Journal of Banking & Finance 68 (2016) 266–278

Contents lists available at ScienceDirect

Journal of Banking & Finance

journal homepage: www.elsevier.com/locate/jbf

# On stability of operational risk estimates by LDA: From causes to approaches

Xiaoping Zhou [a], Antonina V. Durfee [b], Frank J. Fabozzi [c,*]

[a] Citizens Financial Group, United States
[b] Credit Suisse, United States
[c] EDHEC Business School, Nice, France

ARTICLE INFO

*Article history:*
Received 22 September 2014
Accepted 16 January 2016
Available online 7 April 2016

*JEL classification:*
G21
G28

*Keywords:*
Operational risk modeling
Maximum likelihood estimation
Bias correction
Robust estimation
Right-truncated distribution
Probability weighted least squares method

ABSTRACT

The stability of estimates is critical when applying advanced measurement approaches (AMA) such as loss distribution approach (LDA) for operational risk capital modeling. Recent studies have identified issues associated with capital estimates by applying the maximum likelihood estimation (MLE) method for truncated distributions: significant upward mean-bias, considerable uncertainty about the estimates, and non-robustness to both small and large losses. Although alternative estimation approaches have been proposed, there has not been any comprehensive study of how alternative approaches perform compared to the MLE method. This paper is the first comprehensive study on the performance of various potentially promising alternative approaches (including minimum distance approach, quantile distance approach, scaling-based bias correction, upward scaling of lower quantiles, and right-truncated distributions) as compared to MLE with regards to accuracy, precision and robustness. More importantly, based on the properties of each estimator, we propose a right-truncation with probability weighted least squares method, by combining the right-truncated distribution and minimizing a probability weighted distance (i.e., the quadratic upper-tail Anderson–Darling distance), and we find it significantly reduces the bias and volatility of capital estimates and improves the robustness of capital estimates to small losses near the threshold or moving the threshold, demonstrated by both simulation results and real data application.



## 1. Introduction

The Basel II Accord defines operational risk as "the risk of loss resulting from inadequate or failed internal processes, people and systems or from external events". Since 2002, many financial institutions have collected operational loss data, gradually employing more quantitative approaches (referred to as advanced measurement approaches (AMA)[1] under the Basel Accord) to calculate operational risk capital requirements. The loss distribution approach (LDA) is the most common and central framework used for calculating operational risk capital. Under LDA, operational risk loss data are used to estimate a loss frequency distribution and a loss severity distribution. Then they are combined to construct an annual loss distribution. Operational risk regulatory capital or economic capital is then defined as a very high quantile (99.9% for regulatory capital or operational risk exposure, and 99.97% or 99.95% for economic capital) of the annual loss distribution.

Despite the well-known limitations and difficulties of implementing the LDA approach,[2] many financial institutions apply this approach in their operational risk capital planning process with the hope that issues due to data insufficiency will be overcome when there is enough internal loss data (ILD) collected or external loss data (ELD) available. However, given the probable nonhomogeneity of ILD over time[3] and the uncontrollability of ELD quality and consistency, data may never become "enough" in the near future. Although analyzing, filtering, and scaling ELD are very important,[4] it is often difficult for an individual bank to obtain enough information for effectively filtering or scaling ELD such that the data actually reflect

* Corresponding author.
*E-mail addresses:* zxp567@gmail.com (X. Zhou), antoninadurfee@gmail.com (A.V. Durfee), frank.fabozzi@edhec.edu (F.J. Fabozzi).

[1] For introduction to AMA for operational risk modeling, see Cruz (2002), Alexander (2003), Aue and Kalkbrener (2006), and Chernobai et al. (2008).

[2] See Cope et al. (2009) and Ames et al. (2014).

[3] For studies on time-related factors of operational risk events, see Allen and Bali (2007), Chernobai et al. (2012), and Cope et al. (2012).

[4] For studies on scaling ELD, see Cope and Labbi (2008) and Dahen and Dionne (2010).

http://dx.doi.org/10.1016/j.jbankfin.2016.01.014
0378-4266/

a similar risk profile to the bank. For this reason, many financial institutions have relied principally on ILD while using ELD more as a benchmark or for scenario-data design.

Without overlooking the importance of either ILD or ELD in operational risk capital planning, in this paper we analyze the stability issues of operational risk estimates using LDA from causes of instability to approaches for improvement. The root cause of instability of operational risk capital estimates can be still attributed to data insufficiency. This cause may not be overcome in the near future because the 99.9% quantile of annual loss distribution would require thousands of years of data without a parametric assumption for the severity distribution type. However, this is not the only cause of instability in the estimates obtained from LDA. Even under a correct model assumption, as Cope (2011) points out, there is considerable uncertainty with parameter estimates for "truncated severity distributions". Although there have been alternatives such as shifted distributions (see Rozenfeld (2010)), truncated distributions continue to be recommended or required by regulators in certain regions such as in the United States (due to the fear of banks artificially lowering capital estimates by using shifted distributions). However, statistical evidence presented by Cavallo et al. (2012) suggests that truncated and shifted distributions can be equivalently valid or invalid when the true model is unknown. The debate between using truncated or shifted distributions is beyond the scope of this paper.[5] Instead, our focus is on the truncated distribution formulation.

The classical estimation method – maximum likelihood estimation (MLE) – under a truncated parametric distribution assumption for the loss severity distribution would lead to significant uncertainty for the parameter estimates and thus capital estimates (see Cope (2011), Opdyke and Cavallo (2012), and Zhou et al. (2014)). There have been numerous alternative estimation approaches proposed by minimizing alternative distances. These estimation approaches include minimizing Cramer-von Mises (CvM) distance, minimizing Anderson–Darling (AD) distance, and minimizing quantile distance (QD).[6] There are also other alternative approaches proposed by penalizing, constraining or modifying the likelihood function such as penalized likelihood estimators proposed by Cope (2011), optimally-biased robust estimators presented by Horbenko et al. (2010) and Opdyke and Cavallo (2012), and Bayesian estimators suggested by Shevchenko (2010) and Zhou et al. (2014). The motivation of all of these estimators (either using different objective functions or a different parameter space) is to reduce the uncertainty of parameter estimates, thereby improving the stability of capital estimates. However, there has not been a comprehensive study on the sources of instability and how each of the alternative approaches address these instability sources compared to MLE, with quantitative assessment on the bias, volatility and robustness of capital estimates by these approaches.

The contributions of this paper are many-fold. First, we review the sources of instability of operational risk capital estimates from the root causes and nuisance issues. We will review how the proposed alternative approaches address them and the limitations of each approach. Second, we summarize a list of performance measures for evaluating a capital estimator from various perspectives such as bias, volatility and robustness. Third, we conduct a comprehensive comparative study via simulation experiments on the performance of various capital estimators. Specifically, we review the minimum distance approach (minimizing AD, minimizing CvM, minimizing upper-tail AD, and minimizing QD, etc), right-truncated approach, bias-correction by downward scaling (similar to the reduced capital estimator by Opdyke (2014)), and upward-scaling of lower quantiles. Finally, through a systematic comparison, we are able to propose a very promising estimator, which we refer to as the right-truncation with probability weighted least squares (RT-PWLS), by combining the advantages of the right-truncated approach and probability-weighted least squares method. Through simulation experiments for a variety of distributions, we find our proposed estimate significantly improves every perspective of accuracy, precision and robustness compared to MLE or any other approach applied alone. It also has the advantage of simplicity and commonality so that it can be easily implemented and generally applied in practice.

The rest of this paper is structured as follows. In Section 2, we analyze the statistical causes of capital instability in operational risk modeling. In Section 3, we present the evaluation measures for capital estimators. Section 4 compares various alternative approaches to MLE and discusses the merits and limitations of each using examples from a variety of popular operational loss severity distributions. An investigation of the application of RT-PWLS estimators in reducing error and improve robustness of operational risk capital estimates by both simulation and real data examples for a variety of distributions is provided in Section 5. Section 6 concludes the paper and suggests more future directions in operational risk capital modeling under the AMA framework.

## 2. Review of sources of instability and alternative estimators

According to advanced approaches, operational risk exposure (or capital) is often defined as a very high quantile (e.g., 99.9%) of the distribution of a banking organization's potential aggregated operational losses over a one-year horizon. However, the capital stability (or uncertainty) is an important issue and the causes of instability have been uncovered gradually by many researchers.[7] In this section, we summarize and classify the causes of instability (or uncertainty) of operational risk capital estimates from three perspectives: parametric assumption, parameter and capital uncertainty, and non-robustness to data change. Note that there are other causes of capital estimate instability that are also fundamentally important such as the classification of unit of measures and dependence among individual events, which we do not list below. Instead, we focus more on the modeling of a given data set (one unit of measure) with a known reporting threshold under the independent and identically distributed assumption which is the central issue of operational risk modeling.

### 2.1. Parametric model assumption

As shown by Cope et al. (2009), to obtain a reasonable standard error of the quantile of annual loss distribution (even the quantile of severity distribution) requires an impractically large number of observations if based only on the data without extrapolation. In addition, from the single-loss approximation (SLA) formula presented by Böcker and Sprittulla (2006) and its formulation with truncated distribution by Rozenfeld (2012), a major component of a capital estimate can essentially be approximated by an extremely high quantile of the loss severity distribution. For example, to obtain an accurate $\alpha = 99.9\%$ capital estimate for a cell with about $\lambda = 50$ events per year, one needs to have information about the $(1 - \frac{1-\alpha}{\lambda}) \approx 99.998\%$ quantile of the severity distribution, which needs about 50,000 events without a parametric assumption. However, ILD with about 750 events over a 15-year period only provides inference about the 99.87% quantile and any higher percentiles could only rely on extrapolation, which is highly

[5] For a further discussion, see Chernobai et al. (2006), Luo et al. (2009), and Ergashev et al. (2012).

[6] See Ergashev (2008) and Lehérissé and Renaudin (2013).

[7] See Mignola and Ugoccioni (2006), Cope et al. (2009), Cope (2011), Rozenfeld (2012) and Ames et al. (2014).

assumption dependent. We also note that even using ELD, for a 15-year period of sample data, we would need loss data from approximately 1,000/15 ≈ 67 similar banks. However, in practice, because banks often have very different business profiles and risk control levels, ELD cannot directly be pooled together with ILD but instead requires a credible scaling algorithm, which is hardly available at the present time.

To avoid the subjectivity of a parametric assumption, an alternative approach is using nonparametric estimators such as an empirical distribution or a kernel estimation (see Bolance et al. (2012)). Typically nonparametric estimators require a large sample size, otherwise it may result in under-estimation of capital estimates since any extreme quantiles could not exceed the largest loss.

