Local Transformation Kernel Density Estimation of Loss Distributions

Gustafsson, J.; Hagmann, M.; Nielsen, J. P.; Scaillet, Olivier

How to cite

GUSTAFSSON, J. et al. Local Transformation Kernel Density Estimation of Loss Distributions. In: Journal of business & economic statistics, 2009, vol. 27, n° 2, p. 161–175. doi: 10.1198/jbes.2009.0011

This publication URL: https://archive-ouverte.unige.ch/unige:79875

Publication DOI: 10.1198/jbes.2009.0011

Local Transformation Kernel Density Estimation of Loss Distributions

J. Gustafsson[a] M. Hagmann[b] J.P. Nielsen[c] O. Scaillet[d]

Revised version: June 2007

Abstract

We develop a tailor made semiparametric asymmetric kernel density estimator for the estimation of actuarial loss distributions. The estimator is obtained by transforming the data with the generalized Champernowne distribution initially fitted to the data. Then the density of the transformed data is estimated by use of local asymmetric kernel methods to obtain superior estimation properties in the tails. We find in a vast simulation study that the proposed semiparametric estimation procedure performs well relative to alternative estimators. An application to operational loss data illustrates the proposed method.

Key words and phrases: Actuarial loss models, Transformation, Champernowne distribution, asymmetric kernels, local likelihood estimation.

JEL Classification: C13, C14.

[a] Royal&SunAlliance and University of Copenhagen, Copenhagen, Denmark.

[b] Swiss Finance Institute and Concordia Advisors, London, United Kingdom.

[c] CASS Business School, City University, London.

[d] HEC Genève and Swiss Finance Institute, Geneva, Switzerland.

1 Introduction

The specification of a loss distribution is without doubt a key ingredient of any modeling approach in actuarial science and financial risk management. For insurers, a proper assessment of the size of a single claim is of most importance. Loss distributions describe the probability distribution of a payment to the insured. Traditional methods in the actuarial literature use parametric specifications to model single claims. The most popular specifications are the lognormal, Weibull and Pareto distributions. Hogg and Klugman (1984) and Klugman, Panjer and Willmot (1998) describe a set of continuous parametric distributions which can be used for modeling a single claim size. Ideally we would like to benefit from economic guidance in the choice of a parametric loss model. This is sometimes available in other fields. For example Banks, Blundell and Lewbel (1997) consider demand systems that support empirical specification of quadratic Engel curves. Economic justification of empirical loss distribution is still an open question. Besides it is unlikely that something as complex as the generating process of insurance claims can be described by just a few parameters. A wrong parametric specification may lead to an inadequate measurement of the risk contained in the insurance portfolio and consequently to a mispricing of insurance contracts. Such a mispricing can be in some cases rather severe.

A method which does not require the specification of a parametric model is nonparametric kernel smoothing. This method provides valid inference under a much broader class of structures than those imposed by parametric models. Unfortunately, this robustness comes at a price. The convergence rate of nonparametric estimators is slower than the parametric rate, and the bias induced by the smoothing procedure can be substantial even for moderate sample sizes. Since losses are positive variables, the standard kernel estimator proposed by Rosenblatt (1956) has a boundary bias. This boundary bias is due to weight allocation by the fixed symmetric kernel outside the support of the distribution when smoothing close to the boundary is carried out. As a result, the mode close to the boundary typical for loss distributions is often missed. Additionally, standard kernel methods yield wiggly estimation in the tail of the distribution since the mitigation of the boundary bias leads to favor a small bandwidth which prevents pooling enough data. This prevents precise tail measurement of loss distributions which is of primary importance to get appropriate risk measures when designing an efficient risk management system.

We propose a semiparametric estimation framework for the estimation of densities which have support on $[0, \infty)$. We build on transformation based kernel density estimation, which involves nonparametric estimation of a density on $[0, 1]$. Our estimation procedure can deal with all problems of the standard kernel estimator mentioned previously, and this in a single way. Although the parametric models traditionally used in loss modeling may be inaccurate, they can be used in a semiparametric fashion to help to decrease the bias induced by nonparametric smoothing. If the parametric model is accurate, the performance of our semiparametric estimator can be close to pure parametric estimation. Following the idea of Hjort and Glad (1995) (H&G), we start with a parametric estimator of the unknown density instead of a direct single step nonparametric estimation. Then we work with transformed data as in, e.g., Wand et al. (1991). We map the original data within $[0, 1]$ via the estimated parametric start distribution, and correct nonparametrically for possible misspecification. Our approach combines the benefits of using a global parametric start and of using transformation based kernel estimators. We will see that the standard transformation based kernel estimator of Wand et al. (1991) is a special case of our estimator. Since a flexible parametric start helps to mitigate the potential misspecification of the parametric model and gives full access to the potential benefit of bias reduction, we opt for the Champernowne distribution (Champernowne (1936, 1952)) and its modification developed by Buch-Larsen et al. (2005). This distribution has already been successfully applied to the estimation of loss distributions in Buch-Larsen et al. (2005) and Gustafsson et al. (2006). It is especially suitable to model loss data since it has a similar shape close to the boundary as the lognormal model, which is considered to provide a good fit for losses of smaller size. However, unlike the medium-tailed lognormal model, the Champernowne distribution and its modification converge in the tail to the heavy tailed pareto distribution and can therefore also capture the typical thick tailed feature exhibited by empirical loss data. The transformation step can be seen as a type of variance stabilization (denoising) procedure as traditionally used in signal extraction.

To decrease the bias even further, we give some local parametric guidance to this nonparametric correction in the spirit of Hjort and Jones (1996) (H&J). This is achieved by employing either local polynomial or log polynomial models, where the latter method results always in nonnegative density estimates. In contrast with H&J the correction is applied on the transformed data instead of the

original ones. The idea is that in a well specified case the transformed data are uniformly distributed on $[0,1]$ (see Diebold et al. (1998) for use of this property in evaluating density forecasts in financial risk management), and that under slight misspecification they are close to. Then the deviations from the uniform distribution should be easier to handle locally than the true distribution itself. We call this approach *local transformation bias correction*, or LTBC to be short.

We emphasize that appropriate boundary bias correction arising from smoothing on the unit interval is more important in a semiparametric than a pure nonparametric setting. This is because the bias reduction achieved by semiparametric techniques allows us to increase the bandwidth and thus to pool more data. This, however, increases the boundary region where the symmetric kernel allocates weight outside the unit interval. This motivates us to develop LTBC in an asymmetric kernel framework which eliminates the boundary issue completely. Asymmetric kernel estimators for compactly supported data were recently proposed by Brown and Chen (1999) and Chen (1999) as a convenient way to solve the boundary bias problem. The symmetric kernel is replaced by an asymmetric beta kernel which matches the support of the unknown density of the transformed data. Other remedies include the use of particular boundary kernels or bandwidths, see e.g. Rice (1984), Schuster (1985), Jones (1993), Müller (1991) and Jones and Foster (1996). Chen (2000) and Scaillet (2004) discuss some of their disadvantages such as inferior performance on Monte carlo experiments, potential negative density values, as well as complex and slow computational implementation. The beta kernel has a flexible form, is located on the interval $[0,1]$, and produces nonnegative density estimates. Also, it changes the amount of smoothing in a natural way as one moves away from the boundaries. This is particularly attractive when estimating densities which have areas sparse in data because more data points can be pooled. Empirical loss distributions typically have long tails with sparse data, and therefore we think that such a kernel is well suited to solve the boundary problem for the transformed data at the right end of the unit interval. The variance advantage of the asymmetric kernel comes, however, at the cost of a slightly increased bias as we move away from the boundaries compared to symmetric kernels. This highlights the importance of effective bias reduction techniques within the boundaries achieved by local modeling in the spirit of H&J. Another advantage of the beta kernel estimator is its consistency even if the true density of the transformed data is unbounded at $x = 0$ and/or $x = 1$ (see Bouezmarni and Rolin (2004)). In a vast simulation

study, we find that the proposed asymmetric semiparametric density estimation framework exhibits very attractive finite sample performance. Across a range of six different light to heavy tailed test densities our proposed estimator outperforms popular boundary corrected kernel estimators such as the local constant and local linear estimator.

We proceed as follows in this paper: Section 2 explains the semiparametric density estimation framework in the familiar symmetric kernel setting. Section 3 briefly discusses the beta kernel density estimator which is used to extend the proposed semiparametric estimation framework to an asymmetric setting in Section 4. Section 5 briefly discusses the parametric part of our framework given by the generalized Champernowne distribution. Section 6 summarizes the setting and results of a vast simulation study. In Section 7 we provide an application to operational risk data which illustrates the use of the proposed estimation framework. Such a quantitative assessment of operational risk is in line with the new proposal of the Basel Committee on Banking Supervision. Section 8 finally concludes.

2 Local transformation based kernel density estimation

Let $X_1, ..., X_n$ be a random sample from a probability distribution F with an unknown density function $f(x)$ where x has support on $[0, \infty)$. We propose the following local model as a basis to estimate the true density function $f(x)$:

$$m\left(x, \theta_1, \theta_2\left(T_{\theta_1}(x)\right)\right) = T_{\theta_1}^{(1)}(x) \cdot \phi\left(T_{\theta_1}(x), \theta_2\left(T_{\theta_1}(x)\right)\right), \tag{1}$$

where $T_{\theta_1}(x)$ is a parametric family of cumulative distribution functions indexed by the global parameter $\theta_1 = (\theta_{11}, ..., \theta_{1p}) \in \Theta_1 \in \mathbf{R}^p$ and $T_{\theta_1}^{(1)}(x)$ denotes the first derivative of T, e.g. the probability density function in this case. The first part of the local model m consists of the density $T_{\theta_1}^{(1)}(x)$, which serves as a global parametric start and is assumed to provide a meaningful but potentially inaccurate description of the true density $f(x)$. In that sense we follow H&G. The second part of m denoted by $\phi\left(T_{\theta_1}(x), \theta_2\left(T_{\theta_1}(x)\right)\right)$ with $\theta_2\left(T_{\theta_1}(x)\right) = \left(\theta_{21}\left(T_{\theta_1}(x)\right), ..., \theta_{2q}\left(T_{\theta_1}(x)\right)\right) \in \Theta_2 \in \mathbf{R}^q$ serves as the local parametric model for the unknown density function $r(T_{\theta_1}(x))$ of the transformed data $T_{\theta_1}(X_1), ..., T_{\theta_1}(X_1)$ with support in $[0, 1]$. In that sense we follow Wand et al. (1991) by working with transformation based kernel estimators. If T were equal to the true distribution function F,

then r would just be the uniform density. In a misspecified case, however, the role of this "correction density" is, as the name says, to correct the potentially misspecified global start density $T_{\theta_1}^{(1)}(x)$ towards the true density $f(x)$. We call this local transformation bias correction (LTBC), since the approach is inspired by transformation based kernel density estimation but the correction density r is modeled locally by ϕ as inspired by H&J. As already mentioned, the correction density $r(T_{\theta_1}(x))$ is uniformly equal to one if the parametric start is well specified. Hence when the degree of misspecification is not too severe it is intuitively more natural to model the correction density locally than the unknown density itself.

