

Model Uncertainty in Operational Risk Modeling Due to Data Truncation: A Single Risk Case

DAOPING YU¹

University of Central Missouri

VYTARAS BRAZAUSKAS²

University of Wisconsin-Milwaukee

To appear in *Risks*

(*Submitted:* April 25, 2017 *Revised:* August 15, 2017 *Accepted:* September 1, 2017)

Abstract. Over the last decade, researchers, practitioners, and regulators had intense debates about how to treat the data collection threshold in operational risk modeling. For fitting the loss severity distribution, several approaches have been employed: the empirical approach, the “naive” approach, the shifted approach, and the truncated approach. Since each approach is based on a different set of assumptions, different probability models emerge. Thus, model uncertainty arises. The main objective of this paper is to understand the impact of model uncertainty on the value-at-risk (VaR) estimators. To accomplish that, we take the bank’s perspective and study a single risk. Under this simplified scenario we can solve the problem analytically (when the underlying distribution is exponential) and show that it uncovers similar patterns among VaR estimates to those based on the simulation approach (when data follow a Lomax distribution). We demonstrate that for a fixed probability distribution, the choice of the truncated approach yields lowest VaR estimates, which may be viewed as beneficial to the bank, whilst the “naive” and shifted approaches lead to higher estimates of VaR. The advantages and disadvantages of each approach and the probability distributions under study are further investigated using a real data set for legal losses in a business unit (Cruz, 2002).

Keywords & Phrases: Asymptotics; data truncation; delta method; model validation; operational risk; VaR estimation.

¹ Daoping Yu, Ph.D., is an Assistant Professor in the School of Computer Science and Mathematics, University of Central Missouri, Warrensburg, MO 64093, USA. *e-mail:* dyu@ucmo.edu

² CORRESPONDING AUTHOR: Vytautas Brazauskas, Ph.D., ASA, is a Professor in the Department of Mathematical Sciences, University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, WI 53201, USA. *e-mail:* vytaras@uwm.edu

1 Introduction

Basel II/III and *Solvency II*—the leading international regulatory frameworks for banking and insurance industries—mandate that financial institutions build separate capital reserves for operational risk. The Loss Distribution Approach or LDA, within the Advanced Measurement Approach (AMA) framework, is the most sophisticated tool for estimating the operational risk capital. According to LDA, the risk-based capital is an extreme quantile of the annual aggregate loss distribution (e.g., the 99.9th percentile), which is called value-at-risk or VaR. Some recent discussions between the industry and the regulatory community in the United States reveal that the LDA implementation still has a number of “thorny” issues (AMA Group, 2013). One such issue is the treatment of data collection threshold. Here is what is stated on page 3 of the same document: “Although the industry generally accepts the existence of operational losses below the data collection threshold, the appropriate treatment of such losses in the context of capital estimation is still widely debated.”

Various assumptions about the data collection threshold have been considered in the existing literature: known threshold (Baud, Frachot, Roncalli, 2002, and Shevchenko, Temnov, 2009), threshold as unknown parameter (Baud, Frachot, Roncalli, 2002), stochastic threshold whose distribution has to be modeled (Baud, Frachot, Roncalli, 2002, and de Fontnouvelle, DeJesus-Rueff, Jordan, Rosengren, 2006), and time varying threshold that may scale according to inflation and business factors (Shevchenko, Temnov, 2009). In this paper, we will assume the threshold is known. Given (external) operational risk databases, which often collect losses exceeding (for example) \$1 million, such an assumption is appropriate.

Further, the annual aggregate loss variable is a combination of two variables—loss frequency and loss severity—and there are different ways to estimate risk-based capital. One way, for example, is to estimate the untruncated severity and truncation-adjusted frequency and then compute VaR. This approach follows directly from the results described by Brazauskas, Jones, Zitikis (2015). Another way is to estimate the truncated severity and unadjusted frequency to compute VaR. For a comprehensive review of analytic techniques for truncated data in the context of operational risk modeling, see Cruz, Peters, Shevchenko (2015, Section 7.9). Furthermore, as is known in practice, the severity distribution

is a key driver of the capital estimate (Opdyke, 2014). And this is the part of the aggregate model where initial assumptions about the data collection threshold are most influential. A number of authors have examined some aspects of this topic in the past (see, e.g., Moscadelli, Chernobai, Rachev, 2007, Chernobai, Rachev, Fabozzi, 2007, Luo, Schevchenko, Donnelly, 2007, Cavallo, Rosenthal, Wang, Yan, 2012, and Ergashev, Pavlikov, Uryasev, Sekeris, 2016). The modeling approaches they (collectively) considered include: the empirical approach, the “naive” approach, the shifted approach, and the truncated approach. Since each approach is based on a different set of assumptions, different probability models emerge. Thus, model uncertainty arises.

The main objective of this paper is to understand the impact of model uncertainty on risk measurements, and (hopefully) help settle the debate about the treatment of data collection threshold in the context of capital estimation. Solving such a problem under a general setup, i.e., by considering many interdependent risks and multiple stakeholders, is only possible through extensive simulations, but that would not produce much insight. Therefore, we simplify the problem by taking the bank’s perspective and by studying a single risk. Under this simplified scenario we can solve the problem analytically (when the underlying distribution is exponential) and show that it uncovers similar patterns among VaR estimates to those based on the simulation approach (when data follow a Lomax distribution). We demonstrate that for a fixed probability distribution, the choice of the truncated approach yields lowest VaR estimates, which may be viewed as beneficial to the bank, whilst the “naive” and shifted approaches lead to higher estimates of VaR. As for the choice of severity distributions, besides the Lomax distribution which is heavy tailed and hence appropriate in operational risk modeling, we intentionally select the light-tailed exponential distribution to show what happens to VaR estimates when wrong assumptions are made. Moreover, our step-by-step analysis not only shows “what happens” to VaR estimates but it helps understand the questions of “how” and “why” it happens. Also, perhaps surprisingly, our numerical illustrations reveal why the shifted approach is still popular. That is because it is flexible enough to pass standard model validation tests and thus cannot be discarded from practical use based on such tools alone. In summary, this paper contributes to the existing literature by performing an extensive investigation of the impact that model uncertainty has on the VaR estimators, justifies the soundness of the regulatory recommendation (i.e., use the

truncated approach), and paves the way for a number of research problems in this important area.

It is worthwhile to note here that the model uncertainty considered in this paper is an epistemic one, not a random uncertainty. It can be reduced, but not completely eliminated, by employing sound model validation tools and in some cases (e.g., when the shifted approach is used) may require out-of-model knowledge. In a more general context, model uncertainty is an important topic within the model risk governance framework as regulated by the OCC and the Federal Reserve Bank in the U.S. and the Basel Committee on Banking Supervision for the G20 countries (see, e.g., Office of the Comptroller of the Currency, 2011, and Basel Coordination Committee, 2014).

The rest of the paper is structured as follows. In Section 2, we describe how model uncertainty emerges and study its effects on VaR estimates. This is done by employing theoretical results (presented in Appendix) and via Monte Carlo simulations. Next, in Section 3, these explorations are further illustrated using a real data set for legal losses in a business unit. Finally, concluding remarks are offered in Section 4. Also, in Appendix, we provide some technical tools that are essential for analytic treatment of the problem. In particular, key probabilistic features of the generalized Pareto distribution are presented and several asymptotic theorems of mathematical statistics are specified.

2 Model Uncertainty

We start this section by introducing the problem and describing how model uncertainty arises. Then, in Section 2.2, we review several typical models used for estimating VaR. Finally, using the theoretical results of Appendix and Monte Carlo simulations, we finish with two parametric examples, where we evaluate the probability of overestimating true VaR for exponential and Lomax distributions.

2.1 Motivation

In order to fully understand the problem, in this paper we will walk the reader through the entire modeling process and demonstrate how our assumptions affect the end product, which is the estimate of severity VaR. Since the problem involves collected data, initial assumptions, and statistical inference (in this case, point estimation and assessment of estimates' variability), it will be tackled with statistical tools, including theoretical tools (asymptotics), Monte Carlo simulations, and real-data case studies.

Let us briefly discuss data, assumptions, and inference. As noted in Section 1, it is generally agreed that operational losses exist above and below the data collection threshold. Therefore, this implies that choosing a modeling approach is equivalent to deciding on how much probability mass there is below the threshold.