Given the fact that there is a large deviation in extreme quantiles for various distributions but little evidence to differentiate them based on sample information, truncating the upper tail (see Carrillo-Menéndez and Suárez (2012)) may help avoid the economically meaningless estimates and stabilize the capital estimates under different distribution assumptions. One could also taper the right-tail (such as the smoothly truncated stable distributions in Menn and Rachev (2009) and the extended Pareto law (EPL) distribution in Vargas Mendoza (2012)) instead of truncating the right-tail, but the capital estimate will also be highly dependent on the tapering approach and extra instability will be introduced with more parameters.

Another approach recommended by Cope et al. (2009) and Ames et al. (2014) is lowering the quantile and then scaling up by a multiple defined by regulators and probably varying for different loss types. It could be a more feasible approach since it essentially avoids any subjectivity in extrapolating the tails beyond the sample and shifts the focus to improving what is "estimable". An issue with this approach is that the scaling ratio between a higher quantile and a lower quantile is hard to estimate but relies on certain expert information. Under the power-law assumption of the tail behavior, there exists a power scaling between different quantile levels. Degen and Embrechts (2011) further extend the scaling to a "non-constant" version via estimation of a local tail index. However, the estimation of a local tail index may be subject to significant uncertainty since there are very scarce data in the tail portion, especially for small samples.

### *2.2. Uncertainty of parameter estimates and amplified uncertainty of capital estimates*

Apart from model mis-specification risk from various parametric assumptions, the capital estimate can also be sensitive to the parameter estimation method.

The classical method for parameter estimation is the MLE method due to its consistency, asymptotical normality and optimal efficiency (all in parameters' world not capital's world). Nevertheless, the MLE method has a number of issues for small samples or ill-posed problems when the likelihood surface is too flat over the parameter space. Cope (2011) provides an illustration that the likelihood surface is greatly distorted by the denominator term in the density function of truncated distribution, which then significantly increases the volatility of parameter estimates. This issue becomes exacerbated for real data where mis-specification of the severity distribution likely exists. In addition to model mis-specification risk, the behavior of the likelihood function of an assumed distribution for real data often exhibits much worse properties than on simulated data, resulting in much more sensitive capital results for real data than for simulated data.

Due to the poor property of likelihood functions for truncated distributions, there have been alternative distances proposed such as Anderson–Darling distance, Cramér–von Mises (CvM) distance, probability-weighted distance (see Cavestany et al. (2015)) and quantile-distance (QD) (see Ergashev (2008) and Lehérissé and Renaudin (2013)). However, to our knowledge, how efficient they are as compared to MLE and how the factors in defining these distances would impact the performance of capital estimators have not been fully studied. We will provide a review of their performance with comparison to MLE in this paper.

There is another class of estimators that reduce the parameter estimation error by trading off a small bias increase for a larger variance reduction. These include penalized likelihood estimators and Bayesian estimators. Penalized likelihood estimators (see Cope (2011)) are designed to reduce the distortion of likelihood caused by the truncation by adding an additional term to the objective function. However, choosing the penalty parameter could be a challenging problem. Bayesian estimators (see Shevchenko (2010) and Zhou et al. (2014)) can be used to combine prior knowledge of parameters (could be informative or non-informative) and information implied in the sample, which is then used to obtain a posterior distribution of parameters. Researchers have found that Bayesian estimators can be used to reduce the uncertainty of parameter estimates and capital estimates. However, a challenge in applying Bayesian estimators in practice is that the specification of the prior distribution for parameters is not an easy task, especially when the parameters have no intuitive meanings.

Different from the above estimators, the reduced-bias capital estimator (RCE) of Opdyke (2014) focuses directly on capital uncertainty instead of parameter uncertainty. This is based on an observation that the distribution of parameter estimates are translated to a significantly right-skewed distribution of capital estimates. By scaling down the upwardly biased capital estimates obtained by MLE, the bias and variance of capital estimates can be reduced simultaneously. However, a challenging issue in applying the estimator is that a key tuning parameter related to distribution type and sample size requires careful consideration. More importantly, although the estimator is able to shrink the mean-squared error (MSE) and mean-bias, the overall performance such as median-bias and robustness of this estimator are not fully studied. Although we will review the bias-correction framework by proposing a more generally applicable and easy-to-implement simulation-based downward-scaling method to correct the bias, we will point out that these down-scaling methods may address the mean-bias and MSE at the cost of median-bias.

### *2.3. Non-robustness to data change*

In addition to the excessive variance of parameters produced by MLE for truncated distributions, Opdyke and Cavallo (2012) revealed by analyzing the influence functions (Hampel et al. (1986); Huber (2010)) of parameter estimates for truncated distributions that the MLE method exhibits significant non-robustness to data update. Counterintuitively, the operational risk capital estimate can be highly unstable, not only to large loss events but also to small loss events close to the threshold. As far as we know, this non-robustness is not unique to MLE estimators only, but also to some alternative estimators such as minimizing CvM, or minimizing AD etc, which can be demonstrated via easy simulation experiments.

The sensitivity of capital to small losses close to the threshold actually poses another often ignored but potentially significant source of instability of capital estimates: change of threshold. Intuitively, changing the threshold slightly should not lead to a large impact on the capital estimate since moving the threshold upward slightly implies dropping only a very small amount of losses. Given the non-robustness of capital estimates to inclusion or removal of small losses near the threshold, the capital estimates may also be non-robust to the moving threshold. These phenomena can be easily confirmed for various estimators such as MLE, minimizing CvM or minimizing AD. This non-robustness to small losses near the threshold or sensitivity to moving threshold needs to be carefully assessed when designing a capital estimator since changing the threshold might be a usual practice due to an inflation adjustment or the pooling different data sources with various thresholds.

To address the non-robustness issue, Opdyke and Cavallo (2012) introduced an Optimally Bias-Robust Estimator (OBRE) by bounding the influence functions of parameter estimates, and demonstrated through simulation examples that these estimators could improve the robustness of capital estimates significantly. However, the selection of the bounding parameter is not an easy task and requires careful consideration, resulting in this method not generally applied in practice.

There are some more generally applicable approaches such as quantile-distance estimators (with appropriately heavier weight on tail quantiles) or minimizing upper-tail quadratic Anderson–Darling distance that would improve the robustness of the estimator to small loss changes. In addition, the piece-wise distributions (such as extreme value theory (EVT)[8]) isolate the impact on the distribution's body from the impact on the distribution's tail, thus it could also improve the robustness to small loss changes. However there are some drawbacks to the EVT approach. The tail threshold may be selected by visual inspection (such as the Hill plot and mean excess function plot) or automated algorithms (such as the sequential statistical test method proposed by Nguyen and Samorodnitsky (2012)), but the selection of the threshold is often very challenging and leads to considerable instability in capital estimates. Furthermore, the number of loss events above the tail threshold is often very small, which also increases the uncertainty of the estimates for the parameter. In this paper, we do not discuss the EVT or spliced distribution approach.

From the above review of sources of instability and proposed approaches, we note there has not been a systematic comparative study of the performance of all these approaches (or the most promising representatives) in estimating operational risk capital. Before we move forward to the comparison in the following section, we first summarize and define some key performance measures that we consider useful when comprehensively evaluating and comparing various capital estimators.

## 3. Evaluation of capital estimators

To evaluate the performance of an estimator, we need to apply the error measures for capital estimates rather than on parameter estimates. As noted by Opdyke (2014), there are three perspectives that need to be evaluated – accuracy, precision, and robustness. In what follows, we present the mathematical definitions of performance measures explicitly and introduce more measures needed than what have been used in Opdyke (2014). In addition, we present some general qualitative considerations when designing an estimator (simplicity and commonality).

### 3.1. Accuracy (unbiasedness)

The mean-bias (usual so-called "bias") of an estimator is defined as:

$$b(\hat{C}) = E[\hat{C}] - C^*$$

An estimator is downwardly biased if $b(\hat{C}) < 0$, and upwardly biased if $b(\hat{C}) > 0$.

It can also be written in terms of the ratio for positive variables (e.g., capital):

$$\xi(\hat{C}) = \frac{E[\hat{C}]}{C^*}$$

when $\xi(\hat{C}) < 1$, the estimator is downwardly biased, and when $\xi(\hat{C}) > 1$ it is upwardly biased.

Here, the bias is defined through expectation. The bias can also be defined by the deviation in the median estimate (see Lehmann, 1951):

$$medBias(\hat{C}) = median(\hat{C}) - C^*$$

Statistically, the mean-unbiasedness indicates that the average magnitude of over-estimation is equal to the average magnitude of under-estimation, while the median-unbiasedness indicates that the likelihood of over-estimation is equal to the likelihood of under-estimation. For symmetric distributions of $\hat{C}$, the mean-unbiasedness and median-unbiasedness can be achieved at the same time. However, when the distribution of an estimate is skewed, they cannot be reached at the same time.

### 3.2. Precision (estimation error)

The most commonly used measure to evaluate estimation error is the MSE:

$$\mathrm{MSE}(\hat{C}) = E\left[(\hat{C} - C^*)^2\right] = Var(\hat{C}) + Bias(\hat{C})^2$$

Similarly, a relative root MSE can be defined as follows:

$$\mathrm{RRMSE} = \frac{1}{C^*}\sqrt{MSE(\hat{C})}$$

There is another root mean square logarithm error (RMSLE), see Lehérissé and Renaudin (2013), that may be well-suited in operational risk context because it gives more importance to under-estimation errors than to over-estimation errors.

$$\mathrm{RMSLE} = \sqrt{E\left[(\log(\hat{C}) - \log(C^*))^2\right]}$$

This measure is also an often more robust measure than RRMSE by taking the logarithm of the capital estimates.

### 3.3. Robustness

We can evaluate the robustness of a capital estimate by examining its sensitivity to data contamination. For example, the change and relative change of capital estimates for a loss sample $\{x_i\}_{i=1}^n$ with a new loss event of \$x are defined as:

$$\Delta\hat{C}(x) = \hat{C}_{n+1}(x_1, \ldots, x_n, x) - \hat{C}_n(x_1, \ldots, x_n)$$

$$\Delta\hat{C}\%(x) = \frac{\Delta\hat{C}(x)}{\hat{C}_n(x_1, \ldots, x_n)}$$

These measures are useful for assessing the robustness of a capital estimator to different updates to the sample data.

Compared to the robustness to large losses, the robustness to small losses is even more important in operational risk because, intuitively, small loss updates should not lead to significant capital variation. Nevertheless, this is not true for capital estimates obtained from MLE, as revealed by Opdyke and Cavallo (2012) analytically through the derivation of the influence function for truncated distributions.

