Kurtosis	2,217.41
Minimum	1.00
First quartile	1.00
Median	1,002.00
Third quartile	1,861.00
P <sub>95</sub>	13,077.00
P <sub>99</sub>	50,000.00
Maximum	1,870,000.00

In order to design new contracts or calculate fair premiums, the company may be interested in estimating the loss distributions using these data. We will present the density estimates for two groups in section 5. The first group includes claims of insured policyholders that are younger than 30 years old and the second group includes those policyholders that are 30 years old or more. Summary statistics are presented in Table 2 for the two groups.

Table 2. Descriptive statistics for bodily injury payments in automobile insurance.

Statistic	Age < 30 <sup>(a)</sup>	Age ≥ 30 <sup>(b)</sup>
Sample mean	402.70	243.05
Standard deviation	3,952.27	704.54
Skewness	30.42	10.90
Kurtosis	964.86	183.04
Minimum	1.00	1.00
First quartile	22.00	20.00
Median	66.00	68.00
Third quartile	171.00	182.00
P <sub>95</sub>	1,104.00	1,000.00
P <sub>99</sub>	4,600.00	3,000.00
Maximum	126,000.00	17,000.00

Sample sizes: <sup>(a)</sup>1,061 and <sup>(b)</sup> 4,062

It is interesting to note the difference between the two types of policyholders. The summary statistics show that the average paid amount is much higher for young insurers than for the older ones. The percentiles show that the difference occurs at the extreme values, which are more relevant in the subpopulation of young drivers. We would expect the tail to be heavier for policyholders that are under 30 years of age, meaning that it is much more likely to have large claims (extreme values) in this group. Using the loss distribution estimation we will be able to evaluate the difference of shape for the two groups.

## 6 Estimation of loss distributions

In this section we will apply the transformed kernel density estimation method described in section 3. In the first subsection we will detail the steps of the estimation procedure and we will compare the results with the bias correction methods presented in section 2. In the second subsection we will emphasize the use of the transformed kernel density estimation when assessing risk in heterogeneous portfolios.

### 6.1 Automobile collision claims in The Netherlands

The transformed kernel density estimate of the first data set is presented in Figure 1. We have plotted the density estimate in the interval from 1 to 100,000. We would like to point out the smoothed shape, also for the tail and the ability of the density estimate to capture the form of the tail. From the results obtained in the estimation steps, we concluded that the optimal parameter values for the transformation are  $\hat{\lambda}_1 = 863.12$  and  $\hat{\lambda}_2 = -0.59$ . The rule-of-thumb bandwidth is  $\hat{b} = 1,977.07$ . We will discuss the details of the estimation procedure below.

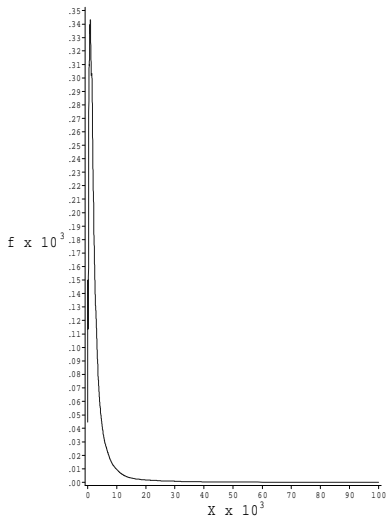


Figure 1. Transformed kernel density estimate of Dutch automobile collision claims.

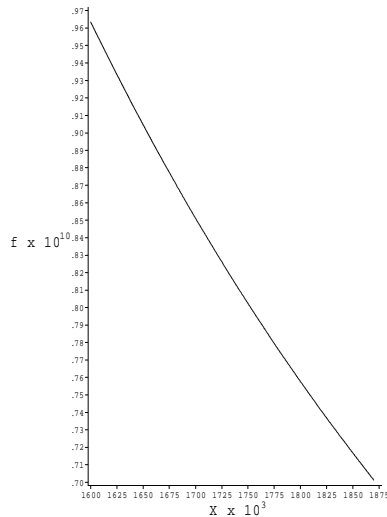


Figure 2. Transformed kernel density estimate of the tail of Dutch automobile collision claims.

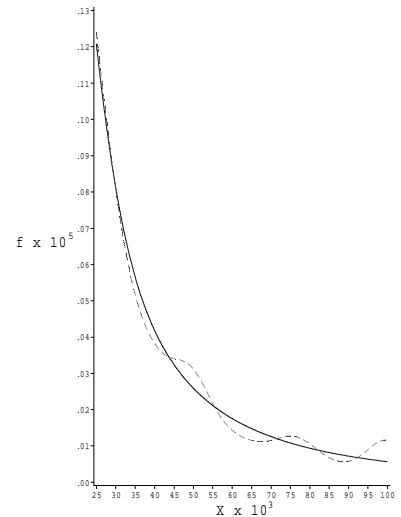


Figure 3. Transformed kernel (solid line) and multiplicative bias correction (dashed line) density estimate of Dutch automobile collision claims.

