Adding Prior Knowledge to Quantitative Operational Risk Models

Catalina Bolancé (Riskcenter, University of Barcelona, Spain)
Montserrat Guillén (Riskcenter, University of Barcelona, Spain) 1
Jim Gustafsson (Ernst and Young, Denmark)
Jens Perch Nielsen (Cass Business School, City University, United Kingdom) 2

Abstract

Our approach is based on the study of the statistical severity distribution of a single loss. We analyze the fundamental issues that arise in practice when modeling operational risk data. We address the statistical problem of estimating an operational risk distribution, both abundant data situations and when our available data is challenged from the inclusion of external data or because of underreporting. Our presentation includes an application to show that failure to account for underreporting may lead to a substantial underestimation of operational risk measures. The use of external data information can easily be incorporated in our modeling approach. The paper builds on methodology developed in Bolance et al. (2012b).

1. Quantifying Operational Risk Guided by Prior Knowledge

Operational risk is one of the risks that are incorporated in the Basel II regulatory framework for financial institutions and in the Solvency II regulatory framework for insurance companies (Gatzert and Wesker, 2012 and Ashby, 2011), hence the importance of the modelization and quantification of this risk. Also, operational risk is important in the context of Enterprise Risk Management (Hoyt and Liebenberg, 2011 and Dhaene et al. 2012). One major issue addressed in Bolance et al (2012b) is how to incorporate prior knowledge into operational risk models. Such prior knowledge can come in many disguises. One being prior knowledge of parametric shapes of distributions, another being prior knowledge of the frequency of underreporting and a third could be prior knowledge arising from external data sources. The fundamental principles of mixing internal and external operational risk data was originally published in this journal in Gustafsson and Nielsen (2008) and Guillen et al. (2008). Bolancé et al. (2012b) take these originally ideas and put them into a broader context, see also the following recent papers proposing alternative methods to quantify operational risk (Cope, E.W., 2012, Cavallo et al., 2012, Feng et al., 2012 and Horbenko et al., 2011).

In this paper we show, with a simple example, the effect of incorporating two different types of prior knowledge into the calculation of Value-at-Risk (VaR) and Tail Value-at Risk (TVaR): external operational risk data and expert information about underreporting probability. We

1 We thank the Spanish Ministry of Science / FEDER grant ECO2010-21787-C0301 and Generalitat de Catalunya SGR 1328.
Corresponding author: jens.nielsen.1@city.ac.uk
2 We thank the Spanish Ministry of Science / FEDER grant ECO2010-21787-C0301 and Generalitat de Catalunya SGR 1328.
Corresponding author: jens.nielsen.1@city.ac.uk
analyze the isolated effect of both types of prior knowledge and the combined effect throughout this work.

Measuring operational risk requires the knowledge of the quantitative tools and the comprehension of financial activities in a very broad sense, both technical and commercial. Our presentation offers a practical perspective that combines statistical analysis and management orientations.

In general, financial and insurance institutions seem to keep historical data on internal operational risk losses, however, given the small probability associated with some operational risks, the companies often lack sufficient sample information to enable them represent the entire domain of the loss distribution associated with operational risk. So, models for operational risk quantification only have limited data to estimate from and validate on. Therefore, complementing internal data with more abundant external data is desirable (Peters and Sisson, 2006, Dahen and Dionne, 2007 or Peters et al., 2009) or to take information from other operational risk lines that the one being estimated (Gustafsson et al., 2006). A suitable framework to overcome the lack of data is getting a compatible sample from a consortium. Individual firms come together to form a consortium, where the population from which the data is drawn is assumed to be similar for every member of the group. We know that consortium data pooling may be rather controversial because not all business are the same, and one can raise questions about whether consortium data can be considered to have been generated by the same process, or whether pooled data do reflect the size of each financial institution transaction volume. Using our method, scaling is corrected during our modeling process, assuming that external data provide prior information to the examination of internal data. So, we will not need to address pre-scaling specifically when combining internal and external data sets.