To evaluate the robustness to small losses, we define a capital sensitivity (CS) measure for truncated distributions with threshold $T$ for a sample $X = \{x_i\}_{i=1}^n$ as following:

[8] See Chavez-Demoulin et al. (2006).

$$CS_\alpha = \sup_{T<H<x_{[n\alpha]}} \left\{ \frac{\hat{C}(\{x_i|x_i > H\})}{\hat{C}(X)} \right\}$$

where $\alpha \in (0,1)$ is a small number (we choose $\alpha$ as 10% in this paper). The measure reflects the maximum relative error of capital estimates when removing the smallest $100\alpha\%$ observations and changing the modeling threshold accordingly.

### 3.4. Simplicity and commonality

In addition to the above quantitative measures, there are also other qualitative considerations when assessing an approach, such as simplicity (whether the approach is simple enough such that it is easy to implement in practice and requires little effort in pre-specifying key inputs) and commonality (whether the approach can be generalized to a more common problem such as a variety of distribution types). These are important factors to decide whether the estimator will become generally applicable and accepted. There have been numerous alternative estimators to MLE proposed, most of which are indeed effective in addressing the uncertainty of parameter estimates or capital estimates. However, many approaches are not generally applicable in practice because the performance significantly depends on some important "prior" inputs or settings that cannot be easily understood or specified for real-world problems. In order for an estimator to be generally applicable, it needs to limit the source of input factors or reduce the sensitivity of the estimator's performance to different inputs. Otherwise, the estimator will introduce significant model risk since the model may be used with inappropriate settings.

## 4. Comparison of various capital estimation methods

In this section, we provide the detailed formulation of various capital (parameter) estimation methods, including a variety of existing approaches in the literature and some extended methods in this paper.

The operational risk capital estimation problem is defined as follows. The annual loss frequency is assumed to follow a Poisson distribution with mean $\lambda$, and the observed loss severity sample over threshold $T$ (denoted as $X = \{x_i\}_{i=1}^n$) are assumed to be from a parametric distribution with density $f(x;\theta)$ and distribution function $F(x;\theta)$. Let the conditional distribution function for $X|X > T$ be $F_T(x;\theta) = \frac{F(x;\theta)-F(T;\theta)}{1-F(T;\theta)}$. The parameter $\theta$ and the capital estimate $VaR_\alpha(L)$ under a confidence level of $100\alpha\%$ are to be estimated such that:

$$P\left(L = \sum_i X_i > VaR_\alpha(L)\right) = 1 - \alpha$$

where $\alpha$ is often an extreme quantile such as 99.9% or 99.97%.

### 4.1. Estimation approaches

The most widely used parameter estimation method is the MLE method due to its consistency, asymptotical normality and optimal efficiency. There are a variety of estimation approaches proposed in literature like minimizing alternative distances such as minimizing CvM distance, minimizing AD distance, and minimizing quantile-distance. We will study their performance based on the evaluation measures from each perspective of accuracy, precision and robustness defined in Section 3, and investigate the factors affecting their performance. Particularly, the new contributions in this section are the following. First, we will analyze the right-truncated approach with detailed formulation on approximation for capital and expected loss estimates, and discuss its attractive properties. Second, we present a simulation-based bias correction method with a downward-scaling of MLE capital estimates (similar to Opdyke (2014)'s RCE). We will also analyze the theory and practical issues regarding a promising approach – scaling up of lower quantiles. Finally, we summarize the properties of each approach regarding its accuracy, precision and robustness, and try to combine their merits when proposing new estimators.

#### 4.1.1. Maximum likelihood estimator

The MLE is obtained as:

$$\arg\min_{\theta\in\Theta} -\sum_{i=1}^n \ln f(x_i;\theta) + n\ln(1 - F(T;\theta))$$

#### 4.1.2. Probability weighted least squares

The probability weighted least squares (PWLS) estimator[9] is obtained as:

$$\arg\min_{\theta\in\Theta} \sum_{i=1}^n \frac{(F_T^*(x_i) - F_T(x_i;\theta))^2}{F_T(x_i;\theta)^p(1 - F_T(x_i;\theta))^m}$$

where $F_T^*(\cdot)$ is the empirical distribution function parameters, and $p$ and $m$ adjust the weight of observations in the left-tail and right-tail, respectively. The PWLS method is related to minimizing Cramer-von Mises (CvM) distance, and quadratic Anderson-Darling distance ($AD^2$) or "upper-tail" Anderson-Darling distance ($AD^2_{up}$) (see Chernobai et al. (2015)) as following:

- Minimizing CvM: $p = 0$, $m = 0$;
- Minimizing $AD^2$: $p = 1$, $m = 1$;
- Minimizing $AD^2_{up}$: $p = 0$, $m = 2$.

There are potentially other types of estimators such as minimizing the AD distance or Kolmogorov–Smirnov (KS) distance. However, based on initial analysis we found they are neither more efficient nor more robust than MLE. In addition, readers can potentially analyze other variety of choices of $p$ and $m$ for the PWLS estimator to further increase the weight of left-tail or right-tail, but we only evaluate the results for the three most typical choices.

#### 4.1.3. Quantile-distance estimation

The quantile-distance estimation approach (see Ergashev (2008) and Lehérissé and Renaudin (2013)) is to estimate the parameters by minimizing the distance between sample quantiles and theoretical quantiles. The quantile-distance estimator is obtained as the following by minimizing a weighted average of difference between the sample quantiles and theoretical quantiles:

$$\arg\min_{\theta\in\Theta} \sum_{k=1}^K w_k\left(\hat{q}(\alpha_k) - \hat{F}^{-1}(\alpha_k;\theta)\right)^2$$

where the specifications used in this paper are defined as follows (based on the original proposal in Ergashev (2008) and discussions in Lehérissé and Renaudin (2013)): $K = \left[\frac{n}{2}\right]$, $w_k = 1/\hat{q}(\alpha_k)^2$, $\alpha_k = c\frac{1-exp(-\frac{mk}{K})}{1-exp(-m)}$, where $m = 2$, $k = 1, 2, \ldots, K$, and $c = 1 - 0.5/n$. The calculation of sample quantiles $\hat{q}(\alpha)$ is a linear interpolation of the empirical quantile to make use of the largest observation in the sample.

[9] The name "probability weighted least squares" is from Cavestany et al. (2015). However, we use the theoretical distributions in the denominator since this is more consistent to the definition of various existing distances like Anderson-Darling distance, and avoids the significant sensitivity of various definitions of empirical distribution function in the denominator.

Readers should be aware that without careful specification of quantile-distance definition, the performance of capital estimates could vary significantly. The selected specifications here seek to reduce the overall uncertainty of capital estimates and the sensitivity to small events near the threshold, which are the typical issues associated with MLE. Briefly, the selection of the number of quantiles $K$ will affect the efficiency of the estimator. The grid of quantiles $\alpha_k$, the weighting of each quantile distance $w_k$, the selection of largest quantile $c$, and even the calculation of the sample quantiles (see Hyndman and Fan (1996) for many variations of sample quantiles) are all important factors in determining which portion of the distribution is emphasized and are all decisive to the performance of capital estimation. Detailed investigation about the impact of each factor for the quantile distance estimation method on capital estimation has been analyzed by the authors but not presented here (will be available upon request) since it is not the focus of this paper.

*4.1.4. Right-truncated approach*

Suppose a reasonable upper bound of a single loss can be estimated as $U$, and losses in the sample $\{x_i\}_{i=1}^n$ are in $[T, U]$, the MLE for $\theta$ is then:

$$\arg\min_{\theta\in\Theta} -\sum_i \ln f(x_i;\theta) + n\ln(F(U;\theta) - F(T;\theta))$$

Let $\lambda_T$ be the conditional frequency above the reporting threshold $T$, then the derived frequency below threshold $\lambda_L$ and unconditional frequency $\lambda$ are respectively $\lambda_L = \frac{F(T;\theta)}{F(U;\theta)-F(T;\theta)}\lambda_T$ and $\lambda = \frac{F(U;\theta)}{F(U;\theta)-F(T;\theta)}\lambda_T$. Let $\tilde{F}(x)$ be the conditional CDF for $X$ over $[T, U]$, i.e., $\tilde{F}(x) = \frac{F(x;\theta)-F(T;\theta)}{F(U;\theta)-F(T;\theta)}$ $(x \in [T, U])$ and $\tilde{F}^{-1}(y) = F^{-1}(F(T;\theta) + y(F(U;\theta) - F(T;\theta)))(y \in (0, 1))$, then a general form of single-loss approximation of capital estimate can be easily represented as:

$$VaR_\alpha(L) \approx \tilde{F}^{-1}\left(1 - \frac{1-\alpha}{\lambda_T}\right) + (\lambda_T - 1)\mu_T + LBT$$

where $\lambda_T = \lambda(F(U;\theta) - F(T;\theta))/F(U;\theta)$ and $\lambda\mu = \lambda_T\mu_T + \lambda_L\mu_L$, where $\lambda_L\mu_L$ can be approximated using empirical information about annual losses below threshold (LBT) and $(\mu_T - \mu)$ is negligible compared to capital estimates.

Empirical experiments show that the derived SLA for right-truncated distributions approximates the true capital estimates very well in many cases when the frequency is small or $U$ is sufficiently large. However, the fast Fourier transform (FFT) method (see Embrechts and Frei (2009)) is used to accurately calculate capital estimates in the case when the frequency is large and the upper bound $U$ is not sufficiently large leading to under-estimated capital estimates.

The right-truncated distribution formulation has a number of attractive properties. First, it always gives rise to a finite conditional severity mean and a finite expected loss (EL) estimate. Under an upper bound $U$ of a single loss, the EL can be represented as:

$$EL = \lambda E[x|x < U] = \lambda\frac{\int_0^U xf(x;\theta)dx}{F(U;\theta)}$$

Since in reality there is no infinite loss, when estimating the EL under an upper bound of a single loss it is important to understand what the "realistic" EL is.

EL can also be decomposed into "expected loss below threshold $T$" and "expected loss above threshold $T$" so that the information about the empirical expected loss below threshold $T$ can be utilized.

$$EL = \lambda_L E[x|x < T] + \lambda_T E[x|T < x < U]$$

$$= \lambda_L\int_0^T x\frac{f(x;\theta)}{F(T;\theta)}dx + \lambda_T\int_T^U x\frac{f(x;\theta)}{F(U;\theta) - F(T;\theta)}dx$$

$$\approx \text{Empirical Annual Loss Below Threshold (LBT)} + \lambda_T\mu_T$$

where $\mu_T = \int_T^U x\frac{f(x;\theta)}{F(U;\theta)-F(T;\theta)}dx$ is the conditional severity mean above threshold, and the unconditional severity mean is $\mu = \frac{\int_0^U xf(x;\theta)dx}{F(U;\theta)}$.