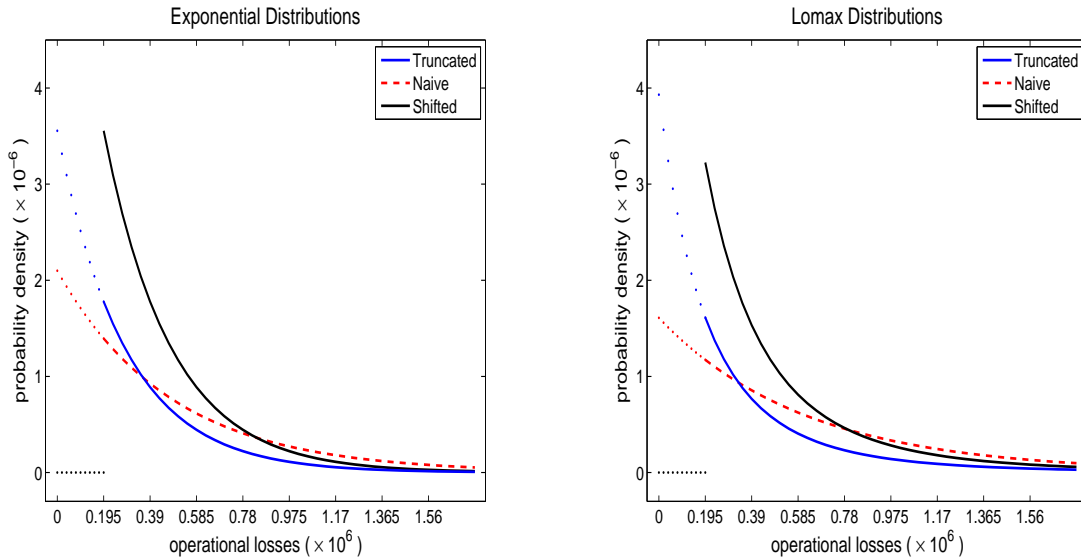


FIGURE 1: Truncated, naive, shifted *Exponential* (σ) and *Lomax* ($\alpha = 3.5, \theta_1$) probability density functions. Data collection threshold $t = 195,000$, with 50% of data unobserved. Parameters σ and θ_1 are chosen to match those in Tables 2 and 3 (see Section 2.3).

In Figure 1, we provide graphs of truncated, naive, and shifted probability density functions of two distributions (studied formally in Section 2.3): *Exponential* which is a light-tailed model, and *Lomax*, with the tail parameter $\alpha = 3.5$, which is a moderately-tailed model (it has three finite moments). We clearly see that those models are quite different below the threshold $t = 195,000$, but in practice that would be unobserved. On the other hand, in the observable range, i.e., above $t = 195,000$, the plotted density functions are similar (note that the vertical axes are in very small units, 10^{-6}) and converge to each other as losses get larger (note how little differentiation there is among the curves when losses exceed 1,000,000). Moreover, it is even hard to spot a difference between the corresponding exponential and Lomax models, though the two distributions possess distinct theoretical properties (e.g., for one all moments are finite, whereas for the other only three). Also, since probability mass below the threshold is one of “known unknowns,” it will have to be estimated from the observed data

(above t). As will be shown in the case study of Section 3, this task may look straightforward, but its outcomes vary and are heavily influenced by the initial assumptions.

To formalize this discussion, suppose that Y_1, \dots, Y_N represent (positive and *i.i.d.*) loss severities resulting from operational risk, and let us denote their pdf, cdf, and qf as f , F , and F^{-1} , respectively. Then, the problem of estimating VaR-based capital is equivalent to finding an estimate of qf at some probability level, say $F^{-1}(\beta)$. The difficulty here is that we observe only those Y_i 's that exceed some known data collection threshold $t \gg 0$. That is, the actually observed variables are X_i 's with

$$X_1 \stackrel{d}{=} Y_{i_1} \mid Y_{i_1} > t, \quad \dots, \quad X_n \stackrel{d}{=} Y_{i_n} \mid Y_{i_n} > t, \quad (2.1)$$

where $\stackrel{d}{=}$ denotes “equal in probability” and $n = \sum_{j=1}^N \mathbf{1}\{Y_j > t\}$. Their cdf F_* , pdf f_* , qf F_*^{-1} are related to F , f , F^{-1} and given by

$$F_*(x) = \frac{F(x) - F(t)}{1 - F(t)}, \quad f_*(x) = \frac{f(x)}{1 - F(t)}, \quad F_*^{-1}(u) = F^{-1}(u + (1 - u)F(t)) \quad (2.2)$$

for $x \geq t$ and $0 < u < 1$, and for $x < t$, $f_*(x) = F_*(x) = 0$.

Further, let us investigate the behavior of $F_*^{-1}(u)$ from a purely mathematical point of view. Since the qf of continuous random variables (which is the case for loss severities) is a strictly increasing function and $(1 - u)F(t) \geq 0$, it follows that

$$F_*^{-1}(u) = F^{-1}(u + (1 - u)F(t)) \geq F^{-1}(u), \quad 0 < u < 1,$$

with the inequality being strict unless $F(t) = 0$. This implies that any quantile of the observable variable X is never below the corresponding quantile of the unobservable variable Y , which is true VaR. This fact is certainly not new (see, e.g., an extensive analysis by Opdyke, 2014, about the effect of Jensen’s inequality in VaR estimation). However, if we now change our perspective from mathematical to statistical and take into account the method of how VaR is estimated, we could augment the above discussion with new insights and improve our understanding.

A review of existing methods shows that, besides estimation of VaR using (2.1) and (2.2) under the truncated distribution framework, there are other parametric methods that employ different strategies, such as the naive and shifted approaches (described in Section 2.2.2). In particular, those two approaches use the data X_1, \dots, X_n and either ignore t or recognize it in some other way than (2.2). Thus, model uncertainty emerges.

2.2 Typical Models

2.2.1 Empirical Model

As mentioned earlier, the empirical model is restricted to the range of observed data. So it uses data from (2.1), but since the empirical estimator $\widehat{F}(t) = 0$, formulas (2.2) simplify to $\widehat{F}_*(x) = \widehat{F}(x)$, $\widehat{f}_*(x) = \widehat{f}(x)$, for $x \geq t$, and $\widehat{F}_*^{-1}(u) = \widehat{F}^{-1}(u)$. Thus, the model cannot take full advantage of (2.2). In this case, the VaR(β) estimator is simply $\widehat{F}^{-1}(\beta) = X_{(\lceil n\beta \rceil)}$, and as follows from Theorem A.1,

$$X_{(\lceil n\beta \rceil)} \text{ is } \mathcal{AN} \left(F_*^{-1}(\beta), \frac{1}{n} \frac{\beta(1-\beta)}{f_*^2(F_*^{-1}(\beta))} \right).$$

We now can evaluate the probability of overestimating true VaR by certain percentage, i.e., we want to study function $H(c) := \mathbf{P}\{X_{(\lceil n\beta \rceil)} > cF^{-1}(\beta)\}$ for $c \geq 1$. Using Z to denote the standard normal random variable and Φ for its cdf, and taking into account (2.2), we proceed as follows:

$$\begin{aligned} H(c) &= \mathbf{P}\{X_{(\lceil n\beta \rceil)} > cF^{-1}(\beta)\} \approx \mathbf{P} \left\{ Z > \left[cF^{-1}(\beta) - F_*^{-1}(\beta) \right] \times \left(\frac{1}{n} \frac{\beta(1-\beta)}{f_*^2(F_*^{-1}(\beta))} \right)^{-1/2} \right\} \\ &= 1 - \Phi \left(\sqrt{\frac{n}{\beta(1-\beta)}} \left[cF^{-1}(\beta) - F^{-1}(\beta + (1-\beta)F(t)) \right] \times \frac{f(F^{-1}(\beta + (1-\beta)F(t)))}{1-F(t)} \right). \end{aligned}$$

From this formula we clearly see that $0.50 \leq H(1) < 1$ with the lower bound being achieved when $F(t) = 0$. Also, at the other extreme, when $c \rightarrow \infty$, we observe $H(c) \rightarrow 0$. Additional numerical illustrations are provided in Table 1.

Several conclusions emerge from the table. First, the case $F(t) = 0$ is a benchmark case that illustrates the behavior of the empirical estimator when data is completely observed (and in that case $X_{(\lceil n\beta \rceil)}$ would be a consistent method for estimating VaR(β)). We see that $H(1) = 0.5$ and then it quickly decreases to 0 as c increases. The decrease is quickest for the light-tailed distribution, exponential($\sigma = 1$), and slowest for the heavy-tailed Lomax($\alpha = 1, \theta_2 = 1$) which has no finite moments. Second, as less data is observed, i.e., as $F(t)$ increases to 0.5 and 0.9, the probability of overestimating true VaR increases for all types of distributions. For example, while the probability of overestimating VaR(0.995) by 20% ($c = 1.2$) for the light-tailed distribution is only 0.226 for $F(t) = 0$, it increases to 0.398 and 0.811 for $F(t) = 0.5$ and 0.9, respectively. If severity follows the heavy-tailed distribution, then $H(1.2)$ is 0.444, 0.612, 0.734 for $F(t) = 0, 0.5, 0.9$, respectively. Finally, in practice,

typical scenarios would be near $F(t) = 0.9$ with moderate- or heavy-tailed severity distributions, which corresponds to quite unfavorable patterns in the table. Indeed, function $H(c)$ declines very slowly and the probability of overestimating $\text{VaR}(0.995)$ by 100% seems like a norm (0.577 and 0.715).

