**DOCUMENT DE TREBALL**

**XREAP2010-3**

# An introduction to parametric and non-parametric models for bivariate positive insurance claim severity distributions

David Pitt
Montserrat Guillén (RFA-IREA)

XARXA DE REFERÈNCIA EN ECONOMIA APLICADA

# An introduction to parametric and non-parametric models for bivariate positive insurance claim severity distributions

David Pitt[1] & Montserrat Guillen[2]
[1] Dept. Economics, University of Melbourne
3010 Victoria, Australia
E-mail: dgpitt@unimelb.edu.au
[2] Dept. Econometrics, University of Barcelona
Diagonal, 690, 08034 Barcelona, Spain
E-mail: mguillen@ub.edu

February 28, 2010

**Abstract**

We present a real data set of claims amounts where costs related to damage are recorded separately from those related to medical expenses. Only claims with positive costs are considered here. Two approaches to density estimation are presented: a classical parametric and a semi-parametric method, based on transformation kernel density estimation. We explore the data set with standard univariate methods. We also propose ways to select the bandwidth and transformation parameters in the univariate case based on Bayesian methods. We indicate how to compare the results of alternative methods both looking at the shape of the overall density domain and exploring the density estimates in the right tail.

## 1 Introduction

We study a set of bivariate positive claims data from motor insurance (property damage and medical expenses costs). The main purpose of the analysis is to explore density estimation procedures, first on a univariate basis and then using a bivariate framework.

Estimation of a suitable bivariate density proves to be our main focus. Fitting an appropriate bivariate density is essential for optimal capital allocation (see, Denault, 2001; Panjer, 2002; Dhaene et al. 2003; Wang, 2002). Some authors have concentrated on deriving explicit forms for the allocation of each line when the loss random vector follows a certain multivariate distribution (see Valdez and Chernih, 2003, for the multivariate elliptical and see Vernic, 2006, for the multivariate skew-normal. Klugman et al. (2008) provide a comprehensive reference on the estimation of univariate and bivariate claims distribution models in insurance. McNeil et al. (2005) provide a review of financial risk measures in the context of estimating claims distributions in non-life insurance and in other settings.

Let us formulate the problem in a multivariate framework. We assume $d$ different types of losses (i.e. guarantees or lines of business). This means the total cost is the aggregate of several type of costs. We denote by $X_m$ the positive loss random variable for the $m$th type at the end of a certain period. Then the total or aggregate loss of the claim is $X = \sum_{i=1}^{d} X_m$. Let us assume that these random variables are continuous and that we are interested in estimating the multivariate probability density function of the random vector $(X_1, ..., X_d)'$, which we denote by $f(x_1, x_2, ..., x_d)$. A good estimate of the multivariate density is need for many actuarial problems, for instance premium calculations. We will denote the joint distribution function of the random vector $(X_1, ..., X_d)'$, $F(x_1, x_2, ..., x_d)$. The domain of this multivariate random variable is $\mathbb{R}^+ \times \cdot \times \mathbb{R}^+$.

Let us denote by $f_m(x)$ the probability density function of the random variable $X_m$ and $F_m(x)$ is its marginal distribution function, where $m = 1, ..., d$.

The marginal density function and the distribution function are unknown and need to be estimated from data. One approach to bivariate claim modelling that has been pursued is to use copulas. Whenever a copula is employed, it is denoted by $C(u_1, ..., u_d)$. The copula corresponding to the joint distribution can be expressed as a function of marginal distribution functions $F(x_1, x_2, ..., x_d) = C(F_1(x_1), ..., F_d(x_d))$.

In our particular case, the data set consists of a sample of claims that include two types of losses: property damage mainly resulting from third party liability and medical expenses that are not included in the Public Health system. Then the total claim cost is the addition of these two components. These data were already used in Bolancé et al. (2008) where both the bivariate skew-normal and normal distributions were fitted. Moreover, given that real data on claim amounts are usually positive and present right skewness, the bivariate lognormal and log-skew-normal distributions were also fitted by Bolancé et al. (2008) and a non-parametric estimation of the joint distribution function using a kernel density estimation method, was also proposed. The claim amounts in the original data set were expressed in thousands of pesetas. To express these in thousands of Euros we used the standard conversion and divided by 166,386.

Here we will start fitting univariate distributions and then we will explore the bivariate case.

# 2 Data set

The claims we considered refer to motor insurance of a major insurer in Spain for accidents that occurred in the year 2000. Data correspond to a random sample of all claims with both costs in property damage and to medical expenses.

Bodily injury is universally covered by the National Health System. This means that

Table 1: Univariate descriptive statistics for the positive claims data set (in 1,000 Euros

| | Mean | Std. Dev. | Skewness | Kurtosis | Min | Max |
|---|---|---|---|---|---|---|
| $X_1$ | 10.984 | 41.276 | 15.652 | 297.142 | 0.078 | 829.012 |
| $X_2$ | 1.706 | 5.188 | 8.037 | 82.019 | 0.006 | 71.250 |

$X_1$ is the cost of property damage and $X_2$ is the cost of medical expenses

medical costs considered here are medical expenses that are not included in the public system such as technical aids, drugs or chiropractic-related. Those expenses have to be paid by the insurer. No compensation for pain and suffering or loss of income are included. Medical expenses contain medical costs related to all those who were injured in the accident. Property damage expenses includes the insured's liability for damages he or she caused to vehicles, property or objects when the accident occurred.

The claims included in our sample are all claims that had already been settled. Although claims for compensations with bodily injury may take a long time to settle, these data were gathered in 2002, so that there has been enough time for the claimant to include most costs, so we consider that these are closed claims.

The sample size contains 518 claims, and for each claim $i$ we observe $X_1$ the cost of property damage and $X_2$ the cost of medical expenses expressed in thousands of euros.

## 2.1 Descriptive statistics

The main empirical characteristics are shown in Table 1.

