Jim Gustafsson (Denmark)

A mixing model incorporating three sources of data for operational risk quantification

Abstract

To meet the new Solvency II Directive for operational risk capital assessment, an insurance company's internal model should make use of internal data and relevant external information. One of the unresolved challenges with operational risk quantification is combining different information sources appropriately (e.g., internal data, consortium data and publicly reported losses). This paper develops a systematic approach that, apart from internal data, incorporates two sources of prior knowledge into internal loss distribution modelling. The standard statistical model resembles the idea with credibility theory and Bayesian methodology, in the sense that the sources of prior knowledge are weighted more when internal data is scarce than when internal data is abundant.

Keywords: operational risk, mixing data sources, actuarial loss models, transformation, multiplicative bias reduction, pre- and post insurance loss distribution.

Introduction

Operational risk refers to the possibility of unexpected events that occur in the normal course of business and includes all things that can happen in a company's daily activities. When we want to quantify how often this can happen and estimate the consequences of its occurrence. Once the phenomenon has been characterized then statistical methods will allow extrapolations[1]. One could, for example, be concerned about what would be a bound for the losses derived from operational risk with a 99.5% probability, i.e. a loss which occurs once every two hundred years. Extrapolating to know your risk for a specific risk tolerance level and then holding the appropriate capital to cover it, bring a level of uncertainty and form the reason for a more sophisticated modelling of operational risk. Calculating loss distributions for operational risk by using only internal data often fails to identify potential risks and unprecedented large losses that have a vast impact on the accuracy of the extrapolation when estimating solvency capital. Conversely, calculating capital by using only external data should intuitively give more information when extrapolation is performed. However, this procedure provides a loss distribution that is not sensitive to internal data. Consequently, the estimated solvency capital will not increase despite the occurrence of a large internal loss, and does not decrease despite the improvements of internal controls.

During the past few years, the interest in and need for useful, accurate and robust methods for operational risk quantification have grown both in the banking and the insurance sector. The general perception is that external data can often be useful in improving the estimation of operational risk loss distributions due to the lack of reliable internal data. The present challenge is to incorporate external information into internal loss distribution modelling by using a systematic approach. Recent methods described in the literature that suggest solutions to this problem are limited, however, the papers by Shevchenko and Wiithrich (2006), Biihlmann, Shevchenko and Wiithrich (2007) and Lam-brigger, Shevchenko and Wiithrich (2007) combine loss data with scenario analysis information via Bayesian inference for the assessment of operational risk. Verrall, Cowell and Khoon (2007) examine Bayesian networks for operational risk assessment. Figini, Guidici, Uberti and Sanyal (2008) develop a method to estimate internal data distribution applying truncated external data originating from the exact same underlying distribution. Wei (2007) applies Bayesian credibility to combine external and internal data, a Bayesian approach is utilized to estimate the frequency distribution and a covariate is introduced to estimate the severity distribution. This framework allows the use of both internal and external data. Gustafsson and Nielsen (2008) develop a systematic approach that incorporates external information into internal loss distribution modelling. The standard statistical model resembles Bayesian methodology and credibility theory as the above references in the sense that prior knowledge (external data) has more weight when internal data is scarce than when internal data is abundant.

Klugman, Panjer and Willmot (1998), McNeil, Frey and Embrechts (2005), Cizek, Hardle and Weron (2005) and Panjer (2006) are important introductions to actuarial estimation techniques of purely parametric loss distributions. Embrechts, Kliippelberg and Mikosch (1999) focus on the tail and offer a broad methodology for estimating these rare events by Extreme Value Theory (EVT). Further, the recent paper by Degen, Embrechts and Lambrigger (2007) discusses some fundamental

[1] To estimate the value of a variable outside a known range by assuming that the estimated value follows logically from the known ones (Wiktionary).

properties of the *g-and-h* distribution and how it is linked to the well documented EVT based methodology. However, if one focuses on the excess function, the link to the Generalized Pareto Distribution (GPD) is an extremely slow convergency and capital estimation for low level of risk tolerance using EVT may lead to inaccurate results if the parametric *g-and-h* distribution is chosen as a model. Also Diebold, Schuermann and Stroughair (2001) and Dutta and Perry (2006) stress the weaknesses of EVT when it comes to real data analysis. This problem may be solved by non- and semi-parametric procedures. A lot of research has focussed on kernel density estimation and regression curves. Estimating probability densities (e.g., Silverman (1986)), regression functions (e.g., Fan and Gijbels (1996)) and higher order kernels (e.g., Jones and Foster (1986)) are perhaps the most popular methods. During the last decade a class of semi-parametric methods was developed and designed to work better than a pure non-parametric approach (see Bolance, Guillen and Nielsen, 2003; Hjort and Glad, 1995; Jones, Linton and Nielsen, 1995; and Clements, Hum and Lindsey, 2003). They showed that non-parametric estimators could be substantially improved in a transformation process and they offer several alternatives for the transformation itself. Here, we develop some of the ideas from the references above. The main contribution of this paper is to propose an estimation method to enable an assessment of a company exposure. The framework allows the use of both internal data, consortium data and publicly reported losses to quantify operational risk, enabling the performance of specific calculation as well as more accurate extrapolation for solvency capital.

This paper proceeds as follows: Sections 1, 2 and 3 give a theoretical overview building the proposed internal loss distribution model in a structural and intuitive manner. Section 1 lays out the theory when only one data source is available, section 2 presents the model by Gustafsson and Nielsen (2008), i.e. when two sources of data could be utilized. In the third section the proposed model is developed and explained. Building on the previous sections, the model will be able to estimate operational risk exposure based on two sources of prior knowledge in combination with a company's internal data. In Section 4, we evaluate the different model characteristics of the proposed model. A number of scenarios are provided for prior knowledge data which should capture different situations a company could face when solvency capital for operational risk is to be settled. In Section 5, an application is provided that illustrates the usefulness of the proposed estimation framework. We should see that the conclusions drawn on the proposed model characteristics in Section 4 are reproduced when using real operational risk data. The study is based on an internal sample set including 89 data points collected over 2 years, a consortium data set with 613 observations with a 5-year collection period, and publicly reported data collected over 9 years and a total of 885 losses. Here all data sets originate from the same event risk category: 'Execution, Delivery and Process Management'.

1. A semi-parametric model incorporating one data source

It has been suggested that a useful approach to model loss data is nonparametric smoothing since this ideally should benefit from a parametric distribution. The most popular nonparametric estimator of an unknown probability density function f is the standard kernel estimator proposed by Rosenblatt (1956).

$$\hat{f}(r) = n^{-1}\sum_{i=1}^{n} K_h(R_i - r). \qquad (1.1)$$

Here, $(R_i)_{1\le i\le n}$ is an independent univariate sample, $h > 0$ is a bandwidth parameter evaluated on $(R_i)_{1\le i\le n}$, $K_h(r) = h^{-1}K(h^{-1}r)$ and $K(\cdot)$ is a unimodal probability density function symmetric about zero with support [-1,1]. The consistency of (1.1) is well documented when the support is unbounded (see, e.g., Silverman, 1986). However, this approach is only suitable when the number of observations is high. The convergency 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 operational risk losses are positive variables, there will exist a boundary bias with the proposed estimator by Rosenblatt (1956). This boundary bias is due to the use of a symmetric kernel K that assigns probability mass outside the support when smoothing is carried out near and at the boundary. To solve the boundary issue a number of solutions have been proposed (see, e.g., Jones and Foster, 1996). A nonparametric estimator with no preferences will work reasonably well for almost any shape. However, during the last decade a class of semi-parametric methods was developed and designed to work better than a pure nonparametric approach. Wand, Marron and Ruppert (1991) established the semi-parametric procedure and showed that a nonparametric approach could be substantially improved with a transformation technique. A slightly adjusted approach could be found in Bolance, Guillen and Nielsen (2003), where they improved the transformation for skewed data. Also Clements, Hum and Lindsey (2003)

followed Wand et al. (1991) but introduced a Mobius-like transformation. In the spirit of Wand et al. (1991) many remedies have been proposed. Incorporating only one data source, Buch-Larsen, Bolance, Guillen and Nielsen (2005) and Gustafsson, Hagmann, Nielsen and Scaillet (2008) found a mapping from the original axis to [0,1] via a parametric start and corrected locally for possible misspecification with non-parametric kernel smoothing. This section lays out the theory presented in the latter reference.