The estimation procedure is as follows: first, estimate the parameter θ_1, which does not depend on x, by maximum likelihood and denote the estimate by $\hat{\theta}_1$. It is well known that when the parametric model $T_{\theta_1}^{(1)}(x)$ is misspecified, $\hat{\theta}_1$ converges in probability to the pseudo true value θ_1^0 which minimizes the Kullback-Leibler distance of $T_{\theta_1}^{(1)}(x)$ from the true $f(x)$, see e.g. White (1982) and Gourieroux, Monfort and Trognon (1984). Second, we estimate the density of the transformed data $\hat{U}_i = T_{\hat{\theta}_1}(X_i)$ at point $\hat{u} = T_{\hat{\theta}_1}(x)$ by using the local likelihood approach of H&J: Choose $\theta_2(\hat{u})$, i.e., $\hat{\theta}_2$, such that

$$\frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_h(\hat{U}_i - \hat{u})v(\hat{u}, \hat{U}_i, \theta_2) - \int \mathcal{K}_h(t-\hat{u})v(\hat{u}, t, \theta_2)\phi(t, \theta_2)\,dt = 0 \tag{2}$$

holds, where $\mathcal{K}_h(z) = (1/h)\mathcal{K}(z/h)$ is a symmetric kernel function, h is the bandwidth parameter and $v(\hat{u}, t, \theta_2)$ is a $q \times 1$ vector of weighting functions. We denote the solution to the above equation by $\hat{\theta}_2(\hat{u})$ and its true theoretical counterpart by $\theta_2^0(u^0)$, where $u^0 = T_{\theta_1^0}(x)$. If we choose the score $\partial \log \phi(u, \theta_2)/\partial\theta_2$ as the weighting function, then Equation (2) is just the first order condition of the local likelihood function given in H&J. In general, the form of the weighting function is driven by the tractability of the implied resulting estimator and is discussed in more detail as we proceed in the paper. The local transformation bias corrected density estimator is finally defined as

$$\hat{f}_T(x) = T_{\hat{\theta}_1}^{(1)}(x)\,\phi\left(T_{\hat{\theta}_1}(x), \hat{\theta}_2\left(T_{\hat{\theta}_1}(x)\right)\right). \tag{3}$$

From the theoretical results concerning bias and variance of the local likelihood estimator given in H&J, it immediately follows that this estimator has the same variance as the standard transformation kernel density estimator introduced by Wand et al. (1991) given by

$$\mathrm{Var}\left(\hat{f}_T(x)\right) = \frac{f(x)}{nh}T_{\theta_1^0}^{(1)}(x)\int K(u)^2\,du + o\left(\frac{1}{nh}\right).$$

The bias is different however. Using the bias expression of the local likelihood estimator as reported in H&J, we can easily show that the bias of the LTBC estimator in case θ_2 is one-dimensional is

$$\text{Bias}\left(\hat{f}_T(x)\right) = \frac{1}{2}h^2\sigma_{\mathcal{K}}^2 T_{\theta_1^0}^{(1)}(x)\left[r^{(2)}\left(T_{\theta_1^0}(x)\right) - \phi^{(2)}\left(T_{\theta_1^0}(x), \theta_2^0\left(T_{\theta_1^0}(x)\right)\right)\right], \tag{4}$$

where $\sigma_{\mathcal{K}}^2 = \int z^2\mathcal{K}(z)\,dz$ and e.g. $r^{(j)}$ denotes going forward the j^{th} derivative of the function r. The magnitude of this bias term depends on how well the correction density can be approximated locally by a suitable parametric model. This is so if the second derivative of the true correction density is small [e.g. $r(u)$ is smooth], or equivalently, if the global parametric start density $T_{\theta_1^0}^{(1)}(x)$ is close to the true density. In general, this bias term depends on the weighting function as well in case only a single local parameter is fitted. This term cancels out however if the weighting function is chosen constant over the support of u. This is the case for all practically relevant cases considered in this paper. For further details we refer to H&J. In the transformation density literature it is common to express the bias in terms of the density function $f(x)$ and the transformation function $T(x)$. We do not follow this convention here to ease the interpretation of the bias term.

Direct local modeling of the density $f(x)$ can be obtained by choosing the transformation function $T_{\theta_1}(x)$ as an improper uniform distribution. Without loss of generality set $T_{\theta_1^0}(x)$ to x. Then the only source of bias reduction is provided by the local model $\phi(x, \theta_2)$ which becomes a local model for f. The bias is $\frac{1}{2}\sigma_{\mathcal{K}}^2 h^2\left[f^{(2)}(x) - \phi^{(2)}(x, \theta_2^0(x))\right]$ as in H&J. If on top of that the local model for ϕ is chosen as a constant, the standard kernel density emerges with a leading bias term given by

$$\frac{1}{2}\sigma_{\mathcal{K}}^2 h^2 f^{(2)}(x). \tag{5}$$

The standard transformation based kernel density estimator of Wand et al. (1991) emerges from choosing the weighting function and the local model both as a constant. From Equations (2) and (3) it follows that the estimator is

$$\hat{f}_T(x) = \frac{1}{\int \mathcal{K}_h(t - T_{\hat{\theta}_1}(x))dt}\frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_h(T_{\hat{\theta}_1}(X_i) - T_{\hat{\theta}_1}(x))T_{\hat{\theta}_1}^{(1)}(x). \tag{6}$$

From (4), the bias is $(1/2)\,\sigma_{\mathcal{K}}^2 h^2 T_0^{(1)}(x)\, r^{(2)}(T_{\theta_1^0}(x))$, which is the bias of the standard transformation kernel density estimator. Assuming that $\mathcal{K}$ has support $[-1, 1]$, the term in the denominator of

Equation (6) integrates to one if u lies in the interior, meaning that both, u/h and $(1-u)/h \to \kappa > 1$. However, close to the boundaries where $0 \leq \kappa < 1$, this integral term normalizes the density estimate and therefore adjusts for the undesirable weight allocation of the symmetric kernel outside the support of the density. This adjustment is not optimal and boundary bias is still of the undesirable order $O(h)$. Like in nonparametric regression, see e.g. Fan and Gijbels (1992), one of the possible boundary bias correction methods which achieves an $O(h^2)$ order is the popular local linear estimator, see Jones (1993) for the density case. To obtain a local linear transformation density estimator, we propose to choose the local model at point u as $\phi(t, \theta_2(u)) = \theta_{21} + \theta_{22}(t-u)$ and the weight functions as 1 and $(t-u)$. The resulting estimator is equivalent to the standard transformation density estimator in the interior of the density. Close to the boundaries it provides however again a correction against weight allocation of the symmetric kernel to areas outside the support of the transformed data. The local linear transformation density estimator is

$$\hat{f}_T(x) = T'_{\hat{\theta}_1}(x) \frac{\hat{g}^0\left(T_{\hat{\theta}_1}(x)\right) - \left[\alpha_1\left(T_{\hat{\theta}_1}(x), h\right)/\alpha_2\left(T_{\hat{\theta}_1}(x), h\right)\right]\hat{g}^1(x)}{\left(\alpha_0\left(T_{\hat{\theta}_1}(x), h\right) - \alpha_1\left(T_{\hat{\theta}_1}(x), h\right)^2/\alpha_2\left(T_{\hat{\theta}_1}(x), h\right)\right)}, \tag{7}$$

where $\hat{g}^j(x)$ is the sample average of $\mathcal{K}_h\left(T_{\hat{\theta}_1}(X_i) - T_{\hat{\theta}_1}(x)\right)\left(T_{\hat{\theta}_1}(X_i) - T_{\hat{\theta}_1}(x)\right)^j$ and

$$\alpha_j(u,h) = h^j \int_{\max\{-1,-u/h\}}^{\min\{1,(1-u)/h\}} \mathcal{K}(v) v^j dv.$$

After presenting the LTBC framework for symmetric kernels, we now turn to an asymmetric kernel version of this above approach. Since the support of these kernels matches the support of the density under consideration, no boundary correction of the type presented above is necessary by construction. We first briefly review the beta kernel density estimator for densities defined on the unit interval introduced by Chen (1999). In Section 4 we will treat the LTBC case.

3 The beta kernel density estimator

The consistency of the standard symmetric kernel estimator (see e.g. Silverman (1986), Haerdle and Linton (1994) or Pagan and Ullah (1999) for an introduction) is well documented when the support of the underlying density is unbounded, i.e. when data live on $(-\infty, +\infty)$. As we have already argued, this symmetric estimator is however no more appropriate in the case the density to be estimated

has bounded support as it is the case for our transformed data located in $[0,1]$ with density r. The symmetric kernel has a boundary bias since it assigns non zero probability outside the support of the distribution when smoothing is carried out near a boundary.

Recently, Chen (1999) has proposed a beta kernel density estimator for densities defined on $[0,1]$. This estimator is based on the asymmetric beta kernel which exhibits two special appealing properties: a flexible form and a location on the unit interval. The kernel shape is allowed to vary according to the data points, thus changing the degree of smoothing in a natural way, and its support matches the support of the probability density function to be estimated. This leads to a larger effective sample size used in the density estimation and usually produces density estimates that have smaller finite-sample variances than other nonparametric estimators. The beta kernel density estimator is simple to implement, free of boundary bias, always non negative, and achieves the optimal rate of convergence ($n^{-4/5}$) for the mean integrated squared error (MISE) within the class of nonnegative kernel density estimators (see Chen (1999) for details). Furthermore, even if the true density is infinite at a boundary, the beta kernel estimator remains consistent (Bouezmarni and Rolin (2004)). This property of the beta kernel density estimator is especially important for the estimation of our transformed loss data: As we have argued earlier on, the density of our transformed loss data is unbounded at the right boundary in case the tail of the used parametric model is lighter than the true one. Similar asymmetric kernel estimators for densities defined on $[0,+\infty)$ have been studied in Chen (2000), Scaillet (2004), and Bouezmarni and Scaillet (2005). These estimators share the same valuable properties as those of the beta kernel estimator used in this paper.