In Figure 1 a plot of $X_1$ and $X_2$ is shown. From Figure 1 it is clear that the data are very bunched with a significant volume of small claims on both the property damage and additional medical expenses. In order to display the features of the data more clearly, we plot the data by transforming both components of the claim costs using natural logarithms. The resulting plot is shown in Figure 2.

We also provide univariate histograms of the individual claim data for both components of the claim costs. These are shown in Figure 3. On each of these histograms we have overlaid a normal probability density function, estimated for the data using the method of maximum likelihood. It is clear that a symmetric distribution, such as the normal, does not provide a good fit to these data. Much of the density under the fitted normal distribution relates to claim sizes smaller than the minimum observed claim value.

As a next step in the modelling, we investigate estimation using the log-normal distribution. Equivalently, we investigate taking the log transforms of each of the two components of our claim data set and fitting normal distributions to the resulting transformed data. Histograms of the log transformed data with overlaid normal density functions are shown in Figure 4. The improvement in fit obtained using the log-normal distribution compared to the normal distribution is apparent.

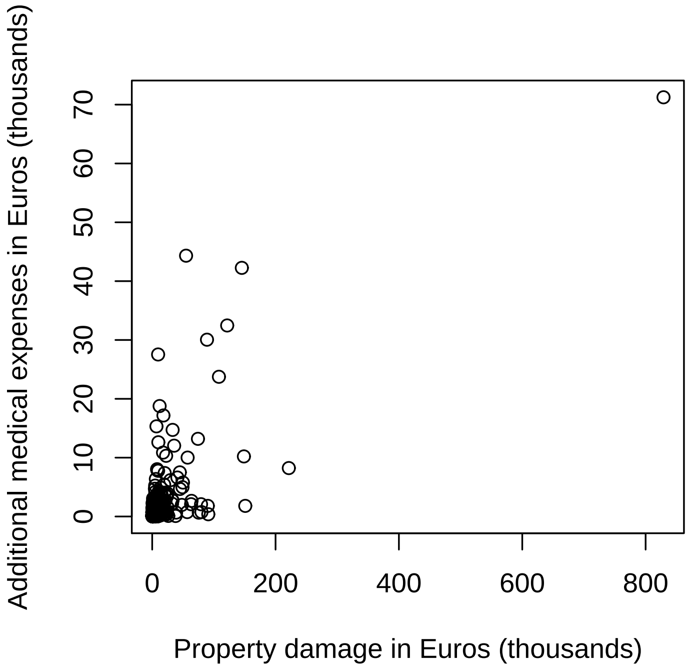



Figure 1: Plot of the positive claims data set

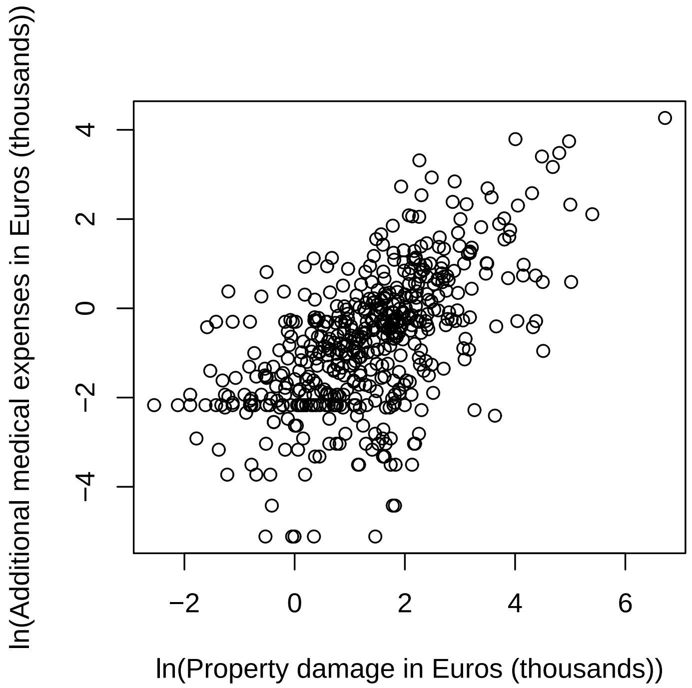
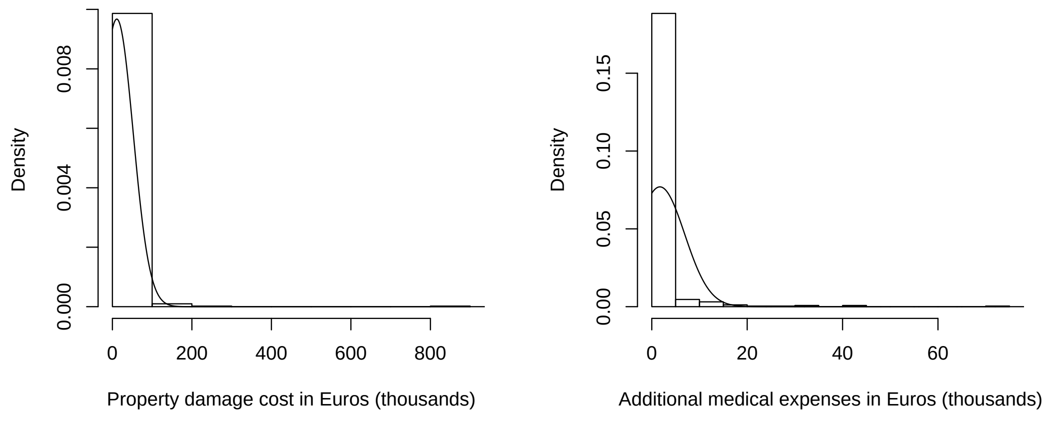



Figure 2: Plot of the logarithm of positive claims data set




(a) (b)

Figure 3: Histogram of univariate positive claims data set with a normal density overlaid