Figure 2 presents the transformed density estimate for the highest part of the domain. It concentrates on costs between 1,600,000 and 1,870,000 Dutch guilders. From this figure, the decreasing smoothed shape of the tail can be observed. Therefore, it can be seen that the transformed kernel density estimate approximate the tail well.

In Figure 3, we present the transformed kernel density estimate for auto claims between 25,000 and 100,000 Dutch guilders. In this figure, we compare the results of the transformed kernel density estimation method with those that are produced by the multiplicative bias correction method of Jones et al. (1995), using a bandwidth equal to  $b = 5,000$ . We conclude that multiplicative bias correction method produces bumps at the tail, due to the presence of scarce data in this interval. On the other hand, we should note that the multiplicative bias

correction method underestimates the main mode of the distribution, which is located at a lower cost value and is not shown in this picture. This means that, a wider window might eliminate the oscillations but would not be able to capture the shape for small claims. The simple bias correction kernel estimation method produces similar results. So, the transformed kernel density estimation method is suitable in this context because it provides a good and smooth estimate over the entire domain of the loss distribution.

The transformed kernel density estimation resulted from of the estimation of the transformation parameters  $\lambda = (\lambda_1, \lambda_2)$  and the selection of the smoothing parameter  $b$ . Figure 4 shows the skewness coefficient as a function of the transformation parameters for these data. It can be seen that there is a subset of parameters that make skewness equal to zero. Figure 5 plots (2) in the logarithm scale as a function of the transformation parameters only for the subset of pairs that make skewness equal to zero. The final transformation parameters were selected so that (2) was minimized.

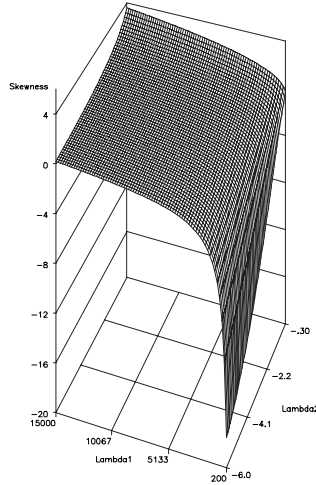


Figure 4. Skewness coefficient  $\hat{\gamma}_y$  as a function of  $\lambda = (\lambda_1, \lambda_2)$  for the Dutch automobile collision claims data.

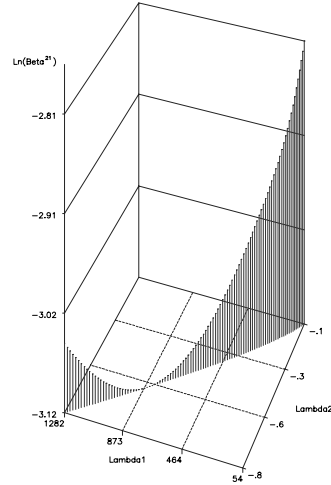


Figure 5.  $\ln(\hat{\beta}_y(f_y'') \times 10^{21})$  as a function of  $\lambda = (\lambda_1, \lambda_2)$  for the subset of pairs that make  $\hat{\gamma}_y = 0$ . Dutch automobile collision claims data.

## 6.2 Automobile collision claims within an insurance company

For the second data set, we present the potential usefulness of the method when applied in practice for the comparison of the loss distribution of two subgroups. For the transformed kernel estimation, we conclude that the estimated parameter values for the transformation are  $\hat{\lambda}_1 = 3.30$  and  $\hat{\lambda}_2 = -0.23$  for young policyholders and  $\hat{\lambda}_1 = 2.45$  and  $\hat{\lambda}_2 = -0.16$  for those policyholders over 30 years. The rule-of-thumb bandwidth is  $\hat{b} = 1,038.96$  for young insurers and  $\hat{b} = 141.62$  for the older ones. Finally, the transformed kernel density estimates for young (solid line) and for older policyholders (dashed line) are presented in Figure 6. In this picture, the plotted x-range concentrates on small claims (up to one millions pesetas), showing that the older drivers have a higher probability for small losses. In Figure 7, we plot the part of the density estimate that corresponds to large claims (those above 5 million pesetas). In the extreme-value region shown in Figure 7, the tail is lighter for the older policyholders and much

heavier for those drivers under 30 years. The methods discussed in section 2 are unable to provide such a smoothed version of the density simultaneously at low cost intervals and for high cost values. In this example, we are able to show that actuaries may use this method to compare the behavior of segments of the portfolio.