Assume $(X_i)_{1\le i\le n} \in \Re_+$ is a sequence of internal random losses from a probability distribution *F(x)* with unknown density *f(x)* that have two continuous derivatives everywhere. Let $T\left(x,\theta_1^x\right)$ be a parametric family of cumulative distribution functions, indexed by the multidimensional global parameter $\theta_1^x = \left\{\theta_{11}^x,\theta_{12}^x,...,\theta_{1p}^x\right\} \in \Theta_1^x \in \Re^p$. As always when estimating loss distributions the aim is to predict the true density *f(x)*. Let the parametric probability density function $T'\left(x,\theta_1^x\right)$ be a first choice, serving as a global parametric start, and assumed to provide a meaningful but potentially inaccurate description of the true density *f(x)*. To create a semi-parametric model which should correct the potentially inaccurate global parametric start $T'\left(x,\theta_1^x\right)$, is to utilize the transformation technique established by Wand, Marron and Ruppert (1991) and transform the internal losses $(X_i)_{1\le i\le n}$ to bounded support [0,1]. With losses transformed to [0,1] the estimation problem is now to find a model for the true density $r(u)\in[0,1]$. For this, let $\phi\left(u,\theta_2(u)\right)$ serve as a local parametric model with the local function $\theta_2(u)=\left\{\theta_{21}(u),\theta_{22}(u),...,\theta_{2q}(u)\right\}\in\Theta_2\in\Re^q$ for the unknown density function $r(u)$. The aim of the model $\phi\left(u,\theta_2(u)\right)$ is to correct the parametric probability density $T'\left(x,\theta_1^x\right)$ locally according to data availability. The interpretation of this is intuitive, in areas where the global parametric start assigns too much (or too small) probability mass according to observed data, the local model captures this and corrects for this misspecification. The semi-parametric model which will form the basis for estimating the true density $f(x)$ is

$$m\left(x,\theta_1^x,\theta_2\left(T\left(x,\theta_1^x\right)\right)\right)= \\ = T'\left(x,\theta_1^x\right)\cdot\phi\left(T\left(x,\theta_1^x\right),\theta_2\left(T\left(x,\theta_1^x\right)\right)\right). \quad (1.2)$$

The process of estimating this semi-parametric model begins with estimate of the global parameter θ_1^x by the maximum likelihood. The maximum likelihood estimator $\hat\theta_1^x$ is employed in $T\left(x,\hat\theta_1^x\right)$ to transform the internal random losses $(X_i)_{1\le i\le n}$ into the support [0,1], and the succeeding estimated transformed data $\hat U_i = T\left(X_i,\hat\theta_1^x\right)$ is used to find an accurate estimated local parametric model $\phi\left(\hat u,\hat\theta_2\left(\hat u\right)\right)$ with $\hat u_i = T\left(X_i,\hat\theta_1^x\right)$ that should correct the global start density $T'\left(x,\hat\theta_1^x\right)$ locally. This lends itself if $\hat\theta_1^x$ is a misspecified approximation of the true density $f(x)$, $\hat\theta_1^x$ should converge in probability to the pseudo true value θ_1^0 which is the least false value according to the Kullback-Leibler (KL) divergency, thereby minimizes the distance between $T'\left(x,\hat\theta_1^x\right)$ and *f(x)*. Consequently, around each $\hat u$ on the transformed data $\left(\hat U_i\right)_{1\le i\le n}$, the best local approximation will be denoted by $\hat\theta_2\left(\hat u\right)$ with theoretical counterpart $\theta_2^0(u)$, with $u=T\left(x,\theta_1^0\right)$, and determined such that

$$n^{-1}\sum_{i=1}^{n} K_h\left(\hat U_i-\hat u\right)\upsilon\left(\hat u,\hat U_i,\theta_2\left(\hat u\right)\right)- \\ -\int_0^1 K_h\left(t-\hat u\right)\upsilon\left(\hat u,t,\theta_2\left(\hat u\right)\right)\phi\left(t,\theta_2\left(\hat u\right)\right)dt=0 \quad (1.3)$$

holds. Here the weight function $\upsilon\left(\hat u,t,\theta_2\left(\hat u\right)\right)$ is a vector of dimension $q \times 1$. If we choose the local model $\phi\left(\hat u,\theta_2\left(\hat u\right)\right)$ as a constant, and the score $\partial\log\phi\left(\hat u,\theta_2\left(\hat u\right)\right)/\partial\theta_2\left(\hat u\right)$ as the weight function, explicitly $\upsilon\left(\hat u,t,\theta_2\left(\hat u\right)\right)=1$, the KL

divergency assigns a semi-parametric estimator that takes the expression

$$m_1\left(x,\hat{\theta}_1^x,\hat{\theta}_2\left(T\left(x,\hat{\theta}_1^x\right)\right)\right)=$$

$$T'\left(x,\hat{\theta}_1^x\right)\cdot\phi\left(T\left(x,\hat{\theta}_1^x\right)\hat{\theta}_2\left(T\left(x,\hat{\theta}_1^x\right)\right)\right)=$$

$$=\frac{T'\left(x,\hat{\theta}_1^x\right)}{n\alpha_{01}\left(T\left(x,\hat{\theta}_1^x\right),h\right)}\sum_{i=1}^{n}K_h\left(T\left(X_i,\hat{\theta}_1^x\right)-T\left(x,\hat{\theta}_1^x\right)\right) \quad (1.4)$$

with

$$\alpha_{ij}(u,h)=\int_{\max\{-1,-u/h\}}^{\min\{1,(1-u)/h\}}(\upsilon h)^i K(\upsilon)^j\,d\upsilon \quad (1.5)$$

defining the asymptotic properties of the kernel density estimator. The indexation on the semiparametric model $m_1(\cdot)$ indicates that the model incorporates one data source. A deeper theoretical exposition can be found in Gustafsson et al. (2008).

The asymptotic bias and variance for (1.4) is well documented in the literature under the constraint as $n\to\infty$, $h\to 0$, and $nh_x\to\infty$. Given that $\hat{\theta}_1^x$ converges with the parametric rate to the pseudo true value θ_1^0, which is faster than a nonparametric rate, it allows us to disregard the noise associated with the estimation of θ_1^0 since it will have no influence on the leading bias and variance terms of $m_1\left(x,\hat{\theta}_1^x,\hat{\theta}_2^x\left(T\left(x,\hat{\theta}_1^x\right)\right)\right)$. Since $u=T\left(x,\theta_1^0\right)$, the transformed random variable $U_i=T\left(X_i,\theta_1^x\right)$ is distributed as follows

$$r(u)=\frac{f\left(T^{\leftarrow}\left(u,\theta_1^0\right)\right)}{\frac{\partial}{\partial\eta}T(\eta,\theta_1^0)\,|_{\eta=T^{\leftarrow}\left(u,\theta_1^0\right)}}.$$

It follows that the bias and variance for (1.4) could be expressed as

$$\mathrm{E}m_1\left(x,\hat{\theta}_1^x,\hat{\theta}_2\left(T\left(x,\hat{\theta}_1^x\right)\right)\right)-f(x)\cong$$

$$\cong\frac{1}{2}\alpha_{21}\left(T\left(x,\theta_1^0\right),h\right)\left(\left(\frac{f(x)}{T'\left(x,\theta_1^0\right)}\right)'\frac{1}{T'\left(x,\theta_1^0\right)}\right)'+o_p\left(h^2\right),$$

$$Vm_1\left(x,\hat{\theta}_1^x,\hat{\theta}_2\left(T\left(x,\hat{\theta}_1^x\right)\right)\right)\cong$$

$$\cong\frac{\alpha_{02}\left(T\left(x,\theta_1^0\right),h\right)}{nh}f(x)T'\left(x,\theta_1^0\right)+o_p\left(\frac{1}{nh}\right).$$

The procedure to obtain the bias and variance expression above could be found in the recent papers of Buch-Larsen et al. (2005) or Gustafsson et al. (2008).

2. A semi-parametric model incorporating two data sources

This section demonstrates how to incorporate prior knowledge into model (1.4), thereby enriching the estimation of an internal loss distribution. The model presented in this section was developed by Gustafsson and Nielsen (2008), where they argue that combining internal data with prior knowledge is the way to get a more accurate extrapolation of a company solvency capital for operational risk. As in model (1.4), assume $(X_i)_{1\le i\le n}\in\Re_+$ is a sequence of internal random losses and extend the information by letting $(Y_j)_{1\le j\le m}\in\Re_+$ be a sample of external losses (e.g., publicly reported losses or consortium data). In addition, let $T\left(x,\theta_1^y\right)$ be a parametric family with density function $T'\left(x,\theta_1^y\right)$, indexed by the global external parameter $\theta_1^y=\left\{\theta_{11}^y,\theta_{12}^y,...,\theta_{1p}^y\right\}\in\Theta_1^y\in\Re^p$. From Gustafsson and Nielsen (2008), we know that this global start should provide prior knowledge of the true density $f(x)$ that is a less inaccurate description than in the situation using θ_1^x as a parametric start. The estimation process begins by estimating θ_1^y on $(Y_j)_{1\le j\le m}$ using maximum likelihood. As in the previous section we should transform the internal losses $(X_i)_{1\le i\le n}$ to the bounded support [0,1], however by considering the transformation $\hat{S}_i=T\left(X_i,\hat{\theta}_1^y\right)$, prior knowledge will be incorporated into the transformation.