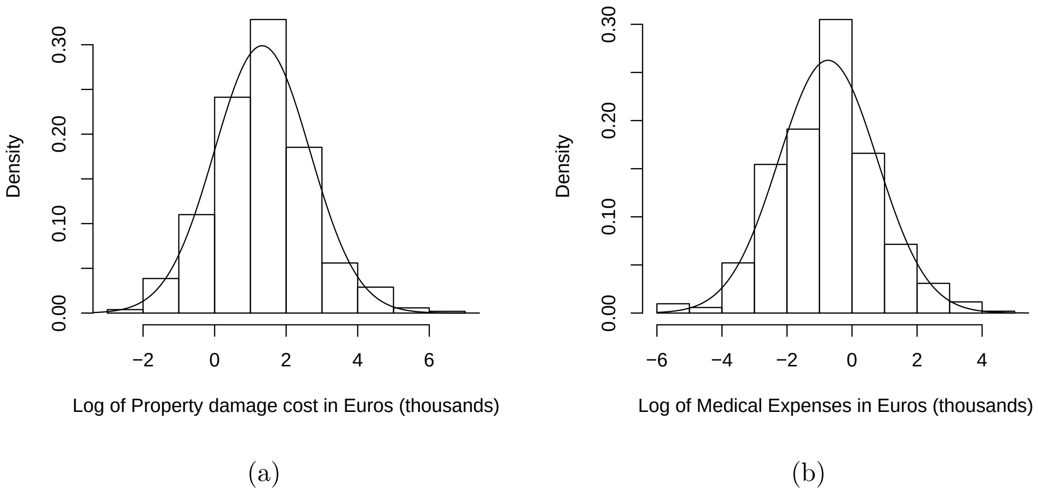



Figure 4: Histogram of univariate log of positive claims data set with a normal density overlaid

# 3 Kernel density estimation

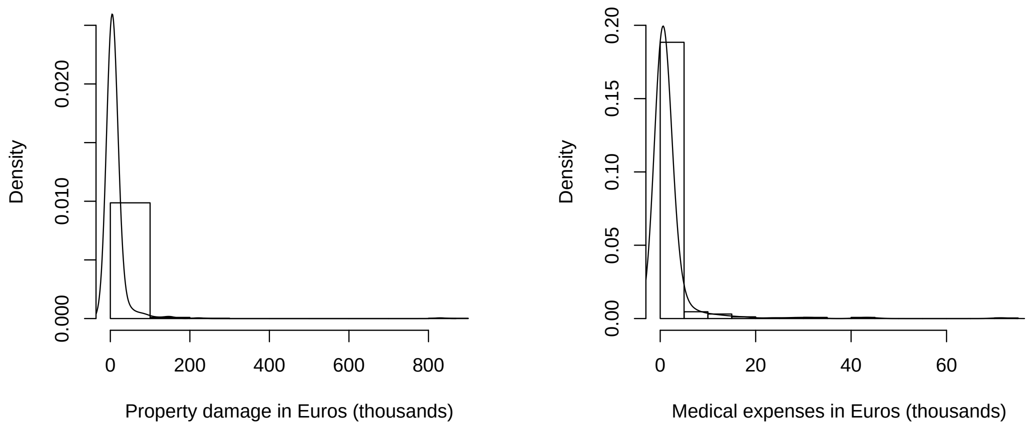
## 3.1 Classical kernel density estimation

For a random sample of $n$ independent and identically distributed observations $x_1, ..., x_n$ of a random variable $X$, the kernel density estimator is:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} k\left(\frac{x - x_i}{h}\right) \tag{1}$$

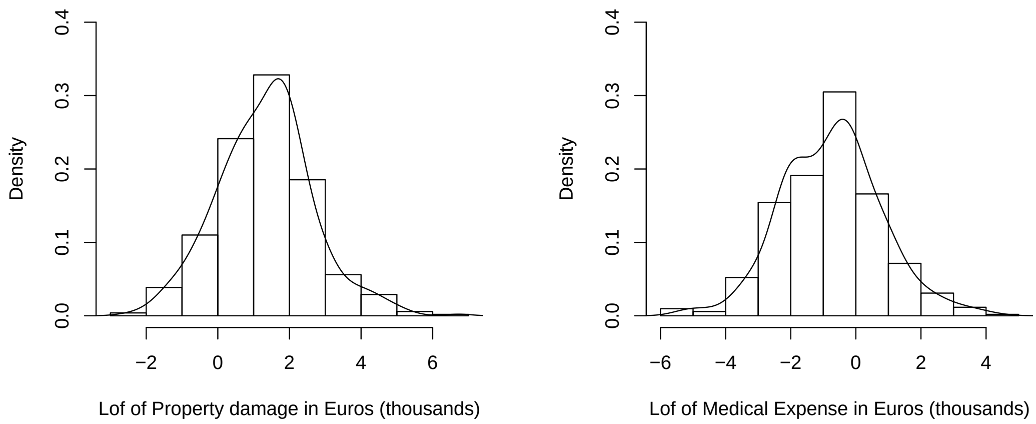
where $h$ is the bandwidth and $k(\cdot)$ is the kernel function. The bandwidth parameter is used to control the amount of smoothing in the estimation so that the greater $h$, the smoother the estimated density curve. The kernel function is usually a symmetric density with zero mean; in this work we will use a Gaussian kernel (see Wand and Jones, 1995). Many methods have been proposed for the selection of the bandwidth parameter in kernel density estimation. In this paper we work with the method of biased cross-validation, also described in Wand and Jones (1995). We provide kernel density estimates for both components of the univariate claims data and also for the log transformation of both components of the claims data. The resulting estimates are shown below in Figures 5 and 6.