TABLE 1: Function $H(c)$ evaluated for various combinations of c , confidence level β , proportion of unobserved data $F(t)$, and severity distributions with varying degrees of tail heaviness ranging from light- and moderate-tailed to heavy-tailed. (The sample size is $n = 100$.)

c	β	$F(t) = 0$			$F(t) = 0.5$			$F(t) = 0.9$		
		<i>Light</i>	<i>Moderate</i>	<i>Heavy</i>	<i>Light</i>	<i>Moderate</i>	<i>Heavy</i>	<i>Light</i>	<i>Moderate</i>	<i>Heavy</i>
1	0.95	0.500	0.500	0.500	0.944	0.925	0.874	1.000	1.000	0.981
	0.995	0.500	0.500	0.500	0.688	0.672	0.638	0.949	0.884	0.738
	0.999	0.500	0.500	0.500	0.587	0.579	0.563	0.767	0.703	0.612
1.2	0.95	0.085	0.178	0.331	0.585	0.753	0.824	1.000	1.000	0.978
	0.995	0.226	0.349	0.444	0.398	0.551	0.612	0.811	0.840	0.734
	0.999	0.331	0.424	0.475	0.414	0.517	0.550	0.615	0.668	0.610
1.5	0.95	0.000	0.010	0.138	0.032	0.326	0.726	0.968	0.996	0.975
	0.995	0.030	0.167	0.362	0.083	0.364	0.571	0.403	0.756	0.727
	0.999	0.137	0.317	0.437	0.191	0.424	0.532	0.358	0.613	0.606
2	0.95	0.000	0.000	0.015	0.000	0.009	0.523	0.056	0.930	0.968
	0.995	0.000	0.026	0.240	0.001	0.127	0.501	0.017	0.577	0.715
	0.999	0.014	0.170	0.376	0.025	0.280	0.500	0.073	0.516	0.600

NOTE: Threshold t is 0 for $F(t) = 0$ and 195,000 for $F(t) = 0.5, 0.9$. Distributions: *Light* = exponential(σ), *Moderate* = Lomax($\alpha = 3.5, \theta_1$), *Heavy* = Lomax($\alpha = 1, \theta_2$). For $F(t) = 0$: $\sigma = \theta_1 = \theta_2 = 1$. For $F(t) = 0.5$: $\sigma = 281,326, \theta_1 = 890,355, \theta_2 = 195,000$. For $F(t) = 0.9$: $\sigma = 84,687, \theta_1 = 209,520, \theta_2 = 21,667$.

2.2.2 Parametric Models

We discuss three parametric approaches: truncated, naive, and shifted.

Truncated Approach: The truncated approach uses the observed data X_1, \dots, X_n and fully recognizes its distributional properties. That is, it takes into account (2.2) and derives MLE values by maximizing the following log-likelihood function:

$$\log \mathcal{L}_T(\theta_1, \dots, \theta_k | X_1, \dots, X_n) = \sum_{i=1}^n \log f_*(X_i) = \sum_{i=1}^n \log \left(\frac{f(X_i)}{1 - F(t)} \right), \quad (2.3)$$

where $\theta_1, \dots, \theta_k$ are the parameters of pdf f . Once parameter MLEs are available, $\text{VaR}(\beta)$ estimate is found by plugging those MLE values into $F^{-1}(\beta)$. \square

Naive Approach: The naive approach uses the observed data X_1, \dots, X_n , but ignores the presence of threshold t . That is, it bypasses (2.2) and derives MLE values by maximizing the following log-likelihood function:

$$\log \mathcal{L}_N(\theta_1, \dots, \theta_k | X_1, \dots, X_n) = \sum_{i=1}^n \log f(X_i). \quad (2.4)$$

Notice that, since $f(X_i) \leq f(X_i)/[1 - F(t)] = f_*(X_i)$ with the inequality being strict for $F(t) > 0$, the log-likelihood of the naive approach will always be less than that of the truncated approach. This in turn implies that parameter MLEs of pdf f derived using the naive approach will always be suboptimal, unless $F(t) = 0$. Finally, $\text{VaR}(\beta)$ estimate is computed by inserting parameter MLEs (the ones found using the naive approach) into $F^{-1}(\beta)$. \square

Shifted Approach: The shifted approach uses the observed data X_1, \dots, X_n and recognizes threshold t by first shifting the observations by t . Then, it derives parameter MLEs by maximizing the following log-likelihood function:

$$\log \mathcal{L}_S(\theta_1, \dots, \theta_k | X_1, \dots, X_n) = \sum_{i=1}^n \log f(X_i - t). \quad (2.5)$$

By comparing (2.4) and (2.5), we can easily see that the naive approach is a special case of the shifted approach (with $t = 0$). Moreover, although this may only be of interest to theoreticians, one could introduce a class of shifted models by considering $f(X_i - s)$, with $0 \leq s \leq t$, and create infinitely many versions of the shifted model. Finally, $\text{VaR}(\beta)$ is estimated by applying parameter MLEs (the ones found using the shifted approach) to $F^{-1}(\beta) + t$. \square

2.3 Parametric VaR Estimation

2.3.1 Example 1: Exponential Distribution

Suppose Y_1, \dots, Y_N are *i.i.d.* and follow an exponential distribution, with pdf, cdf, and qf given by (A.2), (A.1), and (A.4), respectively, with $\gamma = 0$ and $\mu = 0$. However, we observe only variable X whose relation to Y is governed by (2.1) and (2.2). Now, by plugging exponential pdf and/or cdf into

the log-likelihoods (2.3), (2.4), (2.5), we obtain

$$\begin{aligned}\log \mathcal{L}_T(\sigma | X_1, \dots, X_n) &= \sum_{i=1}^n \log \left(\frac{f(X_i)}{1 - F(t)} \right) = \sum_{i=1}^n \log \left(\frac{\sigma^{-1} e^{-X_i/\sigma}}{e^{-t/\sigma}} \right) \\ &= -n \log \sigma + \sum_{i=1}^n \frac{-(X_i - t)}{\sigma},\end{aligned}\tag{2.6}$$

$$\begin{aligned}\log \mathcal{L}_N(\sigma | X_1, \dots, X_n) &= \sum_{i=1}^n \log f(X_i) = \sum_{i=1}^n \log \left(\sigma^{-1} e^{-X_i/\sigma} \right) \\ &= -n \log \sigma + \sum_{i=1}^n \frac{-X_i}{\sigma},\end{aligned}\tag{2.7}$$

$$\begin{aligned}\log \mathcal{L}_S(\sigma | X_1, \dots, X_n) &= \sum_{i=1}^n \log f(X_i - t) = \sum_{i=1}^n \log \left(\sigma^{-1} e^{-(X_i - t)/\sigma} \right) \\ &= -n \log \sigma + \sum_{i=1}^n \frac{-(X_i - t)}{\sigma},\end{aligned}\tag{2.8}$$

where the subscripts T , N , S (for \mathcal{L}) denote “truncated”, “naive”, “shifted”, respectively. Then, by maximizing the log-likelihoods (2.6), (2.7), and (2.8) with respect to σ , we get the following MLE formulas for parameter σ under the truncated, naive, and shifted approaches:

$$\hat{\sigma}_T = \bar{X} - t, \quad \hat{\sigma}_N = \bar{X}, \quad \hat{\sigma}_S = \bar{X} - t,$$

where $\bar{X} = n^{-1} \sum_{i=1}^n X_i$.