The beta kernel estimator of the unknown density r at point u is formally defined as

$$\hat{r}(u) = \frac{1}{n}\sum_{i=1}^{n} K(U_i, u, b), \tag{8}$$

where the asymmetric kernel $K(.)$ is the beta probability density function:

$$K(t,u,b) = \frac{1}{B(u/b+1,(1-u)/b+1)} t^{u/b}(1-t)^{(1-u)/b}, \qquad u \in [0,1],$$

with $B(.)$ denoting the beta function and b being a smoothing parameter, called the bandwidth, such that $b \to 0$ as $n \to \infty$. Chen (1999) shows that the bias and variance of the beta kernel estimator

are given by

$$\begin{aligned}
\text{Bias}\left(\hat{r}\left(u\right)\right) &= \left[\left(1-2u\right)r^{(1)}\left(u\right)+\frac{1}{2}u\left(1-u\right)r^{(2)}\left(u\right)\right]b+o\left(b\right), \qquad (9)\\
\text{Var}\left(\hat{r}\left(u\right)\right) &= \begin{cases} \frac{1}{2\sqrt{\pi}}\frac{n^{-1}b^{-1/2}}{\{u(1-u)\}^{1/2}}\left(r\left(u\right)+O\left(n^{-1}\right)\right) & \text{if } u/b \text{ and } \left(1-u\right)/b\to\infty, \\ \frac{\Gamma(2\kappa+1)}{2^{1+2\kappa}\Gamma^{2}(\kappa+1)}n^{-1}b^{-1}\left(r\left(u\right)+O\left(n^{-1}\right)\right) & \text{if } u/b \text{ or } \left(1-u\right)/b\to\kappa, \end{cases}
\end{aligned}$$

where Γ denotes the gamma function. We note that compared to symmetric kernel density estimators, the bias of the beta kernel density estimator contains the first derivative of r. This may be a problem if r exhibits a substantial slope as in the case of unboundedness at the boundaries. Also the variance is of higher order in the boundary than in the interior. Chen (1999) shows that this has no asymptotic impact on the mean integrated squared error. It is, however, still a caveat of this sort of smoothing. Chen (1999) also proposes an adjustment of beta kernel density estimator which has the first derivative removed in the bias expression in the interior. However, the first derivative term is still present in the boundaries. Based on the local likelihood framework, we will introduce a version of the beta kernel density estimator whose bias expression is free of the first derivative over the whole support. This seems important since our transformed data may have large mass close to the right boundary as mentioned earlier on.

4 The asymmetric LTBC estimator

Apart from being an attractive semiparametric bias reduction framework, LTBC allows us to implement a popular boundary bias reduction by choosing a local linear model to estimate the density of the transformed data. This boundary bias reduction, which is necessary for symmetric kernels, is not required *per se* in the asymmetric beta kernel framework. This because the beta kernel allocates no weight outside the support of the unknown density. The effect of LTBC in an asymmetric beta kernel framework is just to reduce the potentially larger bias for asymmetric kernel techniques.

To obtain an asymmetric version of the LTBC estimator, we just replace in Equation (2) which defines the local likelihood estimator the symmetric kernel $\mathcal{K}$ by the asymmetric beta kernel K: Choose $\theta_2\left(\hat{u}\right)$, i.e., $\tilde{\theta}_2$, such that

$$\frac{1}{n}\sum_{i=1}^{n}K\left(\hat{U}_i,\hat{u},b\right)v(\hat{u},\hat{U}_i,\theta_2)-\int K\left(t,\hat{u},b\right)v(\hat{u},t,\theta_2)\phi\left(t,\theta_2\right)dt=0. \qquad (10)$$

The asymmetric LTBC density estimator is then defined as

$$\hat{f}_{AT}(x) = T^{(1)}_{\hat{\theta}_1}(x)\,\phi\left(T_{\hat{\theta}_1}(x), \tilde{\theta}_2\left(T_{\hat{\theta}_1}(x)\right)\right). \tag{11}$$

Following similar derivation steps as in Hagmann and Scaillet (2007) (H&S) adapted to the asymmetric beta kernel, it is straightforward to develop the bias of this estimator if θ_2 is one dimensional ($q = 1$). We assume again here that the weighting function in this one dimensional local model case is just a constant over the support of u. In general, the bias term again depends on the weighting function, and a lengthy formula can be obtained along the lines of H&J. Using $u^0 = T_{\theta^0_1}(x)$ to shorten notation we have that

$$\text{Bias}\left(\hat{f}_{AT}(x)\right) = \frac{1}{2}T^{(1)}_{\theta^0_1}(x)\left\{\begin{array}{l}(1-2u^0)\left(r^{(1)}(u^0) - \phi^{(1)}(u^0, \theta^0_2(u^0))\right) + \\ u^0(1-u^0)\left(r^{(2)}(u^0) - \phi^{(2)}(u^0, \theta^0_2(u^0))\right)\end{array}\right\} b. \tag{12}$$

Note that an asymmetric version of the standard transformation based kernel density estimator is obtained by choosing the local model for the correction density and the weighting function $v_{\theta^0_1}\left(T_{\theta^0_1}(x)\right)$ as a constant. From Equations (10) and (11) it follows that this estimator takes the very simple form

$$\hat{f}_{AT}(x) = \frac{1}{n}\sum_{t=1}^{n} K\left(T_{\hat{\theta}_1}(X_i), T_{\hat{\theta}_1}(x), b\right) T_{\hat{\theta}_1}(x).$$

Unlike in the symmetric kernel case, no boundary correction terms are necessary. This is because the support of the beta kernel matches the support of the transformed data. Interesting from our perspective is now that the first derivative terms in (12) vanish over the full support as soon as the number of locally fitted parameters is two or larger, we refer to H&S and H&J for more details.

Proposition 1. *The bias of the asymmetric local transformation bias corrected kernel density estimator for $q \geq 2$ is given by*

$$Bias\left(\hat{f}_{AT}(x)\right) = \frac{1}{2}T^{(1)}_{\theta^0_1}(x)\,u^0\left(1-u^0\right)\left[r^{(2)}\left(u^0\right) - \phi^{(2)}\left(u^0, \theta^0_2\left(u^0\right)\right)\right] b + o(b). \tag{13}$$

This result holds true over the whole support. The variance is

$$Var\left(\hat{f}_{AT}(x)\right) = \left\{\begin{array}{l}\frac{1}{2\sqrt{\pi}}\frac{n^{-1}b^{-1/2}}{\{u^0(1-u^0)\}^{1/2}}\left\{T^{(1)}_{\theta^0_1}(x) f(x) + O(n^{-1})\right\} \text{ if } u^0/b \text{ and } (1-u^0)/b \to \infty, \\ \frac{\Gamma(2\kappa+1)n^{-1}b^{-1}}{2^{1+2\kappa}\Gamma^2(\kappa+1)}\left\{T^{(1)}_{\theta^0_1}(x) f(x) + O(n^{-1})\right\} \text{ if } u^0/b \text{ or } (1-u^0)/b \to \kappa.\end{array}\right.$$

There are several worthwhile remarks. First, we obtain the same result as in the symmetric kernel case albeit the underlying proof techniques are quite different in the asymmetric kernel case. The proof follows closely the steps in H&S, and rely on the expansions in Chen (1999) (instead of Chen (2000)) for beta kernel estimators (instead of gamma kernel estimators) based on second order continuous differentiability of the density function. Therefore the proof is omitted. Second, comparing (13) to (9) , the first derivative term vanished and the second derivative $r^{(2)}(u^0)$ in the bias expression has been replaced by $r^{(2)}(u^0) - \phi^{(2)}(u^0, \theta_2^0(u^0))$. So using a local model with more than one parameter performs better than using a simple local constant model if the latter expression is smaller than the former in absolute values. This is the case if the unknown correction density exhibits high local curvature which is the case if the initial transformation $T_{\theta_1}(x)$ has not been close enough to the true distribution function. If the transformation were correct, the LTBC estimator is unbiased up to the order considered. In that case its Mean Squared error (MSE) is subtantially reduced w.r.t. a pure beta kernel estimator since their variances coincide.

Several local models for the correction density r are possible candidates. We propose here to use the local log linear model as the main competitor to the local constant model. This local two-parameter model ensures that the bias of the asymmetric LTBC estimator depends only on the second derivative. Furthermore, the curvature of this local model allows us to diminish the size of $r^{(2)}$ even further as discussed above. A last advantage of this estimator is that unlike the local linear estimator, it always yields positive density estimates.

Example 1. *The popular local log linear density model at point u is* $\phi(t, \theta_2(u)) = \theta_{21} \exp(\theta_{22}(t-u))$. *As weight functions we use the vector* $[1 \;\; (t-u)]$, *which coincides with the score of the local likelihood function. In the Appendix we show that this amounts solving the following system of nonlinear equations for* θ_{21}*:*

$$\begin{aligned} \hat{f}_b(u) &= \theta_{21} \exp(-\theta_{22} u) \psi(\theta_{22}), \\ \hat{g}_b(u) &= \theta_{21} \exp(-\theta_{22} u) \left[\psi^{(1)}(\theta_{22}) - x\psi(\theta_{22})\right], \end{aligned}$$

where $\psi(\cdot)$ *is the moment generating function of the beta probability law. Unfortunately, this has no simple closed form solution and some numerical approximations outlined in the Appendix have to be used (MATLAB or R code which allows simple computation of this estimator can be requested from*

the authors).

We finish this section by comparing the LTBC approach to the Local Multiplicative Bias Correction (LMBC) semiparametric density estimation framework as presented in H&S. Using the same notation as above, these authors consider the local model

$$m\left(x,\theta_1,\theta_2\left(x\right)\right)=T_{\theta_1}^{(1)}\left(x\right)\cdot\varphi\left(x,\theta_2\left(x\right)\right),$$

where θ_1 is estimated in a first step as well globally by maximum likelihood. Then the local likelihood framework is used to choose $\theta_2\left(x\right)$, i.e., $\bar{\theta}_2$, such that

$$\frac{1}{n}\sum_{i=1}^{n}K\left(X_i,x,b\right)v(x,X_i,\theta_2)-\int K\left(t,x,b\right)v(x,t,\theta_2)T_{\hat{\theta}_1}^{(1)}\left(t\right)r\left(t,\theta_2\right)dt=0 \tag{14}$$

holds, where $K\left(X_i,x,b\right)$ denotes here an asymmetric kernel with support on $[0,\infty)$ as for example Chen's gamma kernel. This yields the LMBC estimator $\hat{f}_{LMBC}\left(x\right)=T_{\hat{\theta}_1}^{(1)}\left(x\right)\varphi\left(x,\bar{\theta}_2\left(x\right)\right)$. Concentrating on the interior region where $x/b\rightarrow\infty$, H&S report bias and variance for the gamma kernel based LMBC estimator for $q\geq 2$ as

$$\begin{aligned}\text{Bias}\left(\hat{f}_{LMBC}\left(x\right)\right) &= \frac{1}{2}T_{\theta_1^0}^{(1)}\left(x\right)\left[r_0^{(2)}\left(x\right)-\varphi^{(2)}\left(x,\theta_2^0\left(x\right)\right)\right]xb+o\left(b\right), &(15)\\ \text{Var}\left(\hat{f}_{LMBC}\left(x\right)\right) &= \frac{n^{-1}b^{-1/2}}{2\sqrt{\pi}}\left\{x^{-1/2}f\left(x\right)+O\left(n^{-1}\right)\right\}. &(16)\end{aligned}$$

Comparing Equations (13) and (15) shows that the bias terms of both semiparametric estimation frameworks are very similar: Apart from kernel specific terms, both expressions depend (i) on the global parametric start, and (ii) on the curvature of the correction factor and its distance from the curvature of the local parametric model. The main but small difference is that the correction factor and local model terms are defined on the original $x-$ axis for LMBC, but on the transformed axis for LTBC. Although the variance terms are different, they share a common interpretation. For LTBC the $T_{\theta_1^0}^{(1)}$ term in Proposition 1 reflects the fact that the transformation induces the choice of an implicit location dependent bandwidth $b^*\left(x\right)=b/\left(T_{\theta_1^0}^{(1)}(x)\right)^2$. We note that for the standard beta kernel density estimator this term is not present in the variance expression. Also note that in the context of estimating long tailed densities on $[0,\infty)$, $T_{\theta_1^0}^{(1)}$ will approach zero as x gets large enough, implying that a larger effective bandwidth is chosen when smoothing in the tail region. This is similar to the

LMBC case where the varying shape of the gamma kernel induces an effective bandwidth of $b^* = bx$ which therefore also grows as x gets larger. In summary, from a theoretical perspective we expect both approaches to perform very similarly.