Turning now to the bivariate case, a simple generalization of (1) is performed by means of product kernels (see Scott, 1992, pp. 150-155). More specifically, in the bivariate case, let us consider a random sample of $n$ independent and identically distributed pair observations $(x_{1i}, x_{2i})$, $i = 1, ..., n$, of the random vector $(X_1, X_2)'$. Then the kernel




(a) (b)

Figure 5: Histogram of univariate positive claims data set with kernel density estimate overlaid




(a) (b)

Figure 6: Histogram of univariate log positive claims data set with kernel density estimate overlaid

estimator of the bivariate density function can be expressed as

$$\hat{f}(x_1, x_2) = \frac{1}{nh_1h_2} \sum_{i=1}^{n} \mathbf{k}\left(\frac{x_1 - x_{1i}}{h_1}, \frac{x_2 - x_{2i}}{h_2}\right), \tag{2}$$

where $h_1$ and $h_2$ are bandwidths that, like in the univariate situation, are used to control the degree of smoothing. The function $\mathbf{k}\left(\frac{x_1 - x_{1i}}{h_1}, \frac{x_2 - x_{2i}}{h_2}\right) = k\left(\frac{x_1 - x_{1i}}{h_1}\right) k\left(\frac{x_2 - x_{2i}}{h_2}\right)$ is the product kernel.

## 3.2 Transformation kernel estimation

Classical kernel density estimation does not perform well when the true density is asymmetric. For instance, when one is interested in the density of the claim cost variable, the presence of many small claims produces a concentration of the mass near the low values of the domain and presence of some very large claims causes positive skewness.

The lack of information in the right tail of the domain makes it difficult to obtain a reliable nonparametric estimate of the density in that area. Many authors have worked with heavy-tailed distributions and have adapted kernel estimation methods to this context. Wand, et al. (1991), Clements et al. (2003), Bolancé et al. (2003), Buch-Larsen et al. (2005) and Bolancé et al. (2008) have proposed different transformation kernel density estimation methods, based on parametric families.

Let $T(.)$ be an increasing and monotonous transformation function. If the true density is right skewed, then the chosen transformation $T(.)$ must be a concave function. The transformation kernel estimation method (TKE) consists of transforming the original data so that the new transformed data can be assumed to have been generated from a symmetric random variable, and hence the true density of the transformed variable can easily be reliably approximated using the classical kernel estimation method. Using a change of variable, once the kernel estimation is obtained for the transformed variable, estimation on the original scale is also obtained.

Bolancé et al. (2003) proposed selecting the transformation function from a transformation family that is based on a generalization of the original power family (see Wand, et al., 1991),

$$T_{\lambda_1,\lambda_2}(x) = \begin{cases} (x + \lambda_1)^{\lambda_2} \, sig(\lambda_2) \\ \ln(x + \lambda_1) \end{cases}, \tag{3}$$

with $\lambda_1 \geq -\min(x_i, i = 1, ..., n)$ and $\lambda_2 \leq 1$ for right-skewed data. This parametric family of transformation functions is called the shifted power transformation family. Its main advantage is that it has a simple expression and works particularly well for transformation kernel estimation of asymmetric distributions. In order to estimate the optimal parameters of the shifted power transformation function, we can use the algorithm described by Bolancé et al. (2003).

Let us assume a sample of $n$ independent and identically distributed observations for variable $X_j$ $x_1, ..., x_n$ is available. Here we will omit subscript $j$ to simplify notation since we are only presenting the univariate method. We also assume that a transformation function $T(\cdot)$ is selected, then the data can be transformed so that $y_i = T(x_i)$, $i = 1, ..., n$. We denote the transformed sample by $y_1, ..., y_n$.

So the first step consists of transforming the data set with a function and afterwards estimating the density of the transformed data set using the classical kernel density estimator

$$\widehat{f}(y) = \frac{1}{nb} \sum_{i=1}^{n} K\left(\frac{y - y_i}{b}\right),$$

where $K$ is the kernel function, $b$ is the bandwidth and $y_i$, $i = \{1, ..., n\}$ is the transformed data set. The estimator of the original density is obtained by back-transformation of $\widehat{f}(y)$.

The transformed kernel density estimation method can be formulated as

$$\widehat{f}_T(x) = \frac{T'(x)}{n} \sum_{i=1}^{n} K_b \left\{T(x) - T(x_i)\right\},$$

where, as we mentioned, we have assumed that the transformations are differentiable. The superindex $'$ indicates the first derivative of a function. $K_b(\cdot) = \frac{1}{b} K(\cdot/b)$, where $K$ refers to the kernel function and $b$ is the bandwidth parameter.

## 3.3 Selecting the transformation parameters and the bandwidth

To implement the transformation approach, a method to select the transformation parameters and the bandwidth is necessary. Our approach is: firstly, we restrict ourselves to the set of $\lambda$ parameters that give approximately zero skewness for the transformed data $y_1, .., y_n$ (which have also been scaled to have the same variance as the original sample, see Wand et al., 1991). We define skewness as $\widehat{\gamma}_y = \left\{n^{-1} \sum_{i=1}^{n} (y_i - \overline{y})^3\right\} / \left\{n^{-1} \sum_{i=1}^{n} (y_i - \overline{y})^2\right\}^{\frac{3}{2}}$, where $\overline{y}$ is the sample mean of the transformed data.

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

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

where $\beta_y \left(f_y''\right) = \int_{-\infty}^{+\infty} \left[f_y''(y)\right]^2 dy$ (Wand et al. 1991). Minimizing (4) with respect to the transformation parameters is equivalent to minimizing $\beta_y \left(f_y''\right)$. Hall and Marron (1987) suggested the following estimator:

$$\widehat{\beta}_y \left(f_y''\right) = n^{-1}(n-1)^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} c^{-5} K * K \left\{c^{-1}(y_i - y_j)\right\}, \tag{5}$$

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

$c$ can be estimated by $\widehat{c} = \widehat{\sigma}_x \left( \frac{21}{40\sqrt{2}n^2} \right)^{\frac{1}{13}}$, where $\widehat{\sigma}_x = \sqrt{n^{-1} \sum_{i=1}^{n} (y_i - \overline{y})^2}$ (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, $\widehat{b} = 1.059\widehat{\sigma}_x n^{-\frac{1}{5}}$ and the corresponding transformation estimator will be called $\widehat{f}_T(x, \widehat{\lambda}; \widehat{b})$.