Next, by inserting $\hat{\sigma}_T$, $\hat{\sigma}_N$, and $\hat{\sigma}_S$ into the corresponding qf’s as described in Section 2.2.2, we get the following $\text{VaR}(\beta)$ estimators:

$$\widehat{\text{VaR}}_T(\beta) = -\hat{\sigma}_T \log(1 - \beta), \quad \widehat{\text{VaR}}_N(\beta) = -\hat{\sigma}_N \log(1 - \beta), \quad \widehat{\text{VaR}}_S(\beta) = -\hat{\sigma}_S \log(1 - \beta) + t.$$

Further, a direct application of Theorem A.2 for $\hat{\sigma}_T$ (with obvious adjustment for $\hat{\sigma}_N$), yields that

$$\hat{\sigma}_T \text{ is } \mathcal{N} \left(\sigma, \frac{\sigma^2}{n} \right), \quad \hat{\sigma}_N \text{ is } \mathcal{N} \left(\sigma + t, \frac{\sigma^2}{n} \right), \quad \hat{\sigma}_S \text{ is } \mathcal{N} \left(\sigma, \frac{\sigma^2}{n} \right).$$

Furthermore, having established \mathcal{N} for parameter MLEs, we can apply Theorem A.3 and specify asymptotic distributions for VaR estimators. They are as follows:

$$\widehat{\text{VaR}}_T(\beta) \text{ is } \mathcal{N} \left(-\sigma \log(1 - \beta), \frac{\sigma^2 \log^2(1 - \beta)}{n} \right),$$

$$\widehat{\text{VaR}}_N(\beta) \text{ is } \mathcal{N}\left(-(\sigma + t) \log(1 - \beta), \frac{\sigma^2 \log^2(1 - \beta)}{n}\right),$$

$$\widehat{\text{VaR}}_S(\beta) \text{ is } \mathcal{N}\left(-\sigma \log(1 - \beta) + t, \frac{\sigma^2 \log^2(1 - \beta)}{n}\right).$$

Note that while all three estimators are equivalent in terms of the asymptotic variance, they are centered around different targets. The mean of the truncated estimator is the true quantile of the underlying exponential model (estimating which is the objective of this exercise) and the mean of the other two methods is shifted upwards; in both cases, the shift is a function of threshold t .

Finally, as it was done for the empirical VaR estimator in Section 2.2.1, we now define function $H(c) = \mathbf{P}\{\widehat{\text{VaR}}(\beta) > cF^{-1}(\beta)\}$ for $c \geq 1$, the probability of overestimating the target by $(c - 1)100\%$, for each parametric VaR estimator and study its behavior:

$$H_T(c) \approx 1 - \Phi((c - 1)\sqrt{n}), \quad H_N(c) \approx 1 - \Phi((c - 1)\sqrt{n} - \sqrt{n}(t/\sigma)),$$

$$H_S(c) \approx 1 - \Phi((c - 1)\sqrt{n} + \sqrt{n}(t/\sigma) \log^{-1}(1 - \beta)).$$

Table 2 provides numerical illustrations of functions $H_T(c)$, $H_N(c)$, $H_S(c)$. We select the same parameter values as in the light-tailed cases of Table 1. From Table 2, we see that the case $F(t) = 0$ is special in the sense that all three methods become identical and perform well. For example, the probability of overestimating true VaR by 20% is only 0.023 for all three methods and it is essentially 0 as $c \geq 1.5$. Parametric estimators in this case outperform the empirical estimator (see Table 1) because they are designed for the correct underlying model. However, as proportion of unobserved data increases, i.e., as $F(t)$ increases to 0.5 and 0.9, only the truncated approach maintains its excellent performance. And while the shifted estimator is better than the naive, both methods perform poorly and even rarely improve the empirical estimator. For example, in the extreme case of $F(t) = 0.9$, the naive and shifted methods overestimate true $\text{VaR}(0.95)$ by 50% with probability 1.000 and 0.996, respectively, whereas the corresponding probability for the empirical estimator is 0.968.

TABLE 2: Functions $H_T(c)$, $H_N(c)$, $H_S(c)$ evaluated for various combinations of c , confidence level β , and proportion of unobserved data $F(t)$. (The sample size is $n = 100$.)

c	β	$F(t) = 0$			$F(t) = 0.5$			$F(t) = 0.9$		
		T	N	S	T	N	S	T	N	S
1	0.95	0.500	0.500	0.500	0.500	1.000	0.990	0.500	1.000	1.000
	0.995	0.500	0.500	0.500	0.500	1.000	0.905	0.500	1.000	1.000
	0.999	0.500	0.500	0.500	0.500	1.000	0.842	0.500	1.000	1.000
1.2	0.95	0.023	0.023	0.023	0.023	1.000	0.623	0.023	1.000	1.000
	0.995	0.023	0.023	0.023	0.023	1.000	0.245	0.023	1.000	0.991
	0.999	0.023	0.023	0.023	0.023	1.000	0.159	0.023	1.000	0.909
1.5	0.95	0.000	0.000	0.000	0.000	0.973	0.004	0.000	1.000	0.996
	0.995	0.000	0.000	0.000	0.000	0.973	0.000	0.000	1.000	0.257
	0.999	0.000	0.000	0.000	0.000	0.973	0.000	0.000	1.000	0.048
2	0.95	0.000	0.000	0.000	0.000	0.001	0.000	0.000	1.000	0.010
	0.995	0.000	0.000	0.000	0.000	0.001	0.000	0.000	1.000	0.000
	0.999	0.000	0.000	0.000	0.000	0.001	0.000	0.000	1.000	0.000

NOTE: Threshold t is 0 for $F(t) = 0$ and 195,000 for $F(t) = 0.5, 0.9$. Exponential(σ), with $\sigma = 1$ (for $F(t) = 0$), $\sigma = 281,326$ (for $F(t) = 0.5$), $\sigma = 84,687$ (for $F(t) = 0.9$).

2.3.2 Example 2: Lomax Distribution

Suppose Y_1, \dots, Y_N are *i.i.d.* and follow a Lomax distribution, with pdf, cdf, and qf given by (A.2), (A.1), and (A.4), respectively, with $\alpha = 1/\gamma$, $\theta = \sigma/\gamma$, and $\mu = 0$. However, we observe only variable X whose relation to Y is governed by (2.1) and (2.2). Now, unlike the exponential case, maximization of the log-likelihoods (2.3), (2.4), and (2.5) does not yield explicit formulas for MLEs of a Lomax model. So, in order to evaluate functions $H_T(c)$, $H_N(c)$, $H_S(c)$, we use Monte Carlo simulations to implement the following procedure: (i) generate Lomax-distributed data set according to pre-specified parameters, (ii) numerically evaluate parameters α and θ for each approach, (iii) compute the corresponding estimates of VaR, (iv) check whether the inequality in function $H(c)$ is true for each approach and record the outcomes, and (v) repeat steps (i)–(iv) a large number of times and report the proportion of “true” outcomes in step (iv). To facilitate comparisons with the moderate-tailed scenarios in Table 1, we select simulation parameters as follows:

- Severity distribution Lomax($\alpha = 3.5, \theta_1$): $\theta_1 = 1$ (for $F(t) = 0$), $\theta_1 = 890,355$ (for $F(t) = 0.5$), $\theta_1 = 209,520$ (for $F(t) = 0.9$).

- Threshold: $t = 0$ (for $F(t) = 0$) and $t = 195,000$ (for $F(t) = 0.5, 0.9$).
- Complete sample size: $N = 100$ (for $F(t) = 0$); $N = 200$ (for $F(t) = 0.5$); $N = 1000$ (for $F(t) = 0.9$). The *average* observed sample size is $n = 100$.
- Number of simulation runs: 10,000.

Simulation results are summarized in Table 3, where we again observe similar patterns to those of Tables 1 and 2. This time, however, the entries are more volatile, which is mostly due to the randomness of simulation experiment (e.g., all entries for the T and $c = 1$ cases theoretically should be equal to 0.5, because those cases correspond to the probability of a normal random variable exceeding its mean, but they are slightly off). The $F(t) = 0$ case is where all parametric models perform well, as they should. However, once they leave that comfort zone ($F(t) = 0.5$ and 0.9), only the truncated approach works well, with the naive and shifted estimators performing similarly to the empirical estimator. Since Lomax distributions have heavier tails than exponential, function $H(c)$ under the truncated approach is also affected by that and converges to 0 (as $c \rightarrow \infty$) slower. In other words, for a given choice of model parameters, the coefficient of variation of VaR is larger for the Lomax model than that for the exponential model, thus resulting in larger overestimating probabilities than those in Table 2. The difference between the T entries in Tables 2 and 3 is also influenced by the fact that the numerically found MLE does not often produce very stable or say trustworthy parameter estimates for truncated approach, which is a common technical issue. Nonetheless, the overall message here does not change: we observe certain patterns among functions $H_T(c)$, $H_N(c)$, and $H_S(c)$, which are no different from those of Section 2.3.1 that were found using the theoretical tools.