This information is very important specially for reinsurance, since it indicates that the companies using an excess-of-loss system should be aware that the limit to be set should not be the same for different age-groups. Therefore, the cost of reinsurance may differ for these two groups of insured persons, although one should also take into account the frequency of accidents involving bodily injury costs, because our variable is only measuring its severity. This result is in accordance to common knowledge that young drivers usually have larger claims than older ones. The reasons for this behavior are not unique and have long been discussed in the literature, they point out inexperience and moral hazard.

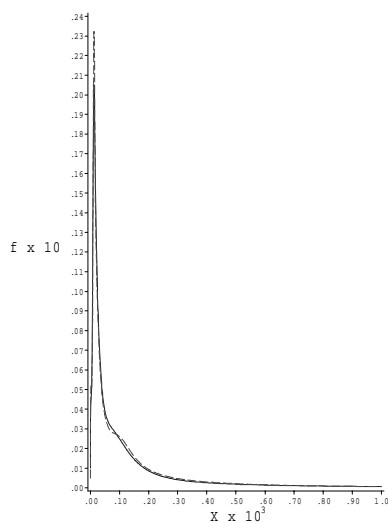


Figure 6. Transformed kernel density estimate of Spanish bodily injury automobile collision claims. For young (solid line) and for older policyholders (dashed line).

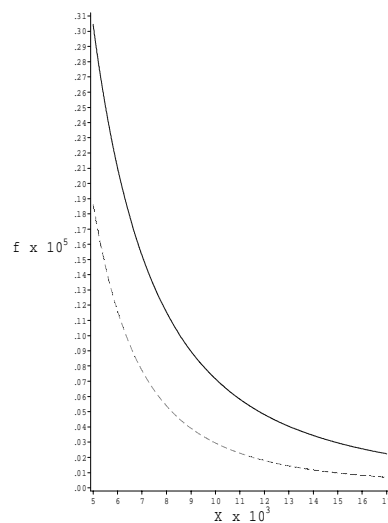


Figure 7. Transformed kernel density estimate of Spanish bodily injury automobile collision claims. For young (solid line) and for older policyholders (dashed line).

## 7 Simulation studies

In this section we present two simulation studies. In the first one, we compare the kernel density estimation methods for different types of distributions that usually arise in actuarial loss models. In the second one, we propose an alternative to the estimation of the tail index and compare our approach with the Hill estimator for the most typical loss distribution shape.

### 7.1 Performance of the kernel density transformation method

We have simulated different densities that are common in actuarial science because they generate skewed and heavy tailed distributions. We have selected three types of distributions: Weibull, lognormal and a mixture of a lognormal and a Pareto (Embrechts et al. 1999). The reason for this choice is twofold: on one hand we want to be exhaustive regarding the extreme value theory, and therefore we select three distributions within the three classes of generalized

extreme value distributions that characterize the behavior of tails. On the other hand, the simulated shapes are very similar to the ones observed in practice when studying financial loss functions. The lognormal hypothesis has often been used as the baseline model as well as the combination of lognormality and Pareto-like forms.

We compare our method with the standard kernel method (with boundary correction):

$$\widehat{f}(x) = n^{-1} \sum_{i=1}^n b^{-1} K_x \{b^{-1}(x - X_i)\} = n^{-1} \sum_{i=1}^n K_{x,b}(x - X_i), \quad (3)$$

where  $b$  is the bandwidth and the kernel  $K_{x,b}$  is the cut-and-normalized kernel (Gasser and Müller 1979). There exist of course more sophisticated ways of correction of a standard kernel estimator at the boundary (see Zhang et al. 1999, for an excellent overview), but we do prefer the simplest possible one for our purposes. This is also in line with the approach in the transformation method where we do not boundary correct at all.

We also consider the multiplicative bias correction estimator of Jones et al. (1995) that is defined as:

$$\widetilde{f}(x) = n^{-1} \sum_{i=1}^n K_{x,b}(x - X_i) \frac{\widehat{f}(x)}{\widehat{f}(X_i)}.$$

This method decreases bias both near the boundary and in interior regions. In the comparative study by Jones and Signorini (1997) this estimator was chosen as the overall best among a range of different bias reducing methods. Our simulation study shows that in the study of actuarial loss functions, the semiparametric transformation technique seems to be superior to the above two standard methods in particular if we concentrate our attention to the tail. As we mentioned in Section 2, this fact is extremely clear when it comes to a density with a heavy Pareto-type tail that is so important for the theory and practice of actuarial loss models.

We measured the performance of the methods by comparing the Integrated Squared Error (ISE) and the Weighted Integrated Squared Error (WISE), as described bellow. These were also used as optimality criteria.