### 3.4 Choosing the transformation for kernel density estimation

One of the challenges in providing kernel density estimates with transformed data is the determination of a suitable transformation. The usual problem of estimating the bandwidth parameter is also present. Bolancé et al. (2008) suggest use of the Box-Cox transformation. Zhang et al. (2006) provide a Bayesian approach to the selection of a suitable bandwidth in multivariate kernel density estimation. We propose here an extension to the Bayesian analysis by Zhang et al. that will simultaneously consider the problem of estimating a suitable bandwidth and also determination of (a) suitable transformation parameter(s). The log transform, described earlier in this paper, will be one of the possible transformations possible under the Box-Cox set of possible transformations. The method described by Zhang et al. (2006) derives a posterior distribution of the bandwidth parameter, conditional on the observed data. Simulations from this posterior distribution, obtained using the method of Metropolis Hastings are obtained. The bandwidth parameter is then estimated as the mean of these posterior distribution simulations. The likelihood function used in this formulation is based on the density of claim costs integrated over the entire range. Given that the focus in non-life insurance is very often on the upper right tail of the distribution of possible outcomes, we will consider likelihood functions where greater weights are given to observations in the upper tail.

## 4 Measuring the goodness of fit

We are interested in evaluating the quality of our density estimates in the whole domain. Let us begin with the log-likelihood function. Since most of our parametric estimates have been found using MLE, then by comparing differences between log-likelihood estimates, we will be able to provide a straightforward measure of the goodness of fit.

Let us first concentrate on the univariate case. Let us assume that we have $\hat{f}(x)$ an estimate of the density for every point $x$ in the domain. Let us assume a sample of $n$ independent and identically distributed observations $x_1, ..., x_n$ is available. Then, we can estimate the log-likelihood function as:

$$ln\hat{L}(\hat{f}(\cdot); x_1, ..., x_n) = \sum_{i=1}^{n} ln\hat{f}(x_i).$$

If a transformation method was used, then instead of $\hat{f}(x_i)$ for $x_1, ..., x_n$, we have a transformed data set $y_1, ..., y_n$, with $y_i = T(x_i)$, $i = 1, ..., n$, where $T(\cdot)$ is the transformation. In this case, the estimated log-likelihood function is:

$$ln\hat{L}(\hat{f}_T(\cdot); T(\cdot); x_1, ..., x_n) = \sum_{i=1}^{n} ln\hat{f}_T(x_i).$$

Note that $\hat{f}_T(\cdot)$ is the transformation estimate of the true density $f(\cdot)$, and it holds that:

$$\hat{f}_T(x_i) = \hat{f}(T(x_i))T'(x_i),$$

where $\hat{f}(T(x_i))$ is the classical kernel estimation on the transformed data set and $T'(\cdot)$ is the first derivative of the transformation function.

We will evaluate improvements in the likelihood, using the difference between $-2ln\hat{L}(\hat{f}_T(\cdot); x_1, ..., x_n)$ for our estimated model and a baseline model.

We also need to formally generalize the previous goodness of fit statistics in two ways. Firstly, we would like to provide a statistic that would give more weight to the tail of the distribution. Secondly, we will generalize this procedure to a two-dimensional case.

## 4.1 Weighted pseudo-log-likelihood: univariate

A weighted pseudo-log-likelihood can be estimated if weights $w_i$, $i = 1, ..., n$ are included preceeding each summation term as:

$$ln_w\hat{L}(\hat{f}_T(\cdot); x_1, ..., x_n) = \sum_{i=1}^{n} w_i ln\hat{f}_T(x_i).$$

If $w_i = 1$, $i = 1, ..., n$, then we would have the same log-likelihood expression, but we can also use some distance as a weight, so that observations that are located close to the centre of the distribution have much less importance than those located in the tail.

We have tried two different expressions for weights. The first one is giving more weight to those observations that are distant from 0. Note that our data are always positive. If we wanted to generalize for both positive and negative values, then we should take absolute values of the data values. The form of the weights is:

$$w_i^{(1)} = \frac{nX_i}{\sum_{i=1}^{n} x_i}.$$

Using these weights in the estimated weighted pseudo-log-likelihood implies that more importance is given to the fit in the tail. Then, since for a given $i$, we have that $ln\hat{f}_T(x_i)$ is negative and it is smaller when $\hat{f}_T(x_i)$ tends to zero (which is exactly what happens in the long tail region) then weighting those summation terms more, means that the $ln_w\hat{L}(\hat{f}_T(\cdot); x_1, ..., x_n)$ is going to be smaller than $ln\hat{L}(\hat{f}_T(\cdot); x_1, ..., x_n)$. Nevertheless, we are going to evaluate goodness of fit by comparing $-2ln_w\hat{L}(\hat{f}_T(\cdot); x_1, ..., x_n)$ for a density estimate and the one obtained for a baseline model, we should not compare this to the estimated log-likelihood where no weights were considered.

The second possible form for the weights is inspired by the same principle as the weighted integrated mean squared error that was proposed in Bolancé et al.(2003), where contributions to the average where weighted with a squared distance. In this case:

$$w_i^{(2)} = \frac{nx_i^2}{\sum_{i=1}^n x_i^2}$$

When a transformation is used, the corresponding expression would be:

$$ln_w \hat{L}(\hat{f}_T(\cdot); T(\cdot); x_1, ..., x_n) = \sum_{i=1}^n w_i ln \hat{f}_T(x_i) = \sum_{i=1}^n w_i ln \left( \hat{f}(T(x_i))T'(x_i) \right),$$

where $w_i$ can ether be equal to $w_i^{(1)}$ or $w_i^{(2)}$.

## 4.2 Weighted pseudo-log-likelihood for the multivariate case

In order to obtain a general expression for bivariate observations $x_i = (x_{1i}, x_{2i})$, $i = 1, ..., n$, we will use a distance measure as a weight. Distance is the eclidean distance to the $(0,0)$ point, therefore, we can define:

$$\mathbf{w}_i^{(1)} = \frac{n\sqrt{x_{1i}^2 + x_{2i}^2}}{\sum_{i=1}^n \sqrt{x_{1i}^2 + x_{2i}^2}}$$

and

$$\mathbf{w}_i^{(2)} = \frac{n(x_{1i}^2 + x_{2i}^2)}{\sum_{i=1}^n (x_{1i}^2 + x_{2i}^2)}.$$

In the bivariate setting, we can define:

$$ln_{\mathbf{w}} \hat{L}(\hat{f}(\cdot); x_1, ..., x_n) = \sum_{i=1}^n \mathbf{w}_i ln \hat{f}(x_{1i}, x_{21})$$

and then use either $\mathbf{w}_i^{(1)}$ or $\mathbf{w}_i^{(2)}$.