As a consequence of this, the estimated local model for the unknown density function $r(s)$ will now refer to $\phi\left(\hat{s},\hat{\theta}_2\left(\hat{s}\right)\right)$, with estimated local function

$$\hat{\theta}_2\left(\hat{s}\right)=\left\{\hat{\theta}_{21}^x\left(\hat{s}\right),\hat{\theta}_{22}^x\left(\hat{s}\right),...,\hat{\theta}_{2q}^x\left(\hat{s}\right)\right\}\in\Theta_2\in\Re^q$$

and $\hat{s}=T\left(x,\hat{\theta}_1^y\right)$. From the previous section we know that if $\hat{\theta}_1^y$ is a misspecified approximation of $f(x)$ a convergency in probability to the pseudo true value θ_1^0 takes place that minimizes the KL distance between $T'\left(x,\hat{\theta}_1^y\right)$ and $f(x)$. Now, choose $\theta_2\left(\hat{s}\right)$ such that

$$n^{-1}\sum_{i=1}^{n}K_h\left(\hat{S}_i-\hat{s}\right)\upsilon\left(\hat{s},\hat{S}_i,\theta_2\left(\hat{s}\right)\right)-$$

$$-\int_0^1 K_h\left(t-\hat{s}\right)\upsilon\left(\hat{s},t,\theta_2\left(\hat{s}\right)\right)\phi\left(\hat{s},\theta_2\left(\hat{s}\right)\right)dt=0 \quad (2.1)$$

holds. Note that the theoretical counterpart is the same as in the scenario with one data source, therefore this model has the same asymptotic bias and variance as (1.4). This is explained in more depth in Gustafsson and Nielsen (2008). Also notable is that the bandwidth is still estimated on the internal sample set. Since we want that as $n\to\infty$ then $h\to 0$, meaning that as internal data increases more correction should be done, so internal data becomes more significant in the internal loss distribution model on the source of prior knowledge. The estimated extended semi-parametric model that incorporates two data sources takes the form

$$m_2\left(x,\hat{\theta}_1^y,\hat{\theta}_2,\left(T\left(x,\hat{\theta}_1^y\right)\right)\right)=$$

$$T'\left(x,\hat{\theta}_1^y\right)\cdot\phi\left(T\left(x,\hat{\theta}_1^y\right)\hat{\theta}_2\left(T\left(x,\hat{\theta}_1^y\right)\right)\right)=$$

$$=\frac{T'\left(x,\hat{\theta}_1^y\right)}{n\alpha_{01}\left(T\left(x,\hat{\theta}_1^y\right),h\right)}\sum_{i=1}^{n}K_h\left(T\left(X_i,\hat{\theta}_1^y\right)-T\left(x,\hat{\theta}_1^y\right)\right). \quad (2.2)$$

The indexation on m_2 implies that the model incorporates two data sources. Note that prior knowledge could be publicly reported losses, consortium data or scenario analysis that unite with the internal sample $\left(X_i\right)_{1\le i\le n}$ in (2.2). The model is also able to handle other combinations if internal data is not available. Examples of such combinations are: publicly reported losses with consortium data, scenario analysis with consortium data, and scenario analysis with publicly reported losses.

3. A semi-parametric model incorporating three data sources

In the situation where different prior knowledge is available to a company, one should take the opportunity to obtain a more accurate and reliable operational risk capital assessment. Certainly, making use of the model which incorporates two data sources, presented in Section 2, and combining the different prior knowledge available with collected internal data could be a flexible way to assess and settle solvency capital for operational risk. Hopefully, the extrapolated capital number from one combination (e.g., publicly reported losses and internal data) is similar to the capital figure from another combination (e.g., consortium data and internal data). However, what if the extrapolated numbers are significantly different? The optimal situation would be if all sources of prior knowledge could be incorporated in one model in a logical and intuitive way. One requirement would be if two sources of prior knowledge are available with more or less the same characteristics, the model should resemble the model presented by Gustafsson and Nielsen (2008). On the other hand, if the second added prior knowledge source differs substantially from the first, the model characteristics should handle this in an obvious and explicable way.

This section develops a model that meets these requirements. The inspiration originates from the bias reduction literature (see Hjort and Glad, 1995; Jones, Linton and Nielsen, 1995; and Jones, Signorini and Hjort, 1999). In these references different kernel density estimators are developed to reduce asymptotic bias from model (1.1), but still contain the same magnitude on the asymptotic variance. In short, the multiplicative bias correction method with underlying pilot estimator (1.1) is presented as

$$\tilde{f}(r)=\tilde{f}(r)\cdot\tilde{g}(r)=$$

$$=\tilde{f}(r)\cdot n^{-1}\sum_{i=1}^{n}\hat{f}\left(R_i\right)^{-1}K_h\left(R_i-r\right). \quad (3.1)$$

Inspired by (3.1), we let $\left(X_i\right)_{1\le i\le n}\in\Re_+$ and $\left(Y_j\right)_{1\le j\le m}\in\Re_+$ be sequences of internal and external random losses (e.g., publicly reported losses) as in the previous section, and extend the information by letting $\left(Z_k\right)_{1\le k\le q}\in\Re_+$ be another sequence of external random losses (e.g., consortium data). As above let $T\left(x,\theta_1^y\right)$ be a parametric family, indexed by the global parameter

$\theta_1^y = \{\theta_{11}^y, \theta_{12}^y, ..., \theta_{1p}^y\} \in \Theta_1^y \in \Re^p$, and let $T(x, \theta_1^z)$ be another parametric family, indexed by the global parameter $\theta_1^z = \{\theta_{11}^z, \theta_{12}^z, ..., \theta_{1p}^z\} \in \Theta_1^z \in \Re^p$. The characteristics of the new model is that the probability density function $T'(x, \theta_1^y)$ will still serve as the global parametric start as in Gustafsson and Nielsen (2008), but now the local parametric model is extended to embrace an added source of prior knowledge. Assume now that this extended correction function $\phi(w, \theta_2(s, w))$, with $s = T(x, \theta_1^y)$ and $\omega = T(x, \theta_1^z)$, is a local parametric model with local function $\theta_2(s, w) = \{\theta_{21}(s, w), \theta_{22}(s, w), ..., \theta_{2q}(s, w)\} \in \theta_2 \in \Re^q$ that should correct the global parametric start density $T'(x, \theta_1^y)$. Then, the presented semi-parametric model enriched with three different sources of data could be formulated as

$$m(x, \theta_1^y, \theta_2(T(x, \theta_1^y), T(x, \theta_1^z))) =$$
$$m(x, \theta_1^y, \theta_2(T(x, \theta_1^y))) \cdot \phi(T(x, \theta_1^y), \theta_2(T(x, \theta_1^y), T(x, \theta_1^z)))$$
$$= T'(x, \theta_1^y)$$
$$\cdot \phi(T(x, \theta_1^y), \theta_2(T(x, \theta_1^y))) \cdot \phi(T(x, \theta_1^y), \theta_2(T(x, \theta_1^y) T(x, \theta_1^z))).$$

The process of estimating (3.2) begins by estimating θ_1^y on $(Y_j)_{1 \le j \le m}$ and θ_1^z on $(Z_k)_{1 \le k \le q}$ using maximum likelihood estimation. As a consequence, different prior knowledge will be incorporated in the transformed data sets $\hat{S}_i = T(X_i, \hat{\theta}_1^y), \hat{s} = T(x, \hat{\theta}_1^y)$, and $\hat{W}_i = T(X_i, \hat{\theta}_1^z), \hat{w} = T(x, \hat{\theta}_1^z)$. Now, let $\phi(\hat{s}, \hat{\theta}_2(\hat{s}))$ be the same estimated local model as in (2.2) that incorporates two data sources, and let $\phi(\hat{s}, \theta_2(\hat{s}, \hat{w}))$ be the extended local model that incorporates an additional source of prior knowledge. The estimation of the local model is achieved by choosing the local function $\theta_2(\hat{s}, \hat{w})$ such that

$$n^{-1} \sum_{i=1}^{n} K_h(\hat{S}_i - \hat{s}) \upsilon(\hat{s}, \hat{S}_i, \theta_2^x(\hat{s}, \hat{w})) -$$
$$- \int_0^1 K_h(t - \hat{s}) \upsilon(\hat{s}, t, \theta_2^x(\hat{s}, \hat{w})) \cdot$$
$$\cdot m_2(x, \hat{\theta}_1^y, \hat{\theta}_2^x(\hat{s})) \phi(\hat{s}, \theta_2(\hat{s}, \hat{w})) dt = 0 \qquad (3.3)$$

holds. Note that equation (3.3) includes the estimated model (2.2). For simplicity, if we use the abbreviations $\hat{S}_i$, $\hat{W}_i$, $\hat{s}$ and $\hat{w}$ denned above, the estimated version of (3.2) takes the following form

$$m_3(x, \hat{\theta}_1^y, \hat{\theta}_2(T(x, \hat{\theta}_1^y)), T(x, \hat{\theta}_1^z)) =$$
$$= m_2(x, \hat{\theta}_1^y, \hat{\theta}_2^x(T(x, \hat{\theta}_1^y))) \cdot \phi(T(x, \hat{\theta}_1^y), \hat{\theta}_2(T(x, \hat{\theta}_1^y)), T(x, \hat{\theta}_1^z)) =$$
$$= m_2(x, \hat{\theta}_1^y, \hat{\theta}_2(\hat{s})) \cdot \phi(\hat{s}, \hat{\theta}_2(\hat{s}, \hat{w})) =$$
$$= T'(x, \hat{\theta}_1^y) \cdot \phi(\hat{s}, \hat{\theta}_2(\hat{s})) \cdot \phi(\hat{s}, \hat{\theta}_2(\hat{s}, \hat{w})) \qquad (3.4)$$

with

$$\phi(\hat{s}, \hat{\theta}_2(\hat{s})) = (n \alpha_{01}(\hat{s}, h))^{-1} \sum_{i=1}^{n} K_h(\hat{S}_i - \hat{s}),$$

$$\phi(\hat{s}, \hat{\theta}_2(\hat{s}, \hat{w})) = (n \alpha_{01}(\hat{s}, h))^{-1} \sum_{i=1}^{n} \frac{K_h(\hat{W}_i - \hat{s})}{\phi(\hat{S}_i, \hat{\theta}_2(\hat{s}))}.$$

The indexation on m_3 implies that this model incorporates three sources of data. The asymptotic bias and variance of (3.2) is presented in Theorem 1. As in the previous sections we know that $\hat{\theta}_1^y$ and $\hat{\theta}_1^z$ converge with a parametric rate to the pseudo true value θ_1^0, faster than the nonparametric rate, which allows us to disregard the noise associated with the estimation of θ_1^0. In the Theorem we make the following shortening to make place $m_3(x, \hat{\theta}_1^y, \hat{\theta}_2^x(T(x, \hat{\theta}_1^y), T(x, \hat{\theta}_1^z))) = \hat{m}_3$.