TABLE 3: Functions $H_T(c)$, $H_N(c)$, $H_S(c)$ evaluated for various combinations of c , confidence level β , and proportion of unobserved data $F(t)$. (The *average* sample size is $n = 100$.)

c	β	$F(t) = 0$			$F(t) = 0.5$			$F(t) = 0.9$		
		T	N	S	T	N	S	T	N	S
1	0.95	0.453	0.453	0.453	0.459	0.951	0.982	0.547	0.908	1.000
	0.995	0.433	0.433	0.433	0.435	0.692	0.734	0.444	0.891	0.998
	0.999	0.426	0.426	0.426	0.437	0.149	0.624	0.331	0.867	0.944
1.2	0.95	0.131	0.131	0.131	0.095	0.945	0.791	0.356	0.904	0.999
	0.995	0.247	0.247	0.247	0.184	0.208	0.518	0.170	0.889	0.993
	0.999	0.297	0.297	0.297	0.272	0.059	0.484	0.121	0.845	0.864
1.5	0.95	0.009	0.009	0.009	0.002	0.626	0.270	0.112	0.879	0.998
	0.995	0.097	0.097	0.097	0.044	0.044	0.278	0.021	0.875	0.872
	0.999	0.178	0.178	0.178	0.123	0.016	0.313	0.019	0.843	0.708
2	0.95	0.000	0.000	0.000	0.000	0.032	0.010	0.002	0.865	0.984
	0.995	0.025	0.025	0.025	0.004	0.004	0.090	0.000	0.851	0.563
	0.999	0.075	0.075	0.075	0.032	0.002	0.147	0.001	0.224	0.459

NOTE: Threshold t is 0 for $F(t) = 0$ and 195,000 for $F(t) = 0.5, 0.9$. Lomax($\alpha = 3.5, \theta_1$), with $\theta_1 = 1$ (for $F(t) = 0$), $\theta_1 = 890,355$ (for $F(t) = 0.5$), $\theta_1 = 209,520$ (for $F(t) = 0.9$).

3 Real-Data Example

In this section we illustrate how all the modeling approaches considered in this paper (empirical and three parametric) perform on real data. We go step-by-step through the entire modeling process, starting with model fitting and validation, continuing with VaR estimation, and completing the example with model-based predictions for quantities below the data collection threshold. Note that for the parametric approaches we employ both exponential and Lomax models, although exponential is clearly not a viable model for operational risk data (because its tail is too light for such data). However, the exponential distribution is a model for which all relevant formulas are explicit and can be easily verified by the reader. Moreover, the data analysis exercise also serves as an example on how to identify inappropriate models (e.g., exponential) and, if the model validation step is ignored, to illustrate how wrong the predictions based on such models can be.

3.1 Data

We will use the data set from Cruz (2002, p. 57), which has 75 observations and represents the cost of legal events for a business unit. The cost is measured in the U.S. dollars. To illustrate the impact of data collection threshold on the selected models, we split the data set into two parts: losses that are *at least* \$195,000, which will be treated as observed and used for model building and VaR estimation, and losses that are *below* \$195,000, which will be used at the end of the exercise to assess the quality of model-based predictions. This data-splitting scenario implies that there are 54 observed losses. A quick exploratory analysis of the observed data shows that it is right-skewed and potentially heavy-tailed, with the first quartile 248,342, median 355,000, and the third quartile 630,200; its mean is 546,021, standard deviation 602,912, and skewness 3.8.

3.2 Model Fitting

We fit two models to the observed data, exponential and Lomax, and use three parametric approaches, truncated, naive, and shifted. The truncation threshold is $t = 195,000$. For the exponential model, MLE formulas for σ are available in Section 2.3.1. For the Lomax distribution, we perform numerical maximization of the log-likelihoods (2.3), (2.4), and (2.5) to compute parameter values. For the data set under consideration, the resulting MLE values are reported in Table 4. Also, the corresponding estimates for parameter variances and covariances were computed using Theorem A.3.

TABLE 4: Parameter MLEs (with variance and covariance estimates in parentheses) of the exponential and Lomax models, using truncated, naive, and shifted approaches.

<i>Model</i>	<i>Truncated</i>	<i>Naive</i>	<i>Shifted</i>
Exponential	$\hat{\sigma} = 351,021 (2.28 \times 10^9)$	$\hat{\sigma} = 546,021 (5.52 \times 10^9)$	$\hat{\sigma} = 351,021 (2.28 \times 10^9)$
Lomax	$\hat{\alpha} = 1.91 (0.569)$ $\hat{\theta} = 151,234 (3.84 \times 10^{10})$ $(\widehat{cov}(\hat{\alpha}, \hat{\theta}) = 138,934)$	$\hat{\alpha} = 22.51 (5,189.86)$ $\hat{\theta} = 11,735,899 (1.54 \times 10^{15})$ $(\widehat{cov}(\hat{\alpha}, \hat{\theta}) = 2.82 \times 10^9)$	$\hat{\alpha} = 1.91 (0.569)$ $\hat{\theta} = 346,234 (3.84 \times 10^{10})$ $(\widehat{cov}(\hat{\alpha}, \hat{\theta}) = 138,934)$

3.3 Model Validation

To validate the fitted models we employ quantile-quantile plots (QQ-plots) and two goodness-of-fit statistics, Kolmogorov-Smirnov (KS) and Anderson-Darling (AD).

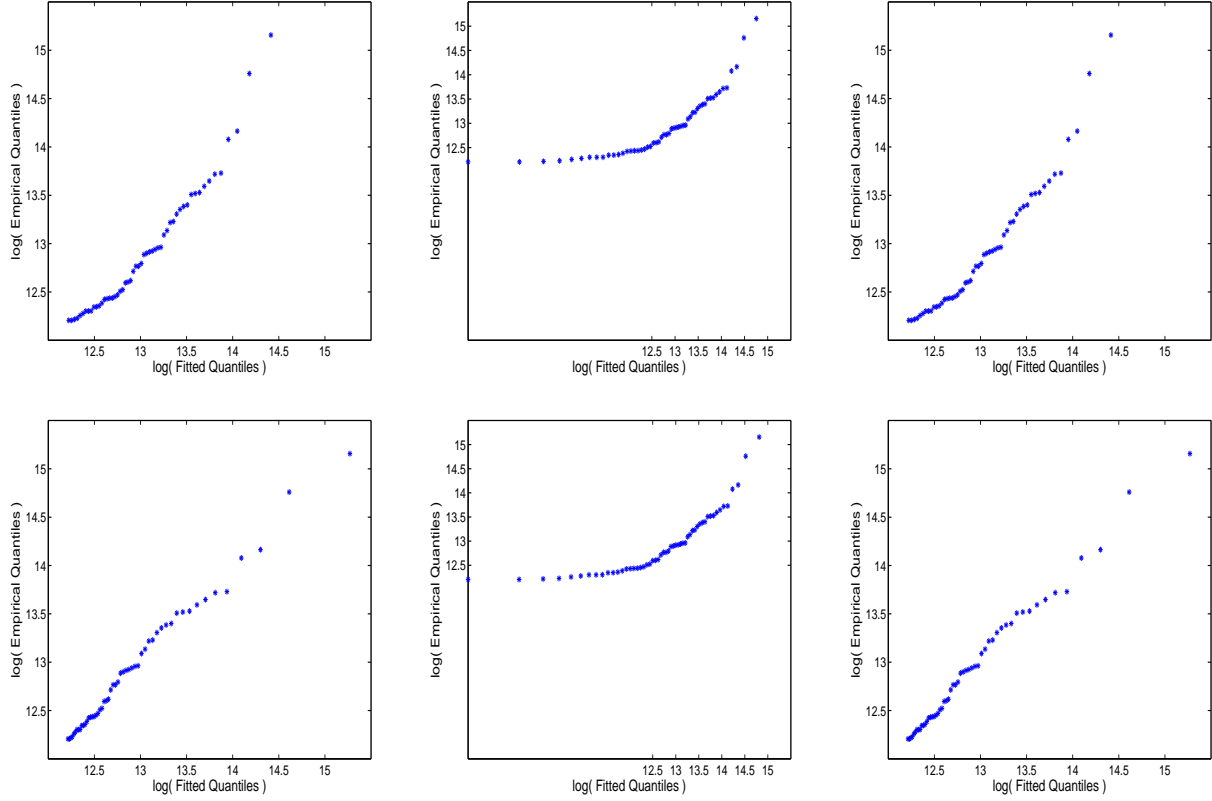


FIGURE 2: Fitted-versus-observed \log -losses for exponential (top row) and Lomax (bottom row) distributions, using truncated (left), naive (middle), and shifted (right) approaches.