When a transformation is used in the bivariate setting.

Suppose $(Y_1, Y_2)' = T(X_1, X_2)'$, then

$$ln_{\mathbf{w}} \hat{L}(\hat{f}_T(\cdot); T(\cdot); x_1, ..., x_n) = \sum_{i=1}^n \mathbf{w}_i ln \hat{f}_T(x_{1i}, x_{2i}) = \tag{6}$$

$$\sum_{i=1}^n \mathbf{w}_i ln f(T(x_{1i}, x_{2i})) \begin{vmatrix} \frac{\partial Y_1}{\partial x_1}(x_{1i}, x_{2i}) & \frac{\partial Y_1}{\partial x_2}(x_{1i}, x_{2i}) \\ \frac{\partial Y_2}{\partial x_1}(x_{1i}, x_{2i}) & \frac{\partial Y_2}{\partial x_2}(x_{1i}, x_{2i}) \end{vmatrix} \tag{7}$$

# 5 Conclusions

In this paper we fitted several univariate distributions and a kernel density to a real data set from motor insurance.

The kernel estimation approach provides a smoothed version of the empirical distribution. We also provided details of goodness of fit criteria based on standard likelihood theory and also using weighted likelihoods where greater weight is given to density estimation in the right tail of the distribution. This is going to be further developed in the transformation kernel density estimation for the multivariate case.

# References


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

[2] Bolancé, C., Guillén, M., Pelican, E. and Vernic, R., 2008. Skewed bivariate models and nonparametric estimation for the CTE risk measure. Insurance: Mathematics and Economics 43, 3, 386-393.

[3] Bolancé, C., Guillén, M. and Nielsen, J.P., 2008. Inverse Beta transformation in kernel density estimation. Statistics & Probability Letters, in press.

[4] 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.

[5] 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.

[6] Denault, M., 2001. Coherent allocation of risk capital. Working paper. Ecole des H.E.C., Montreal.

[7] Dhaene, J., Goovaerts, M.J. and Kaas, R., 2003. Economic capital allocation derived from risk measures. North American Actuarial Journal 7, 44-59.

[8] Panjer, H.H., 2002. Measurement of risk, solvency requirements and allocation of capital within financial conglomerates. 27th International Congress of Actuaries, Cancun 2002. (see also http://www.actuaries.org/events/congresses/Cancun/afir_subject/afir_14_panjer.pdf)

[9] Klugman, S.A., Panjer, H.H, and Willmot, G.E., 2008. Loss models: from data to decisions. 3rd Edition. John Wiley & Sons, New Jersey.

[10] McNeil, A.J., Frey. R. and Embrechts, P., 2005. Quantitative risk management: concepts, techniques and tools. Princeton University Press., Princeton.

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

[12] Scott, D.W., 1992. Multivaraite Density Estimation. Theory, Practice and Visualization. John Wiley & Sons, Inc.

[13] Valdez, E.A. and Chernih, A., 2003. Wang's capital allocation formula for elliptically contoured distributions. Insurance: Mathematics and Economics 33, 517-532.

[14] Vernic, R. 2006. Multivariate Skew-Normal distributions with applications in insurance. Insurance: Mathematics and Economics 38, 413-426.

[15] Wand, M.P. and Jones, M.C., 1995. Kernel Smoothing. Chapman & Hall.

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

[17] Wang, S., 2002. A set of new methods and tools for enterprise risk capital management and portfolio optimization. Working paper. SCOR reinsurance company (http://www.casact.com/pubs/forum/02sforum/02sf043.pdf).

[18] 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.

[19] Zhang, X., King, M.L. and Hyndman, R.J., 2006. A Bayesian approach to bandwidth selection for multivariate kernel density estimation. Computational Statistics & Data Analysis 50, 3009-3031.

## 2006

**CREAP2006-01**
**Matas, A.** (GEAP)**; Raymond, J.Ll.** (GEAP)
"Economic development and changes in car ownership patterns"
(Juny 2006)

**CREAP2006-02**
**Trillas, F.** (IEB)**; Montolio, D.** (IEB)**; Duch, N.** (IEB)
"Productive efficiency and regulatory reform: The case of Vehicle Inspection Services"
(Setembre 2006)

**CREAP2006-03**
**Bel, G.** (PPRE-IREA)**; Fageda, X.** (PPRE-IREA)
"Factors explaining local privatization: A meta-regression analysis"
(Octubre 2006)

**CREAP2006-04**
**Fernàndez-Villadangos, L.** (PPRE-IREA)
"Are two-part tariffs efficient when consumers plan ahead?: An empirical study"
(Octubre 2006)

**CREAP2006-05**
**Artís, M.** (AQR-IREA)**; Ramos, R.** (AQR-IREA)**; Suriñach, J.** (AQR-IREA)
"Job losses, outsourcing and relocation: Empirical evidence using microdata"
(Octubre 2006)

**CREAP2006-06**
**Alcañiz, M.** (RISC-IREA)**; Costa, A.; Guillén, M.** (RISC-IREA)**; Luna, C.; Rovira, C.**
"Calculation of the variance in surveys of the economic climate"
(Novembre 2006)

**CREAP2006-07**
**Albalate, D.** (PPRE-IREA)
"Lowering blood alcohol content levels to save lives: The European Experience"
(Desembre 2006)

**CREAP2006-08**
**Garrido, A.** (IEB)**; Arqué, P.** (IEB)
"The choice of banking firm: Are the interest rate a significant criteria?"
(Desembre 2006)