Second, the right-truncated distribution formulation provides a finite SLA formula for capital estimates for infinite-mean distributions. The mean-corrected SLA proposed by Böcker and Sprittulla (2006) have demonstrated great accuracy in approximating the capital estimates and have been widely used in practice as a computationally efficient approach to calculate capital estimates. However, there has been difficulty in using it for infinite-mean distributions due to the infinite severity mean. To deal with this issue, Opdyke (2014) proposed some interpolation-based method based on the correction terms for the SLA introduced by Degen (2010). Instead, the right-truncated distribution can be employed to avoid the infinite-mean issue and truncate the unreasonably high tails even for a finite-mean distribution.

*4.1.5. Bias-correction by downward scaling*

Note that the capital estimates based on an unbiased estimator for the parameters is always upwardly biased (see Opdyke (2014)). In classical statistics, shrinkage estimators can be applied to reduce the MSE for an unbiased estimator by trading a small bias for more reduction in the variance (e.g., James–Stein estimator (James and Stein, 1961) in estimating the mean of Gaussian random vectors, and Ledoit-Wolf estimator (Ledoit and Wolf, 2004) in estimating the covariance matrix). In the case of capital estimates (where the ML estimator is itself biased upward), we can actually see that the correction of the bias and reduction of the variance can be achieved at the same time.

Let $\hat{C}_B$ be the bias-corrected estimate which is written as a scaled version of $\hat{C}$ :[10]

$$\hat{C}_B = \beta\hat{C}$$

Then the bias of the estimate is: $b(\hat{C}_B) = \beta E[\hat{C}] - C^*$. Let $\xi = \frac{E[\hat{C}]}{C^*}$ be the ratio of expectation of estimate to the true estimate. It is easy to see that when $\beta = 1/\xi$, $\hat{C}_B$ is unbiased. The key to bias correction is to estimate the bias ratio $\xi$ for $\hat{C}$.

The MSE of the bias-corrected estimator $\hat{C}_B$ can be written as: $\mathrm{MSE}(\hat{C}_B) = Var(\hat{C}_B) + b(\hat{C}_B)^2 = \beta^2 Var(\hat{C}) + (\beta\xi - 1)^2 C^{*2}$. By comparing this to the MSE of the MLE $\hat{C}$: $\mathrm{MSE}(\hat{C}) = Var(\hat{C}) + b(\hat{C})^2 = Var(\hat{C}) + (\xi - 1)^2 C^{*2}$, we can see that when $\xi > 1$ (as it often is for MLE) and $\beta = \frac{1}{\xi}$, the MSE of bias-corrected estimator $\hat{C}_B$ is strictly smaller than that of $\hat{C}$.

$$\mathrm{MSE}(\hat{C}_B) = \beta^2 Var(\hat{C}) < \mathrm{MSE}(\hat{C}).$$

This means the bias correction in the capital estimate is not only to correct the bias, but will also reduce the estimation variance. This makes the "scaling" estimator not only favorable in terms of improving estimation stability but also necessary since it reduces the bias rather than introducing bias as the shrinkage estimators do in other applications.

[10] Kim and Hardy (2007) presented a shifting bias correction instead of scaling bias correction using bootstrapping. However, the downward-shifting approach may result in negative capital estimates and does not have the benefit of shrinking variance as the scaling method does. There are ratio-based bias correction techniques in other statistical applications (such as Sprugel (1983), Newman (1993), and Snowdon (1991)).

The key in making the scaling estimator useful is to accurately estimate the bias ratio $\xi$. In the following, we provide a simulation-based approach based on parametric bootstrapping to obtain an estimate $\hat{\xi}$ for the bias ratio.

1. Estimate parameter $\hat{\theta}_0$ from the sample, and calculate the capital $\hat{C}_0$ using $\hat{\theta}_0$;
2. Simulate $M$ random samples from the estimated distribution with parameter $\hat{\theta}_0$;
3. Estimate parameters $\{\hat{\theta}_i\}$ and calculate the capital estimate $\{\hat{C}_i\}$ for each sample $i = 1, 2, \ldots, M$.
4. Calculate the estimate $\hat{\xi}$ as the ratio:

$$\hat{\xi} = \frac{\mathrm{mean}(\{\hat{C}_i\})}{\hat{C}_0} \approx \frac{\mathrm{mean}(\{\hat{C}_i\})}{\mathrm{median}(\{\hat{C}_i\})}$$

The basis for the second "≈" is based on the fact that the median of capital estimates is often preserved in parametric bootstrapping: $\mathrm{median}(\{\hat{C}_i\}) \approx \hat{C}_0$. The ratio between the mean and median is estimated by utilizing a simulation experiment. This simulation-based bias correction technique is similar to the reduced capital estimator (RCE) proposed by Opdyke (2014) in the sense that both are based on downward scaling of MLE capital estimates. However, we provide a more general simulation-based framework here without resorting to the analytical Fisher information and tuning the power-scaling ratios for various sample size and distribution types as described in Opdyke (2014).

#### 4.1.6. Scaling up lower quantiles

Since the estimation of the extremely high quantile has significant variability, Cope et al. (2009) and Ames et al. (2014) both suggest estimating a lower quantile and then scaling up by a multiple defined by regulators. Theoretically, according to the SLA of capital estimates, the ratio of capital estimates under two confidence levels $\alpha_1$ and $\alpha_2$ ($\alpha_1 < \alpha_2$, but both reasonably close to the tail) can be represented as:

$$\frac{VaR_{\alpha_2}}{VaR_{\alpha_1}} \approx \frac{\tilde{F}^{-1}\left(1 - \frac{1-\alpha_2}{\lambda_T}\right)}{\tilde{F}^{-1}\left(1 - \frac{1-\alpha_1}{\lambda_T}\right)}$$

In EVT, there have been attempts to estimate the high quantiles based on the tail behavior assumption. When the distribution is a power-law, for example, $1 - F(x) \sim a x^{-\frac{1}{\gamma}}$, it is easy to see that $\frac{VaR_{\alpha_2}}{VaR_{\alpha_1}} \approx \left(\frac{1-\alpha_1}{1-\alpha_2}\right)^{\gamma}$, which is known as the EVT scaling law. The parameter $\gamma$ is called the extreme value index and can be estimated using the Hill estimator when the distribution has a Pareto-type tail. Degen and Embrechts (2011) further extends the scaling to a "local" slope version for extended regular variation class: $\frac{VaR_{\alpha_2}}{VaR_{\alpha_1}} \approx \left(\frac{1-\alpha_1}{1-\alpha_2}\right)^{\gamma(t)}$ (where $t = \frac{1}{1-\alpha_1}$ and $\gamma(t)$ is estimated via local regression methods), which is called *penultimate* approximation as opposed to the *ultimate* approximation for the power-law. These approaches are effective to address the model mis-specification issue, since in the real world the true parametric form of the severity distribution is unknown. However, under a known distribution form, they are less efficient than the MLE since only a small subset of the data is used when estimating the scaling coefficient. In this paper, we mainly focus on the uncertainty of the capital estimate under a known distribution type, and will use the true scaling coefficient to scale up a lower quantile in order to produce a benchmark of targeting performance when the scaling coefficient is accurate.

### 4.2. Simulation experiments

In this section, we comprehensively review the performance of the capital estimates that we previously described: (1) MLE; (2) minimizing CvM; (3) minimizing $AD^2$; (4) minimizing $AD^2_{up}$; (5) minimizing QD; (6) right-truncated MLE; (7) bias correction by scaling down MLE capital; (8) scaling up lower quantiles (90%) with a known scaling ratio, and; (9) the new estimator that we propose, the RT-PWLS by combining the right-truncated distribution with minimizing a version of PWLS ($AD^2_{up}$) to improve both estimation error and robustness to small losses over MLE. We will compare the performance of capital estimators using the measures described in Section 3: (1) mean bias; (2) median bias; (3) RRMSE; (4) RMSLE, and; (5) $CS_\alpha(\alpha = 10\%)$.

We will investigate the performance for a variety of different distributions[11] as an illustration for a threshold $T$ equal to \$10,000 and $\lambda_T$ equal to 20:

| Distribution | True capital $VaR_{99.9\%}$ | True capital $VaR_{99.97\%}$ |
|---|---|---|
| Truncated lognormal (10.5, 2.5) | \$778 million | \$1,535 million |
| Truncated loggamma (27, 0.38) | \$525 million | \$1,226 million |
| Truncated loglogistic (11.5, 0.85) | \$482 million | \$1,325 million |

To investigate the consistency of various estimators and the performance under various lengths of sample, we vary the sample size from 100 to 400 with increment 100. For each sample size, we simulate $M$ = 1,000 samples from the assumed distribution, and estimate the parameters using MLE and calculate capital estimates by SLA.

For the right-truncated approach, the selection of the upper bound $U$ of a single loss should be chosen such that it is sufficiently large but also needs to be reasonably tight to be useful in reducing the volatility of the capital estimates. We select it as \$5 billion here, which is 3–4 times the true estimate of $VaR_{99.97\%}$ and is a very conservative estimate that does not need strong expert information. In addition, to avoid a possible error of SLA (although it does not seem to be material here when the upper bound is large and frequency is small), we use the FFT method to calculate the capital estimate for the right-truncated approach to ensure accuracy.

For the approach of scaling up the lower quantile, we will estimate the capital level of 90%, and then scale up according to the true ratio $VaR_{99.9\%}/VaR_{90\%}$, and $VaR_{99.97\%}/VaR_{90\%}$. Also, the capital estimate for the lower quantile is obtained by applying the FFT method to increase accuracy.

For the scaling-down approach, in each experiment ($m$ = 1, 2,… $M$), we use the estimated parameters to simulate $MM$ = 500 samples, and scale the estimated capital by the estimated bias ratio (mean to median of the capital estimated for the $MM$ samples).