Theorem 1. *Let* $(X_i)_{1 \le i \le n}$, $(Y_j)_{1 \le j \le m}$ *and* $(Z_k)_{1 \le k \le q}$ *be iid random variables with density* f. *Suppose that* f *has four continuous derivatives everywhere, and that as* $n \to \infty$, $h \to 0$, *and* $nh \to \infty$. *Then the asymptotic bias and variance for* m_3 *is presented as*

$$\mathrm{E}\hat{m}_3 - f(x) \cong -\frac{h^4}{4} \alpha_{21}(T(x, \theta_1^0), h) \cdot$$
$$\cdot f(x) \left(\alpha_{21}(T(x, \theta_1^0) h \left(\left(\frac{f(x)}{T'(x, \theta_1^0)} \right)' \frac{1}{T'(x, \theta_1^0)} \right)' \frac{1}{f(x)} \right)'' +$$
$$+ o_p(h^4)$$

$$V\hat{m}_3 \cong \frac{T'(x, \theta_1^0) f(x)}{nh} \int \tilde{K}(t) dt + o_p\left(\frac{1}{nh}\right)$$

with $\tilde{K}(t) \cong 2K(t) - K * K(t)$ *and* $K * \tilde{K}(t) = \int K(t - \upsilon) K(\upsilon) d\upsilon$ *is the convolution.*

The calculation of the asymptotic bias and variance can be found in Appendix A.

4. Characteristics of the models

In this section, we evaluate the characteristics of the models considered in the previous sections. The mixing models (2.2) and (3.4) are interpreted and compared to (1.4) by adding different types of prior knowledge. Throughout the study, we assume that the internal sample $(X_i)_{1\le i\le n}$ is a sequence of random losses drawn from a $\log N(\mu_x,\sigma_x)$ distribution with sample size n = 200, and true density function

$$f(x)\frac{e^{-(\log\{x\}-\mu_x)^2/(2\sigma_x^2)}}{x\sigma_x\sqrt{2\pi}},\ x\in\Re_+. \qquad (4.1)$$

The true density is parameterized in terms of location and scale parameters such that $\{\mu_x,\sigma_x\}=\{1,1\}$.

4.1. Evaluation of the characteristics by incorporating one source of prior knowledge. Assume $(Y_j)_{1\le j\le m}$ is a sequence of external random losses (e.g., publicly reported losses) with m = 1000 drawn from (4.1) but with location and scale parameters $\{\mu_y,\sigma_y\}$. The study commences by evaluating different scenarios for different parameters $\{\mu_y,\sigma_y\}$ and intepreting the effect on model (2.2) compared to model (1.4). This allows us to see how mixing model (2.2) peforms when different external data characteristics are available. In total, eight different scenarios will be studied between models (2.2) and (1.4) and are summarized in Table 1.

Table 1. Different scenarios considered on the parametrization between models (2.2) and (1.4)

	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6	Scenario 7	Scenario 8
$\{\mu_y,\sigma_y\}$	{1, 1}	{1, 1}	{1, 1}	{1, 1}	{1, 1}	{1, 1}	{1, 1}	{1, 1}
$\{\mu_x,\sigma_x\}$	$\left\{\mu_x,\frac{3}{4}\sigma_x\right\}$	$\{\mu_x,\sigma_x\}$	$\left\{\mu_x,\frac{5}{4}\sigma_x\right\}$	$\left\{\mu_x,\frac{3}{2}\sigma_x\right\}$	$\left\{\frac{3}{4}\mu_x,\sigma_x\right\}$	$\{\mu_x,\sigma_x\}$	$\left\{\frac{5}{4}\mu_x,\sigma_x\right\}$	$\left\{\frac{3}{2}\mu_x,\sigma_x\right\}$

As seen in Table 1, the eight scenarios differ in that the first four situations utilize the same location parameter on the external sample as the internal sample, i.e. $\mu_x=\mu_y$, and changes are made on the scale parameter. For the remaining four scenarios, the opposite is studied, holding the scale parameters equal and changing the location on the external sample.

The underlying distribution assumption made for models (2.2), (3.4) and (1.4) is the Generalized Champernowne Distribution GCD $\theta_{11},\theta_{12},\theta_{13}$ with density function

$$T'(x,\theta_1)=\frac{\theta_{11}(x+\theta_{13})^{(\theta_{11}-1)}\left((\theta_{12}+\theta_{13})^{\theta_{11}}-(\theta_{13})^{\theta_{11}}\right)}{\left((x+\theta_{13})^{\theta_{11}}+(\theta_{12}+\theta_{13})^{\theta_{11}}-2(\theta_{13})^{\theta_{11}}\right)^2},$$
$$x\in\Re_+, \qquad (4.2)$$

where the parameters $\theta_{11}>0$, $\theta_{12}>0$ and $\theta_{13}>0$ are defined by the global parameter $\theta_1=\{\theta_{11},\theta_{12},\theta_{13}\}\in\Theta_1\in\Re^3$. Here θ_{11} is a tail parameter, θ_{12} controls the body of the distribution, and θ_{13} is a shift parameter that controls the domain close to zero. The global parameter θ_1 in (4.2) is estimated by maximum likelihood on each pseudo scenario sample $(X_i)_{1\le i\le 200}$ and $(Y_j)_{1\le j\le 1000}$, resulting in $\hat{\theta}_1^x$ and $\hat{\theta}_1^y$. The kernel function K is chosen as the Epanechnikov function and the bandwidth h is estimated on the transformed axes using Silverman's plug-in bandwidth (see Silverman, 1986).

Consequently, models (2.2) and (1.4) are estimated and evaluated in Figure 1 (Appendix B).

The top row in Figure 1 represents the first four scenarios considered in Table 1. Here, an identical location parameter is assumed between the samples, while scenarios change from a low to a high volatile data on the external sample. The first graph indicates that when prior knowledge data is less volatile than collected internal data, model (2.2) places more probability mass in the beginning of the domain. Thereby, model (2.2) proves to have light tail characteristics compared to model (1.4). As 12 the external data becomes more volatile, the mode remains the same but with less probability mass. Thus, the tail becomes more extreme for (2.2). The final four scenarios in Table 1 are given by the bottom row in Figure 1 (Appendix B). Here, the standard deviation is maintained, while we find the mode shifting to the right when increasing the location parameter.

4.2. Evaluation of the characteristics by incorporating two sources of prior knowledge. We continue with investigating the characteristics of the developed model (3.4). Assuming the same internal data and publicly reported data as above, we extend the analysis by adding an extra source of prior knowledge. Let $(Z_k)_{1\le k\le q}$ with q = 1000, be

a sequence of random losses (e.g., consortium data) drawn from (4.1) with location and scale parameters $\{\mu_z, \sigma_z\}$. In this section, we consider sixteen different scenarios: the first eight present the same information on the prior knowledge sources $(Y_j)_{1\le j\le 1000}$ and $(Z_k)_{1\le k\le 1000}$ while the remaining eight scenarios consider different information.The first eight scenarios that will be investigated between models (3.4) and (1.4) are summarized in Table 2.

Table 2. Different scenarios considered on the parametrization between models (3.4) and (1.4)

	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6	Scenario 7	Scenario 8
$\{\mu_x, \sigma_x\}$	{1, 1}	{1, 1}	{1, 1}	{1, 1}	{1, 1}	{1, 1}	{1, 1}	{1, 1}
$\{\mu_y, \sigma_y\}$	$\{\mu_x, \frac{3}{4}\sigma_x\}$	$\{\mu_x, \sigma_x\}$	$\{\mu_x, \frac{5}{4}\sigma_x\}$	$\{\mu_x, \frac{3}{2}\sigma_x\}$	$\{\frac{3}{4}\mu_x, \sigma_x\}$	$\{\mu_x, \sigma_x\}$	$\{\frac{5}{4}\mu_x, \sigma_x\}$	$\{\frac{3}{2}\mu_x, \sigma_x\}$
$\{\mu_z, \sigma_z\}$	$\{\mu_x, \frac{3}{4}\sigma_x\}$	$\{\mu_x, \sigma_x\}$	$\{\mu_x, \frac{5}{4}\sigma_x\}$	$\{\mu_x, \frac{3}{2}\sigma_x\}$	$\{\frac{3}{4}\mu_x, \sigma_x\}$	$\{\mu_x, \sigma_x\}$	$\{\frac{5}{4}\mu_x, \sigma_x\}$	$\{\frac{3}{2}\mu_x, \sigma_x\}$

Figure 2 (Appendix B) presents the results for the eight scenarios. As an appealing feature model (3.4) behaves similarly to model (2.2) when adding similar data. This is encouraging since if the information from the new data source have similar characteristics as the source of prior knowledge that already is incorporated, model (3.4) falls down on model (2.2). Logically, this is how it should be, since the exposure should not change if no different information is added to the model. In this situation, the local parametric model $\phi\left(\hat{s}, \hat{\theta}_2(\hat{s}, \hat{w})\right)$ in (3.4) is close to one, and the remaining functions in (3.4) will be multiplied with something close to one.