**CREAP2006-09**
**Segarra, A.** (GRIT)**; Teruel-Carrizosa, M.** (GRIT)
"Productivity growth and competition in spanish manufacturing firms:
What has happened in recent years?"
(Desembre 2006)

**CREAP2006-10**
**Andonova, V.; Díaz-Serrano, Luis.** (CREB)
"Political institutions and the development of telecommunications"
(Desembre 2006)

**CREAP2006-11**
**Raymond, J.L.**(GEAP)**; Roig, J.L..** (GEAP)
"Capital humano: un análisis comparativo Catalunya-España"
(Desembre 2006)

**CREAP2006-12**
**Rodríguez, M.**(CREB)**; Stoyanova, A.** (CREB)
"Changes in the demand for private medical insurance following a shift in tax incentives"
(Desembre 2006)

**CREAP2006-13**
**Royuela, V.** (AQR-IREA)**; Lambiri, D.**; **Biagi**, **B.**
"Economía urbana y calidad de vida. Una revisión del estado del conocimiento en España"
(Desembre 2006)

**CREAP2006-14**
**Camarero, M.; Carrion-i-Silvestre, J.LL.** (AQR-IREA)**.**;**Tamarit**, **C.**
"New evidence of the real interest rate parity for OECD countries using panel unit root tests with breaks"
(Desembre 2006)

**CREAP2006-15**
**Karanassou, M.; Sala, H.** (GEAP)**.**;**Snower** , **D. J.**
"The macroeconomics of the labor market: Three fundamental views"
(Desembre 2006)

**2007**

**XREAP2007-01**
**Castany, L** (AQR-IREA)**; López-Bazo, E.** (AQR-IREA)**.**;**Moreno** , **R.** (AQR-IREA)
"Decomposing differences in total factor productivity across firm size"
(Març 2007)

**XREAP2007-02**
**Raymond, J. Ll.** (GEAP)**; Roig, J. Ll.** (GEAP)
"Una propuesta de evaluación de las externalidades de capital humano en la empresa"
(Abril 2007)

**XREAP2007-03**
**Durán, J. M.** (IEB)**; Esteller, A.** (IEB)
"An empirical analysis of wealth taxation: Equity vs. Tax compliance"
(Juny 2007)

**XREAP2007-04**
**Matas, A.** (GEAP)**; Raymond, J.Ll.** (GEAP)
"Cross-section data, disequilibrium situations and estimated coefficients: evidence from car ownership demand"
(Juny 2007)

**XREAP2007-05**
**Jofre-Montseny, J.** (IEB)**; Solé-Ollé, A.** (IEB)
"Tax differentials and agglomeration economies in intraregional firm location"
(Juny 2007)

**XREAP2007-06**
**Álvarez-Albelo, C.** (CREB)**; Hernández-Martín, R.**
"Explaining high economic growth in small tourism countries with a dynamic general equilibrium model"
(Juliol 2007)

**XREAP2007-07**
**Duch, N.** (IEB)**; Montolio, D.** (IEB); **Mediavilla, M.**
"Evaluating the impact of public subsidies on a firm's performance: a quasi-experimental approach"
(Juliol 2007)

**XREAP2007-08**
**Segarra-Blasco, A.** (GRIT)
"Innovation sources and productivity: a quantile regression analysis"
(Octubre 2007)

**XREAP2007-09**
**Albalate, D.** (PPRE-IREA)
"Shifting death to their Alternatives: The case of Toll Motorways"
(Octubre 2007)

**XREAP2007-10**
**Segarra-Blasco, A.** (GRIT); **Garcia-Quevedo, J.** (IEB); **Teruel-Carrizosa, M.** (GRIT)
"Barriers to innovation and public policy in catalonia"
(Novembre 2007)

**XREAP2007-11**
**Bel, G.** (PPRE-IREA); **Foote, J.**
"Comparison of recent toll road concession transactions in the United States and France"
(Novembre 2007)

**XREAP2007-12**
**Segarra-Blasco, A.** (GRIT);
"Innovation, R&D spillovers and productivity: the role of knowledge-intensive services"
(Novembre 2007)

**XREAP2007-13**
**Bermúdez Morata, Ll.** (RFA-IREA); **Guillén Estany, M.** (RFA-IREA), **Solé Auró, A**. (RFA-IREA)
"Impacto de la inmigración sobre la esperanza de vida en salud y en discapacidad de la población española"
(Novembre 2007)

**XREAP2007-14**
**Calaeys, P.** (AQR-IREA); **Ramos, R.** (AQR-IREA), **Suriñach, J**. (AQR-IREA)
"Fiscal sustainability across government tiers"
(Desembre 2007)

**XREAP2007-15**
**Sánchez Hugalbe, A.** (IEB)
"Influencia de la inmigración en la elección escolar"
(Desembre 2007)

**2008**

**XREAP2008-01**
**Durán Weitkamp, C.** (GRIT)**; Martín Bofarull, M.** (GRIT) ; **Pablo Martí, F.**
"Economic effects of road accessibility in the Pyrenees: User perspective"
(Gener 2008)

**XREAP2008-02**
**Díaz-Serrano, L.; Stoyanova, A. P.** (CREB)
"The Causal Relationship between Individual's Choice Behavior and Self-Reported Satisfaction: the Case of Residential Mobility in the EU"
(Març 2008)

**XREAP2008-03**
**Matas, A.** (GEAP)**; Raymond, J. L.** (GEAP)**; Roig, J. L.** (GEAP)
"Car ownership and access to jobs in Spain"
(Abril 2008)

**XREAP2008-04**
**Bel, G.** (PPRE-IREA) **; Fageda, X.** (PPRE-IREA)
"Privatization and competition in the delivery of local services: An empirical examination of the dual market hypothesis"
(Abril 2008)

**XREAP2008-05**
**Matas, A.** (GEAP); **Raymond, J. L.** (GEAP); **Roig, J. L.** (GEAP)
"**Job accessibility and employment probability**"
(Maig 2008)