[11] The density functions for the distributions are

$$\text{Lognormal}: \; f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left\{-\frac{(\ln\, x-\mu)^2}{2\sigma^2}\right\}, x > 0, \sigma > 0$$

$$\text{Loggamma}: \; f(x;\kappa,\theta) = \frac{1}{\theta^{\kappa}\Gamma(\kappa)}\left(\frac{(\ln x)^{\kappa-1}}{x}\right)\exp\left(-\frac{\ln x}{\theta}\right), x > 1, \kappa > 0, \theta > 0$$

$$\text{Loglogistic}: \; f(x;\mu,\sigma) = \frac{\frac{1}{\sigma}x^{\frac{1}{\sigma}-1}}{e^{\frac{\mu}{\sigma}}\left(1+\left(\frac{x}{e^{\mu}}\right)^{\frac{1}{\sigma}}\right)^2}, x > 0, \sigma > 0$$

Note that to improve the robustness of the ratio and reduce the number of trials in the simulation, when calculating the mean capital we applied a trimming with $\alpha = 10\%$ in order to exclude the top and bottom 10%. This percentage depends on heavy tailedness of the distribution and how many trials are simulated. It acts as a tuning parameter – the higher $\alpha$ is, the less the scaling is. Generally, we found 10% trimming is a good choice to significantly reduce the bias after scaling across a variety of distributions and different sample sizes.

### 4.3. Results and discussion

Table 1 presents the results for the truncated lognormal distribution, including the mean-bias, median-bias, RRMSE, RMSLE, and CS(10%) of both 99.9% and 99.97% capital by using the nine approaches. To conserve space and because the conclusions are unchanged, the results for the loggamma and loglogistic distributions are not reported here but are available from the authors.

The approach of "Scaling up lower quantile with a known scaling ratio" serves as the target performance benchmark since it utilizes the additional information beyond the data. Based on the simulation results of performance for various estimation approaches, we can see the following from Table 1:

(1) Minimizing the usual alternative distances such as CvM or two-sided $AD^2$ distance does not improve estimation accuracy, precision or robustness at all. While the median biases of both approaches are close to zero like MLE, the mean-bias and RRMSE are significantly higher than MLE. In addition, when removing the smallest 10% number of losses, the capital estimates by these alternative distances would change significantly (more than 50% for a sample of 100 events), which is even worse than MLE (over 30% for a sample of 100 events).

(2) We observe that the bias-correction method, although significantly reducing the mean-bias and variance, introduces higher negative median-bias. This is unfavorable because although the average magnitude of under-estimation is close to the average magnitude of over-estimation, the likelihood of under-estimation is much higher than the likelihood of over-estimation. In addition, we note that this approach provides almost no improvement over MLE in the robustness to small losses.

**Table 1**
Comparison of performance of various capital estimators (truncated lognormal distribution).

| Truncated lognormal (10.5, 2.5), $T$ = \$10,000, $\lambda$ = 20 | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Method | Sample size | 99.9% capital (true estimate = \$778 Mm) | | | | | 99.97% capital (true estimate = \$1,535 Mm) | | | | |
| | | Mean bias | Median bias | RRMSE | RMSLE | CS(10%) | Mean bias | Median bias | RRMSE | RMSLE | CS(10%) |
| MLE | 100 | 85% | −4% | 304% | 108% | 32% | 113% | −4% | 413% | 118% | 36% |
| | 200 | 37% | −2% | 131% | 76% | 25% | 46% | −1% | 160% | 83% | 29% |
| | 300 | 24% | −1% | 88% | 62% | 21% | 29% | −1% | 103% | 68% | 23% |
| | 400 | 10% | −7% | 63% | 51% | 18% | 13% | −7% | 72% | 56% | 21% |
| Min CvM | 100 | 283% | −12% | 1359% | 142% | 52% | 446% | −13% | 2356% | 157% | 58% |
| | 200 | 101% | −2% | 390% | 103% | 40% | 138% | −2% | 555% | 114% | 45% |
| | 300 | 51% | −4% | 167% | 83% | 33% | 64% | −4% | 208% | 91% | 36% |
| | 400 | 23% | −4% | 103% | 66% | 31% | 29% | −5% | 126% | 72% | 34% |
| Min AD2 | 100 | 243% | 13% | 865% | 132% | 63% | 353% | 16% | 1391% | 146% | 69% |
| | 200 | 104% | 9% | 370% | 95% | 44% | 138% | 11% | 528% | 104% | 49% |
| | 300 | 49% | 4% | 143% | 74% | 34% | 61% | 6% | 175% | 82% | 38% |
| | 400 | 22% | −2% | 90% | 58% | 30% | 28% | −3% | 108% | 64% | 34% |
| Min AD2up | 100 | 204% | −5% | 944% | 127% | 14% | 307% | −6% | 1656% | 140% | 16% |
| | 200 | 68% | 1% | 204% | 91% | 9% | 87% | 1% | 258% | 101% | 11% |
| | 300 | 51% | 9% | 153% | 78% | 7% | 65% | 11% | 192% | 86% | 8% |
| | 400 | 28% | 1% | 101% | 63% | 6% | 35% | 1% | 122% | 70% | 7% |
| Min QD | 100 | −10% | −56% | 137% | 125% | 12% | −4% | −60% | 168% | 137% | 13% |
| | 200 | −15% | −42% | 86% | 93% | 9% | −14% | −44% | 100% | 103% | 11% |
| | 300 | −20% | −35% | 58% | 74% | 8% | −21% | −38% | 64% | 82% | 9% |
| | 400 | −20% | −31% | 52% | 64% | 7% | −20% | −33% | 57% | 70% | 8% |
| RT-MLE ($U$ = \$5Bln) | 100 | 31% | −5% | 121% | 92% | 28% | 9% | −9% | 82% | 90% | 27% |
| | 200 | 20% | −3% | 84% | 68% | 22% | 7% | −6% | 64% | 66% | 21% |
| | 300 | 14% | −3% | 66% | 57% | 18% | 5% | −7% | 54% | 55% | 18% |
| | 400 | 5% | −9% | 51% | 47% | 17% | −2% | −12% | 44% | 48% | 17% |
| Bias Correction | 100 | −10% | −43% | 105% | 100% | 28% | −12% | −48% | 116% | 110% | 32% |
| | 200 | −3% | −25% | 71% | 73% | 22% | −3% | −28% | 78% | 80% | 24% |
| | 300 | −2% | −16% | 61% | 56% | 20% | −2% | −18% | 69% | 62% | 22% |
| | 400 | 2% | −7% | 53% | 52% | 17% | 2% | −10% | 58% | 57% | 19% |
| Scaling-up | 100 | 9% | −9% | 65% | 52% | 12% | 9% | −9% | 65% | 52% | 12% |
| ($\alpha$ = 90% with known ratio) | 200 | 5% | −3% | 42% | 39% | 9% | 5% | −3% | 42% | 39% | 9% |
| | 300 | 2% | −3% | 32% | 30% | 7% | 2% | −3% | 32% | 30% | 7% |
| | 400 | 4% | 0% | 29% | 28% | 6% | 4% | 0% | 29% | 28% | 6% |
| RT-Min AD2up ($U$ = \$5Bln) | 100 | 47% | −6% | 146% | 100% | 11% | 18% | −12% | 93% | 96% | 10% |
| | 200 | 34% | −1% | 107% | 79% | 8% | 15% | −4% | 76% | 75% | 7% |
| | 300 | 29% | 6% | 91% | 67% | 6% | 14% | 3% | 67% | 64% | 6% |
| | 400 | 16% | −1% | 70% | 57% | 6% | 6% | −4% | 56% | 55% | 5% |

*Note:* The table shows the performance measures (mean bias, median-bias, RRMSE, RMSLE, CS(10%)) for both 99.9% and 99.97% capital estimated using the nine approaches. The simulation experiments are repeated for $M$ = 1000 severity samples drawn from truncated lognormal (10.5, 2.5) with threshold \$10,000 at each sample size (100, 200, 300 and 400). The annual frequency is assumed to be known from a Poisson distribution with mean $\lambda = 20$.

(3) The quantile-distance estimation method (with proper definitions of weighting and quantile grids) could significantly reduce the mean-bias and variance, as well as improve the robustness to small losses. However, the results suggest that it may introduce significant median-bias as well. In addition, there are a number of factors that would affect the performance of the quantile-distance approach significantly, such as how to choose the quantiles and how the sample quantiles are defined. Simply speaking, the quantile-distance estimators should not be taken for granted as a "panacea" to MLE and practitioners need to understand the impact of various factors. As such, even when the approach may perform well in some particular cases, we feel this method may not be a generally applicable approach since there are too many important sensitivity sources in defining the distance.

(4) Minimizing $AD^2_{up}$ may improve the robustness to small losses. However, this approach introduces more bias and variance into the estimates than MLE.

(5) The right-truncated approach using the MLE as parameter estimation method significantly reduces the bias and variance of capital estimates. This suggests that specifying an upper bound for a single loss (even when it is very high and not so tight to the individual loss experience) would significantly reduce the uncertainty of capital estimates. However, this approach does not improve much the robustness to small losses.

(6) Based on the above two observations, a natural idea is to see whether combining the right-truncated approach with the minimizing $AD^2_{up}$ would help to achieve a more balanced performance in estimation error and robustness. From the results of "RT-PWLS($m = 2$)", we can see that for the small sample size or the higher quantile cases, the mean-bias, RRMSE and robustness to small threshold are all significantly improved compared to the MLE or Min $AD^2_{up}$ approach without right truncation.