Many different issues arise in practice when financial managers analyze their operational risk data. We will focus on four of them:

(i) A statistical distributional assumption for operational losses has to be fitted and it must be flexible enough.
(ii) Sample size of internal information is rather small.
(iii) External data can be used to complement internal information, but pooling the data requires a mixing model that combines internal and external information, and possibly a pre-scaling process.
(iv) Data used to assess operational risk are sometimes underreported, which means that losses do happen that are unknown or hidden. Data on operational losses are often not recorded below a given threshold due to existing habits or organizational reasons, which means that information on losses below a given level is not collected. One should not confuse underreporting with un-recording because the latter is a voluntary action not to collect all those losses below a given level.

There are many reasons for operational risk being extremely difficult to quantify: sparse data, underreporting, extreme values, dependence between events, granularity (day/month/yearly time scale...) and frequency and severity joint modeling (Gzyl, 2011). In this paper we focus on three of the cited reasons: sparse data, underreporting, extreme values, we review a methodology that addresses these three aspects together.
Based on market experience, one cannot just do a standard statistical analysis of available data and hand results into the supervisory authorities. It is necessary to add extra information. We call extra information “prior knowledge.” While this term originates from Bayesian statistics, we use it in a broader sense: prior knowledge is any extra information that helps in increasing the accuracy of estimating the statistical distribution of internal data. In this article, we consider three types of prior knowledge, and one is the prior knowledge coming from external data. First, one has data that originate from other companies or even other branches (for instance, insurance companies have so few observations that they need to borrow information from banks, see Bolancé, et al., 2012b for more information on this) that are related to the internal operational risk data of interest—but only related; they are not internal data and should not be treated as such. Second, one can get a good idea of the underreporting pattern by asking experts a set of relevant questions that in the end make it possible to quantify the underreporting pattern. Third, we take advantage of prior knowledge of relevant parametric models. It is clear that if you have some good reason to assume some particular parametric model of your operational risk statistical distribution, then this makes the estimation of this distribution a lot easier.

Prior knowledge on a parametric assumption can be built up from experience with fitting parametric models to many related data sets, and in the end one can get some confidence that one distribution seems more likely to fit than another distribution.

We have so far come up with the prior knowledge that often the generalized Champernowne distribution is a good candidate for operational risk modeling. This distribution was initially used to model income, so it adapts well to situations where the majority of observations correspond to small values, whereas a few observations correspond to extremely high values (Buch-Larsen et al., 2005). We also consider the art of using this prior knowledge exactly to the extent it deserves. This is, when we have very little internal data at our disposal, we have to rely heavily on prior information.

Our approach based on kernel smoothing has the appealing property that small data sets imply that a lot of smoothing is necessary when estimating the statistical probability density function, and this again implies that we end up not differing too much from our prior knowledge. When data becomes more abundant, less smoothing is optimal, which implies that we have sufficient data to reveal the individual distributional properties of our internal data downplaying the importance of prior knowledge. Another key result based on real data is that our transformation approach also has a robustifying effect when prior knowledge is uncertain (Bolancé et al., 2012b). For example, in a real data study where three very different parametric models were used, very different conclusions of the quantification of operational risk were given. After our transformation approach, similar results were obtained, independent or almost independent of the parametric assumption used as prior knowledge. Therefore, our recommendation is to “Find as much prior knowledge as possible and then adjust this prior knowledge with the real observations according to how many real observations are available”, see Bolancé et al. (2012b).

The structure of this article is as follows Section 2 presents briefly the semiparametric model approach for modeling operational risk severities. Sections 3 and 4 present the method to combine internal and external data and a procedure to take into account underreporting, i.e. the fact that some small losses may not have been recorded and so the density is somehow underestimated in the intervals corresponding to small losses. Section 5 presents an application
and the results are discusses. We present this illustration in order to show the steps towards quantification of operational risk under the premises of data scarcity, presence of external data sources and underreporting. A final section concludes.

2 Density Models for Operational Risk Severities

Traditional methods for studying severities or loss distributions use parametric models. Two of the most shapes used as building blocks are based on the lognormal distribution and the Pareto distribution (see also Carrillo et al., 2012 and Pigeon and Denuit, 2011), see also recent studies suggesting different skewness distributions (Eling, 2012, Bolancé et al., 2008). Traditionally, a parametric fit can easily be obtained, but this is not the only possible approach. In fact, the main reason why the parametric approach can be controversial is that a certain distribution is imposed on the data. This assumption leads to a conservative risk measurement, which is sometimes unrealistic and dangerous.