**XREAP2008-06**
**Basher, S. A.; Carrión, J. Ll.** (AQR-IREA)
Deconstructing Shocks and Persistence in OECD Real Exchange Rates
(Juny 2008)

**XREAP2008-07**
**Sanromá, E.** (IEB)**; Ramos, R.** (AQR-IREA); Simón, H.
Portabilidad del capital humano y asimilación de los inmigrantes. Evidencia para España
(Juliol 2008)

**XREAP2008-08**
**Basher, S. A.; Carrión, J. Ll.** (AQR-IREA)
Price level convergence, purchasing power parity and multiple structural breaks: An application to US cities
(Juliol 2008)

**XREAP2008-09**
**Bermúdez, Ll.** (RFA-IREA)
A priori ratemaking using bivariate poisson regression models
(Juliol 2008)

**XREAP2008-10**
**Solé-Ollé, A.** (IEB), **Hortas Rico, M.** (IEB)
Does urban sprawl increase the costs of providing local public services? Evidence from Spanish municipalities
(Novembre 2008)

**XREAP2008-11**
**Teruel-Carrizosa, M.** (GRIT), **Segarra-Blasco, A.** (GRIT)
Immigration and Firm Growth: Evidence from Spanish cities
(Novembre 2008)

**XREAP2008-12**
**Duch-Brown, N.** (IEB), **García-Quevedo, J.** (IEB), **Montolio, D.** (IEB)
Assessing the assignation of public subsidies: Do the experts choose the most efficient R&D projects?
(Novembre 2008)

**XREAP2008-13**
**Bilotkach, V.**, **Fageda, X.** (PPRE-IREA), **Flores-Fillol, R.**
Scheduled service versus personal transportation: the role of distance
(Desembre 2008)

**XREAP2008-14**
**Albalate, D.** (PPRE-IREA), **Gel, G.** (PPRE-IREA)
Tourism and urban transport: Holding demand pressure under supply constraints
(Desembre 2008)

## 2009

**XREAP2009-01**
**Calonge, S.** (CREB)**; Tejada, O.**
"A theoretical and practical study on linear reforms of dual taxes"
(Febrer 2009)

**XREAP2009-02**
**Albalate, D.** (PPRE-IREA)**; Fernández-Villadangos, L.** (PPRE-IREA)
"Exploring Determinants of Urban Motorcycle Accident Severity: The Case of Barcelona"
(Març 2009)

**XREAP2009-03**
**Borrell, J. R.** (PPRE-IREA)**; Fernández-Villadangos, L.** (PPRE-IREA)
"Assessing excess profits from different entry regulations"
(Abril 2009)

**XREAP2009-04**
**Sanromá, E.** (IEB)**; Ramos, R.** (AQR-IREA), **Simon, H.**
"Los salarios de los inmigrantes en el mercado de trabajo español. ¿Importa el origen del capital humano?"
(Abril 2009)

**XREAP2009-05**
**Jiménez, J. L.; Perdiguero, J.** (PPRE-IREA)
"(No)competition in the Spanish retailing gasoline market: a variance filter approach"
(Maig 2009)

**XREAP2009-06**
**Álvarez-Albelo,C. D.** (CREB)**, Manresa, A.** (CREB)**, Pigem-Vigo, M.** (CREB)
"International trade as the sole engine of growth for an economy"
(Juny 2009)

**XREAP2009-07**
**Callejón, M.** (PPRE-IREA)**, Ortún V, M.**
"The Black Box of Business Dynamics"
(Setembre 2009)

**XREAP2009-08**
**Lucena, A.** (CREB)
"The antecedents and innovation consequences of organizational search: empirical evidence for Spain"
(Octubre 2009)

**XREAP2009-09**
**Domènech Campmajó, L.** (PPRE-IREA)
"Competition between TV Platforms"
(Octubre 2009)

**XREAP2009-10**
**Solé-Auró, A.** (RFA-IREA),**Guillén, M.** (RFA-IREA)**, Crimmins, E. M.**
"Health care utilization among immigrants and native-born populations in 11 European countries. Results from the Survey of Health, Ageing and Retirement in Europe"
(Octubre 2009)

**XREAP2009-11**
**Segarra, A.** (GRIT)**, Teruel, M.** (GRIT)
"Small firms, growth and financial constraints"
(Octubre 2009)

**XREAP2009-12**
**Matas, A.** (GEAP)**, Raymond, J.Ll.** (GEAP), **Ruiz, A.** (GEAP)
"Traffic forecasts under uncertainty and capacity constraints"
(Novembre 2009)

**XREAP2009-13**
**Sole-Ollé, A.** (IEB)
"Inter-regional redistribution through infrastructure investment: tactical or programmatic?"
(Novembre 2009)

**XREAP2009-14**
**Del Barrio-Castro, T.**, **García-Quevedo, J.** (IEB)
"The determinants of university patenting: Do incentives matter?"
(Novembre 2009)

**XREAP2009-15**
**Ramos, R.** (AQR-IREA), **Suriñach, J.** (AQR-IREA)**, Artís, M.** (AQR-IREA)
"Human capital spillovers, productivity and regional convergence in Spain"
(Novembre 2009)

**XREAP2009-16**
**Álvarez-Albelo, C. D.** (CREB), **Hernández-Martín, R.**
"The commons and anti-commons problems in the tourism economy"
(Desembre 2009)

**2010**

**XREAP2010-01**
**García-López, M. A.** (GEAP)
"The Accessibility City. When Transport Infrastructure Matters in Urban Spatial Structure"
(Febrer 2010)

**XREAP2010-02**
**García-Quevedo, J.** (IEB), **Mas-Verdú, F.** (IEB), **Polo-Otero, J.** (IEB)
"Which firms want PhDs? The effect of the university-industry relationship on the PhD labour market"
(Març 2010)

**XREAP2010-03**
**Pitt, D.**, **Guillén, M.** (RFA-IREA)
"An introduction to parametric and non-parametric models for bivariate positive insurance claim severity distributions"
(Març 2010)

xreap@pcb.ub.es