From an implementation perspective we note that Equations (10) and (14) indicate that the local likelihood criterion in the LMBC framework differs from the LTBC criterion in two respects: (i) integration takes place over the original x-scale, and (ii) the LMBC criterion involves the global parametric start as an integration term. This implies that for LMBC, the computational implementation of the local likelihood step varies with the chosen parametric start. This maybe considered as a disadvantage compared to LTBC where the local likelihood step is independent of the chosen parametric start. On the other hand, it has to be mentioned that whereas LTBC always has to rely on (although the same) numerical approximation procedures, H&S show that the LMBC framework allows us to develop simple closed form solutions for certain kernel and parametric start combinations. A further advantage is that in the LTBC framework the correction factor is not defined as a ratio as this is the case for LMBC (equal to $f/T^{(1)}_{\theta^0_1}$). Some clipping maybe therefore necessary in the latter approach if the denominator gets too small in rare cases, we refer to Hjort and Glad (1995) for more details.

5 The generalized Champernowne distribution

We choose the generalized Champernowne distribution as a flexible parametric start for our semiparametric density estimation framework. Flexibility is especially important here since we lack economic guidance to favor one loss parametric model over another one. The generalized Champernowne distribution can take a large range of shapes relevant to fit the empirical features of loss data. As evaluated in a study by Gustafsson and Thuring (2006), the generalized Champernowne distribution is preferable to other common choices to model loss data such as the Pareto, Weibull, and lognormal distribution and is therefore especially useful for our empirical study. We use the three parameter generalized Champernowne distribution as defined in Buch-Larsen et al. (2005):

Definition 1. *The generalized Champernowne cdf is defined for* $y \geq 0$ *and has the form*

$$F_{\alpha,M,c}(y) = \frac{(y+c)^\alpha - c^\alpha}{(y+c)^\alpha + (M+c)^\alpha - 2c^\alpha} \ , \ \forall y \in \mathbb{R}_+$$

with parameter $\alpha > 0$, $M > 0$ *and* $c \geq 0$. *Its density is*

$$f_{\alpha,M,c}(y) = \frac{\alpha\,(y+c)^{\alpha-1}\,((M+c)^\alpha - c^\alpha)}{\left((y+c)^\alpha + (M+c)^\alpha - 2c^\alpha\right)^2} \ , \ \forall y \in \mathbb{R}_+.$$

For $c = 0$ this parametric distribution function is a special case of the one suggested by Champernowne (1936, 1952). The standard Champernowne distribution has also been used by Clements et al. (2003) in their approach to transformed density estimation. In this paper, we prefer using the generalized version as a parametric start for our semiparametric density estimator. The extensive simulation study carried out by Buch-Larsen et al. (2005) shows that the flexibility of the generalized Champernowne distribution is worthwhile. It outweighs stability advantages in estimating the cdf when the parameter c is simply set to zero. This parameter increases greatly shape flexibility at the zero boundary. We refer to Buch-Larsen et al. (2005) for a comprehensive description of the flexibility of this distribution. Estimation of the parameter (α, M, c) is obtained by maximizing the log–likelihood function

$$\begin{aligned} l(\alpha, M, c) \quad = \quad & n\left(\log(\alpha) + \log\left((M+c)^\alpha - c^\alpha\right)\right) + (\alpha - 1)\sum_{i=1}^{n} \log(Y_i + c) - \\ & 2\sum_{i=1}^{n} \log\left((Y_i+c)^\alpha + (M+c)^\alpha - 2c^\alpha\right). \end{aligned}$$

Following Buch-Larsen et al. (2005), we first replace M by the empirical median estimate $\hat{M}$ and then choose the parameters (α, c) by maximizing $l(\alpha, \hat{M}, c)$. This procedure is not fully efficient, but simplifies the estimation procedure considerably. It is motivated by the fact that $F_{\alpha,M,c}(M) = 0.5$. Also Lehmann (1991) pointed out that the median is a robust estimator especially for heavy tailed distributions.

The next chapter examines the properties of the LTBC estimator using the generalized Champernowne distribution as a parametric start in small and moderate samples.

6 Monte Carlo study

In this section we evaluate the finite sample performance of the semiparametric transformation based estimators considered in the previous sections. We present the different test densities on which the Monte Carlo study is carried out and remind the reader of the semiparametric estimators we compare in this paper. After presenting various statistical performance measures, we discuss the obtained results. Our design of Monte Carlo study is identical to the design of Buch-Larsen et al. (2005) that concluded that the transformation approach based on the generalized Champernowne distribution outperforms the recent advances in transfomation based loss estimation, Bolance, Guillen and Nielsen (2003) and Clements, Hurn and Lindsay (2003) as well as the classical transformation approach of Wand, Ruppert and Marron (1991). Therefore, when we conclude below that our new transformation estimators outperform the estimator of Buch-Larsen et al. (2005), this implies that our new transformation estimators also outperform the estimators of Wand, Ruppert and Marron (1991), Bolance, Guillen and Nielsen (2003) and Clement, Hurn and Lindsay (2003) for actuarial loss distributions.

6.1 Considered estimators and test densities

All semiparametric density estimators considered in this Monte Carlo study use the generalized Champernowne distribution function as their parametric start. The estimators are different however in modeling the correction function. We consider the following estimators for the density of the transformed data defined earlier as:

- the symmetric kernel based standard local constant kernel estimator (LC) as provided in (6),
- the symmetric kernel based local linear estimator (LL) as given in (7),
- the local constant beta kernel density estimator (LCB) as provided in (8),
- the local log linear beta kernel estimator (LLLB) as defined in (11).

The Epanechnikov kernel is used throughout for estimators involving symmetric kernels. Furthermore, we also consider the generalized Champernowne distribution (CH) as a purely parametric

competitor to the above semiparametric estimators. We compare the listed estimators on different test densities as summarized below:

1. Weibull, $f(x) = \gamma x^{(\gamma-1)} e^{-x^\gamma}, \qquad \gamma = 1.5$
2. Normal, $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \qquad (\mu, \sigma) = (5, 1)$
3. Truncated Normal, $f(x) = \frac{2}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \qquad (\mu, \sigma) = (0, 1)$
4. Truncated Logistic, $f(x) = \frac{2}{s} e^{\frac{x}{s}} \left(1 + e^{\frac{x}{s}}\right)^2, \qquad s = 1$
5. Lognormal, $f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\log(x)-\mu}{\sigma}\right)^2}, \qquad (\mu, \sigma) = (0, 0.5)$
6. Lognormal and Pareto with mixing probability p,

$$f(x) = \frac{p}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\log(x)-\mu}{\sigma}\right)^2} + \frac{(1-p)\rho\lambda^\rho}{(x+\lambda)^{(\rho+1)}}, \qquad (p, \mu, \sigma, \lambda, \rho, c) = \begin{cases} (0.7, 0, 1, 1, 1, -1) \text{ or} \\ (0.3, 0, 1, 1, 1, -1) \end{cases}.$$

These six test distributions can be categorized as light (1-4)-, medium (5)- and heavy tailed (6). They are the same as the ones used by Buch-Larsen et al. (2005). For the heavy tail distribution, we utilize two different weights in the mixture to vary the amount of heavy tailness.

We treat the simple local constant estimator using the Epanechnikov kernel as our benchmark in what follows. More complicated estimators should outperform this benchmark such that their practical application is worthwhile.

6.2 Design of the Monte Carlo study

The performance measures we consider are the integrated absolute and squared error (IAD and ISE respectively). These performance statistics measure the error between the true and the estimated density with equal weight across the support of the density. To be able to focus on the fit in the tails we also consider the weighted integrated squared error (WISE) of the various estimators. Before presenting these measures, we apply the same substitution rule as in Clements, Hurn and Lindsay (2003), namely $y = \dfrac{x - M}{x + M}$, where M denotes the median. The error calculation is then restricted to the interval $[-1, 1]$ and therefore avoids integration to infinity. Denoting the true test density by $f(x)$ and its estimate by $\hat{f}(x)$, we can write the statistical performance measures as

$$
\begin{aligned}
\text{IAD} &= \int_0^{\infty} \left| \hat{f}\left(x\right) - f(x) \right| dx = \\
& \quad 2M \int_{-1}^{1} (1-y)^{-2} \left| \hat{f}\left(\frac{M(1+y)}{1-y} \right) - f\left(\frac{M(1+y)}{1-y} \right) \right| dy \\
\text{ISE} &= \left(\int_0^{\infty} \left(\hat{f}(x) - f(x) \right)^2 dx \right)^{1/2} = \\
& \quad \left(2M \int_{-1}^{1} (1-y)^{-2} \left(\hat{f}\left(\frac{M(1+y)}{1-y} \right) - f\left(\frac{M(1+y)}{1-y} \right) \right)^2 dy \right)^{1/2} \\
\text{WISE} &= \left(\int_0^{\infty} \left(\hat{f}(x) - f(x) \right)^2 x^2 dx \right)^{1/2} = \\
& \quad \left(2M \int_{-1}^{1} \frac{(1+y)^2}{(1-y)^4} \left(\hat{f}\left(\frac{M(1+y)}{1-y} \right) - f\left(\frac{M(1+y)}{1-y} \right) \right)^2 dy \right)^{1/2}.
\end{aligned}
$$

The experiments are based on 2,000 random samples of length $n = 50$, $n = 100$, $n = 500$ and $n = 1,000$. Concerning the choice of bandwidth, we consider in this Monte Carlo study a very simple bandwidth selection procedure which is often used in practice: the Silverman (1986) rule of thumb, also termed as a normal scaled bandwidth. This simple rule assumes that the transformed data are normally distributed. This quick rule is known to give nice results in practice when data are symmetric and unimodal (cf. Section 3 of Wand and Jones (1995)). In our case it is a good candidate when the parametric model is not correct, but close too. If the parametric model is correct the transformed data should be uniformly distributed. In principle we could exploit this feature to select the bandwidth. However since the second derivative of the density of a uniform is nil, the bandwidth minimizing the asymptotic MISE is not defined since it would be infinite. Using an Epanechnikov kernel, it can be shown by straightforward algebra (see e.g. Gustafsson et al. (2006)) that the normal scaled bandwidth is given by $h = \left(\frac{40\sqrt{\pi}}{n} \right)^{1/5} \hat{\sigma}$, where $\hat{\sigma}$ is the empirical standard deviation. For estimators involving the asymmetric beta kernel we follow Renault and Scaillet (2004) and take $b = \hat{\sigma} n^{-2/5}$. We note that these bandwidths procedures are fairly simple and reported results

concerning the performance of our estimators can therefore be considered as being conservative.