Operational risk analysts are interested in having good estimates of all the values in the domain range: small losses because they are very frequent, medium losses causing a dramatic increase of expenses (demanding liquidity), and large losses that may be the most serious, because they can cause solvency issues. We study all severity sizes at the same time, and we do not break the domain into separate intervals. Actually, defining subintervals of the domain and fitting a distribution in every subinterval is an obvious approach when a unique parametric shape cannot be fitted across the whole domain range.

As we have mentioned, practitioners often decide to analyze large and small losses separately because no single, classical parametric model fits all claim sizes. This approach leaves some important challenges: choosing the appropriate parametric model, identifying the best way of estimating the parameters, and determining the threshold level between large and small losses. In this section we present a systematic approach to the estimation of loss distributions that is suitable for heavy-tailed situations.

Alternatively, nonparametric smoothing is suitable for modeling operational loss data because it allows for more flexible forms for the density of the random variable that generates the information. The versatility of a nonparametric approach is convenient in our context because the shape of the density can be very different for small and large losses. Our approach is based on a slightly adjusted version of the semiparametric transformed kernel density estimation method. For the transformation, we use the generalized Champernowne distribution function, which has three parameters and adapts very well to the presence of small and large losses simultaneously. The combination of a nonparametric estimation with a previous parametric transformation of the data is the reason why we use the term semiparametric (Bolancé et al., 2003, 2008 and 2012a and Buch-Larsen et al., 2005, and in a multivariate context Buch Kromann, et al., 2011 and Buch Kromann, et al., 2012).

We show that the estimation is easy to implement (including the often complicated question of choosing the amount of smoothing) and that the new method provides a good and smooth estimate of the density in the upper tail (see Bolancé et al., 2012b, for a guided example chapter with programs and routines ready for implementation). Parametric models have often been
justified due to their simplicity and their ability to solve the problem of lack of smoothness of the empirical approach, so we provide the practitioner with a nonparametric alternative to traditional parametric estimation techniques.

Classical kernel estimation is substantially improved if data are transformed as a preliminary step. The method is called semiparametric because we use a parametric transformation function and then nonparametric kernel estimation. We introduce a transformation method with a parametric cumulative distribution function, but it could also be implemented with other types of functions, and we advocate the semiparametric transformation method because it behaves very well when it comes to estimating severity distributions.

Loss distributions, that is, densities associated to the severity of operational risk events, have typically one mode for the low loss values and then a long heavy tail. Existing results based on simulation studies have shown that the transformation kernel estimation method is able to estimate all three possible kinds of tails, namely, the Fréchet type, the Weibull type, and the Gumbel type (Embrechts et al., 2006). This makes the semiparametric transformation approach method extremely useful for operational risk analysis. A specific feature of the semiparametric transformation principle is that it is also able to estimate the risk of a heavy-tailed distribution beyond the data.

Let us assume that a sample of \( n \) independent and identically distributed observations \( X_1, \ldots, X_n \) is available, they correspond to random variable \( X \). We also assume that a transformation function \( T(\cdot) \) is selected, and then the data can be transformed so that \( Y_i = T(X_i) \) for \( i = 1, \ldots, n \). We denote the transformed sample by \( Y_1, \ldots, Y_n \) and the corresponding random variable is \( Y \). The first step consists of transforming the data set with a function \( T(\cdot) \) and afterwards estimating the density of the transformed data set using the classical kernel density estimator. The estimated density of transformed random variable at point \( y = T(x) \) is:

\[
\hat{f}(y) = \frac{1}{bn} \sum_{i=1}^{n} K\left(\frac{y-Y_i}{b}\right),
\]

where \( K(\cdot) \) is the kernel function, \( b \) is the bandwidth, and \( Y_i \) for \( i = 1, \ldots, n \) is the transformed data set. Buch-Larsen et al. (2005) propose to use Epanechnikov kernel and the rule-of-thumb bandwidth of Silverman (1986). The estimator of the original density is obtained by a back-transformation to the original scale and it is called transformation density estimation:

\[
f(x) = T'(x)\hat{f}(y)
\]

and the transformation is the generalized Champernowne distribution function:

\[
T(x) = \frac{(x+c)^{\delta-c}\delta}{(x+c)^{\delta}+(M+c)^{\delta}-2c^\delta},
\]