In Figure 2, we present plots of the fitted-versus-observed quantiles for the six models of Section 3.2. In order to avoid visual distortions due to large spacings between the most extreme observations, both axes in all the plots are measured on the logarithmic scale. That is, the points plotted in those graphs are the following pairs:

$$\left(\log \left(\widehat{G}^{-1}(u_i) \right), \log \left(X_{(i)} \right) \right), \quad i = 1, \dots, 54,$$

where \widehat{G}^{-1} is the estimated parametric qf, $X_{(1)} \leq \dots \leq X_{(54)}$ denote the ordered losses, and $u_i = (i-0.5)/54$ is the quantile level. For the truncated approach, $\widehat{G}^{-1}(u_i) = \widehat{F}^{-1}(u_i + \widehat{F}(195,000)(1-u_i))$; for the naive approach, $\widehat{G}^{-1}(u_i) = \widehat{F}^{-1}(u_i)$; for the shifted approach, $\widehat{G}^{-1}(u_i) = \widehat{F}^{-1}(u_i) + 195,000$. Also, the corresponding cdf and qf functions were evaluated using the MLE values from Table 4.

We can see from Figure 2 that Lomax models show a better overall fit than exponential models, and especially in the extreme right tail. That is, most of the points in those plots do not deviate from

the 45° line. The naive approach seems off, but the truncated and shifted approaches do a reasonably good job for both distributions, with Lomax models exhibiting slightly better fits.

The KS and AD goodness-of-fit statistics measure, respectively, the maximum absolute distance and the cumulative weighted quadratic distance (with more weight on the tails) between the empirical cdf $\widehat{F}_n(x) = n^{-1} \sum_{i=1}^n \mathbf{1}\{X_i \leq x\}$ and the parametrically estimated cdf $\widehat{G}(x)$. Their respective computational formulas are given by

$$\text{KS}_n = \max_{1 \leq i \leq n} \left\{ \left| \widehat{G}(X_{(i)}) - \frac{i-1}{n} \right|, \left| \widehat{G}(X_{(i)}) - \frac{i}{n} \right| \right\}$$

and

$$\text{AD}_n = -n + n \sum_{i=1}^n (i/n)^2 \log \left(\frac{\widehat{G}(X_{(i+1)})}{\widehat{G}(X_{(i)})} \right) - n \sum_{i=0}^{n-1} (1 - i/n)^2 \log \left(\frac{1 - \widehat{G}(X_{(i+1)})}{1 - \widehat{G}(X_{(i)})} \right),$$

where $195,000 = X_{(0)} \leq X_{(1)} \leq \dots \leq X_{(n)} \leq X_{(n+1)} = \infty$ denote the ordered claim severities. Also, $\widehat{G}(X_{(i)}) = \widehat{F}_*(X_{(i)})$ for the truncated approach, $\widehat{G}(X_{(i)}) = \widehat{F}(X_{(i)})$ for the naive approach, and $\widehat{G}(X_{(i)}) = \widehat{F}(X_{(i)} - 195,000)$ for the shifted approach. Note that $n = 54$ and the corresponding cdf's were evaluated using the MLE values from Table 4. Also, the p -values of the KS and AD tests were computed using parametric bootstrap with 10,000 simulation runs. For a brief description of the parametric bootstrap procedure, see, e.g., Klugman, Panjer, Willmot (2012, Section 20.4.5).

TABLE 5: Values of KS and AD statistics (with p -values in parentheses) for the fitted models, using truncated, naive, and shifted approaches.

<i>Model</i>	Kolmogorov-Smirnov			Anderson-Darling		
	<i>Truncated</i>	<i>Naive</i>	<i>Shifted</i>	<i>Truncated</i>	<i>Naive</i>	<i>Shifted</i>
Exponential	0.186 (0.004)	0.307 (0.000)	0.186 (0.004)	3.398 (0.000)	4.509 (0.000)	3.398 (0.000)
Lomax	0.072 (0.632)	0.316 (0.000)	0.072 (0.631)	0.272 (0.671)	4.696 (0.000)	0.272 (0.678)

As the results of Table 5 suggest, both naive models are strongly rejected by the KS and AD tests, which is consistent with the conclusions based on QQ-plots. The truncated and shifted exponential models are also rejected, which strengthens our “weak” decisions based on QQ-plots. Unfortunately, for this data set, neither KS nor the AD test can help us with differentiating between the truncated and shifted Lomax models as both of them fit the data very well.

3.4 VaR Estimates

Having fitted and validated the models, we now compute several point and interval estimates of $\text{VaR}(\beta)$ for all six models. The purpose of calculating $\text{VaR}(\beta)$ estimates for all, “good” and “bad,” models is to see the impact that model fit (which is driven by the initial assumptions) has on the capital estimates. The results are summarized in Table 6, where, for completeness, empirical estimates of $\text{VaR}(\beta)$ are also reported. The confidence intervals for the exponential models are derived using Theorem A.3 and based on the variance estimates from Table 4. For the Lomax models, the confidence intervals are obtained using parametric bootstrap with 10,000 simulation runs.

TABLE 6: $\text{VaR}(\beta)$ estimates (with 95% confidence intervals in parentheses), measured in millions and based on the fitted models, using truncated, naive, and shifted approaches.

<i>Model</i>	β	<i>Truncated</i>	<i>Naive</i>	<i>Shifted</i>
Exponential	0.95	1.052 (0.771; 1.332)	1.636 (1.199; 2.072)	1.247 (0.966; 1.527)
	0.995	1.860 (1.364; 2.356)	2.893 (2.121; 3.665)	2.055 (1.559; 2.551)
	0.999	2.425 (1.778; 3.071)	3.772 (2.766; 4.778)	2.620 (1.973; 3.266)
Lomax	0.95	0.576 (0.071; 1.160)	1.670 (1.134; 2.206)	1.514 (0.978; 2.755)
	0.995	2.281 (0.413; 4.758)	3.114 (2.257; 5.023)	5.417 (2.213; 20.604)
	0.999	5.504 (1.100; 13.627)	4.214 (3.019; 8.586)	12.797 (3.649; 89.992)

Empirical estimates of $\text{VaR}(\beta)$: 1.416 (for $\beta = 0.95$) and 3.822 (for $\beta = 0.995$ and 0.999).

We see from the table that the $\text{VaR}(\beta)$ estimates based on the naive approach significantly differ from the rest. The difference between truncated and shifted estimates at the exponential model is $t = 195,000$. For the Lomax model, these two approaches, which exhibited nearly perfect fits to data, produce substantially different estimates, especially at the very extreme tail. Finally, in view of such large differences between parametric estimates (which resulted from models with excellent fits), the empirical estimates do not seem completely off.

3.5 Model Predictions

As the final test of our models, we check their out-of-sample predictive power. Table 7 provides the “unobserved” legal losses, which will be used to verify how accurate are our model-based predictions. To start with, we note that the empirical and shifted models are not able to produce meaningful predictions because they assume that such data were impossible to occur (i.e., $\widehat{F}(195,000) = 0$ for

these two approaches). So now we work only with the truncated and naive models.

TABLE 7: Unobserved costs of legal events (below \$195,000).

142,774.19	146,875.00	151,000.00	160,000.00	176,000.00	182,435.12	191,070.31
143,000.00	150,411.29	153,592.54	165,000.00	176,000.00	185,000.00	192,806.74
145,500.50	150,930.39	157,083.00	165,000.00	180,000.00	186,330.00	193,500.00

SOURCE: Cruz (2002), page 57.

First of all, we report the estimated probabilities of losses below the data collection threshold, $\hat{F}(195,000)$. For the exponential models it is 0.300 (naive) and 0.426 (truncated). For the Lomax models it is 0.310 (naive) and 0.794 (truncated). Secondly, using these probabilities we can estimate the total, observed and unobserved, number of losses. For the exponential models $\hat{N} = 77.2 \approx 77$ (naive) and $\hat{N} = 94.1 \approx 94$ (truncated). For the Lomax models $\hat{N} = 78.3 \approx 78$ (naive) and $\hat{N} = 262.1 \approx 262$ (truncated). Note how different from the rest is the estimate of the truncated Lomax model. By the way, one should not forget that this model exhibited the best statistical fit for the observed data.

For predictions that are verifiable, in Table 8 we report model-based estimates of the number of losses, the average loss, and the total loss in the interval $[150,000; 175,000]$. We also provide the corresponding 95% confidence intervals for the predictions. The intervals were constructed by using the variance and covariance estimates of Table 4 in conjunction with Theorem A.3. Notice that using the data points from Table 7 it is straightforward to verify that the actual number of losses is 8, the average loss is 156,627, and the total loss is 1,253,017. We see from Table 8 that, except for the average loss measure, there are big disparities in predictions between different approaches. This has mostly to do with the quality of model fit for the given data set, which is good for the truncated Lomax model but bad for the other models and/or approaches. As a consequence, 95% confidence intervals based on the truncated Lomax model cover the actual values of two important measures—number of losses (8) and total loss (1,253,017)—but those based on the truncated exponential model do not. Moreover, both naive models fit the data poorly and produce point and interval predictions that are even further from their respective targets than those of the truncated exponential model. In addition, if one chose to ignore the model validation step and proceeded directly to predictions based on the naive models, they would be (falsely) reassured by the consistency of such predictions (number of losses: 2.6 and 2.7; total loss: 426,197 and 441,155).