The final eight scenarios which we will evaluate are summarized in Table 3. Here, different characteristics are assumed on the two sources of prior knowledge.

Table 3. Different scenarios considered on the parametrization between models (3.4) and (1.4)

	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6	Scenario 7	Scenario 8
$\{\mu_x, \sigma_x\}$	{1, 1}	{1, 1}	{1, 1}	{1, 1}	{1, 1}	{1, 1}	{1, 1}	{1, 1}
$\{\mu_y, \sigma_y\}$	$\{\mu_x, \frac{3}{4}\sigma_x\}$	$\{\mu_x, \sigma_x\}$	$\{\mu_x, \frac{5}{4}\sigma_x\}$	$\{\mu_x, \frac{3}{2}\sigma_x\}$	$\{\frac{3}{4}\mu_x, \sigma_x\}$	$\{\mu_x, \sigma_x\}$	$\{\frac{5}{4}\mu_x, \sigma_x\}$	$\{\frac{3}{2}\mu_x, \sigma_x\}$
$\{\mu_z, \sigma_z\}$	$\{\mu_x, \frac{3}{2}\sigma_x\}$	$\{\mu_x, \frac{5}{4}\sigma_x\}$	$\{\mu_x, \sigma_x\}$	$\{\mu_x, \frac{3}{4}\sigma_x\}$	$\{\frac{3}{2}\mu_x, \sigma_x\}$	$\{\frac{5}{4}\mu_x, \sigma_x\}$	$\{\mu_x, \sigma_x\}$	$\{\frac{3}{4}\mu_x, \sigma_x\}$

Figure 3 (Appendix B) tells us that if an extra prior knowledge source is added which has the same or larger standard deviation than the external data source added first, the model takes that into account and presents a heavier tail. If we add a less volatile source than already presented, the model decreases in the tail.

5. Pre- and post-insurance loss distribution

This section demonstrates the usefulness of the proposed model in an empirical study. The data considered are an internal sample set $(X_i)_{1\le i\le n}$, publicly reported losses $(Y_j)_{1\le j\le m}$ filtered for relevance, and insurance consortium data $(Z_k)_{1\le k\le q}$, where all sources originate from the event risk category: 'Execution, Delivery and Process Management'. The objective is to evaluate next years' operational risk exposure by estimating total loss distributions on the presented data with pre- and post-insurance recoveries, then employ the risk measure Value-at-Risk (VaR) for different levels of risk tolerance. The developed model (3.4) will be evaluated against five benchmark models in a Monte Carlo simulation, where all six models lean on the same frequency assumption. Table 4 shows the descriptive statistics of the three data sets included in the study.

Table 4. Statistics for event risk category execution, delivery and process management

	Number of losses	Maximum loss (m£)	Sample mean (m£)	Sample median (m£)	Standard deviation (m£)	Time horizon (T)
Internal data, $(X_i)_{1\le i\le n}$	89	1.05	0.08	0.03	0.17	2
Consortium data, $(Z_k)_{1\le k\le q}$	613	31.4	0.44	0.03	2.25	5
Publicly reported data, $(Y_j)_{1\le j\le m}$	885	106.3	1.51	0.04	6.61	9

It should be noted that the mean is larger than the median in all cases, consistent with right skewed data. Further, the number of losses, the maximum loss, the standard deviation and the collection period are different. Also remarkable are the wide differences in the maximum loss.

5.1. The severity models. The global parametric start for the severity models is estimated by the maximum likelihood. Here, we assume that the underlying distributions $T\left(x,\theta_1^x\right)$, $T\left(x,\theta_1^y\right)$ and $T\left(x,\theta_1^z\right)$ originate from a GCD defined by (4.2) and the respective estimated global parameters are provided in Table 5.

Table 5. Maximum likelihood estimation of the global parameters for a GCD

	$\hat{\theta}_{11}$	$\hat{\theta}_{12}$	$\hat{\theta}_{13}$
$\hat{\theta}_1^x$ on the internal data $(X_i)_{1\le i\le n}$	1.714	0.032	0
$\hat{\theta}_1^z$ on the consortium data $(Z_k)_{1\le k\le q}$	1.224	0.033	0
$\hat{\theta}_1^y$ on the publicly reported data $(Y_j)_{1\le j\le m}$	0.440	0.044	0.067

It should be noted that the estimated tail parameter $\hat{\theta}_{11}$ for the different data sets is widely different. The publicly reported data provide a very heavy tail since the corresponding low value of $\hat{\theta}_{11}$. The internal data set appears not to be significantly heavy tailed, and the consortium data offer a moderate tail for the GCD. For the sake of simplicity, we introduce the abbreviations $\hat{F}_1,\hat{F}_2,...,\hat{F}_6$ for the estimated severity models considered in the Monte Carlo analysis. A detailed description of each abbreviation will be given below.

$\hat{F}_1$: Pure parametric GCD. Model (4.2) is estimated on the internal data $(X_i)_{1\le i\le 89}$, consistent with $T\left(x,\hat{\theta}_1^x\right)$.

$\hat{F}_2$: Pure parametric GCD. Model (4.2) is estimated on consortium data $(Z_k)_{1\le k\le 613}$, consistent with $T\left(x,\hat{\theta}_1^z\right)$.

$\hat{F}_3$: Pure parametric GCD. Model (4.2) is estimated on publicly reported data $(Y_j)_{1\le j\le 885}$, consistent with $T\left(x,\hat{\theta}_1^y\right)$.

$\hat{F}_4$: A semi-parametric model. Model (1.4) is estimated using collected internal data $(X_i)_{1\le i\le 89}$, consistent with the estimator $m_1\left(x,\hat{\theta}_1^x,\hat{\theta}_2^x\left(T\left(x,\hat{\theta}_1^x\right)\right)\right)$.

$\hat{F}_5$: A semi-parametric model. Model (2.2) is estimated using collected internal data $(X_i)_{1\le i\le 89}$ and publicly reported losses $(Y_j)_{1\le j\le 885}$ consistent with the estimator $m_2\left(x,\hat{\theta}_1^y,\hat{\theta}_2^x\left(T\left(x,\hat{\theta}_1^y\right)\right)\right)$.

$\hat{F}_6$: A semi-parametric model. Model (3.4) is estimated using collected internal data $(X_i)_{1\le i\le 89}$, publicly reported losses $(Y_j)_{1\le j\le 885}$ and consortium data $(Z_k)_{1\le k\le 613}$, consistent with the estimator $m_3\left(x,\hat{\theta}_1^y,\hat{\theta}_2\left(T\left(x,\hat{\theta}_1^y\right)\right),T\left(x,\hat{\theta}_1^z\right)\right)$.

5.2. The frequency model and the Monte Carlo simulation. For the annual frequency model we assume that $N(t)$ is an independent homogeneous Poisson process denoted as $N(t)\in Po(\lambda t)$, where the intensity $\lambda>0$. The maximum likelihood estimator of the annual intensity of internal losses is $\hat{\lambda}$ where n = 89 and T = 2 from Table 2, which gives $\hat{\lambda}\cong 45$. The Monte Carlo simulation begins by simulating the Poisson frequency R times with estimated intensity $\hat{\lambda}$. Denote the annual simulated frequencies by $\hat{\lambda}_r$ with $r=1,...,R$ with number of simulations R = 100000. For each $\hat{\lambda}_r$ we draw randomly uniform distributed samples and combine these with loss sizes taken from the inverse function of the severity distribution models $\hat{F}_\delta,\delta=1,2,...,6$. By using Monte Carlo simulation with the severity and frequency assumptions we could create a simulated one year operational risk loss distribution for each model. Let the abbreviations $M_\delta,\delta=1,2,...,6$, be the total loss distributions obtained through the Monte Carlo simulation and expressed as

$$M_{\delta r}=\sum_{k=1}^{\hat{\lambda}_r}\hat{F}_\delta^{\leftarrow}(u_{rk}),r=1,...,R$$

with $u_{rk}\in U(0,1)$ for $k=1,...,\hat{\lambda}_r$, and

$$\hat{F}_\delta(u_{rk})=\int_0^{u_{rk}}\hat{f}(\xi)d\xi,$$

where $\hat{F}_\delta$ is the compared severity estimators described above.