where \( \delta \) is a shape parameter, \( M \) is a scale parameter and \( c \) has some scale parameters properties. See more details in Buch-Larsen et al. (2005) and Bolancé et al. (2012a,b). The choice of bandwidth is an essential point when using nonparametric smoothing techniques. The bandwidth itself could be thought of as a scaling factor that determines the width of the kernel function, and thereby it controls how wide a probability mass is provided in the neighborhood of each data point. The rule-of-thumb bandwidth used in Buch-Larsen et al. (2005) is a good and easy approach to estimate the smoothing amount.
In combination with the transformation, the bandwidth excerpt a very different effect in the original scale. For instance, a small bandwidth that is applied to a transformation that has compressed the original data quite considerably would clearly cover a wide interval in the original situation (untransformed scale). Conversely, a large bandwidth in a transformation that has widened a subdomain in the original scale may be equivalent to a small bandwidth that is used in the original scale. So, a constant bandwidth applied after the transformation to obtain a kernel estimate has the same effect as implementing a variable bandwidth in the original scale with no transformation.

3 Mixing Internal and External Data

For simplicity, we introduce superindex I to indicate an internal data source, and we use superindex E to denote an external source. The transformed data set will have a density equal to the true internal density divided by the estimated external density distribution transformed such that the support is on \([0,1]\). When the external data distribution is a close approximation to the internal data distribution, we have simplified the problem to estimating something close to a uniform distribution, see Bolancé et al. (2012b) for more details. When this estimation problem has been solved through kernel smoothing, we backtransform to the original axis and find our final mixing estimator.

Let \(X_1^I, X_2^I, \ldots, X_n^I\) be a sequence of \(n^I\) collected internal losses, and let \(X_1^E, X_2^E, \ldots, X_n^E\) be a sequence of \(n^E\) external reported losses. Then external reported losses represent prior knowledge that should enrich a limited internal data set. We assume that scales are comparable and a filtration has been made, so that the external information has already been prepared to be directly mixed with the internal data.

The internal data sample \(X_1^I, X_2^I, \ldots, X_n^I\) is then transformed with the estimated parameters of Champernowne cumulative distribution function using external data:

\[
T(X_i^I) = \frac{(X_i^I + c^E)^{\delta^E} - (c^E)^{\delta^E}}{(X_i^I + c^E)^{\delta^E} + (M^E + c^E)^{\delta^E} - 2(c^E)^{\delta^E}}.
\]

Also, to avoid scale differences between internal and external data, we can transformed internal data using parameters \(\delta^E\) and \(c^E\) estimated with external data and scale parameter \(M^E\) estimated with internal data.

The whole or partial parametric estimator of transformation based on the external data set offers prior knowledge from the general financial industry. The correction of this prior knowledge is based on the transformation density estimation of the internal data points that is used as explained in section 2. This leads to a consistent and fully nonparametric estimator of the density of the internal data.

4 The Underreporting Function and Underreporting Correction
When modeling a publicly reported database, one needs to fit the operational risk exposure to a company’s internal situation. This is done by introducing a function into the model, which could be thought of as a scaling function explaining the likelihood of a loss being reported at different threshold levels.

A very simple interpretation of how the function affects the distribution is that, for small and medium-sized losses, the density up-weighted to include more probability mass in that domain, while larger losses are down-weighted. This can be called an underreporting function.

The underreporting function should, as well as being continuous, pass exactly through the predetermined reporting level values. The requirement is mathematically termed as interpolating the report levels, which is a prediction made through and between the report levels. A further requirement is to have a continuous and well-defined function after the last observed report level. Mathematically, this is defined as extrapolation and is a prediction made beyond the last observed report level. Of course, a prediction based upon extrapolation is potentially less reliable than a prediction between report levels since one must assume that a historical pattern will continue outside the reported value domain. Table 1 shows an example of probabilities of reporting, these probabilities have been estimated by risk experts from a major international insurance company. The underreporting function is obtained interpolating the probabilities of reporting.