6.3 Monte Carlo results

Table I reports average obtained statistical performance measures and their standard deviations across all test densities. Careful analysis of this table reveals a simple summary result: The asymetric kernel based and locally guided LLLB estimator is the only estimator which consistently adds value over the symmetric kernel based LC benchmark across all test densities, statistical performance measures and considered sample sizes. The LLLB estimator never performs worse than the benchmark, but adds in most cases substantial value. In the heavy tailed test density with mixing parameter $p = 0.3$, the WISE is reduced by 15-31% for the WISE depending on sample size, whereas the ISE is even reduced up to 48%.

Although Chen's (1999) estimator does sometimes slightly outperform its locally guided version LLLB, its performance is overall worse than the one of the LLLB estimator since it can substantially underperform the provided benchmark. So overall, our proposed local log linear beta kernel density estimator adds value over its simple LCB version. Finally, the local linear estimator shows a rather unattractive performance. In cases where this estimator beats the LC benchmark, the advantage is typically small. However, the underperformance can be rather severe as the simulation results for the heavy tailed test densities show. In this case the considered statistical performance measures are up to 20% worse than those of the simple benchmark. Finally we note that our favourite LLLB estimator outperforms the CH estimator across all test densities and considered sample sizes.

*** Table 1 about here***

7 Application to operational loss data

In this section we demonstrate the practical relevance of the above presented estimation methods by applying them on operational risk modelling using publicly reported external loss data. The SAS® OpRisk Global dataset consists of loss events collected mainly from large financial institutes, scaled such that it fits an insurance company's exposure to operational risk. Using an external database is a natural step when internal data suffers from (i) a limited collection period, (ii) incomplete coverage of

all relevant event risk categories (ERC), or (iii) the company risk experts have reason to believe that one or more internal business lines have not been exposed to large losses (their empirical distribution gives an incomplete assessment of the risk involved). Furthermore, regulators suggest in the Advanced Measurement Approach (AMA) that external data sources should complement company internal data sources to calculate an adequate capital requirement, see International Convergence of Capital Measurement and Capital Standard, Basel Committee on Banking Supervision for details.

7.1 The dataset

The scaled dataset delivers information on 1379 loss events in excess of US$ 1 million, splitted into four main event risk categories; 1. Fraud, 2. Employment Practices and Workplace Safety, 3. Business Disruption and System Failures, and 4. Execution, Delivery & Process Management. Table 2 reports summary statistics (reported in million US dollars) for each risk group.

	Number of Losses (n_i)	Maximum Loss (US$M)	Sample Mean (US$M)	Sample Median (US$M)	Standard Deviation (US$M)	Time Horizon (η_i)
ERC.1	538	1703	29.12	4	130.3	9
ERC.2	721	415	14.67	3.7	37.5	8
ERC.3	45	220	42.01	11	62.2	6
ERC.4	75	195.6	13.82	2.91	33.1	3

TABLE 2: Statistics for the external reported operational risk exposure

Event Risk categories 3 & 4 have very few reported losses which is partly due to the fact that these categories have been collected over a shorter time period only (3 to 6 years). The loss distributions in all categories exhibit clear right skewed features as the mean loss is much higher than the corresponding median loss in each risk group. Event Risk Category 1 (Fraud) reports the most significant single loss. Not surprisingly, this category also reports the highest standard deviation of reported losses. To gain further information on the considered dataset, Figure 1 offers a visual inspection of the loss arrival for each event risk category over time.

*** Figure 1 about here ***

7.2 Severity distribution across risk categories

Estimation of the severity distributions across the four presented risk categories (or business lines from an insurer perspective) proceeds as described in the previous Sections. In a first step we estimate for each risk category the parameters of the generalized Champernowne distribution and in parentheses its bootstrap approximated standard errors with 10.000 iterations, which are summarized in Table 3. We remind the reader that in our robust estimation approach M is just estimated by the empirical sample median.

	ERC.1	ERC.2	ERC.3	ERC.4
$\hat{\alpha}_i$	1.08 (0.088)	1.33 (0.045)	0.80 (0.182)	1.09 (0.176)
$\hat{M}_i$	4.00 (0.165)	3.70 (0.113)	11.00 (3.230)	2.91 (0.406)
$\hat{c}_i$	0 (1.104)	0 (0)	0 (4.463)	0 (0.995)

TABLE 3: Estimated parameters and standard errors for each ERC.

It is worth being noted that the scaling parameter c is estimated to be zero in all four risk groups. This implies that for our data the original version of the generalized Champernowne distribution provides the best empirical fit.

Having estimated the parameters for our parametric start model we proceed by transforming the data into [0, 1] and estimate the correction factor within each risk group. We do this using all the estimators analyzed in our simulation study to examine their differences on a concrete empirical dataset. Figure 2 plots the different correction factors which are all located around one, indicating that the used parametric model provides a reasonable first fit to the loss data under consideration. When the correction factor is estimated to be below (above) one at point x, then this implies that the generalized Champernowne density does not provide an adequate fit of the data at this point and the final density estimate will take smaller (larger) values at this point. An apparent feature of the estimated correction factors in Figure 2 is that the estimators (6) and (7) produce similar

results. The asymmetric beta kernel based estimators (8) and (11) are similar in the interior, but (11) exhibits higher curvature in the boundary domain.

*** Figure 2 about here ***

Figure 3 shows the final semiparametric density estimates for the four risk categories in comparison to their fitted Champernowne distribution. These estimates are obtained by multiplying the parametric start by the estimated correction factors as described in Formula (3).

*** Figure 3 about here ***

7.3 The total loss distribution

In this section we perform a Monte-Carlo simulation to obtain the total (or aggregated) operational risk loss distribution based on the four risk event categories. We again provide a comparison across the four different severity distribution estimators considered in the previous section to highlight their empirical differences. To be able to compute the total loss distribution, we need a model for loss arrivals. We assume that loss arrivals can be described by a discrete stochastic process in continuous time, commonly used to model the arrival of operational risk events. We denote by $N_i(t)$ the random number of observed losses in risk category i over a fixed time period $(0, t]$, with $N_i(0) = 0$. We assume that we have i independent homogenous Poisson processes denoted as $N_i(t) \in Po(\lambda_i t)$, where $\lambda_i > 0$ is the intensity parameter in the i^{th} risk category. The maximum likelihood estimator of the annual intensity of losses is $\hat{\lambda}_i = N_i/\eta_i$, where η_i is the observed number of losses in risk group i (or business line) as reported in Table 2. We proceed as follows to simulate the total loss distribution across all categories:

- Conditional on $N_i = n_i \; \forall i$, we simulate our poisson process of events that occur over a one year horizon through all lines of operational risk. The simulated annual frequencies are denoted by $\hat{\lambda}_{ir}$, with $i = 1, ..., 4$ and $r = 1, ..., R$, where we choose $R = 20.000$.

- Then for each $\{i, r\}$ we draw randomly $\hat{\lambda}_{ir}$ samples from a uniform distribution, i.e

$$u_{irk} \in U(0, 1), \qquad k = 1, ..., \hat{\lambda}_{ir}.$$

- These random number are then used to compute R simulated total annual losses across all categories, using the formula

$$S_r = \sum_{i=1}^{4} \sum_{k=1}^{\hat{\lambda}_{ir}} \hat{F}_{T_i}^{\leftarrow}(u_{irk}) \quad \text{for } r = 1, ..., R,$$

where $F_{T_i}^{\leftarrow}(\cdot)$ is the inverse cumulative loss distribution function for risk group i as defined by our presented semiparametric estimation framework through

$$\hat{F}_{T_i}(x) = \int T_{\hat{\theta}_{i1}}^{(1)}(x)\phi(T_{\hat{\theta}_{1i}}(x), \hat{\theta}_{2i}(T_{\hat{\theta}_{1i}}(x)))dx.$$

We perform this process for all four considered semiparametric severity distribution estimators. Figure 4 shows that there are significant differences in the exposure between estimators. Clearly, both the models including correction function (11) and (8) present a heavier tail than the others.

*** Figure 4 about here ***

Table 4 collects summary statistics for the simulated total loss distribution across all models as listed in Section 6.1. Among the usual summary statistics we also report the extreme quantiles at level 0.95 and 0.995, which we denote as Value-at-Risk (VaR) measures following the naming practice pursued by regulators and the banking industry. For comparison purposes, we also computed results for three parametric benchmark distributions, namely the lognormal-, Pareto- and generalized Champernowne distribution (LN, PA, CH respectively). The two former models whose parametric definitions are reported in Section 6 were chosen as they are widely used in the insurance industry when calculating capital requirement and reserves, the latter because it presents our parametric start model for the severity distributions. Parameter estimates were obtained in all three cases by maximum likelihood.

Model Assumption	Maximum Loss (US$M)	Sample Mean (US$M)	Sample Median (US$M)	Standard Deviation (US$M)	VaR 95% (US$M)	VaR 99.5% (US$M)
LC	8168	2813	2728	708	4093	5087
LL	7652	2864	2771	751	4212	5253
LCB	12840	4272	4108	1228	6502	8393
LLLB	12160	3562	3414	1078	5527	7065
CH	6973	2442	2367	604	3535	4372
LN	5041	2490	2453	450	3281	3836
PA	12010	8479	8458	816	9856	10747

TABLE 4: Statistics for the total loss distribution obtained by 7 different model assumptions.

All models have in common to produce a right skewed total loss distribution as the sample mean is larger than the sample median. Comparing the median and the two VaR measures for the total loss distribution as derived from the purely parametric PA and LN models produces interesting insights. The Pareto model predicts clearly much higher extreme losses than the lognormal model, but at the same time produces a median of the distribution which appears very large in comparison to inferences drawn from all other model solutions. On the other hand, the parametric LN model predicts the lowest losses of all considered model alternatives. Although a firm conclusion can not be drawn as the true model generating the data is unknown, this empirical evidence is in line with the fact that actuaries consider the lognormal model to provide a good fit for small losses, but favour the Pareto model to draw inferences in the tail of the loss distribution. The generalized Champernowne model seems in our empirical application to combine the advantages of those two popular models and predicts slightly higher extreme losses than the lognormal model, but a similar median total loss. It is this nice feature of the generalized Champernowne model which makes it very attractive as a parametric start for our semiparametric estimators. We note that all semiparametric models indicate higher extreme losses than those predicted by the generalized Champernowne model. This is especially true for the asymmetric kernel based estimators which performed well in our simulation study. These results clearly highlight the advantages of our semiparametric risk measurement

approach compared to a pure parametric one which might lead to misleading risk measurement.