5.3. Insurance recoveries. To make the study more realistic for a company, each of the individual events $\hat{F}_\delta^{\leftarrow}(u_{rk})$ in the total loss distributions $(M_{\delta r})_{1\le r\le R}$ will go through an insurance filter. For

each simulated amount $\hat{F}_{\delta}^{\leftarrow}(u_{rk})$ the payable insurance amount is determined. Since specific operational risk reinsurance contracts are unusual, expert opinions could be taken into account to assess whether operational risk losses for the considered event risk category are covered under other reinsurance contracts by the company. The procedure begins by determining whether the individual events are covered by the insurance. A random process takes place using a Bernoulli distribution with probability p_1 to ascertain whether the event is insured. If a loss event is insured, the methodology then seeks to determine which insurance cover type the event falls within. This is done by generating a Multinomial distribution with a given probability p_{2c}, for cover type c. If a loss event is insured, falls within one of the cover types, one determines if the loss event is honored or not. Once again, a Bernoulli trial takes place with probability p_3. Finally, if a loss event is insured and honored, a random process should determine whether the event is paid in the next financial year. Once again a Bernoulli trial takes place with probability p_4. Note, if the amount $\hat{F}_{\delta}^{\leftarrow}(u_{rk})$ does not pass through all steps in the insurance filtration, the simulated amount is seen as an uninsured event. If a loss event has passed through the filtration, the calculation of insurance payout follows a non-proportional reinsurance contract (excess-of-loss) with a deductible of D million pounds sterling and limit of L million pounds sterling. Table 6 summarizes the input values for the insurance cover filtration.

Table 6. Expert opinions on insurance recoveries information for the Event Risk Category (ERC): Execution, Delivery and Process Management

	Proportion of insured p1	Cover type	Cover type probability {p21, p22}	Probability honoring p3	Probability payment p4	Deductible D	Limit L
		Directors & officers	15%	90%	80%		
ERC	90%					5	20
		Professional indemnity	85%	80%	90%		

The individual events for the total of six pre-insurance loss distributions $M_{\delta r}$ are filtered according to Table 6, and the post-insurance total loss distributions are generated for each δ as

$$M_{\delta r}^{*} = \sum_{k=1}^{\hat{\lambda}_r} \hat{F}_{\delta}^{*\leftarrow}(u_{rk}),\ r = 1,\ldots,R\ .$$

Note that the following equalities will always hold, $M_{\delta r}^{*} \le M_{\delta r}$.

5.4. Value-at-Risk. For each of the twelve models $(M_{\delta})_{1\le\delta\le6}$ and $(M_{\delta}^{*})_{1\le\delta\le6}$ the risk measure Value-at-Risk (VaR) is estimated. For the pre-insured loss distributions we have

$$VaR_{\alpha}(M_{\delta r}) = \sup\{m \in \Re \mid \mathrm{P}(M_{\delta r} \le m) \le \alpha\}$$

and for the post-insured distributions we use

$$VaR_{\alpha}(M_{\delta r}^{*}) = \sup\{m \in \Re \mid \mathrm{P}(M_{\delta r}^{*} \le m) \le \alpha\}.$$

If we compare each models VaR outcome separately for different risk tolerances α, and subsequently evaluate the insurance recovery effect, we see in Figure 4 (Appendix B) that some models are not affected by the insurance filter.

Interpreting Figure 4, we find that models M_1 and M_4 are identical to their post-insurance loss distributions M_1^{*} and M_4^{*}, respectively. This is not remarkable since these loss distributions have severity estimators that only incorporate internal collected data in their estimation, therefore the limit level L million pounds sterling was never reached in the simulation. Focus on the second model M_2, after risk tolerance level α = 99%, insurance recoveries occur. The remaining three post-insurance loss distributions M_3^{*}, M_5^{*} and M_6^{*} have all been affected with insurance recoveries for low risk tolerance levels. This is not extraordinary since all three models' severity estimators include the publicly reported data set that is characterized by extremely large losses.

Table 7 focuses in more detail on the tail for the outcomes presented in Figure 4.

Table 7. Value-at-Risk with focus on the tail for the different loss distributions

	$VaR_{95\%}$	$VaR_{99\%}$	$VaR_{99.5\%}$	$VaR_{99.9\%}$
M_1	2.90	3.28	3.44	3.74
M_1^{*}	2.90	3.28	3.44	3.74
M_2	4.81	5.63	5.96	6.64
M_2^{*}	4.81	5.14	5.50	6.30
M_3	57.48	76.98	84.78	101.72
M_3^{*}	42.67	64.32	72.48	91.68
M_4	3.29	3.78	3.97	4.32
M_4^{*}	3.29	3.78	3.97	4.32

Table 7 (cont.). Value-at-Risk with focus on the tail for the different loss distributions

	$VaR_{95\%}$	$VaR_{99\%}$	$VaR_{99.5\%}$	$VaR_{99.9\%}$
M_5	46.70	56.69	67.58	79.69
M_5^*	25.09	35.09	46.11	58.07
M_6	15.89	20.61	22.53	26.58
M_6^*	12.31	17.63	19.57	23.95

If we compare the results for risk tolerance level α = 95% we see that model M_3 extrapolates the highest value. This model only takes into account the publicly reported losses when estimating the severity distribution. Further, mixing model M_5 presented by Gustafsson and Nielsen (2008) is much closer to model M_3 than model M_1. The interpretation of this is that more weight is assigned to the publicly reported data than to the internal sample. If we look at the return period one in two hundred years (99.5%), it is interesting to see that model M_5 reduces the solvency capital amount with 21.47 million pounds sterling by including insurance recoveries. Model M_6 shows lower values than M_5 for all α-values. Since the extra added source of prior knowledge is less volatile and different in mean, the model corrects M_5 to be less heavy tailed.

Conclusions

This paper has explored the possibility of applying a model that incorporates three sources of data for operational risk capital assessment. The model displays desirable characteristics for operational risk modelling. In a scenario setup we worked out different scenarios on the added external sources and the performance of the outcome from the developed model is appealing. The model is receptive to extreme events introduced by a prior knowledge source, but the statistical equalities from internal data and the second added source of prior knowledge play an essential role in the correction of the model.

We examine operational risk modelling using only internal data. By considering $VaR_{99.5\%}$ for model M_1 we find a total number of 3.44 million pounds sterling. However, by adding an extra source of prior knowledge, the publicly reported loss data set, the same return period results in a value which is almost 20 times higher. As we can see, model M_5 is close to M_3, meaning that for these data sets model M_5 provides a flat local model. The presented model M_6 takes the consortium data into account. We find that $VaR_{99.5\%}$ is three times smaller for model M_6 than M_5, but the effect from the publicly reported data is still present since $VaR_{99.5\%}$ for M_6 is six times higher than the outcome when using only internal data, and more than three times higher compared to using only consortium data. The model also has appealing features when more abundant internal losses will be collected and taken into account and thereby a greater correction on the prior knowledge will be made.

References

1. Bolance C, Guillen, M., and Nielsen, J.P. (2003). Kernel density estimation of actuarial loss functions. *Insurance, Mathematics and Economics*, Vol. 32, 19-36.
2. 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.
3. Buch-Kromann, T., Englund, M., Gustafsson, J., Nielsen, J.P. and Thuring, F. (2007). Non-parametric estimation of operational risk losses adjusted for under-reporting. *Scandinavian Actuarial Journal*, Vol. 4, 293-304.
4. Cizek P., Hardle, W. and Weron, R. (2005). *Statistical Tools for Finance and Insurance*. Springer-Verlag Berlin Heidelberg.
5. Clements, A.E., Hurn, A.S. and Lindsay, K.A. (2003). Mobius-like mappings and their use in kernel density estimation. *Journal of the Americal Statistical Association*, Vol. 98, 993-1000.
6. Degen, M., Embrechts, P. and Lambrigger, D.D. (2007). The Quantitative Modeling of Operational Risk: Between G-and-H and EVT. *Astin Bulletin*, Vol. 37, No. 2, 265-292.
7. Diebold, F., Schuermann, T. and Stroughair, J. (2001). Pitfalls and oppurtunities in the use of extreme value theory in risk mangement. In: Refenes, A.-P., Moody, J. and Burgess, A. (Eds.), *Advances in Computational Finance,* Kluwer Academic Press, Amsterdam, pp. 3-12, Reprinted from: *Journal of Risk Finance*, Vol. 1, 30-36 (Winter 2000).
8. Dutta, K. and Perry, J. (2006). A tale of tails: an empirical analysis of loss distribution models for estimating operational risk capital. Federal Reserve Bank of Boston, Working Paper No 06-13.
9. Embrechts, P., Kluppelberg, C and Mikosch, T. (1999). *Modeling Extremal Events for Insurance and Finance*. Springer.
10. Figini, S., Giudici, P., Uberti, P. and Sanyal, A. (2008). A statistical method to optimize the combination of internal and external data in operational risk measurement. *The Journal of Operational Risk*, Vol. 2, No. 4, 69-78.
11. Guillen, M., Gustafsson, J., Nielsen, J.P. andPritchard, P. (2007). Using external data in operational risk. *The Geneva Papers*, Vol. 32, 178-189.