The value of ISE is the  $L^2$  distance between the estimated density ( $\overline{f}(x)$ ) and the simulated density ( $f(x)$ ), namely:

$$ISE = \int_{-\infty}^{+\infty} \{\overline{f}(x) - f(x)\}^2 dx.$$

We also used the weighted measure of distance, the WISE, which is a generalization of the ISE.

$$WISE = \int_{-\infty}^{+\infty} \{\overline{f}(x) - f(x)\}^2 x^2 dx.$$

In our situation, the weighting function is  $x^2$ , therefore WISE is giving more importance to deviations in the tails. This also has a practical interpretation, since the  $x$ -values correspond to payments to be made by the insurer. This measure penalizes deviancies of large amounts.

One difficulty when comparing the different methods is the choice of the bandwidth. As in Jones and Signorini (1997) we will present a "best case" analysis. In this way, the best possible performance of each estimator is investigated. In every situation, we study the optimal bandwidth in terms of the value of  $b$  that minimizes the ISE and also the WISE. Nevertheless, the optimal bandwidth will not probably be achieved, so that the results will not hold in practice. For the transformation method, we also study the rule-of-thumb bandwidth choice. This is a

quick way to choose the smoothing parameter. We denote this transformation estimator by  $\hat{f}(x, \hat{\lambda}; \hat{b})$ , whereas  $\hat{f}(x, \hat{\lambda})$  denotes the transformation estimator with the in-obtainable optimal bandwidth.

Our simulation study covers three sample sizes: 100, 200 and 1,000. Every situation is replicated 100 times. For the lognormal case, we also simulate two different shapes, namely two parameter values. The four generated distributions are: a Weibull distribution with parameter  $c = 1.5$  and positive domain, a lognormal distribution with  $\mu = 0$  and  $\sigma = 0.5$ , a lognormal distribution with  $\mu = 0$  and  $\sigma = 1$ , and positive domain and, a mixture of a lognormal distribution with parameters  $\mu = 0$  and  $\sigma = 1$  and domain on  $[1, +\infty)$  and a Pareto distribution with parameter  $a = 1$ , also defined on  $[1, +\infty)$ . The mixture was 70% lognormal and 30% Pareto. As presented in Figure 8, these four distributions show increasingly heavier tails (Johnson, Kotz and Bulakrishnan 1994).

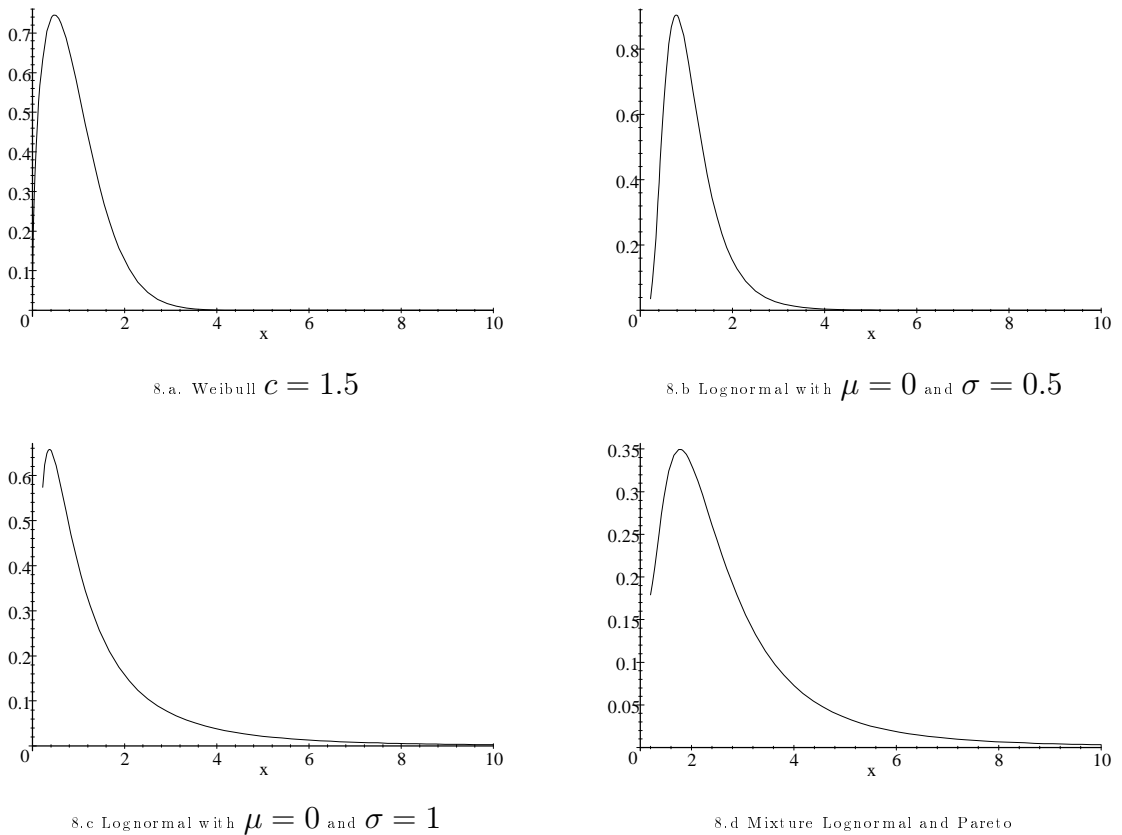


Figure 8. Simulated densities.