8 Conclusion

In this paper we have considered semiparametric estimators based on symmetric and asymmetric kernels for the estimation of densities on the nonnegative real line. This framework allows us to use a flexible parametric start given by the generalized Champernowne distribution, which is then corrected in a nonparametric fashion. To improve the efficiency of this estimator even further, a local guidance in terms of a local log linear model was proposed, which in connection with the asymmetric beta kernel yielded attractive performance in our vast simulation study. The attractiveness of the results more than justify the slight numerical burden exhibited by this asymmetric local log linear estimator. The approach should therefore be useful in applied work in economics, finance and actuarial science involving non- and semiparametric techniques. This point has already been demonstrated with an empirical application to operational loss data.

9 Acknowledgements

The second and fourth authors received support by the Swiss National Science Foundation through the National Center of Competence in Research: Financial Valuation and Risk Management (NCCR FINRISK). We would like to thank Arthur Lewbel, an associate editor and two referees for constructive criticism and comments which have led to a substantial improvement over the previous version.

10 Appendix

In this appendix we briefly discuss the derivation of the log linear version of the asymmetric LTBC estimator and its computation. We use the local model $\phi(t, \theta_2(u)) = \theta_{21} \exp(\theta_{22}(t-u))$. From

Equation (10) we have to solve

$$\frac{1}{n}\sum_{i=1}^{n} K\left(U_i,u,b\right)\begin{pmatrix}1\\ U_i-u\end{pmatrix}-\theta_{21}\int_0^1 K\left(t,u,b\right)\begin{pmatrix}1\\ t-u\end{pmatrix}\exp\left(\theta_{22}\left(t-u\right)\right)dt$$

which amounts to solve the equations

$$\hat{f}_b(u) = \theta_{21}\exp\left(-\theta_{22}u\right)\psi\left(\theta_{22}\right), \tag{17}$$

$$\hat{g}_b(u) = \theta_{21}\exp\left(-\theta_{22}u\right)\left[\psi^{(1)}\left(\theta_{22}\right)-x\psi\left(\theta_{22}\right)\right], \tag{18}$$

where $\psi\left(\theta_2\right)$ is the moment generating function (m.g.f.) of the beta distribution. Note that we need to solve

$$\frac{\hat{g}_b(x)}{\hat{f}_b(x)}=\frac{\left[\psi^{(1)}\left(\theta_{22}\right)-x\psi\left(\theta_{22}\right)\right]}{\psi\left(\theta_{22}\right)} \tag{19}$$

for θ_{22}. The beta m.g.f is

$$\psi\left(\theta_{22}\right)=1+\sum_{k=1}^{\infty}\left(\prod_{r=0}^{k-1}\frac{\alpha+r}{\alpha+\beta+r}\right)\frac{\theta_{22}^k}{k!}.$$

This expression and its first derivative can be rewritten as

$$\psi\left(\theta_{22}\right) = 1+\frac{\Gamma\left(\alpha+\beta\right)}{\Gamma\left(\alpha\right)}\sum_{k=1}^{\infty}\frac{\Gamma\left(k+\alpha\right)}{\Gamma\left(k+\alpha+\beta\right)}\frac{\theta_{22}^k}{k!},$$

$$\psi^{(1)}\left(\theta_{22}\right) = \frac{\Gamma\left(\alpha+\beta\right)}{\Gamma\left(\alpha\right)}\sum_{k=1}^{\infty}\frac{\Gamma\left(k+\alpha\right)}{\Gamma\left(k+\alpha+\beta\right)}\frac{\theta_{22}^{k-1}}{(k-1)!}.$$

To compute the local log linear estimator, we numerically solve (19) for θ_{22} by approximating the above sum expressions by the first J terms. The estimate $\hat{\theta}_{21}$ can then be found from (17) to get the log linear version of the asymmetric LTBC estimator.

References

[1] Banks, J., Blundell, R. and Lewbel, A. (1997). Quadratic Engel curves and Consumer Demand. Review of Economics and Statistics, 79, 527-539.

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

[3] Bouezmarni, T. and Rolin, J.M. (2004). Consistency of Beta Kernel Density Function Estimator. Canadian Journal of Statistics 31, 89-98.

[4] Bouezmarni, T. and Scaillet, O. (2005). Consistency of Asymmetric Kernel Density Estimators and Smoothed Histograms with Application to Income Data. Econometric Theory 21, 390-412.

[5] Brown, B.M. and Chen, S.X. (1999). Beta-Bernstein Smoothing for Regression Curves with Compact Supports. Scandinavian Journal of Statistics 26, 47-59.

[6] Buch-Larsen, T. (2006). Claims Cost Estimation of Large Insurance Losses. Journal of Actuarial Practice 13, pages 191-211.

[7] Buch-Larsen, T., Nielsen, J.P., Guillen, M. and Bolance, C. (2005). Kernel Density Estimation for Heavy-tailed Distributions Using the Champernowne Transformation. Statistics, Vol. 39, No. 6, 503-518.

[8] Champernowne, D.G. (1936). The Oxford Meeting, September 25-29, 1936. Econometrica, Vol. 5, No. 4, October 1937.

[9] Champernowne, D.G. (1952). The Graduation of Income Distributions. Econometrica, 20, 591-615.

[10] Chen, S.X. (1999). A Beta Kernel Estimator for Density Functions. Computational Statistics and Data Analysis 31, 131-145.

[11] Chen, S.X. (2000). Probability Density Function Estimation Using Gamma Kernels. Annals of the Institute of Statistical Mathematics 52, 471-480.

[12] Clements, A.E., Hurn, A.S. and Lindsay, K.A. (2003). Möbius-like Mappings and their Use in Kernel Density Estimation. Journal of the Americal Statistical Association 98, 993-1000.

[13] Diebold, F.X., Gunther, H. and Tay, C. (1998). Evaluating Density Forecasts with Applications to Financial Risk Management. International Economic Review, 39, 863-883.

[14] Fan, J. and Gijbels, I.(1992). Variable Bandwidth and Local Linear Regression Smoothers. The Annals of Statistics 20, 2008-2036.

[15] Gouriéroux, Ch., Monfort, A. and Trognon, A. (1984). Pseudo Maximum Likelihood Methods: Theory. Econometrica 52, 680-700.

[16] Gustafsson, J., Nielsen, J.P., Pritchard, P. and Roberts, D. (2006). Quantifying Operational Risk Guided by Kernel Smoothing and Continuous credibility: A Practioner's view. The Journal of Operational Risk, Vol. 1, No. 1, 43-56.

[17] Gustafsson, J. and Thuring, F. (2006). Best Distribution for Severity Modelling: A Computational Evaluation of Different Loss Distribution Assumptions. Working Paper.

[18] Hagmann, M. and Scaillet, O. (2007). Local Multiplicative Bias Correction for Asymmetric Kernel Density Estimators. Forthcoming in Journal of Econometrics.

[19] Hjort, N.L. and Glad, I.K. (1995). Nonparametric Density Estimation with a Parametric Start. The Annals of Statistics 24, 882-904.

[20] Hjort, N.L. and Jones, M.C. (1996). Locally Parametric Nonparametric Density Estimation. The Annals of Statistics 24, 1619-1647.

[21] Hogg, R. and Klugman, S.A. (1984). Loss Distributions, New York: John Wiley & Sons, Inc.

[22] Härdle, W. and Linton, O. (1994). Applied Nonparametric Methods. Handbook of Econometrics IV, Eds Engle, R. and D. McFadden, North-Holland, Amsterdam, 2295-2339.

[23] Jones, M.C. (1993). Simple Boundary Correction for Kernel Density Estimation. Statistics and Computing, 3, 135-146.

[24] Jones, M.C. and Foster, P. (1996). A Simple Nonnegative Boundary Correction Method for Kernel Density Estimation. Statistica Sinica 6, 1005-1013.

[25] Klugman, S.A., Panjer H.A. and Willmot, G.E. (1998). Loss Models: From Data to Decisions. New York: John Wiley & Sons, Inc.

[26] Lehmann, E.L. (1991). Theory of Point Estimation. Wadsworth & Brooks/Cole.

[27] Müller, H. (1991). Smooth Optimum Kernel Estimators near Endpoints. Biometrika 78, 521-520.

[28] Pagan, A. and Ullah, A. (1999). Nonparametric Econometrics. Cambridge University Press, Cambridge.

[29] Renault, O. and Scaillet, O. (2004). On the Way to Recovery: A Nonparametric Bias Free Estimation of Recovery Rate Densities. Journal of Banking and Finance, 28, 2915-2931.

[30] Rice, J. (1984). Boundary Modification for Kernel Regression. Communications in Statistics: Theory and Methods 13, 893-900.

[31] Rosenblatt, M. (1956). Remarks on Some Nonparametric Estimates of a Density Function. The Annals of Mathematical Statistics, 27, 642-669.

[32] Scaillet, O. (2004). Density Estimation Using Inverse and Reciprocal Inverse Gaussian Kernels. Journal of Nonparametric Statistics, 16, 217-226.

[33] Schuster, E. (1985). Incorporating Support Constraints into Nonparametric Estimators of Densities. Communications in Statistics: Theory and Methods, 1123-1136.

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

[35] Wand, M.P. and Jones, M.C. (1995). Kernel Smoothing. New York: Chapman & Hall.

[36] Wand, M.P., Marron, J.S. and Ruppert, D. (1991). Transformation in Density Estimation (with comments). Journal of the American Statistical Association 94, 1231-1241.

[37] White, H. 1982, Maximum Likelihood Estimation of Misspecified Models. Econometrica 50, 1-26.