12. 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.
13. Gustafsson, J., Hagmann, M., Nielsen, J.P. and Scaillet, O. (2008). Local Transformation Kernel Density Estimation of Loss Distributions. *To Appear In The Journal of Business and Economic Statistics.*
14. Gustafsson, J. and Nielsen, J.P. (2008). A Mixing Model for Operational Risk. *To Appear In The Journal of Operational Risk.*
15. Hjort, N.L. and Glad, IK. (1995). Nonparametric Density Estimation with a Parametric Start. *The Annals of Statistics*, Vol. 24, 882-904.
16. Jones, M.C. and Foster P.J. (1996). Generalised Jackknifing and Higher Order Kernels. *Nonparametric Statistics,* Vol. 3, 81-94.
17. Jones, M.C, Linton O. and Nielsen, J.P. (1995). A simple bias reduction method for density estimation. *Biometrica*, Vol. 82, No. 2, 327-338.
18. Jones, M.C, Signorini D.F. and Hjort, N.L. (1999). On multiplicative bias correction in kernel density estimation. *The Indian Journal of Statistics*, Vol. 61, No 2, 422-430.
19. Klugman, S.A., Panjer, H.A. and Willmot, G.E. (1998). *Loss Models: From Data to Decisions.* New York: John Wiley & Sons, Inc.
20. McNeil, A.J., Frey, R. and Embrechts, P. (2005). *Quantitative Risk Management.* Princeton Series in Finance.
21. Panjer, H.H. (2006). *Operational Risk: Modeling Analytics.* NewYork: John Wiley & Sons, Inc.
22. Rosenblatt, M. (1956). Remarks on Some Nonparametric Estimates of a Density Function. The Annals of Mathematical Statistics, 27, 642-669.
23. Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman & Hall, London.
24. Wand, M.P., Marron, J.S. and Ruppert, D. (1991). Transformation in Density Estimation (with comments). *Journal of the American Statistical Association*, Vol. 94, 1231-1241.
25. Wei, R. (2007). Quantifcation of operational losses using firm specified information and external database). *Journal of Operational Risk*, Vol. 1, No. 4, 3-34

Appendix A

In order to shorten the proof we introduce the notation $\xi\left(\hat{s},\hat{\theta}_2\right)=\phi\left(\hat{s},\hat{\theta}_2(\hat{s})\right)\cdot\phi\left(\hat{s},\hat{\theta}_2(\hat{s},\hat{w})\right)$ with the theoretical counterpart $\xi\left(s,\theta_2^0\right)$ with $s=T\left(x,\theta_1^0\right)$. From Jones, Signorini and Hjort (1999) the calculation of the bias and variance for $\xi\left(\hat{s},\hat{\theta}_2\right)$ is given by:

$$\mathrm{E}\xi\left(\hat{s},\hat{\theta}_2\right)-r(s)\cong-\frac{h^4}{4}\alpha_{21}(s,h)\left(r(s)-\xi\left(s,\theta_2^0\right)\right)\left(\alpha_{21}(s,h)\frac{r^{(2)}(s)-\xi^{(2)}\left(s,\theta_2^0\right)}{r(s)-\xi\left(s,\theta_2^0\right)}\right)+o_p\left(h^4\right)$$

$$\cong-\frac{h^4}{4}\alpha_{21}(s,h)r(s)\left(\alpha_{21}(s,h)\frac{r^{(2)}(s)}{r\ (s)}\right)+o_p\left(h^4\right),$$

$$V\xi\left(\hat{s},\hat{\theta}_2\right)\cong\frac{r(s)-\xi\left(s,\theta_2^0\right)}{nh}\int\widetilde{K}(t)dt+o_p\left(\frac{1}{nh}\right)$$

$$\cong\frac{r(s)}{nh}\int\widetilde{K}(t)dt+o_p\left(\frac{1}{nh}\right).$$

By back-transforming to original axes, we need to consider the following equalities:

$$\begin{cases} r\left(T\left(x,\hat{\theta}_1^0\right)\right)=\dfrac{f(x)}{T^{(1)}\left(x,\theta_1^0\right)}, \\ \dfrac{\partial r\left(T\left(x,\theta_1^0\right)\right)}{\partial T\left(x,\hat{\theta}_1^0\right)}=r^{(1)}\left(T\left(x,\theta_1^0\right)\right)=\left(\dfrac{f(x)}{T^{(1)}\left(x,\theta_1^0\right)}\right)^{(1)}\dfrac{1}{T^{(1)}\left(x,\theta_1^0\right)}, \\ \dfrac{\partial^2 r\left(T\left(x,\theta_1^0\right)\right)}{\partial T\left(x,\theta_1^0\right)^2}=r^{(2)}\left(T\left(x,\theta_1^0\right)\right)=\left(\left(\dfrac{f(x)}{T^{(1)}\left(x,\theta_1^0\right)}\right)^{(1)}\dfrac{1}{T^{(1)}\left(x,\theta_1^0\right)}\right)^{(1)}\dfrac{1}{T^{(1)}\left(x,\theta_1^0\right)}. \end{cases}$$

As a consequence, inserting these expressions in the bias and variance formulas, we obtain

$$\mathrm{E}m_3 - f(x) = \mathrm{E}T^{(1)}\left(x,\hat{\theta}_1^y\right)\cdot\xi\left(T\left(x,\hat{\theta}_1^y\right),\hat{\theta}_2\right)$$

$$\cong \frac{-T^{(1)}\left(x,\theta_1^0\right)h^4\alpha_{21}(s,h)f(x)}{4T^{(1)}\left(x,\theta_1^0\right)}\left(\alpha_{21}(s,h)\left(\left(\frac{f(x)}{T^{(1)}\left(x,\theta_1^0\right)}\right)^{(1)}\frac{1}{T^{(1)}\left(x,\theta_1^0\right)}\right)^{(1)}\frac{T^{(1)}\left(x,\theta_1^0\right)}{f(x)T^{(1)}\left(x,\theta_1^0\right)}\right)^{(2)}$$

$$\cong -\frac{h^4}{4}\alpha_{21}\left(T\left(x,\theta_1^0\right),h\right)f(x)\left(\alpha_{21}\left(T\left(x,\theta_1^0\right),h\right)\left(\left(\frac{f(x)}{T^{(1)}\left(x,\theta_1^0\right)}\right)^{(1)}\frac{1}{T^{(1)}\left(x,\theta_1^0\right)}\right)^{(1)}\frac{1}{f(x)}\right)^{(2)}$$

$$V_3 = VT^{(1))}\left(x,\hat{\theta}_1^y\right)\cdot\xi\left(T\left(x,\hat{\theta}_1^y\right),\hat{\theta}_2\right)\cong\frac{T^{(1)}\left(x,\theta_1^0\right)f(x)}{nhT^{(1)}\left(x,\theta_1^0\right)}\int\widetilde{K}(t)dt\cong\frac{T^{(1)}\left(x,\theta_1^0\right)f(x)}{nh}\int\widetilde{K}(t)dt\,.$$

Appendix B

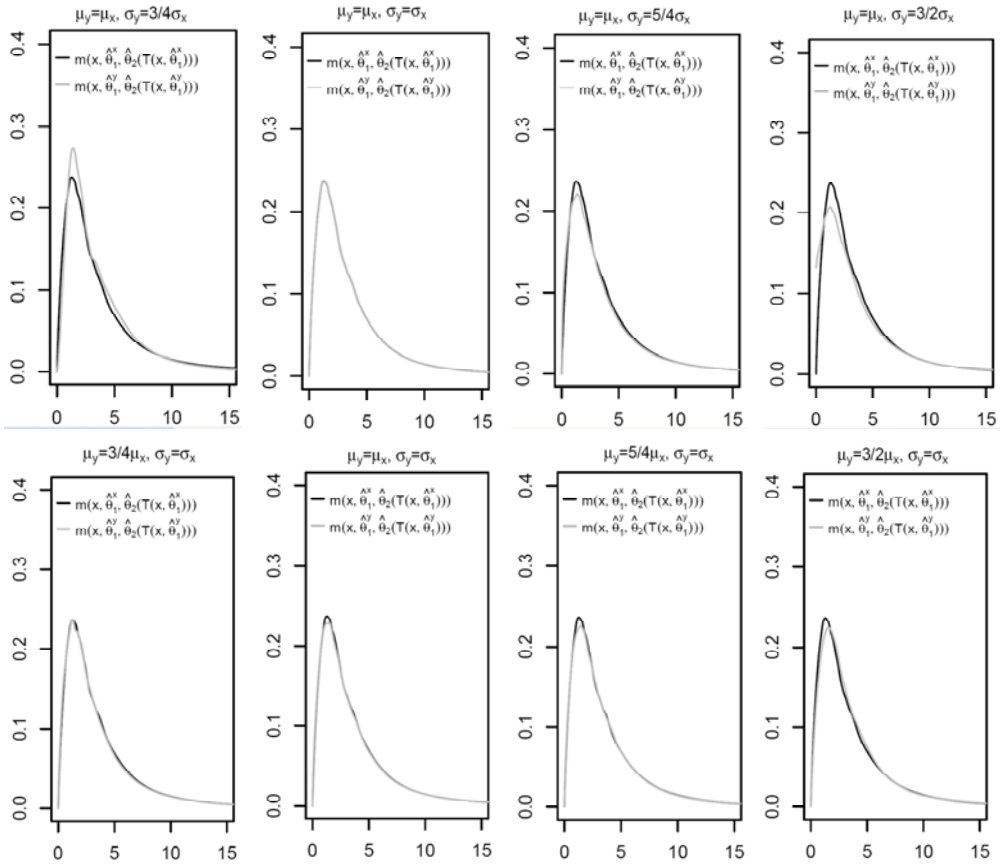

Fig. 1. Eight scenarios that evaluate the characteristics of model (2.2) compared to (1.4)