For each density estimator, we have calculated the mean, the median and the standard error of the ISE and the WISE. The results are given in tables 3 to 6, multiplied by  $10^3$ .

The transformation method with optimal bandwidth choice provides the best results both in terms of the ISE and the WISE. In general, the greatest advantage of our method is in the heavy tail situation.

In the Weibull case shown in Table 3, where the tail is not very heavy, the advantage of the transformation method is not so big. Nevertheless, for a large sample size, the  $\hat{f}(x, \hat{\lambda}; \hat{b})$

estimator is clearly better than estimators  $\hat{f}(x)$  and  $\tilde{f}(x)$  with the in-obtainable optimal bandwidths. A quite powerful result in favor of the transformation method. For the lognormal densities in Tables 4 and 5, we are looking at the shapes that have traditionally been used by actuaries for loss models. Again, we see that our estimator has the best performance and that the rule-of-thumb bandwidth estimator is a good choice. The estimators  $\hat{f}(x)$  and  $\tilde{f}(x)$  perform very similarly to one another for all sample sizes and they are inferior to the transformation method even they have the best possible in-obtainable bandwidth compared to the transformation method with a quick data-adaptive bandwidth choice.

When we consider the Pareto tail behavior in Table 6, we see that the transformation estimator  $\hat{f}(x, \hat{\lambda})$  has the greatest improvements with sample size, in particular if we look at the WISE. This means that if we give more importance to the tail estimation, including the weights, the results are even more favorable to the transformation method. As a matter of fact, the transformation method is almost a factor 10 better than the standard kernel methods. Such an improvement is quite substantial and way beyond the improvements normally obtained from bias reducing estimators (Jones and Signorini 1997).

As a general comment, for small sample sizes, the superiority of the transformation methods becomes more evident when the tail becomes heavier and that even when the rule-of-thumb bandwidth is chosen in the transformation method and the best possible bandwidth is selected for the estimators  $\hat{f}(x)$  and  $\tilde{f}(x)$  then the later cannot improve the results of  $\hat{f}(x, \hat{\lambda}; \hat{b})$ .

For all the densities and large samples sizes, the mean of the ISE and the mean of the WISE for  $\hat{f}(x, \hat{\lambda}; \hat{b})$  are very close to the ones obtained for  $\hat{f}(x, \hat{\lambda})$  when selecting the optimal bandwidth. For the estimators  $\hat{f}(x)$  and  $\tilde{f}(x)$  we must keep in mind that the bandwidth choice is a large and difficult topic, that will make their practical performance considerable worse. We therefore think that our results demonstrate that  $\hat{f}(x, \hat{\lambda}; \hat{b})$  is an interesting and easy-to-do proposal that behaves very satisfactorily - in particular in the heavy tail situation.

Table 3. Summary results of the simulation study for the Weibull ( $c = 1.5$ ).

		ISE			WISE		
		Mean	Med	Std	Mean	Med	Std
n=100	$\hat{f}(x)$	8.158	6.769	4.568	2.883	2.059	2.368
	$\tilde{f}(x)$	10.036	8.817	5.118	2.932	2.391	1.626
	$\hat{f}(x, \hat{\lambda})$	6.964	4.096	8.319	2.449	1.936	1.905
	$\hat{f}(x, \hat{\lambda}; \hat{b})$	8.327	5.569	8.648	3.256	2.772	2.320
n=200	$\hat{f}(x)$	4.929	4.089	2.708	1.629	1.089	1.250
	$\tilde{f}(x)$	6.327	5.471	3.068	1.838	1.521	1.158
	$\hat{f}(x, \hat{\lambda})$	3.750	2.521	3.641	1.495	0.950	1.305
	$\hat{f}(x, \hat{\lambda}; \hat{b})$	4.506	2.872	4.180	1.814	1.299	1.507
n=1000	$\hat{f}(x)$	1.734	1.662	0.689	0.561	0.452	0.375
	$\tilde{f}(x)$	2.256	2.112	0.779	0.649	0.610	0.326
	$\hat{f}(x, \hat{\lambda})$	0.849	0.680	0.676	0.404	0.300	0.302
	$\hat{f}(x, \hat{\lambda}; \hat{b})$	0.974	0.859	0.689	0.453	0.357	0.307



Table 4. Summary results of the simulation study for the lognormal ( $\sigma = 0.5$ ).