TABLE 8: Model-based predictions (with 95% confidence intervals in parentheses) of several statistics for the unobserved losses between \$150,000 and \$175,000.

<i>Model</i>	<i>Truncated</i>			<i>Naive</i>		
	number of losses	average loss	total loss	number of losses	average loss	total loss
Exponential	4.2 (3.0; 5.5)	162,352 (162,312; 162,391)	685,108 (452,840; 917,376)	2.6 (1.9; 3.4)	162,405 (162,379; 162,430)	426,197 (141,592; 710,802)
Lomax	9.9 (3.3; 16.5)	162,017 (161,647; 162,388)	1,609,649 (543,017; 2,676,281)	2.7 (1.8; 3.7)	162,397 (162,343; 162,451)	441,155 (288,324; 593,985)

4 Concluding Remarks

In this paper, we have studied the problem of model uncertainty in operational risk modeling, which arises due to different (seemingly plausible) model assumptions. We have focused on the statistical aspects of the problem by utilizing asymptotic theorems of mathematical statistics, Monte Carlo simulations, and real-data examples. Similar to other authors who have studied some aspects of this topic before, we conclude that:

- The naive and empirical approaches are inappropriate for determining VaR estimates.
- The shifted approach, although fundamentally flawed (simply because it assumes that operational losses below the data collection threshold are impossible), has the flexibility to adapt to data well and successfully pass standard model validation tests.
- The truncated approach is theoretically sound, when appropriate fits data well, and (in our examples) produces lower VaR-based capital estimates than those of the shifted approach.

The research presented in this paper invites follow-up studies in several directions. For example, as the first and most obvious direction, one may choose to explore these issues for other, perhaps more popular in practice, distributions such as lognormal or loggamma. If the chosen model lends itself to analytic investigations, then our Example 1 (in Section 2.3) is a blueprint for analysis. Otherwise, one may follow our Example 2 for a simulations-based approach. Second, VaR can be replaced by a different risk measure. For instance, the Expected Shortfall (also known as Tail-VaR or Conditional Tail Expectation) has some theoretical advantages over VaR (e.g., it is a coherent risk measure)

and is a recommended measure in *Swiss Solvency Test*. Third, due to the theoretical soundness of the truncated approach, one may try to develop model-selection strategies for truncated (but not necessarily nested) models. This line of work, however, may be quite challenging due to “flatness” of the truncated likelihoods, a phenomenon frequently encountered in practice (see Cope, 2011). The fourth venue of research that may also help with the latter problem is robust model fitting. There are several excellent contributions to this topic in the operational risk literature (see, e.g., Horbenko, Ruckdeschel, Bae, 2011, Opdyke and Cavallo, 2012, and Chau, 2013), but more work can be done.

Acknowledgements

The authors are very appreciative of valuable insights and useful comments provided by two anonymous referees, which helped to substantially improve the paper.

References

- [1] AMA Group (2013). *AMA Quantification Challenges: AMAG Range of Practice and Observations on “The Thorny LDA Topics”*. Risk Management Association.
- [2] Arnold, B.C. (2015). *Pareto Distributions*, 2nd edition. Chapman & Hall.
- [3] Basel Coordination Committee (2014). Supervisory guidance for data, modeling, and model risk management under the operational risk advanced measurement approaches. *Basel Coordination Committee Bulletin* **14**(1), 1–17.
- [4] Baud, N., Frachot, A., Roncalli, T. (2002). Internal data, external data and consortium data for operational risk measurement: How to pool data properly? *Working Paper*, Groupe de Recherche Opérationnelle, Crédit Lyonnais, France.
- [5] Brazauskas, V., Jones, B., Zitikis, R. (2015). Trends in disguise. *Annals of Actuarial Science* **9**(1), 58–71.
- [6] 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), 3–38.
- [7] Chau, J. (2013). *Robust Estimation in Operational Risk Modeling*. Master’s Thesis, Department of Mathematics, Utrecht University.
- [8] Chernobai, A.S., Rachev, S.T., Fabozzi, F.J. (2007). *Operational Risk: A Guide to Basel II Capital Requirements, Models, and Analysis*. Wiley.

- [9] Cope, E. (2011). Penalized likelihood estimators for truncated data. *Journal of Statistical Planning and Inference* **141**(1), 345–358.
- [10] Cruz, M.G. (2002). *Modeling, Measuring and Hedging Operational Risk*. Wiley.
- [11] Cruz, M.G., Peters, G.W., Shevchenko, P.V. (2015). *Fundamental Aspects of Operational Risk and Insurance Analytics: A Handbook of Operational Risk*, Wiley.
- [12] de Fontnouvelle, P., DeJesus-Rueff, V., Jordan, J., Rosengren, E. (2006). Capital and risk: New evidence on implications of large operational losses. *Journal of Money, Credit, and Banking* **38**(7), 1819–1846.
- [13] Ergashev, B., Pavlikov, K., Uryasev, S., Sekeris, E. (2016). Estimation of truncated data samples in operational risk modeling. *Journal of Risk and Insurance* **83**(3), 613–640.
- [14] Horbenko, N., Ruckdeschel, P., Bae, T. (2011). Robust estimation of operational risk. *Journal of Operational Risk* **6**(2), 3–30.
- [15] Klugman, S.A., Panjer, H.H., Willmot, G.E. (2012). *Loss Models: From Data to Decisions*, 4th edition. Wiley.
- [16] Luo, X., Shevchenko, P.V., Donnelly, J.B. (2007). Addressing the impact of data truncation and parameter uncertainty on operational risk estimates. *Journal of Operational Risk* **2**(4), 3–26.
- [17] Moscadelli, M., Chernobai, A., Rachev, S.T. (2005). Treatment of missing data in the field of operational risk: The impacts on parameter estimates, EL and UL figures. *Operational Risk* **6**(6), 28–34.
- [18] Office of the Comptroller of the Currency (2011). Supervisory guidance on model risk management. *SR Letter* **11**(7), 1–21.
- [19] Opdyke, J.D. (2014). Estimating operational risk capital with greater accuracy, precision, and robustness. *Journal of Operational Risk* **9**(4), 3–79.
- [20] 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.
- [21] Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley.
- [22] Shevchenko, P., Temnov, G. (2009). Modeling operational risk data reported above a time-varying threshold. *Journal of Operational Risk* **4**(2), 19–42.

Appendix

In this appendix, we provide some theoretical results that are key to analytic derivations in the paper. Specifically, in Section A.1, the generalized Pareto distribution (GPD) is introduced and a few of its

special and limiting cases are discussed. In Section A.2, the asymptotic normality theorems for sample quantiles (equivalently, value-at-risk or VaR) and the maximum likelihood estimators (MLE) of model parameters are presented. The well-known delta method is also provided in this section.

A.1 Generalized Pareto Distribution

The cumulative distribution function (cdf) of the three-parameter GPD is given by

$$F_{\text{GPD}(\mu, \sigma, \gamma)}(x) = \begin{cases} 1 - (1 + \gamma(x - \mu)/\sigma)^{-1/\gamma}, & \gamma \neq 0, \\ 1 - \exp(-(x - \mu)/\sigma), & \gamma = 0, \end{cases} \quad (\text{A.1})$$

and the probability density function (pdf) by

$$f_{\text{GPD}(\mu, \sigma, \gamma)}(x) = \begin{cases} \sigma^{-1} (1 + \gamma(x - \mu)/\sigma)^{-1/\gamma-1}, & \gamma \neq 0, \\ \sigma^{-1} \exp(-(x - \mu)/\sigma), & \gamma = 0, \end{cases} \quad (\text{A.2})$$

where the pdf is positive for $x \geq \mu$, when $\gamma \geq 0$, or for $\mu \leq x \leq \mu - \sigma/\gamma$, when $\gamma < 0$. The parameters $-\infty < \mu < \infty$, $\sigma > 0$, and $-\infty < \gamma < \infty$ control the location, scale, and shape of the distribution, respectively. Note that when $\gamma = 0$ and $\gamma = -1$, the GPD reduces to the shifted exponential distribution (with location μ and scale σ) and the uniform distribution on $[\mu; \mu + \sigma]$, respectively. If $\gamma > 0$, then the Pareto-type distributions are obtained. In particular:

- Choosing $1/\gamma = \alpha$, $\sigma/\gamma = \theta$, and $\mu = \theta$ leads to what actuaries call a single-parameter Pareto distribution, with the scale parameter $\theta > 0$ (usually treated as known *deductible*) and shape $\alpha > 0$.
- Choosing $1/\gamma = \alpha$, $\sigma/\gamma = \theta$, and $\mu = 0$ yields the Lomax distribution with the scale parameter $\theta > 0$ and shape $\alpha > 0$. (This is also known as a Pareto II distribution.)