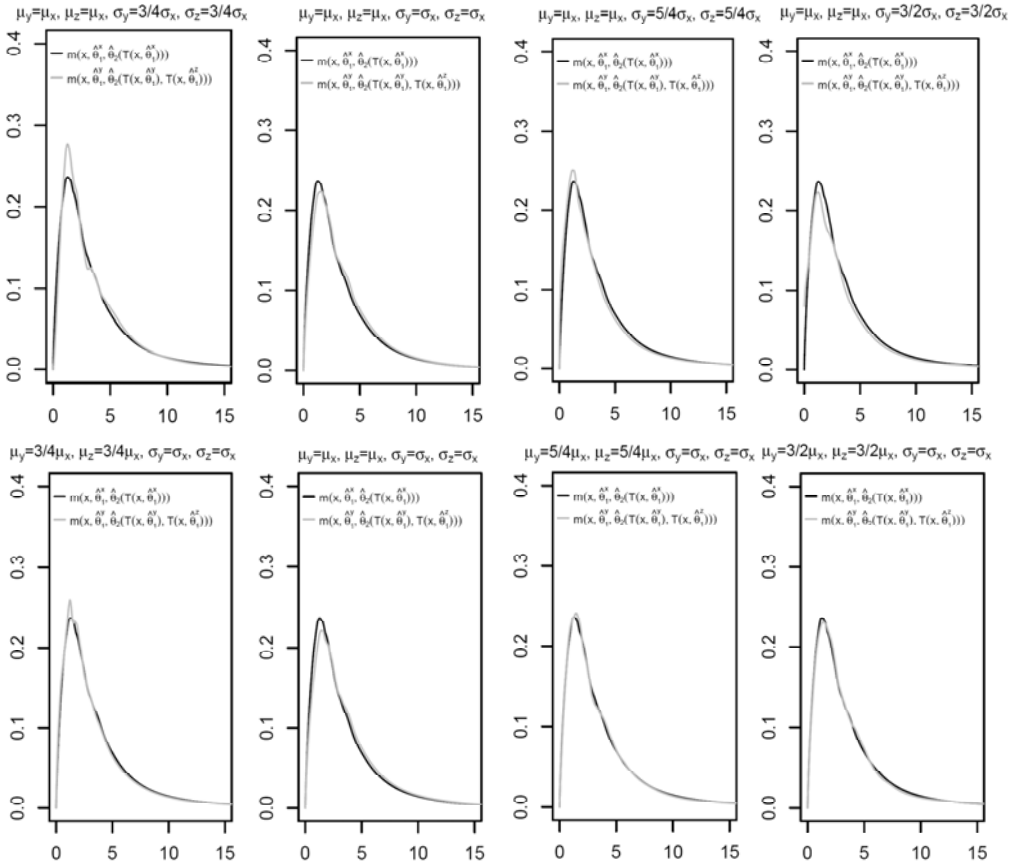

Fig. 2. Eight scenarios that evaluate the characteristics of model (3.4) compared to (1.4) where the prior knowledge data sources are similar in their appearance

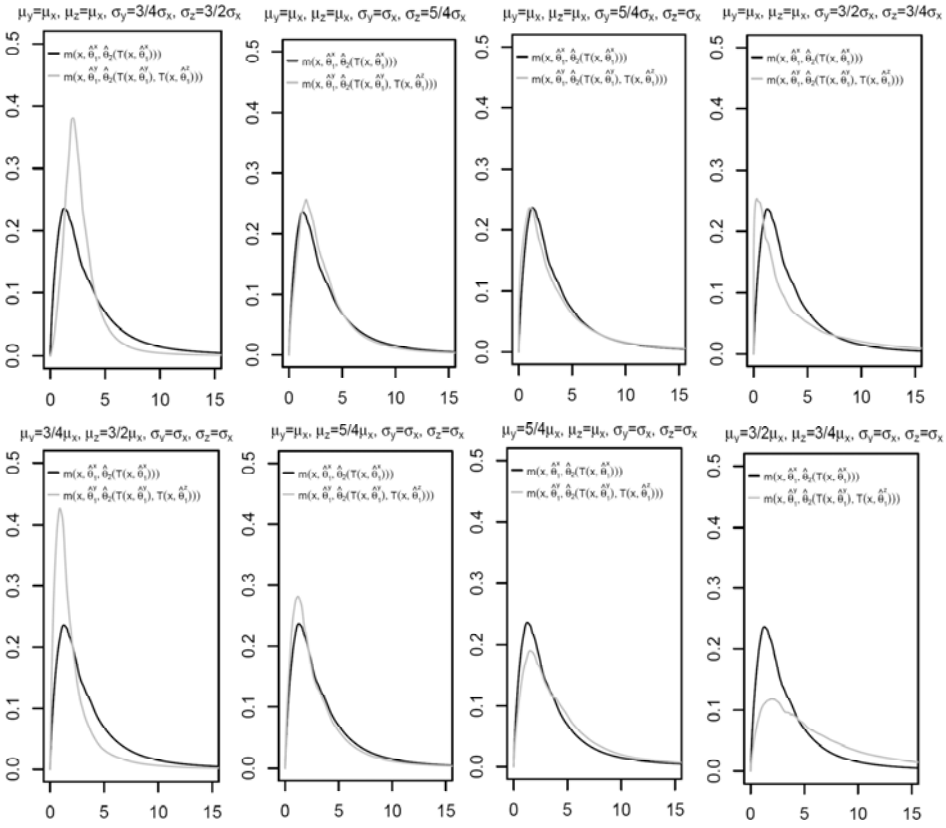

Fig. 3. Eight scenarios that evaluate the characteristics of model (3.4) compared to (1.4) where the prior knowledge data sources are different in their appearance

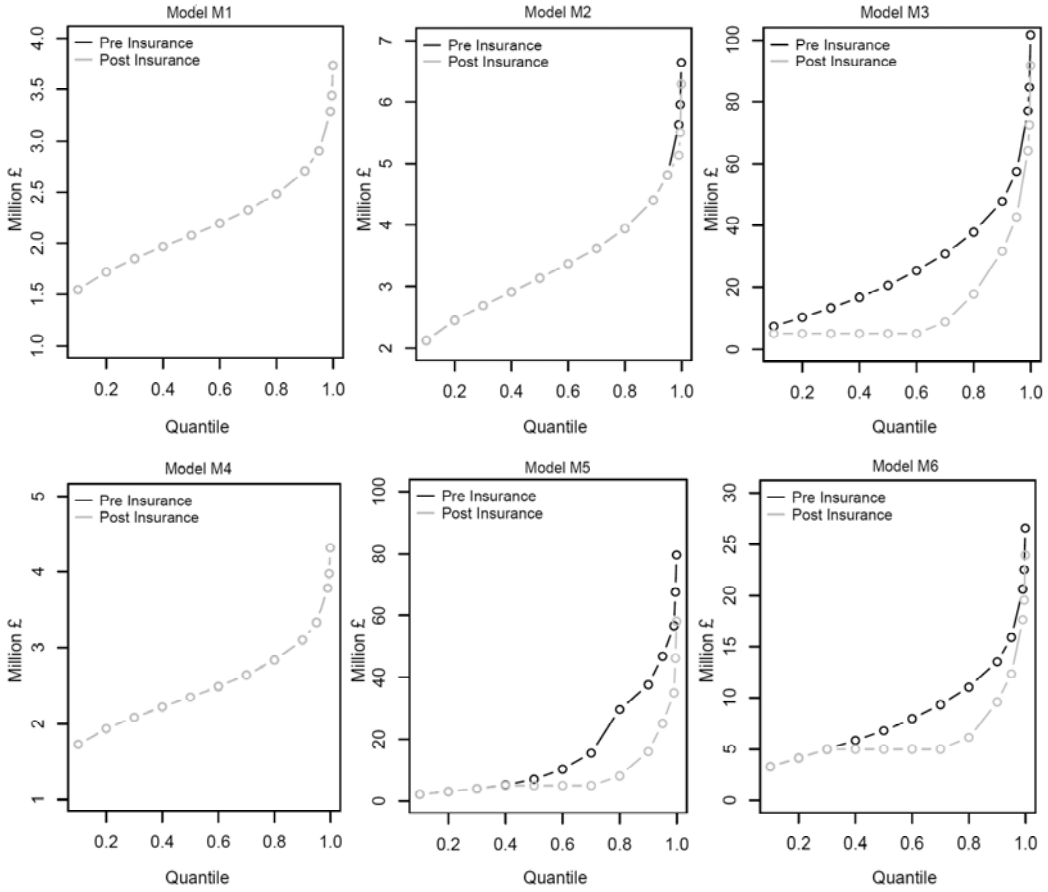

Fig. 4. Estimated Value-at-Risk for different risk tolerance for the six proposed models with and without insurance cover