		ISE			WISE		
		Mean	Med	Std	Mean	Med	Std
n=100	$\hat{f}(x)$	7.164	5.656	6.252	7.508	5.673	6.747
	$\tilde{f}(x)$	7.210	5.830	6.441	7.085	5.111	6.104
	$\hat{f}(x, \hat{\lambda})$	7.780	6.863	6.118	5.603	5.039	4.090
	$\hat{f}(x, \hat{\lambda}; \hat{b})$	9.092	7.605	6.465	6.767	5.727	5.222
n=200	$\hat{f}(x)$	3.894	3.261	2.608	3.829	3.130	2.549
	$\tilde{f}(x)$	4.359	3.786	2.964	3.729	3.251	2.260
	$\hat{f}(x, \hat{\lambda})$	3.044	2.327	2.337	2.430	1.848	1.828
	$\hat{f}(x, \hat{\lambda}; \hat{b})$	3.836	3.293	2.421	3.021	2.504	1.884
n=1000	$\hat{f}(x)$	1.660	1.435	1.064	1.428	1.215	0.863
	$\tilde{f}(x)$	1.636	1.248	1.071	1.348	1.153	0.804
	$\hat{f}(x, \hat{\lambda})$	1.300	0.879	0.959	0.988	0.755	0.758
	$\hat{f}(x, \hat{\lambda}; \hat{b})$	1.541	1.133	0.991	1.180	0.940	0.785

Table 5. Summary results of the simulation study for the lognormal ( $\sigma = 1$ ).

		ISE			WISE		
		Mean	Med	Std	Mean	Med	Std
n=100	$\hat{f}(x)$	5.929	5.336	2.716	3.590	3.069	2.061
	$\tilde{f}(x)$	6.522	6.205	2.450	2.936	2.593	1.455
	$\hat{f}(x, \hat{\lambda})$	4.859	2.942	6.761	1.311	1.105	1.016
	$\hat{f}(x, \hat{\lambda}; \hat{b})$	5.981	3.801	7.515	1.720	1.340	1.335
n=200	$\hat{f}(x)$	4.624	4.303	1.977	2.649	2.339	1.459
	$\tilde{f}(x)$	5.527	5.361	1.674	2.420	2.338	0.901
	$\hat{f}(x, \hat{\lambda})$	2.295	1.655	2.167	0.753	0.594	0.585
	$\hat{f}(x, \hat{\lambda}; \hat{b})$	2.838	2.029	2.409	1.038	0.900	0.733
n=1000	$\hat{f}(x)$	1.487	1.294	0.665	0.670	0.613	0.266
	$\tilde{f}(x)$	1.947	1.840	0.733	0.829	0.780	0.278
	$\hat{f}(x, \hat{\lambda})$	0.615	0.453	0.490	0.197	0.162	0.127
	$\hat{f}(x, \hat{\lambda}; \hat{b})$	0.717	0.572	0.506	0.233	0.191	0.140

Table 6. Summary results of the simulation study for the Mixture ( $\sigma = 1, a = 1$ ).

		ISE			WISE		
		Mean	Med	Std	Mean	Med	Std
n=100	$\hat{f}(x)$	4.853	4.356	3.219	36.203	26.826	35.827
	$\tilde{f}(x)$	6.048	5.379	3.465	33.879	29.347	28.307
	$\hat{f}(x, \hat{\lambda})$	2.509	2.048	2.132	5.383	4.221	4.376
	$\hat{f}(x, \hat{\lambda}; \hat{b})$	3.308	2.416	2.559	7.351	5.665	6.070
n=200	$\hat{f}(x)$	2.765	2.251	1.942	25.888	16.790	29.224
	$\tilde{f}(x)$	3.767	3.073	2.132	24.128	17.834	21.315
	$\hat{f}(x, \hat{\lambda})$	1.810	1.338	1.619	3.767	2.914	3.042
	$\hat{f}(x, \hat{\lambda}; \hat{b})$	2.106	1.584	1.758	4.364	3.135	3.248
n=1000	$\hat{f}(x)$	0.994	0.867	0.532	10.654	5.606	15.337
	$\tilde{f}(x)$	1.513	1.333	0.584	10.260	6.534	11.508
	$\hat{f}(x, \hat{\lambda})$	0.548	0.352	0.596	1.170	0.810	1.098
	$\hat{f}(x, \hat{\lambda}; \hat{b})$	0.687	0.518	0.623	1.402	1.090	1.134

## 7.2 An alternative measure to the tail index

In this part, we want to study if the kernel density transformation estimation method can be useful to estimate the tail index ( $\alpha$ ), which is the shape parameter of the extreme value distribution (see, Reiss and Thomas 1997 and Dress, de Haan and Resnick 2000). We propose to estimate  $\alpha$  by minimizing the WISE of our density estimate to a Pareto with domain on  $[1, +\infty)$ .