<table>
<thead>
<tr>
<th>Event risk category:</th>
<th>Loss amount</th>
<th>0,001</th>
<th>0,01</th>
<th>0,1</th>
<th>0,5</th>
<th>1</th>
<th>10</th>
<th>50</th>
<th>100</th>
<th>110</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal Fraud</td>
<td></td>
<td>0,01</td>
<td>0,1</td>
<td>0,42</td>
<td>0,55</td>
<td>0,6</td>
<td>0,85</td>
<td>0,93</td>
<td>0,99</td>
<td>1</td>
</tr>
<tr>
<td>External Fraud</td>
<td></td>
<td>0,01</td>
<td>0,1</td>
<td>0,42</td>
<td>0,55</td>
<td>0,6</td>
<td>0,85</td>
<td>0,93</td>
<td>0,99</td>
<td>1</td>
</tr>
<tr>
<td>Employment Practices and Workplace Safety</td>
<td></td>
<td>0,05</td>
<td>0,1</td>
<td>0,32</td>
<td>0,43</td>
<td>0,5</td>
<td>0,73</td>
<td>0,88</td>
<td>0,99</td>
<td>1</td>
</tr>
<tr>
<td>Business Disruption</td>
<td></td>
<td>0,05</td>
<td>0,1</td>
<td>0,18</td>
<td>0,27</td>
<td>0,4</td>
<td>0,67</td>
<td>0,8</td>
<td>0,95</td>
<td>1</td>
</tr>
<tr>
<td>Clients, Products, and System Failures</td>
<td></td>
<td>0,15</td>
<td>0,2</td>
<td>0,3</td>
<td>0,45</td>
<td>0,5</td>
<td>0,75</td>
<td>0,9</td>
<td>0,95</td>
<td>1</td>
</tr>
<tr>
<td>Damage to Physical Assets</td>
<td></td>
<td>0,01</td>
<td>0,05</td>
<td>0,17</td>
<td>0,22</td>
<td>0,5</td>
<td>0,58</td>
<td>0,78</td>
<td>0,85</td>
<td>1</td>
</tr>
<tr>
<td>Execution, Delivery, and Process Management</td>
<td></td>
<td>0,01</td>
<td>0,05</td>
<td>0,06</td>
<td>0,08</td>
<td>0,1</td>
<td>0,25</td>
<td>0,5</td>
<td>0,75</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 1 The table shows the probability of reporting a loss by size and event risk categories.

Table 1 presents the estimated probability of reporting for each risk category. It can be seen that the event risk category number seven (which corresponds to Execution, Delivery, and Process Management) has the lowest reporting likelihood, which means that losses for this kind of operational risk are likely to be underreported. On the other hand, smaller losses that belong to category number five (which corresponds to Clients, Products, and Systems Failure) have a much higher probability of being reported than small losses for other types of risks. It is also noteworthy that, for all event risk categories, expert opinion assesses that the probability of reporting a loss is smaller or equal than 99%, even for very large losses. In fact, this acts as a
bound, and it implies that there is at least a 1% chance that a very large loss occurs and is not reported.

The estimation of the underreporting function is essential in our approach, but unfortunately, very little is known about the magnitude of underreporting and even less about the validity of experts’ estimation. We think that further research needs to be done on the stability of these estimates, and we believe that here we provide a starting point for discussion.

Let \( u(\cdot) \) the underreporting function, if \( \hat{f}(x) \) is transformation kernel density (in section 2) of reported claim, then the estimated density function of occurred claims is:

\[
\hat{f}_0(x) = \frac{\hat{f}(x)(u(x))^{-1}}{\int_0^\infty \hat{f}(t)(u(t))^{-1} dt},
\]

where we approximate the integral in the denominator numerically.

Underreporting is corrected using the definition of a the probability density of the statistical random variable and re-scaling it in order to take into account that sample is not fully representative of the true losses. Underreporting correction is equivalent to weighting the data in order that the sample correctly represents the original distribution. This technique is well-known and it is widely used in statistical sampling theory. Each observation is weighted with the inverse of the probability of selection. In this case, the probability of selection is given by the underreporting function. More details on the above underreporting approach is to be found in Bolance et al. (2012b) and Gustafsson et al. (2006).

5 An Operational Risk Application Combining Mixing and Underreporting

In this section we show that a company should combine internal and external data for internal loss distribution modeling, when the sample of internal data seems insufficient to represent all the domain accurately. Here we take into account both minimum collection threshold as well as underreporting. The mixing model should utilize the external data as prior knowledge. Then, a nonparametric smoothing technique adjustment should be applied on the prior knowledge according to internal data availability.