TABLE 1: Statistical performance Measures obtained from Monte Carlo Study

In this table we summarize the Monte Carlo results for the parametric generalized Champernowne distribution and the symmetric kernel based local constant and local linear, as well as the asymmetric beta local constant and local log linear estimator. Performance measures are the integrated absolute deviation (IAD), the integrated squared error (ISE) and the weighted integrated squared error (WISE). Simulation standard errors are small printed below main results.

	n=50			n=100			n=500			n=1000		
Estimator	IAD	ISE	WISE	IAD	ISE	WISE	IAD	ISE	WISE	IAD	ISE	WISE
							Weibull					
LC	1825	1401	1156	1396	1084	891	817	661	524	695	573	446
	(61.1)	(36.3)	(25.2)	(28.8)	(16.3)	(12.5)	(3.8)	(1.9)	(1.9)	(1.8)	(0.8)	(0.9)
LL	1786	1394	1136	1375	1080	880	802	637	518	675	534	440
	(52.1)	(33.5)	(22.5)	(25.1)	(15.8)	(11.7)	(4.3)	(2.6)	(2.2)	(2.4)	(1.4)	(1.2)
LCB	1758	1321	1088	1345	1037	848	830	661	533	729	587	467
	(54.3)	(30.3)	(23.8)	(27.5)	(14.9)	(12.3)	(3.7)	(1.7)	(2.1)	(1.9)	(0.8)	(1.0)
LLLB	1717	1301	1070	1311	1020	825	762	619	484	647	537	411
	(50.2)	(28.8)	(22.5)	(26.7)	(14.9)	(11.8)	(4.2)	(1.8)	(1.7)	(2.3)	(0.9)	(0.8)
CH	2264	1621	1193	1501	1113	907	915	722	563	726	599	471
	(43.2)	(21.6)	(15.2)	(8.7)	(6.2)	(2.4)	(0.9)	(0.7)	(0.6)	(0.1)	(0.1)	(0.1)

TABLE 1: (Continued)

	n=50			n=100			n=500			n=1000		
Estimator	IAD	ISE	WISE	IAD	ISE	WISE	IAD	ISE	WISE	IAD	ISE	WISE
						Normal						
LC	1797	906	4646	1384	694	3607	832	397	2153	696	324	1794
	(54.7)	(17.6)	(421.3)	(26.5)	(8.3)	(201.6)	(4.2)	(1.4)	(32.9)	(1.9)	(0.7)	(15.4)
LL	1761	899	4618	1345	691	3596	794	402	2148	655	328	1777
	(48.7)	(15.5)	(386.7)	(24.4)	(7.6)	(197.7)	(4.6)	(1.6)	(42.7)	(2.4)	(0.8)	(23.5)
LCB	1474	818	4205	1291	637	3315	818	378	2051	706	316	1764
	(52.8)	(14.8)	(352.6)	(26.0)	(7.7)	(184.5)	(4.4)	(1.4)	(33.2)	(2.3)	(0.7)	(11.6)
LLLB	1453	807	4118	1245	617	3198	743	345	1872	627	284	1576
	(48.9)	(14.0)	(331.5)	(24.8)	(7.6)	(179.0)	(4.9)	(1.6)	(35.7)	(2.6)	(0.9)	(19.0)
CH	1810	917	5826	1416	706	3994	841	413	2261	714	382	1806
	(32.6)	(8.6)	(116.4)	(17.6)	(3.2)	(90.7)	(1.3)	(0.9)	(12.4)	(0.6)	(0.1)	(9.2)
						Truncated Normal						
LC	1913	1496	1164	1446	1100	916	926	666	642	774	534	574
	(60.1)	(48.7)	(17.6)	(23.0)	(18.9)	(6.8)	(4.5)	(4.0)	(1.2)	(2.1)	(1.9)	(0.6)
LL	1846	1494	1122	1400	1105	881	875	643	610	718	511	531
	(57.0)	(49.9)	(18.0)	(24.0)	(21.0)	(8.2)	(5.6)	(4.4)	(2.3)	(2.9)	(2.2)	(1.4)
LCB	1871	1464	1098	1448	1121	889	1032	780	665	924	683	616
	(52.9)	(42.6)	(17.8)	(20.0)	(17.7)	(7.1)	(3.9)	(4.5)	(1.4)	(1.9)	(2.2)	(0.7)
LLLB	1793	1396	1061	1356	1026	836	875	605	591	754	494	534
	(56.2)	(48.4)	(16.8)	(23.1)	(20.1)	(6.2)	(5.3)	(4.6)	(1.1)	(2.9)	(2.3)	(0.6)
CH	2113	1502	1175	1496	1149	942	1088	797	682	944	691	634
	(47.2)	(39.6)	(12.4)	(15.6)	(12.2)	(5.8)	(1.9)	(1.6)	(0.9)	(0.8)	(0.7)	(0.1)

TABLE 1: (Continued)

	n=50			n=100			n=500			n=1000		
Estimator	IAD	ISE	WISE	IAD	ISE	WISE	IAD	ISE	WISE	IAD	ISE	WISE
						Truncated Logistic						
LC	1745	1080	1280	1307	795	990	743	434	602	606	342	511
	(64.7)	(31.3)	(30.0)	(28.3)	(13.7)	(13.9)	(4.9)	(2.5)	(2.4)	(2.1)	(1.2)	(1.0)
LL	1729	1116	1261	1305	834	975	742	452	593	602	354	498
	(56.6)	(31.1)	(28.0)	(25.3)	(13.2)	(14.4)	(5.3)	(2.8)	(3.3)	(2.5)	(1.3)	(1.7)
LCB	1669	1029	1225	1249	765	925	744	445	588	631	372	511
	(57.2)	(28.1)	(26.1)	(28.0)	(13.9)	(13.4)	(4.7)	(2.8)	(2.5)	(2.1)	(1.5)	(1.1)
LLLB	1647	1027	1224	1212	741	910	675	387	542	548	301	457
	(55.9)	(31.3)	(25.1)	(26.6)	(13.5)	(13.2)	(5.3)	(2.8)	(2.2)	(2.5)	(1.3)	(0.9)
CH	1826	1127	1340	1396	812	1008	770	496	621	631	383	527
	(42.6)	(19.4)	(18.7)	(21.6)	(8.2)	(8.6)	(2.8)	(1.4)	(1.7)	(1.9)	(0.9)	(0.8)
						Lognormal						
LC	1413	1149	1082	1363	1048	1045	796	595	593	664	484	484
	(22.4)	(15.0)	(14.4)	(28.2)	(19.8)	(19.4)	(3.7)	(2.9)	(2.8)	(1.9)	(1.4)	(1.4)
LL	1382	1138	1071	1332	1035	1027	783	598	596	652	488	489
	(20.4)	(14.7)	(13.0)	(23.4)	(17.3)	(16.5)	(3.8)	(3.2)	(3.0)	(2.2)	(1.8)	(1.7)
LCB	1354	1083	1030	1315	984	968	781	570	566	678	485	483
	(22.5)	(14.7)	(14.6)	(26.2)	(17.1)	(17.2)	(3.6)	(2.8)	(2.8)	(1.9)	(1.5)	(1.5)
LLLB	1340	1077	1015	1271	944	932	721	523	521	608	433	433
	(22.7)	(15.5)	(15.1)	(24.2)	(15.6)	(16.9)	(3.9)	(2.8)	(2.7)	(2.2)	(1.5)	(1.5)
CH	1477	1165	1099	1381	1056	1059	803	602	608	683	501	507
	(18.6)	(12.2)	(11.7)	(17.2)	(11.6)	(9.3)	(2.9)	(1.8)	(2.1)	(0.9)	(0.8)	(0.7)

TABLE 1: (Continued)

	n=50			n=100			n=500			n=1000		
Estimator	IAD	ISE	WISE	IAD	ISE	WISE	IAD	ISE	WISE	IAD	ISE	WISE
	70% Lognormal - 30% Pareto											
LC	1330	897	884	1119	770	716	889	615	569	782	481	472
	(26.5)	(14.7)	(15.8)	(15.2)	(7.1)	(7.1)	(5.6)	(2.9)	(2.4)	(4.0)	(1.8)	(1.8)
LL	1335	935	903	1198	821	778	1001	701	634	891	519	506
	(27.0)	(15.3)	(17.5)	(15.8)	(7.9)	(8.3)	(5.4)	(2.9)	(2.6)	(3.7)	(1.9)	(2.1)
LCB	1124	754	771	947	656	596	888	589	551	779	460	431
	(20.3)	(9.5)	(8.2)	(12.4)	(5.9)	(5.3)	(7.1)	(3.4)	(2.8)	(4.1)	(2.4)	(2.6)
LLLB	1149	769	790	985	679	621	892	591	554	780	461	431
	(19.5)	(9.6)	(8.6)	(12.7)	(5.9)	(5.6)	(6.9)	(3.3)	(2.7)	(4.0)	(2.2)	(2.5)
CH	1496	978	1026	1263	848	912	1175	717	738	956	589	596
	(19.7)	(8.3)	(6.2)	(11.8)	(6.3)	(6.1)	(7.6)	(2.4)	(2.9)	(3.6)	(1.7)	(1.9)
	30% Lognormal - 70% Pareto											
LC	1342	985	922	1146	785	779	674	416	485	496	324	286
	(26.6)	(11.5)	(10.5)	(17.1)	(9.3)	(8.6)	(4.2)	(4.1)	(1.9)	(3.5)	(1.0)	(0.9)
LL	1396	1020	1069	1336	932	936	739	492	521	517	350	336
	(29.5)	(26.5)	(26.1)	(28.6)	(17.5)	(15.0)	(6.3)	(5.4)	(2.9)	(5.5)	(1.3)	(1.0)
LCB	1107	778	874	911	608	630	509	268	387	371	168	199
	(14.5)	(12.8)	(12.9)	(9.2)	(5.6)	(4.8)	(2.3)	(2.9)	(2.9)	(1.1)	(0.8)	(0.7)
LLLB	1126	817	895	967	635	662	509	268	389	369	168	197
	(16.9)	(14.3)	(13.6)	(10.5)	(6.5)	(5.3)	(2.2)	(2.9)	(1.1)	(1.9)	(0.8)	(0.6)
CH	1567	1244	1206	1476	1192	1026	1101	737	804	879	635	711
	(27.3)	(16.8)	(15.2)	(18.4)	(12.6)	(13.3)	(11.2)	(8.7)	(9.1)	(5.2)	(4.6)	(5.1)

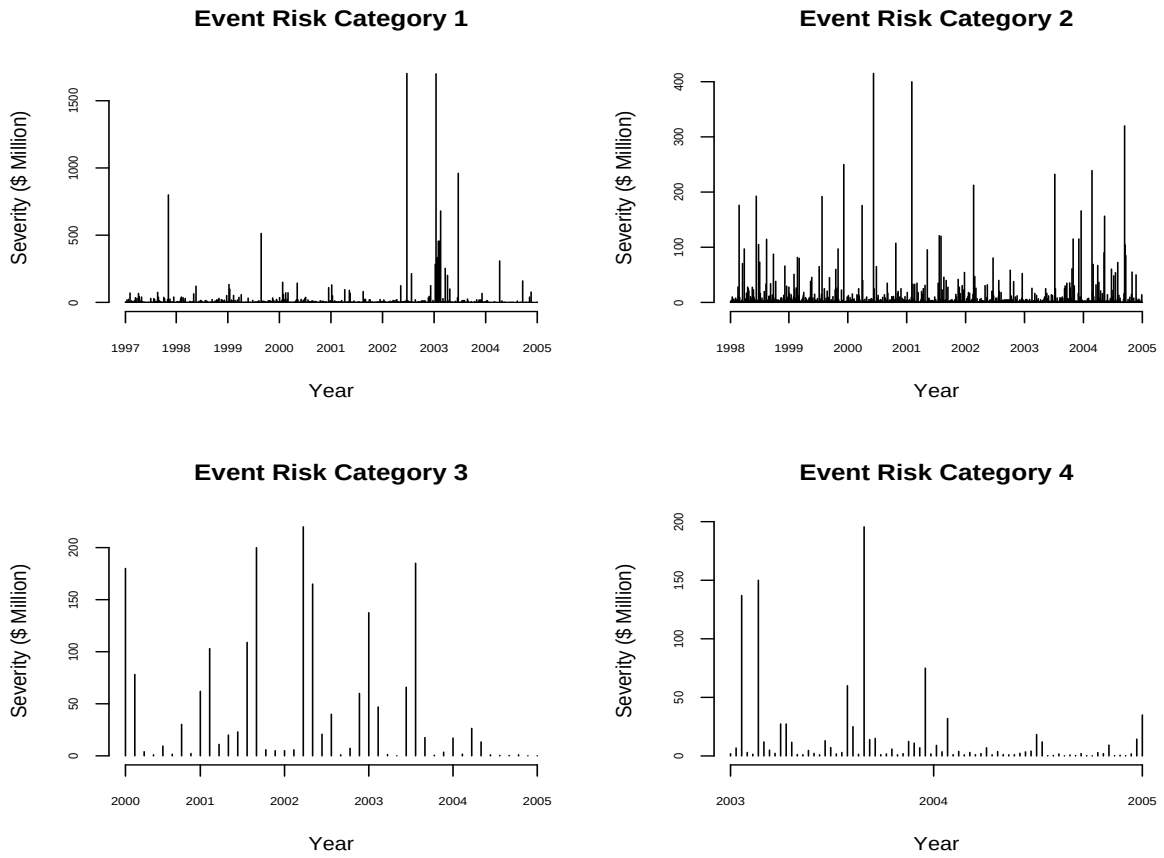

FIGURE 1: Operational risk losses for each event risk category over the collection period.