**Table 2**
Comparison of performance of various capital estimation approaches and limitations.

| Method | Accuracy | Precision | Robustness | Other limitations |
|---|---|---|---|---|
| MLE | High mean-bias | High RRMSE | Not robust | |
| Min CvM | Significant mean-bias | Significant RRMSE | Not robust | |
| Min $AD^2$ | Significant mean-bias | Significant RRMSE | Not robust | |
| Min $AD^2_{up}$ | Significant mean-bias | Significant RRMSE | Robust to small losses | |
| Min QD | Low mean-bias (possible), but depends on definition of QD | Low RRMSE (possible) | Robust (possible) | Too many important factors in definition |
| RT-MLE | Low mean-bias and median-bias | Low RRMSE | Not robust to small losses | |
| Bias correction | Low mean bias but high negative median-bias | Low RRMSE | Not robust | |
| Scaling up | Low mean-bias and median-bias | Low RRMSE | Robust | Scaling ratio unknown |
| RT-PWLS ($m = 2$) | Low mean-bias and median-bias | Low RRMSE | Robust to small losses | |

**Table 3**
Summary of important inputs, limitations and merits of various recently proposed alternative approaches to MLE.

| Method | Input | Limitations | Merits |
|---|---|---|---|
| Bayesian approach | Prior distribution of parameters; choice of posterior estimates (e.g., MMSE or MAP) | Hard to specify prior for "parameters"; slow calculation | Flexible framework for parameters uncertainty and incorporation of expert information |
| Quantile-distance | Definition of quantile distance (selection of number and grid of quantiles, weighting, calculation of sample quantile) | There are too many factors (some are not intuitive) affecting the performance | Can improve robustness to small events, and reduce capital estimation volatility (but needs caution in use) |
| Scaling | Level of lower quantiles; scaling ratio | The ratio is hard to estimate and depends on tail behavior | Improves robustness and reduce volatility by avoiding directly estimating extreme quantile |
| EVT or spliced distributions | The selection of tail threshold; the distribution type of tail portion | Use of small subset of data; large sensitivity | Robust to small events |
| OBRE | Selection of bounding factor | Not intuitive for capital estimation; influence of extreme losses is also bounded | More robust to both small and large events |
| RCE | The selection of quantile grids over parameters space; selection of scaling parameter "c" for various distributions and sample sizes | High quantiles of parameter space around initial parameter estimates may lead to undefined capital; need to derive the Fisher information for each distribution type; need to pre-calibrate the scaling parameter "c" carefully; may lead to significant median-bias | Reduce bias and RRMSE |
| Bias correction | Selection of trimming percentage when calculating the mean of capital estimates | Sensitive to simulation error and choice of trimming in calculating the mean of capital estimates; low speed | Easy to implement; reduce bias and RRMSE |
| RT-PWLS ($m = 2$) | Selection of upper bound of a single loss | | The input of upper bound is intuitive; reduce volatility and improves robustness to small losses; consistent to "finite loss" exposure in real world |

Table 2 summarizes the above comparison on the performance and limitations of each approach discussed. For the review of general applicability, we also summarize in Table 3 the sensitivity sources or important input factors of each alternative approach to MLE so that modelers need to pay close attention to when applying them.

## 5. Application of RT-PWLS estimators

From the results reported in Section 4, we have seen that the right-truncated probability weighted least squares (RT-PWLS ($m = 2$)) method may be very useful for improving the estimation accuracy while also improving the robustness of the capital estimators. In this section, we investigate further the application of RT-PWLS estimators in reducing error and improving robustness of operational risk capital estimates by both simulation and real data examples for a variety of distributions.

### 5.1. Simulation results

In order to demonstrate that the merits of RT-PWLS estimators are generally applicable, we conduct a simulation experiment for a variety of parameters and distributions that are frequently used for operational loss severity modeling, such as lognormal, loggamma, and loglogistic distributions. Also, we will assess the sensitivity of the performance to the selection of right-truncation upper bound $U$.

From Table 1 we note that the estimation bias and error are mostly significant for small sample sizes. Therefore, we will focus on the comparison of MLE and RT-PWLS estimators with respect to their performance for small sample sizes. We analyze the performance of these estimators for a loss sample over a 15-year period with the frequency from the Poisson distribution (where the mean of annual frequency $\lambda = 10$ and $\lambda = 20$) and severity from the lognormal, loggamma, and loglogistic distributions. Different from our previous experiments that in the previous section for various sample sizes where we focus on the sensitivity from the loss severity distribution only because of the need to validate the consistency of each estimator, here we also allow for the sensitivity from the loss frequency distribution such that it is more similar to the true problem. The right-truncation upper bounds are selected as three times and five times the true 99.9% capital estimates, which are quite conservative even for some very heavy-tailed distributions.

Table 4 shows the results for the lognormal distribution. The results for the loggamma and loglogistic distributions are not reported here but are available from the authors. The conclusions are the same as for the lognormal distribution. From Table 4 we can see that the RT-PWLS estimator would significantly reduce the mean-bias, RRMSE, and capital sensitivity to small losses even with a very high upper bound for an individual loss. We need to point out that the specification of the upper bound is not to simply put a cap on the capital since the estimated capital is far less than the chosen upper bound, but is to reduce the unnecessary volatility of parameter and capital estimation caused by the infinite upper bound. In addition, the PWLS estimator is effective for reducing the sensitivity of capital to small losses near the threshold since higher weights are assigned to larger losses. The experiments in this section further confirm that the benefit of RT-PWLS estimators is not a coincidence for a particular distribution but are generally applicable for other distributions with various parameters.

### 5.2. Empirical application

Here we apply the RT-PWLS estimator to a real operational loss data set in a particular Basel loss event type from a US commercial bank from 2003/01/01 to 2015/06/30. We will compare its performance with respect to the stability of expected loss and capital estimates to MLE. Note that the methods investigated in this paper are general methods so that they can be easily applied to any other severity distributions.

The unit of measure we analyze is a typical small sample size cell with only 94 events above $10,000 over a 12.5-year period,

**Table 4**
Comparison of MLE with RT-PWLS estimators on capital estimation performance (Lognormal).

| Lognormal | True | MLE | | | | | RT-PWLS ($U = 3*True 99.9% capital) | | | | | RT-PWLS ($U = 5*True 99.9% capital) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Parameters | 99.9% capital ($M) | Mean-bias | Median-bias | RRMSE | RMSLE | CS (10%) | Mean-Bias | Median-bias | RRMSE | RMSLE | CS (10%) | Mean-bias | Median-bias | RRMSE | RMSLE | CS (10%) |
| *lambda = 10; sample period = 15 years; threshold = $10 K* | | | | | | | | | | | | | | | | |
| (9.5,2) | 32 | 17% | −12% | 96% | 69% | 20% | 0% | −13% | 58% | 63% | 6% | 11% | −11% | 77% | 69% | 7% |
| (9.5,2.5) | 218 | 39% | −12% | 162% | 88% | 25% | 1% | −18% | 66% | 77% | 7% | 19% | −15% | 96% | 84% | 8% |
| (9.5,3) | 1503 | 96% | −4% | 342% | 110% | 33% | 5% | −11% | 70% | 85% | 8% | 29% | −5% | 107% | 94% | 9% |
| (10,2) | 49 | 18% | −4% | 81% | 62% | 23% | 8% | −4% | 58% | 59% | 6% | 20% | −1% | 78% | 65% | 8% |
| (10,2.5) | 330 | 40% | −5% | 148% | 84% | 29% | 9% | −8% | 68% | 73% | 6% | 29% | −4% | 101% | 82% | 7% |
| (10,3) | 2277 | 74% | −8% | 266% | 107% | 28% | 0% | −12% | 68% | 88% | 7% | 22% | −6% | 101% | 96% | 8% |
| (10.5,2) | 76 | 23% | −5% | 101% | 64% | 20% | 6% | −6% | 59% | 60% | 6% | 18% | −4% | 81% | 66% | 7% |
| (10.5,2.5) | 508 | 42% | −8% | 156% | 85% | 24% | 4% | −8% | 64% | 72% | 7% | 22% | −3% | 94% | 80% | 8% |
| (10.5,3) | 3489 | 64% | −11% | 241% | 102% | 31% | −2% | −19% | 68% | 89% | 6% | 19% | −13% | 100% | 96% | 7% |
| *lambda = 20; sample period = 15 years; threshold = $10 K* | | | | | | | | | | | | | | | | |
| (9.5,2) | 46 | 6% | −7% | 53% | 48% | 14% | 5% | −4% | 48% | 49% | 4% | 12% | −2% | 61% | 53% | 4% |
| (9.5,2.5) | 332 | 20% | −7% | 91% | 63% | 21% | 4% | −5% | 55% | 57% | 5% | 18% | −1% | 77% | 64% | 6% |
| (9.5,3) | 2479 | 41% | −6% | 151% | 83% | 27% | 3% | −11% | 64% | 73% | 6% | 24% | −5% | 94% | 80% | 7% |
| (10,2) | 70 | 15% | −4% | 68% | 52% | 16% | 10% | −1% | 54% | 49% | 5% | 20% | 1% | 72% | 55% | 5% |
| (10,2.5) | 506 | 23% | −5% | 99% | 64% | 20% | 5% | −6% | 54% | 56% | 4% | 19% | −1% | 76% | 63% | 5% |
| (10,3) | 3768 | 45% | 5% | 149% | 78% | 27% | 5% | −3% | 59% | 66% | 6% | 24% | 5% | 86% | 73% | 8% |
| (10.5,2) | 108 | 11% | −3% | 57% | 45% | 17% | 8% | −1% | 49% | 46% | 5% | 16% | 1% | 63% | 51% | 5% |
| (10.5,2.5) | 780 | 22% | −2% | 88% | 59% | 18% | 6% | −2% | 53% | 54% | 5% | 19% | 3% | 76% | 61% | 6% |
| (10.5,3) | 5789 | 26% | −8% | 113% | 74% | 27% | −1% | −10% | 55% | 67% | 5% | 15% | −4% | 77% | 73% | 6% |

*Note:* The table shows the performance measures (mean bias, median-bias, RRMSE, RMSLE, CS(10%)) for 99.9% capital estimated using the three approaches: (1) MLE; (2) RT-PWLS (with $U = 3$*true 99.9% capital); and (3) RT-PWLS (with $U = 5$*true 99.9% capital). The simulation experiments are repeated for $M = 1000$ times for a compound Poisson distribution where the severity distribution is truncated lognormal (10.5,2.5) with threshold $10,000 and frequency is from a Poisson distribution with annual mean $\lambda = 20$ over a 15-year period.