We want to compare our estimates with the results given by the Hill estimator in the case where we simulate a mixture of a lognormal and a Pareto distribution, where the true tail exponent is  $\alpha = 1$ . We have used the 100 samples of size 100, 200 and 1,000 that were used in the simulation of the previous subsection. Then, we have taken the  $k\%$  of the observations in the upper tail,  $k \geq 5$  (if the sample size is 1,000, we have taken  $k \geq 0.5$ ). For every fixed  $k$ , we have estimated the tail index using the Hill's method. Figures 9, 10 and 11 show the mean squared error of the Hill estimates as a function of  $k$ , for the three sample sizes. The horizontal dashed line indicates the level of the mean squared error of the estimates when our approach is used.

As we can see, our approach to tail index estimation provides results that are much more accurate than the Hill estimation for all sample sizes. Another advantage is that our method eliminates the problem of selecting  $k$ .

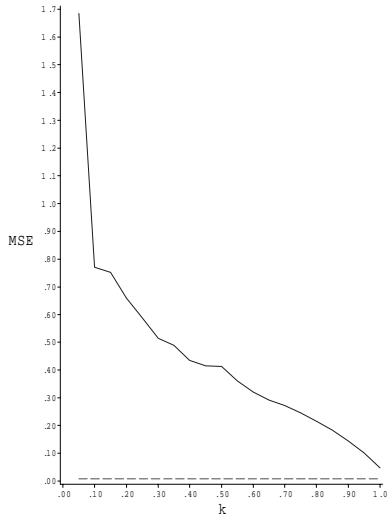


Figure 9. Mean Squared Error of the Hill estimator of  $\alpha$  as a function of the k% of the upper tail used (solid line) and Mean Squared Error of our approach (dashed line),  $n=100$ .

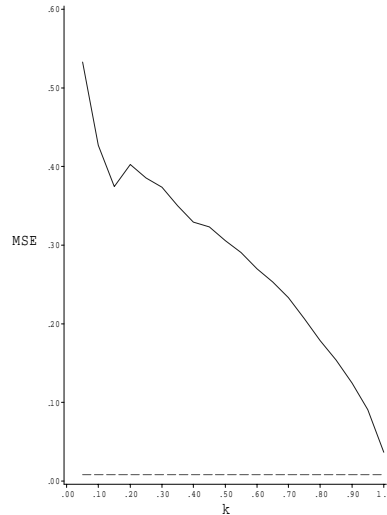


Figure 10. Mean Squared Error of the Hill estimator of  $\alpha$  as a function of the k% of the upper tail used (solid line) and Mean Squared Error of our approach (dashed line),  $n=200$ .

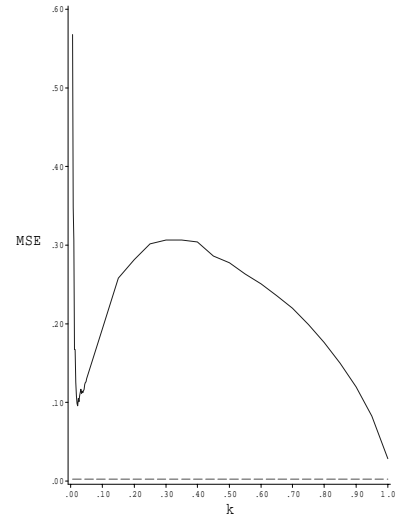


Figure 11. Mean Squared Error of the Hill estimator of  $\alpha$  as a function of the k% of the upper tail used (solid line) and Mean Squared Error of our approach (dashed line),  $n=1,000$ .

## 8 Conclusion

In this paper we have shown how the transformed kernel estimation method is used to estimate actuarial loss functions. We study the properties of this transformation estimator and we find a relationship to the properties of the multiplicative bias corrected kernel estimator with a variable bandwidth. We have compared our method with other kernel estimation approaches in this context using two data sets of automobile collision claims. We have also presented a simulation study that shows on the one hand that the method can successfully be implemented using the direct estimation of the bandwidth parameter and, on the other hand, that it can also be used as an alternative to the Hill estimator of the tail index. The importance of this methodology has also been emphasized. In the applications, showing that the transformation method is very useful for actuarial loss distribution analysis. The problems of bandwidth selection and transformation parameter estimation have also been overcome. As Jones and Signorini (1997) already pointed out, the semiparametric methods like ours seem to be a good direction for research.

## References

- Bolancé, C. (1999), *Estimación núcleo transformada en el análisis del coste de reclamaciones en el seguro del automóvil*, University of Barcelona, Ph.D. Thesis.
- Danielson, J., and Vries, C. G. (1997), “Tail index and quantile estimation with very high frequency data,” *Journal of Empirical Finance*, 4, 241-257.
- Daykin, C. D., Pentikäinen, T., and Pesonen, M. (1994), *Practical risk theory for actuaries*, London: Chapman & Hall.