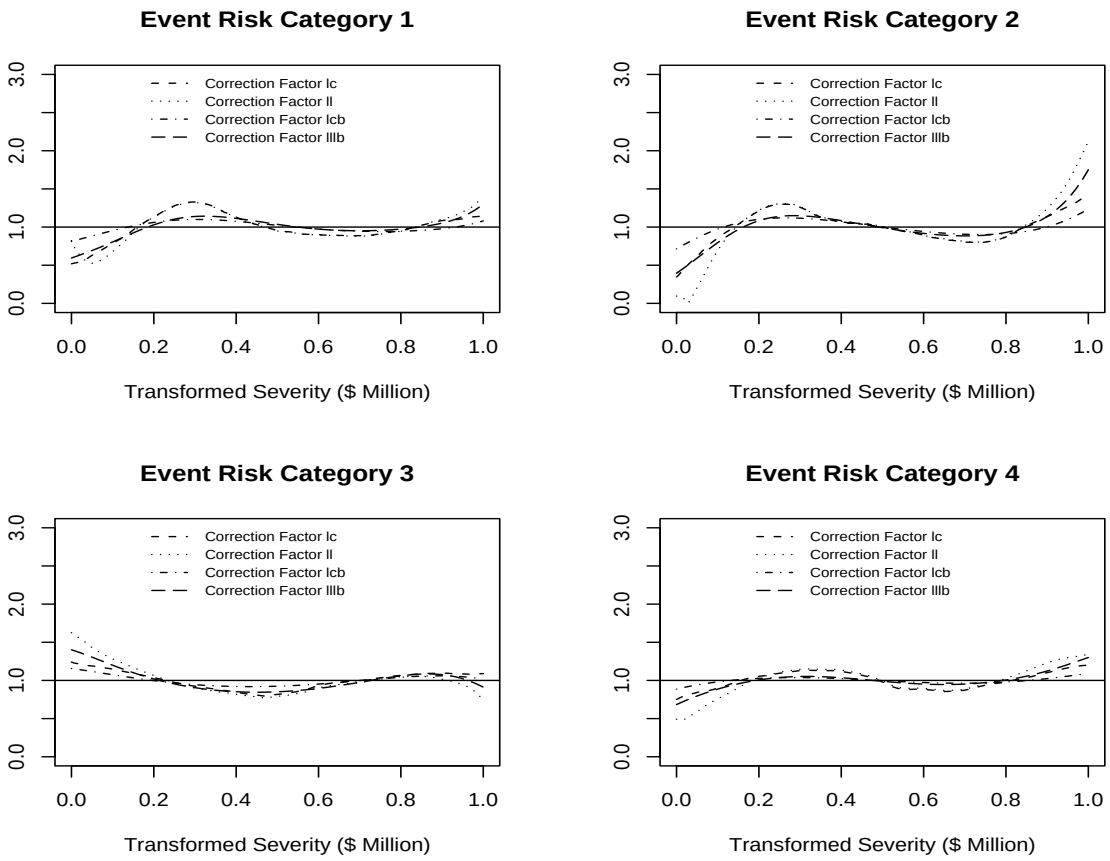

FIGURE 2: The different correction factors estimated on transformed operational risk losses for each event risk category.

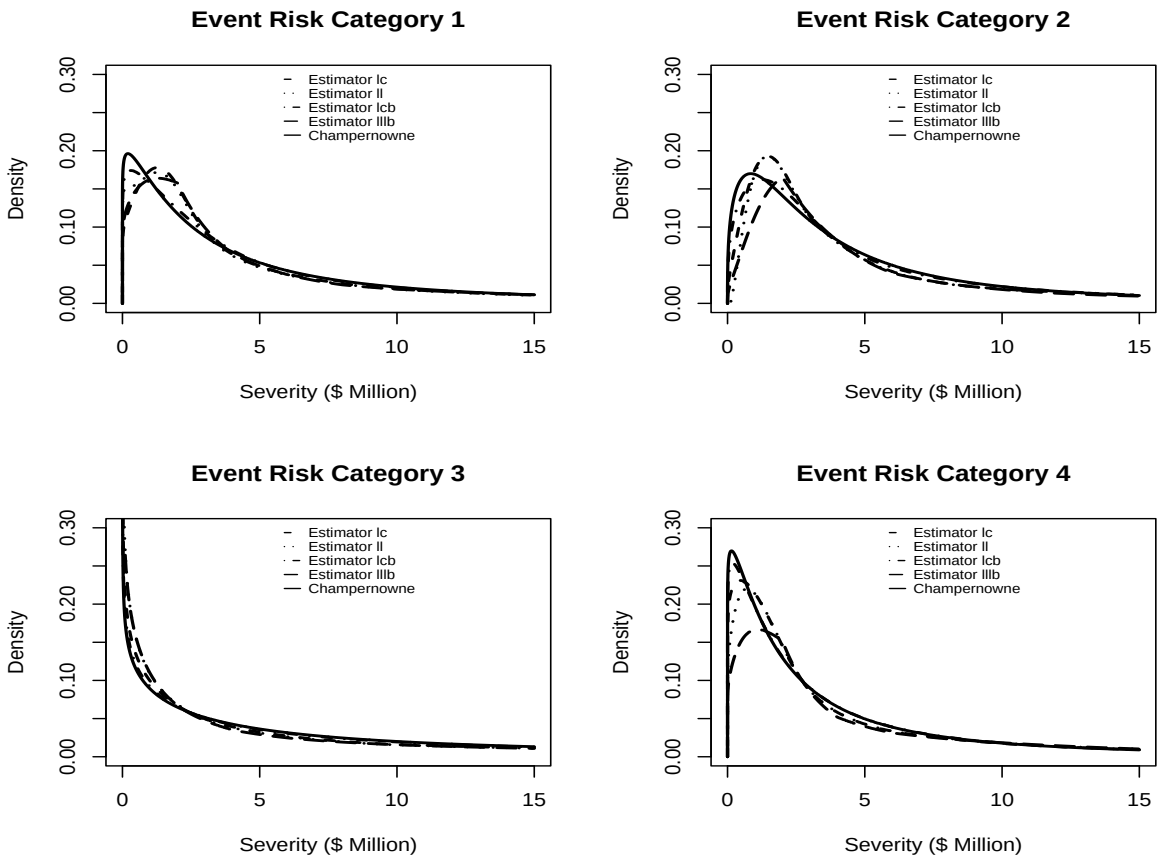

FIGURE 3: Semiparametric density plots obtained by correcting the parametric Champernowne start density by four different nonparametric correction factor estimates.

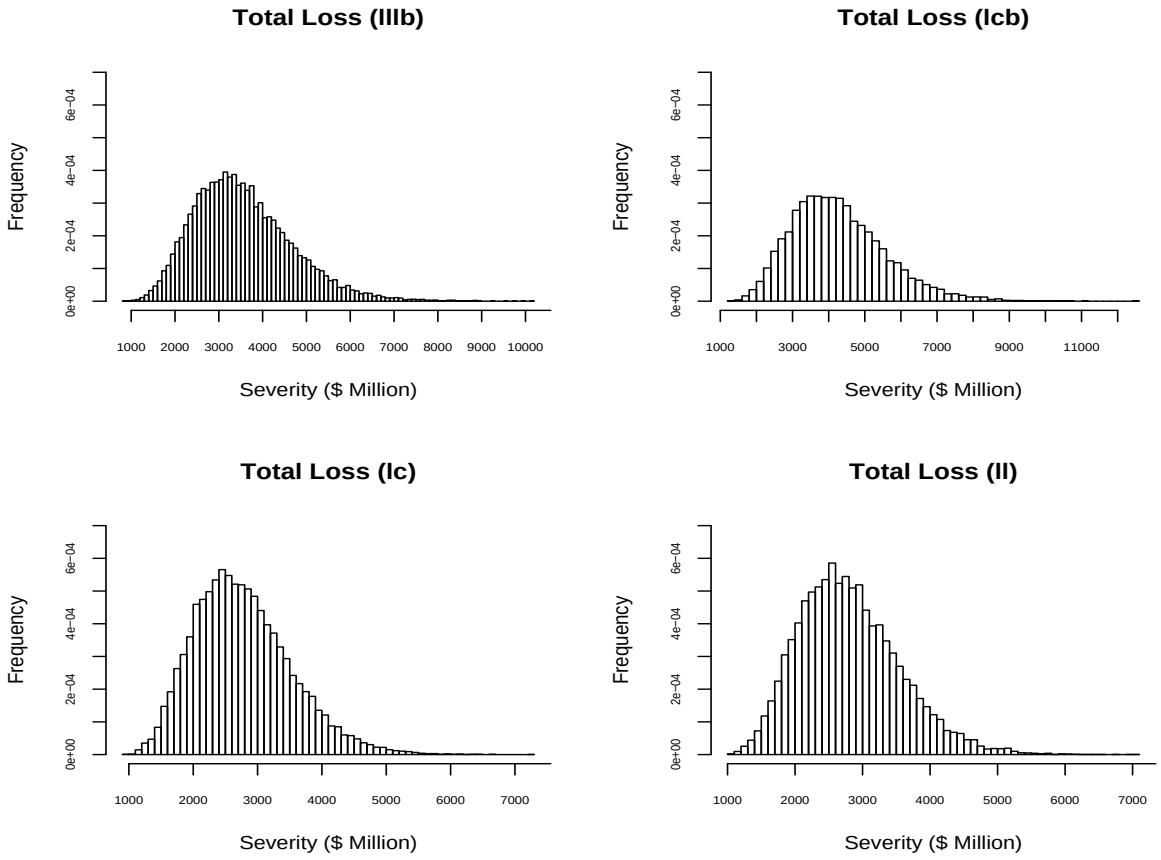

FIGURE 4: Total loss distribution frequencies for the four different models obtained through 20,000 simulations.