**Table 5**
Comparison of parameters, expected loss and capital results by various estimators on real data set.

| Number of observations: 94 | | Annual average loss: $0.65 M | | | | | Annual frequency: 7.52 | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Estimation method | Distribution name | Parameters | | AD test p value | KS test p value | AIC | Expected loss ($M) | 99.9% capital ($M) | CS (10%) | Distribution of 99.9% capital by bootstrap | | | |
| | | | | | | | | | | 10% ($M) | 90% ($M) | Skewness | Kurtosis |
| MLE | Lognormal | 7.75 | 2.17 | 0.65 | 0.69 | 2211.35 | 0.68 | 13.87 | 21% | 4.49 | 52.07 | 2.13 | 8.74 |
| | Loggamma | 28.89 | 0.31 | 0.65 | 0.70 | 2211.41 | 0.73 | 20.00 | 27% | 6.89 | 96.74 | 4.93 | 30.74 |
| | Loglogistic | 9.21 | 0.94 | 0.49 | 0.71 | 2211.89 | 2.26 | 83.67 | 35% | 16.74 | 429.39 | 4.59 | 35.50 |
| | GPD | 0.90 | 9823.72 | 0.54 | 0.67 | 2211.81 | 1.50 | 65.44 | 34% | 6.54 | 313.03 | 3.52 | 16.14 |
| RT-MLE ($U$ = 20LL) | Lognormal | 7.74 | 2.17 | 0.74 | 0.75 | 2211.35 | 0.68 | 13.30 | 19% | 4.49 | 30.18 | 0.40 | 1.99 |
| | Loggamma | 28.77 | 0.31 | 0.65 | 0.71 | 2211.40 | 0.71 | 17.40 | 21% | 6.84 | 33.44 | 0.47 | 2.08 |
| | Loglogistic | 9.19 | 0.94 | 0.46 | 0.63 | 2211.83 | 0.87 | 29.78 | 12% | 14.00 | 37.66 | -0.72 | 2.77 |
| | GPD | 0.91 | 9701.06 | 0.53 | 0.68 | 2211.77 | 0.83 | 28.09 | 13% | 6.40 | 36.40 | -0.36 | 2.02 |
| RT-PWLS ($U$ = 20LL) | Lognormal | 8.08 | 2.04 | 0.62 | 0.67 | 2211.42 | 0.62 | 10.56 | 6% | 3.75 | 30.84 | 0.49 | 2.04 |
| | Loggamma | 33.06 | 0.27 | 0.61 | 0.84 | 2211.48 | 0.65 | 13.70 | 7% | 4.64 | 35.15 | 0.53 | 2.23 |
| | Loglogistic | 9.47 | 0.87 | 0.31 | 0.44 | 2212.08 | 0.76 | 25.54 | 4% | 9.66 | 37.46 | -0.38 | 2.04 |
| | GPD | 0.84 | 11539.19 | 0.37 | 0.48 | 2211.90 | 0.73 | 23.81 | 6% | 4.14 | 36.09 | -0.12 | 1.82 |
| RT-MLE ($U$ = 50LL) | Lognormal | 7.75 | 2.17 | 0.75 | 0.71 | 2211.35 | 0.68 | 13.80 | 21% | 4.49 | 44.16 | 1.09 | 3.46 |
| | Loggamma | 28.86 | 0.31 | 0.65 | 0.69 | 2211.40 | 0.72 | 19.40 | 25% | 6.88 | 57.31 | 1.27 | 3.79 |
| | Loglogistic | 9.20 | 0.94 | 0.43 | 0.69 | 2211.87 | 0.95 | 49.22 | 21% | 15.81 | 79.58 | -0.01 | 2.00 |
| | GPD | 0.90 | 9773.14 | 0.47 | 0.70 | 2211.79 | 0.89 | 43.74 | 22% | 6.51 | 75.12 | 0.35 | 2.05 |
| RT-PWLS ($U$ = 50LL) | Lognormal | 8.08 | 2.04 | 0.70 | 0.69 | 2211.42 | 0.62 | 10.75 | 7% | 3.75 | 49.51 | 1.21 | 3.64 |
| | Loggamma | 33.08 | 0.27 | 0.64 | 0.85 | 2211.49 | 0.65 | 14.53 | 8% | 4.65 | 65.56 | 1.39 | 4.04 |
| | Loglogistic | 9.47 | 0.87 | 0.33 | 0.37 | 2212.11 | 0.81 | 37.31 | 7% | 10.17 | 78.91 | 0.25 | 1.90 |
| | GPD | 0.83 | 11567.26 | 0.50 | 0.50 | 2211.92 | 0.77 | 32.94 | 9% | 4.15 | 73.76 | 0.57 | 2.17 |

*Note:* The table shows the estimated parameters, goodness-of-fit tests results (AD test and KS test $p$ values obtained by simulation), Akaike information criteria (AIC), expected loss, 99.9% capital, capital sensitivity to removal of 10% smallest losses, 10-th and 90-th percentile and skewness and kurtosis of the bootstrapped 99.9% capital distribution by using the five different estimation approaches (1) MLE; (2) RT-MLE (with $U = 20^*$largest loss); (3) RT-PWLS (with $U = 20^*$largest loss); (4) RT-MLE (with $U = 50^*$largest loss); and (5) RT-PWLS (with $U = 50^*$largest loss) for four different severity distributions (lognormal, loggamma, loglogistic and GPD).

and the average annual loss is about $0.65 million. In Table 5, we fit the truncated lognormal, loggamma, loglogistic and generalized Pareto distributions[12] to this cell. We will investigate the following approaches in estimating parameters: (1) MLE; (2) right-truncated and MLE ($U = 20$ times the largest loss); (3) right-truncated and PWLS ($U = 20$ times the largest loss); (4) right-truncated and MLE ($U = 50$ times the largest loss), and; (5) right-truncated and PWLS ($U = 50$ times the largest loss).

From the results by MLE, we can see that all of them pass the AD and KS tests with similar $p$-values and the differences between AIC values are very small. This indicates that they have very similar statistical quality: although the lognormal distribution is preferred since it has the lowest AIC, the relative likelihoods of the other three models (loggamma, loglogistic, and GPD) to the lognormal model are respectively 0.97, 0.76 and 0.80, according to the approximation formula $exp(\frac{\Delta AIC_i}{2})$.[13] However, the estimated expected loss and 99.9% capital from these distributions using the MLE method may vary significantly under different distribution assumptions. From Table 5 it can be seen that although there is almost no difference in the goodness-of-fit with other distributions, the loglogistic distribution leads to the highest capital ($84 million) and expected loss ($2.26 million), as well as the widest 80% confidence interval for capital by bootstrap ([$17 million, $429 million]). In addition, the capital is highly sensitive to the small losses near the threshold: removing 10% small losses would lead to up to a 35% change in capital. These results suggest the presence of significant model risk and non-robustness of the capital estimates by MLE.

When truncating the losses at 20 times the largest loss, the estimates of expected loss from various distributions become much more similar and closer to the empirical average annual loss, and the capital estimates from various distributions become less volatile. This demonstrates that specifying an upper bound for a single loss would significantly reduce the volatility of both expected loss and capital estimates. Also, it will slightly improve the robustness to small losses.

When truncating the losses at 20 times the largest loss and estimating the parameters using PWLS, we can see that both the volatility of capital by bootstrap and sensitivity of capital to the removal of small losses are significantly reduced. The skewness and kurtosis of bootstrapped capital distribution by right-truncation are significantly lower than without right-truncation.

Finally, we assess the choice of the upper bound of single loss by varying the upper bound from 20 times the largest loss to 50 times the largest loss. As expected, the volatility of capital estimates will become higher with a looser upper bound. However, we can still see that they are significantly lower than without any right-truncation. What is more important is that using the PWLS estimator leads to more stable capital estimates with respect to the movement of small losses, an extremely important finding for practitioners.

The comparison of capital sensitivity to a moving threshold (removal of up to 10% of the smallest losses) by MLE, RT-MLE ($U$ = 20LL) and RT-PWLS ($U$ = 20LL) are shown in Fig. 1. It can be seen that the RT-PWLS acquires the advantages of both the right-truncation in reducing capital volatility in re-sampling and PWLS estimators in capital sensitivity to small losses near the threshold.

It is more inspiring that the upper bounds we selected are actually very conservative in comparison to the losses historically experienced. Even with such conservative bounds, we can achieve such significant improvement. In addition, the upper bound has very intuitive meanings, making this approach more attractive and practical than other approaches that require a special and deep understanding of the tuning parameters' impact on model performance such as the Bayesian or bias-correction methods. In

[12] Generalized Pareto density function: $f(x;\kappa,\sigma) = \frac{1}{\sigma}(1+\frac{\kappa x}{\sigma})^{-(1+\frac{1}{\kappa})}, x > 0, \kappa > 0, \sigma > 0$.

[13] See Rozenfeld (2014) for discussions about model selection and model averaging in operational risk, and Burnham and Anderson (2002) for more about general model selection and multimodel inference theory.

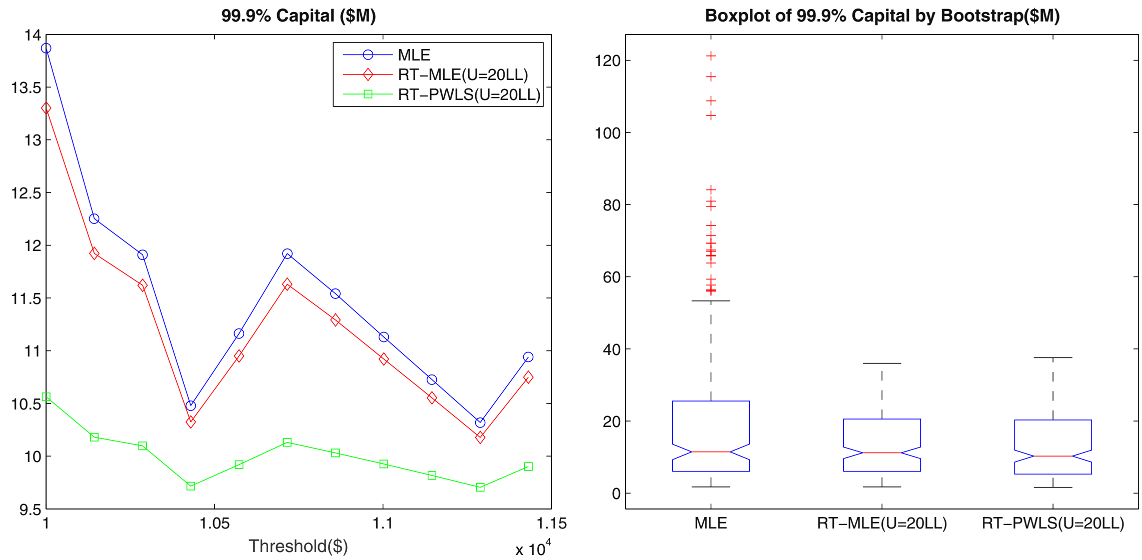


**Fig. 1.** Capital sensitivity to moving threshold and capital box-plot by bootstrap for various approaches (Lognormal). Note: The left plot shows the sensitivity of 99.9% capital estimates to the moving threshold from \$10,000 to the 10-th percentile of historical loss severity sample (\$11,432) for the three approaches (MLE, RT-MLE ($U = 20^*$largest loss) and RT-PWLS ($U = 20^*$largest loss)) under the lognormal distribution assumption. The right plot shows the box-plot of bootstrapped distribution of 99.9% capital estimates by the three approaches.

practice, the upper bound may be selected by risk managers by considering the scenario analysis data and external environment.