For a comprehensive treatment of Pareto distributions, the reader may be referred to Arnold (2015), and for their applications to loss modeling in insurance, see Klugman, Panjer, Willmot (2012).

A useful property for modeling operational risk with the GPD is that the truncated cdf of excess values remains a GPD (with the same shape parameter γ), and it is given by

$$\mathbf{P}\{X \leq x \mid X > t\} = \frac{\mathbf{P}\{t < X \leq x\}}{\mathbf{P}\{X > t\}} = 1 - \left(1 + \gamma \frac{x - t}{\sigma + \gamma(t - \mu)}\right)^{-1/\gamma}, \quad x > t, \quad (\text{A.3})$$

where the second equality follows by applying (A.1) to the numerator and denominator of the ratio.

In addition, besides functional simplicity of its cdf and pdf, another attractive feature of the GPD is that its quantile function (qf) has an explicit formula. This is especially useful for model diagnostics (e.g., quantile-quantile plots) and for risk evaluations based on VaR measures. Specifically, for $0 < u < 1$, the qf is found by inverting (A.1) and given by

$$F_{\text{GPD}(\mu, \sigma, \gamma)}^{-1}(u) = \begin{cases} \mu + (\sigma/\gamma) \left((1-u)^{-\gamma} - 1 \right), & \gamma \neq 0, \\ \mu - \sigma \log(1-u), & \gamma = 0. \end{cases} \quad (\text{A.4})$$

A.2 Asymptotic Theorems

Suppose X_1, \dots, X_n represent a sample of *independent and identically distributed (i.i.d.)* continuous random variables with cdf G , pdf g , and qf G^{-1} , and let $X_{(1)} \leq \dots \leq X_{(n)}$ denote the ordered sample values. We will assume that g satisfies all the regularity conditions that usually accompany theorems such as the ones formulated below. (For more details on this topic, see, e.g., Serfling, 1980, Sections 2.3.3 and 4.2.2.) Note that a review of modeling practices in the U.S. financial service industry (see AMA Group, 2013) suggests that practically all the severity distributions in current use would satisfy the regularity assumptions mentioned above. In view of this, we will formulate “user friendly” versions of the most general theorems, making them easier to work with. Also, throughout the paper the notation \mathcal{AN} is used to denote “asymptotically normal.”

Since VaR measure is defined as a population quantile, say $G^{-1}(\beta)$, its empirical estimator is the corresponding sample quantile $X_{(\lceil n\beta \rceil)}$, where $\lceil \cdot \rceil$ denotes the “rounding up” operation. We start with the asymptotic normality result for sample quantiles. Proofs and complete technical details are available in Section 2.3.3 of Serfling (1980).

Theorem A.1 [ASYMPTOTIC NORMALITY OF SAMPLE QUANTILES]

Let $0 < \beta_1 < \dots < \beta_k < 1$, with $k > 1$, and suppose that pdf g is continuous, as discussed above. Then the k -variate vector of sample quantiles $(X_{(\lceil n\beta_1 \rceil)}, \dots, X_{(\lceil n\beta_k \rceil)})$ is \mathcal{AN} with the mean vector $(G^{-1}(\beta_1), \dots, G^{-1}(\beta_k))$ and the covariance-variance matrix $[\sigma_{ij}^2]_{i,j=1}^k$ with the entries

$$\sigma_{ij}^2 = \frac{1}{n} \frac{\beta_i(1-\beta_j)}{g(G^{-1}(\beta_i))g(G^{-1}(\beta_j))}.$$

In the univariate case ($k = 1$), the sample quantile

$$X_{(\lceil n\beta \rceil)} \text{ is } \mathcal{AN} \left(G^{-1}(\beta), \frac{1}{n} \frac{\beta(1-\beta)}{g^2(G^{-1}(\beta))} \right).$$

Clearly, in many practical situations the univariate result will suffice, but Theorem A.1 is more general and may be used, for example, to analyze business decisions that combine a set of VaR estimates.

The main drawback of statistical inference based on the empirical model is that it is restricted to the range of observed data. For the problems encountered in operational risk modeling, this is a major limitation. Therefore, a more appropriate alternative is to estimate VaR parametrically, which first requires estimates of the distribution parameters and then those values are applied to formula of $G^{-1}(\beta)$ to find an estimate of VaR. The most common technique for parameter estimation is MLE. The following theorem summarizes its asymptotic distribution. (Description of the method, proofs and complete technical details are available in Section 4.2 of Serfling, 1980.)

Theorem A.2 [ASYMPTOTIC NORMALITY OF MLES]

Suppose pdf g is indexed by k unknown parameters, $(\theta_1, \dots, \theta_k)$, and let $(\hat{\theta}_1, \dots, \hat{\theta}_k)$ denote the MLE of those parameters. Then, under the regularity conditions mentioned above,

$$(\hat{\theta}_1, \dots, \hat{\theta}_k) \text{ is } \mathcal{AN} \left((\theta_1, \dots, \theta_k), \frac{1}{n} \mathbf{I}^{-1} \right),$$

where $\mathbf{I} = [I_{ij}]_{i,j=1}^k$ is the Fisher information matrix, with the entries given by

$$I_{ij} = \mathbf{E} \left[\frac{\partial \log g(X)}{\partial \theta_i} \cdot \frac{\partial \log g(X)}{\partial \theta_j} \right].$$

In the univariate case ($k = 1$),

$$\hat{\theta} \text{ is } \mathcal{AN} \left(\theta, \frac{1}{n} \frac{1}{\mathbf{E} \left[\left(\frac{\partial \log g(X)}{\partial \theta} \right)^2 \right]} \right).$$

Having parameter MLEs, $(\hat{\theta}_1, \dots, \hat{\theta}_k)$, and knowing their asymptotic distribution is useful. Our ultimate goal, however, is to estimate VaR—a function of $(\hat{\theta}_1, \dots, \hat{\theta}_k)$ —and to evaluate its properties. For this we need a theorem that would specify asymptotic distribution of functions of asymptotically

normal vectors. The *delta method* is a technical tool for establishing asymptotic normality of *smoothly* transformed asymptotically normal random variables. Here we will present it as a direct application to Theorem A.2. For the general theorem and complete technical details, see Serfling (1980, Section 3.3).

Theorem A.3 [THE DELTA METHOD]

Suppose that $(\widehat{\theta}_1, \dots, \widehat{\theta}_k)$ is \mathcal{AN} with the parameters specified in Theorem A.2. Let the real-valued functions $h_1(\theta_1, \dots, \theta_k), \dots, h_m(\theta_1, \dots, \theta_k)$ represent m different risk measures, tail probabilities or other functions of model parameters. Then, under some smoothness conditions on functions h_1, \dots, h_m , the vector of MLE-based estimators

$$\left(h_1(\widehat{\theta}_1, \dots, \widehat{\theta}_k), \dots, h_m(\widehat{\theta}_1, \dots, \widehat{\theta}_k) \right) \text{ is } \mathcal{AN}\left((h_1(\theta_1, \dots, \theta_k), \dots, h_m(\theta_1, \dots, \theta_k)), \frac{1}{n} \mathbf{D} \mathbf{I}^{-1} \mathbf{D}' \right),$$

where $\mathbf{D} = [d_{ij}]_{m \times k}$ is the Jacobian of the transformations h_1, \dots, h_m evaluated at $(\theta_1, \dots, \theta_k)$, that is, $d_{ij} = \partial h_i / \partial \theta_j \Big|_{(\theta_1, \dots, \theta_k)}$. In the univariate case ($m = 1$), the parametric estimator

$$h(\widehat{\theta}_1, \dots, \widehat{\theta}_k) \text{ is } \mathcal{AN}\left(h(\theta_1, \dots, \theta_k), \frac{1}{n} \mathbf{d} \mathbf{I}^{-1} \mathbf{d}' \right),$$

where $\mathbf{d} = \left(\partial h / \partial \theta_1, \dots, \partial h / \partial \theta_k \right) \Big|_{(\theta_1, \dots, \theta_k)}$.