In our example, the mixing model is applied as a first step. In the second step, the underreporting correction is introduced. We assume that the same underreporting applied both for the internal and the external data.

We use two data sets which correspond to operational risk data from internal and external sources. These data have been generated for illustrative purposes from a particular company and type of operational risk. We will call these data the internal data set and the external data set, respectively. In Table 2 we show the basic descriptive statistics of our two data sets. Results in Table 2 show that the size of losses is very different in the internal sample and the external sample. The reason for the different scale is that data from the external sample have been obtained from companies that may be larger than the company from which internal data were obtained. Therefore with mixing transformation techniques we also provide a rescaling procedure that would solve the problem of combining both data sets. Also, we note that quantile
99% and maximum in internal data are the same, this evidences that for large losses internal data are scarce.

Table 2. Descriptive statistics of the two operational risk data sets

<table>
<thead>
<tr>
<th></th>
<th>Number of observations</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Median</th>
<th>Quantile 95%</th>
<th>Quantile 99%</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal data set</td>
<td>75</td>
<td>0.176</td>
<td>0.278</td>
<td>0.081</td>
<td>0.780</td>
<td>1.773</td>
<td>1.773</td>
</tr>
<tr>
<td>External data set</td>
<td>700</td>
<td>0.679</td>
<td>4.094</td>
<td>0.033</td>
<td>1.672</td>
<td>19.439</td>
<td>52.130</td>
</tr>
</tbody>
</table>

Source: Own and Bolancé et al. (2012b)

In Figure 1 we show the results of the semiparametric model described in section 2 on the tail of pdf for the internal and the external data set, respectively. This figure shows as distribution of external data is heavier tailed than one of internal data.

Figure 1. Extreme internal and external data.

Source: Bolancé et al. (2012b)

With the aim to estimate Value-at-Risk (VaR) and Tail-Value-at-Risk, we estimate four models for the density function of the losses associated with operational risk, named M1, M2, M3 and M4. Model M1 denotes the semiparametric model described in section 2 and will be used for the internal and the external data set, respectively. Models M2 and M3 are mixing models in section 3. In M2 we use transformation parameters estimated with external data to obtain transformed kernel density estimation with internal data. The rescaling procedure is introduced in model M3. This model is similar to M2, but in M3 we use shape transformation parameters estimated with external data and the scale transformation parameter estimated with internal data.
Finally, model M4 is the benchmark model with a Weibull assumption on the severity estimated with internal data.

From the severity density function, we estimate the distribution of total operational risk. We assume that the annual number of internal losses that have been reported has mean equal 30, then we can assume that the annual number of internal losses follows a Poisson process with maximum likelihood estimated parameter $\lambda = 30$. We denote the annual simulated reported frequencies by $\lambda_r$ with $r = 1, \ldots, R$ and number of simulations $R = 10,000$. After correcting for underreporting, we estimate the expected number of occurred claims as $\lambda_o = \lambda / P_o$, where $P_o = \int_0^\infty \tilde{f}_o(t)u(t)dt$ and we denote the annual simulated occurred frequencies by $\lambda_o$. The value of $\lambda_o$ for each estimated model is around 126 occurred losses.

For each annual simulated reported frequencies we draw $\lambda_r$ random uniformly distributed samples $u_{rk} \in U(0,1)$ for $k = 1, \ldots, \lambda_r$ and obtain loss sizes taken from the inverse estimated cumulative distribution function of the severity distribution of reported claims $\tilde{F}(u_{rk})$, this function is obtained integrating the estimated densities function calculate with the different models: M1, M2, M3 and M4. The outcome are the simulated reported annual total losses denoted by $S_r = \sum_{k=1}^{\lambda_r} \tilde{F}^{-1}(u_{rk}), r = 1, \ldots, R$, where $\tilde{F}^{-1}(u_{rk})$ is the inverse of the cumulative distribution function.