## 6. Conclusion

The instability of operational risk estimates by MLE is a significant issue puzzling many risk modelers in financial institutions. While the research on alternative approaches for estimating the operational risk capital has become significantly diverse (some are too complicated to be understood by risk managers), in this paper we focus on the central modeling issue of fitting truncated distributions and estimating extreme quantiles. We analyze the causes of the instability from model assumption, estimation accuracy and robustness, and review the performance of recently proposed approaches for rectifying each. Based on comparisons from our simulation results and application to real world data, we find that the right-truncation formulation with probability-weighted least squares estimation method that we introduce in the paper is not only helpful for reducing the uncertainty of capital and expected loss estimates but also for improving the robustness of estimates to small data change. This approach also has the advantage that specification of its input will be very intuitive to risk managers so that it is generally applicable in practice as opposed to many existing methods which require many inputs that are hardly understandable by risk managers.

Finally, we provide connections from the central modeling issue we discussed in this paper to some recent research work on alternative frameworks other than LDA. There have been active discussions on whether advanced measurement approaches (typically the loss distribution approach) shall be continued or not for modeling operational risk given its significant instability and lack of forward-looking information. For example, the Basel Committee has proposed approaches to calculate the operational risk regulatory capital in a simpler way, such as the revised standardized approach (RSA, see BCBS (2014)) based on a business indicator and the standardized measurement approach (SMA, see BCBS (2016)) which is mostly built upon the RSA but incorporates the business indicator component with a simple loss component adjustment. These methods could serve as a benchmark capital estimate at the enterprise level, but may not be helpful for institutions to understand the source of operational risk since they do not provide any insight about the risk profile in different types of loss events or in different business units. Rozenfeld (2014) provides a model averaging framework for operational risk since in the real world the loss generating mechanism can be very complex and may not be captured by a single model. Using model averaging may reduce the risk of model mis-specification. In addition, there has been work on developing more explanatory (factor-based) operational risk models such as Chavez-Demoulin et al. (2015) where the parameters of the loss frequency and severity distributions are explained by covariates. However, when exploring more complex approaches in estimating capital, it is always essential to address the fundamental problem – stability and robustness in estimating an extreme quantile of a distribution. The comparative study of various alternative approaches and proposed estimators on the central modeling issue in this paper should also shed some light on designing future capital estimators for these extensions of LDA.

## References

Alexander, C. (Ed.), 2003. Operational Risk: Regulation, Analysis and Management. Financial Times Prentice Hall, London.
Allen, L., Bali, T.G., 2007. Cyclicality in catastrophic and operational risk measurements. Journal of Banking & Finance 31 (4), 1191–1235.
Ames, M., Schuermann, T., Scott, H.S., 2014. Bank Capital for Operational Risk: A Tale of Fragility and Instability. Available at SSRN 2396046.
Aue, F., Kalkbrener, M., 2006. LDA at work: Deutsche Bank's approach to quantifying operational risk. Journal of Operational Risk 1 (4), 49–93.
Basel Committee on Banking Supervision, 2014. Operational Risk – Revisions to the Simpler Approaches. Available at <http://www.bis.org/publ/bcbs291.htm>.
Basel Committee on Banking Supervision, 2016. Standardised Measurement Approach for Operational Risk. Available at <http://www.bis.org/bcbs/publ/d355.htm>.
Böcker, K., Sprittulla, J., 2006. Operational VAR: meaningful means. Risk Magazine 12, 96–98.
Bolance, C., Ayuso, M., Guillen, M., 2012. A nonparametric approach to analyzing operational risk with an application to insurance fraud. Journal of Operational Risk 7 (1), 1–16.
Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. Springer-Verlag, New York.
Carrillo-Menéndez, S., Suárez, A., 2012. Robust quantification of the exposure to operational risk: bringing economic sense to economic capital. Computers & Operations Research 39 (4), 792–804.

Cavallo, A., Rosenthal, B., Wang, X., Yan, J., 2012. Treatment of the data collection threshold in operational risk: a case study with the lognormal distribution. Journal of Operational Risk 7 (1), 3–18.
Cavestany, R., Boultwood, B., Escudero, L., 2015. Operational Risk Capital Models. Risk Books, London.
Chavez-Demoulin, V., Embrechts, P., Nešlehová, J., 2006. Quantitative models for operational risk: extremes, dependence and aggregation. Journal of Banking & Finance 30 (10), 2635–2658.
Chavez-Demoulin, V., Embrechts, P., Hofert, M., 2015. An extreme value approach for modeling operational risk losses depending on covariates. Journal of Risk and Insurance.
Chernobai, A., Menn, C., Rachev, S.T., Truck, S., Moscadelli, M., 2006. Treatment of incomplete data in the field of operational risk: the effects on parameter estimates, EL and UL figures. In: The Advanced Measurement Approach to Operational Risk. Risk Books, London, pp. 145–168.
Chernobai, A.S., Rachev, S.T., Fabozzi, F.J., 2008. Operational Risk: A Guide to Basel II Capital Requirements, Models, and Analysis. John Wiley & Sons, Hoboken, NJ.
Chernobai, A., Rachev, S.T., Fabozzi, F.J., 2015. Composite goodness-of-fit tests for left-truncated loss samples. In: Lee, C.F., Lee, J.C. (Eds.), Handbook of Financial Econometrics and Statistics. Springer-Verlag, New York, pp. 575–596.
Chernobai, A., Jorion, P., Yu, F., 2012. The determinants of operational risk in US financial institutions. Journal of Financial and Quantitative Analysis 46 (6), 1683–1725.
Cope, E.W., Mignola, G., Antonini, G., Ugoccioni, R., 2009. Challenges and pitfalls in measuring operational risk from loss data. Journal of Operational Risk 4 (4), 3–27.
Cope, E., Labbi, A., 2008. Operational loss scaling by exposure indicators: evidence from the ORX database. Journal of Operational Risk 3 (4), 25–45.
Cope, E.W., 2011. Penalized likelihood estimators for truncated data. Journal of Statistical Planning and Inference 141 (1), 345–358.
Cope, E.W., Piche, M.T., Walter, J.S., 2012. Macroenvironmental determinants of operational loss severity. Journal of Banking & Finance 36 (5), 1362–1380.
Cruz, M.G., 2002. Modeling, Measuring and Hedging Operational Risk. John Wiley & Sons, New York.
Dahen, H., Dionne, G., 2010. Scaling models for the severity and frequency of external operational loss data. Journal of Banking & Finance 34 (7), 1484–1496.
Degen, M., 2010. The calculation of minimum regulatory capital using single-loss approximations. Journal of Operational Risk 5 (4), 1–15.
Degen, M., Embrechts, P., 2011. Scaling of high-quantile estimators. Journal of Applied Probability 48 (4), 968–983.
Embrechts, P., Frei, M., 2009. Panjer recursion versus FFT for compound distributions. Mathematical Methods of Operations Research 69 (3), 497–508.
Ergashev, B., 2008. Should Risk Managers Rely on Maximum Likelihood Estimation Method While Quantifying Operational Risk? Available at SSRN 1075249.
Ergashev, B., Pavlikov, K., Uryasev, S., Sekeris, E., 2012. Estimation of Truncated Data Samples in Operational Risk Modeling. Available at SSRN 2193493.
Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A., 1986. Robust Statistics. John Wiley & Sons, New York, NY.
Horbenko, N., Ruckdeschel, P., Bae, T., 2010. Robust Estimation of Operational Risk, arXiv preprint arXiv:1012.0249.
Huber, S., 2010. (Non-) robustness of maximum likelihood estimators for operational risk severity distributions. Quantitative Finance 10 (8), 871–882.
Hyndman, R.J., Fan, Y., 1996. Sample quantiles in statistical packages. The American Statistician 50 (4), 361–365.
James, W., Stein, C., 1961. Estimation with quadratic loss. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, pp. 361–379, No. 1961.
Kim, J.H.T., Hardy, M.R., 2007. Quantifying and correcting the bias in estimated risk measures. Astin Bulletin 37 (2), 365.
Ledoit, O., Wolf, M., 2004. A well-conditioned estimator for large-dimensional covariance matrices. Journal of Multivariate Analysis 88 (2), 365–411.
Lehérissé, V., Renaudin, A., 2013. Quantile distance estimation for operational risk: a practical application. Journal of Operational Risk 8 (2), 73–102.
Lehmann, E.L., 1951. A general concept of unbiasedness. Annals of Mathematical Statistics 22 (3), 587–592.
Luo, X., Shevchenko, P.V., Donnelly, J.B., 2009. Addressing the Impact of Data Truncation and Parameter Uncertainty on Operational Risk Estimates, arXiv preprint arXiv:0904.2910.
Menn, C., Rachev, S.T., 2009. Smoothly truncated stable distributions, GARCH-models, and option pricing. Mathematical Methods of Operations Research 69 (3), 411–438.
Mignola, G., Ugoccioni, R., 2006. Sources of uncertainty in modeling operational risk losses. Journal of Operational Risk 1 (2), 33–50.
Newman, M.C., 1993. Regression analysis of log-transformed data: statistical bias and its correction. Environmental Toxicology and Chemistry 12 (6), 1129–1133.
Nguyen, T., Samorodnitsky, G., 2012. Tail inference: where does the tail begin? Extremes 15 (4), 437–461.
Opdyke, J.D., Cavallo, A., 2012. Estimating operational risk capital: the challenges of truncation, the hazards of maximum likelihood estimation, and the promise of robust statistics. Journal of Operational Risk 7 (3), 3–90.
Opdyke, J.D., 2014. Estimating operational risk capital with greater accuracy, precision and robustness. Journal of Operational Risk 9 (4), 3–79.
Rozenfeld, I., 2010. Using Shifted Distributions in Computing Operational Risk Capital. Available at SSRN 1596268.
Rozenfeld, I., 2012. Estimating operational risk capital from truncated data with constrained optimization. In: Presented at ABA Operational Risk Modeling Forum. Washington DC, July 18–20, 2012.
Rozenfeld, I., 2014. Statistical model selection in operational risk. In: Presented at ABA Operational Risk Modeling Forum. Chicago, IL, Oct 14–16, 2014.
Shevchenko, P., 2010. Modelling Operational Risk Using Bayesian Inference. Springer Verlag, New York.
Snowdon, P., 1991. A ratio estimator for bias correction in logarithmic regressions. Canadian Journal of Forest Research 21 (5), 720–724.
Sprugel, D.G., 1983. Correcting for bias in log-transformed allometric equations. Ecology 64 (1), 209–210.
Vargas Mendoza, A., 2012. Essays on Financial Economics Doctoral dissertation. Massachusetts Institute of Technology.
Zhou, X., Giacometti, R., Fabozzi, F.J., Tucker, A.H., 2014. Bayesian estimation of truncated data with applications to operational risk measurement. Quantitative Finance 14 (5), 863–888.