- Drees, H., de Haan, L., and S. Resnick (in press), "How to make a Hill plot," *The Annals of Statistics*.
- Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997), *Modelling extremal events for Insurance and Finance*, Berlin: Springer.
- Feuerverger, A., and P. Hall (1999), "Estimating a tail exponent by modelling departure from a Pareto distribution," *The Annals of Statistics*, 27, 2, 760-781.
- Gasser, T., and Müller, H. G. (1979), "Kernel estimation of regression functions", in *Smoothing Techniques for Curve Estimation*, T. Gasser and M. Rosenblatt (eds), Berlin: Springer, 23-68.
- Hall, P. (1990), "Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems," *Journal of Multivariate Analysis*, 32, 177-203.
- Hall, P., and Marron, J. S. (1987), "Estimation of integrated squared density derivatives," *Statistics & Probability Letters*, 6, 407-413.
- Hill, B. M. (1975), "A simple general approach to inference about the tail of the distribution," *The Annals of Statistics*, 3, 1163-1174.
- Hjort, N. L., and Glad, I. K. (1995), "Non-parametric density estimation with a parametric start," *The Annals of Statistics*, 23, 882-904.
- Hoog, R., and Klugman, S. A. (1984), *Loss distributions*, New York: Wiley.
- Hössjer, O., and Ruppert, D. (1995), "Asymptotics for the transformation kernel density estimator," *The Annals of Statistics*, 23, 1198-1222.
- Johnson, N. L., Kotz, S., and Bulakrishnan, N. (1994), *Continuous Univariate Distributions*, Volume 1, New York: John Wiley and Sons, Inc.
- Jones, M. C. (1990), "Variable kernel density estimates and variable kernel density estimates." *Australian Journal of Statistics*. 32, 361-371.
- Jones, M. C., Linton, O. B., and Nielsen, J. P. (1995), "A simple and effective bias reduction method for density estimation," *Biometrika*, 82, 93-101.
- Jones, M. C., and Signorini, J. P. (1997), "A comparison of higher-order bias kernel density estimators," *Journal of the American Statistical Association*, 92, 1063-1073.
- Kalb, R. J. K., Kofman, P., and Vorst T. C. F. (1995), "Mixtures of tails in clustered automobile collision claims," *Insurance: Mathematics & Economics*, 18, 89-107.
- Klugman, S. A., Panjer H. A., and Willmot, G. E. (1998), *Loss models. From data to decisions*, New York: John Wiley & Sons, Inc.
- Panjer, H. (1981), "Recursive evaluation of a family of compound distributions" *Astin Bulletin*, 12, 22-26.

- Park, U. B., and Marron, J. S. (1990), "Comparison of data-driven bandwidth selectors," *Journal of the American Statistical Association*, 85, 66-72.
- Park, B. U., Kim, W. C., Ruppert, D., Jones, M. C., Signorini, D. F., and Kohn, R. (1997), "Simple transformation techniques for improved non-parametric regression," *Scandinavian Journal of Statistics*, 24, 145-163.
- Reiss, R. D., and Thomas, M. (1997), *Statistical Analysis of Extreme Values*, Basel: Birkhäuser Verlag.
- Rupper, D., and Cline, D. B. H. (1994), "Bias reduction in kernel density estimation by smoothed empirical transformations," *The Annals of Statistics*, 22, 185-210.
- Ruppert, D., and Wand M. P. (1992), "Correcting for kurtosis in density estimation," *Australian Journal of Statistics*, 34, 19-29.
- Silverman, B. W. (1986), *Density estimation for statistics and data analysis*, London: Chapman & Hall.
- Terrell, G. R. (1990), "The maximal smoothing principle in density estimation," *Journal of the American Statistical Association*, 85, 270-277.
- Terrell, G. R., and Scott, D. W. (1985), "Oversmoothed nonparametric density estimation," *Journal of the American Statistical Association*, 80, 209-214.
- Wand, M. P., and Jones, M. C. (1995), *Kernel Smoothing*, London: Chapman & Hall.
- Wand, M. P., Marron, J. S, and Ruppert, D. (1991), "Transformations in density estimation" (with comments), *Journal of the American Statistical Association*, 86, 343-361.
- Yang, L., and Marron, J. S. (1999), "Iterated Transformation-Kernel Density Estimation," *Journal of the American Statistical Association*, 94, 580-589.
- Zhang, S., Karunamuni, R. J., and Jones, M. C. (1999), "An improved estimator of the density function at the boundary," *Journal of the American Statistical Association*, 94, 1231-1241.