We do the same using the annual simulated occurred frequencies $\hat{\lambda}_o$ and cumulative distribution function with underreporting correction. The simulated occurred annual total losses are $S_{oR} = \sum_{k=1}^{\hat{\lambda}_o} \tilde{F}_o^{-1}(u_{rk}), r = 1, \ldots, R$, with $u_{rk} \in U(0,1)$ for $k = 1, \ldots, \hat{\lambda}_o$ and $\tilde{F}_o^{-1}(u_{rk})$ is the inverse of the cumulative distribution function obtained integrating the estimated densities from the four proposed models with underreporting correction, as we describe in section 4.

Table 3 we report VaR and TVaR for risk tolerance $\alpha = \{0.95, 0.99, 0.995\}$ for the simulated $S_r$ total loss distribution for each model. Results for models M1, M2, M3 and M4 reveals that when we include external information (models M2 and M3) the estimate for the operational risk measure is larger than it is for other models.

In Table 4 we show the ratio between results in Table 3 and the same results including underreporting. We can see that results without underreporting subestimate risk.

<table>
<thead>
<tr>
<th>Model</th>
<th>Var95%</th>
<th>Var99%</th>
<th>Var99.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>6.53</td>
<td>7.65</td>
<td>8.84</td>
</tr>
<tr>
<td>M2</td>
<td>7.11</td>
<td>8.38</td>
<td>9.89</td>
</tr>
<tr>
<td>M3</td>
<td>7.60</td>
<td>8.93</td>
<td>10.68</td>
</tr>
<tr>
<td>M4</td>
<td>7.76</td>
<td>8.99</td>
<td>10.37</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>TVar95%</th>
<th>TVar99%</th>
<th>TVar99.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>7.24</td>
<td>8.27</td>
<td>10.23</td>
</tr>
<tr>
<td>M2</td>
<td>7.90</td>
<td>9.16</td>
<td>11.45</td>
</tr>
<tr>
<td>M3</td>
<td>8.50</td>
<td>10.06</td>
<td>12.39</td>
</tr>
<tr>
<td>M4</td>
<td>8.52</td>
<td>9.72</td>
<td>11.89</td>
</tr>
</tbody>
</table>

Source: Own and Bolancé et al. (2012b)
Table 4. Ratio between risk measures with underreporting and without underreporting.

<table>
<thead>
<tr>
<th>Model</th>
<th>Var_{95%}</th>
<th>Var_{99%}</th>
<th>Var_{99.5%}</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>2.41</td>
<td>2.92</td>
<td>4.91</td>
</tr>
<tr>
<td>M2</td>
<td>3.24</td>
<td>5.38</td>
<td>9.18</td>
</tr>
<tr>
<td>M3</td>
<td>3.38</td>
<td>5.64</td>
<td>8.81</td>
</tr>
<tr>
<td>M4</td>
<td>2.03</td>
<td>1.94</td>
<td>1.89</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>TVar_{95%}</th>
<th>TVar_{99%}</th>
<th>TVar_{99.5%}</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>2.83</td>
<td>3.84</td>
<td>6.78</td>
</tr>
<tr>
<td>M2</td>
<td>4.64</td>
<td>7.02</td>
<td>9.58</td>
</tr>
<tr>
<td>M3</td>
<td>4.75</td>
<td>6.96</td>
<td>9.13</td>
</tr>
<tr>
<td>M4</td>
<td>1.98</td>
<td>1.91</td>
<td>2.02</td>
</tr>
</tbody>
</table>

6 Discussion

Our recommendation is to pay attention to underreporting whereas model selection for combining internal and external data is especially relevant only when the tolerance level is high, which is when the quantile for assessing risk if high.

Our application handles the problem underreporting together with the mixing of internal and external data. It clearly shows that failure to account for underreporting may lead to a substantial underestimation of operational risk. The use of external data information can easily be incorporated in our modeling approach. So, even if the underlying theoretical foundation is technically difficult, the intuition and the practical implementation of operational risk models is very straightforward.

Our method also addresses the statistical problem of estimating a density function with extreme values and then how to combine data sources, while accounting for operational risk underreporting. Underreporting, which is a feature that practitioners admit is the most dangerous problem of operational risk data quality, has generally been much ignored in the specialized literature. More details on the underlying methodology are to be found in Bolance et al. (2012b), but the applications presented there had not shown the combined effect of mixing internal and external data and the issue of underreporting at the same time.

References


