May 2016

Statistical Contributions to Operational Risk Modeling

Daoping Yu
University of Wisconsin-Milwaukee

Statistical Contributions to Operational Risk Modeling

by

Daoping Yu

A Dissertation Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in Mathematics

at

The University of Wisconsin–Milwaukee

May 2016

ABSTRACT

Statistical Contributions to Operational Risk Modeling

by

Daoping Yu

University of Wisconsin-Milwaukee, 2016
Under the Supervision of Professor Vytaras Brazauskas

In this dissertation, we focus on statistical aspects of operational risk modeling. Specifically, we are interested in understanding the effects of model uncertainty on capital reserves due to data truncation and in developing better model selection tools for truncated and shifted parametric distributions. We first investigate the model uncertainty question which has been unanswered for many years because researchers, practitioners, and regulators could not agree on how to treat the data collection threshold in operational risk modeling. There are several approaches under consideration—the empirical approach, the "naive" approach, the shifted approach, and the truncated approach—for fitting the loss severity distribution. Since each approach is based on a different set of assumptions, different probability models emerge. Thus, model uncertainty arises. When possible we investigate such model uncertainty analytically using asymptotic theorems of mathematical statistics and several parametric distributions commonly used for operational risk modeling, otherwise we rely on Monte Carlo simulations. The effect of model uncertainty on risk measurements is quantified by evaluating the probability of each approach producing conservative capital allocations based on the value-at-risk measure. These explorations are further illustrated using a real data set for legal losses in a business unit. After clarifying some prevailing misconceptions around the model uncertainty issue in operational risk modeling, we then employ standard (Akaike Information Criterion, AIC, and Bayesian Information Criterion, BIC) and introduce new model selection tools for truncated and shifted parametric models. We find that the new criteria, which are based on information complexity and asymptotic mean curvature of the model likelihood, are more effective at distinguishing between the competing models than AIC and BIC.

Table of Contents

List of Figures

List of Tables

ACKNOWLEDGMENTS

First of all, I would like to express my deep and sincere gratitude to my PhD advisor Professor Vytaras Brazauskas, who has introduced me into the fascinating field of operational risk modeling analytics. Without the consistent patience and encouragement throughout his guidance, I would have not found that an actuarial academic career would be the one I would love to pursue with enthusiasm and persistence. Professor Brazauskas has taught me how to design the structure of whatever it is, a presentation, a proposal or a dissertation. He has trained me to become a down-to-earth detail-oriented researcher yet being able to have a clear understanding of the big picture. Professor Brazauskas has contributed numerous ideas in the establishment of our research problems and objectives, from analytical derivations to numerical simulations and how to polish writing and communicate effectively in conferences.

With all my heart I would like to thank other members of my dissertation committee: Professors Jay Beder, Jugal Ghorai, Hans Volkmer and Wei Wei. I appreciate their constant support and valuable comments regarding my research.

I am very grateful for the care and love from Dr. Yang Ho, Clinical Professor and Co-Director of the actuarial science program. Dr. Ho has influenced me in many ways, from teaching actuarial students financial mathematics to building up sustainable passion in actuarial-related research topics.

I would like to cordially thank the probability seminar faculty: Professors Eric Key, Richard Stockbridge, Chao Zhu. The seminar talks have given me much inspiration in my own study. I extremely value the experiences to give my own talks in the probability seminar, where seminar participants helped me to better understand the topics I presented.

Special thanks go to Dr. Runhuan Feng, now Interim Director of the actuarial science program at the University of Illinois at Urbana-Champaign, who has motivated me to keep up with good work in interdisciplinary areas across pure mathematics and more practical applications in insurance and finance.

I am grateful to Professors Daniel Gervini and Dexuan Xie, Senior Lecturers Ms. Ruslana Korotkaya, Mr. Larry Schultz and Ms. Judy Tolkan. It is a joy to take courses from such nice professors and work with such professional passionate teachers.

I would like to express my sincere gratitude to the department chair Professor Kyle Swanson, and the department's graduate program coordinator Professor Yi-Ming Zou. I would like to acknowledge the graduate committee for consistent financial support including teaching assistantship and summer research excellence awards, conference travel grants and chancellor graduate student awards.

My thanks also go to the administrative staff in the department of mathematical sciences at the University of Wisconsin-Milwaukee, for the friendly and energetic atmosphere in the department.

I would like to thank my fellow graduate student Ranadeera Gamage Madhuka Samanthi with whom I have experienced the common joy, tears and struggle in passing actuarial exams and completing our PhD degrees. I am also grateful for many other fellow graduate students and friends: Lijuan Bi, Jingchun Chen, Sami Cheong, Hua Huang, Yi Jiang, Rui Jie, Jiao Li, Xiaoying Lin, Shenyan Pan, Chudamani Poudyal, Noriyuki Sugiyama, Martin Vieten, Zaihao Wang, Linjiao Wu, Qiuyi Wu, Wen Yang, Jinyong Ying, Qian Zhao, and Xin Zhao in the mathematics department; Xin Li in the engineering department; Hong Qi in the physics department; Hui Ge, Shuwen Wang, and Xu Zhang in the business school.

I could on no account forget to mention my dear friends Sharan Mohanadoss, Chris Beimborn and John Bleidorn whom I met upon coming to Milwaukee, and through whom I have got to know many nice people. We have had a lot of fun activities in the Greater Milwaukee area.

Last but not least, I would like to give the highest credit for my success to my family, for their unconditional love and everlasting support since I was born.

Daoping Yu
Milwaukee, Wisconsin
May 2016

Chapter 1

Introduction

1.1 Operational Risk

The topology of financial risks recognizes three top risk categories: market risk, credit risk, and operational risk. In recent years, operational risk has become one of the major risk factors a financial services organization faces, and a serious threat to the financial stability of the global economy. Distinguished from the other two categories, operational risk is, in large part, a firm-specific and non-systemic risk. To understand what this risk entails, one can think of the following examples: embezzlement of insiders, fraud from outsiders, system failure of hardware and software, cyber attacks, power outages and communication interruptions, earthquakes, floods, fires, terrorist attacks and so on. In short, events of this kind will result in so-called operational losses for banks and insurance companies. It is worthwhile noting here that this risk is not exclusive to the financial service sector; other public and private organizations, businesses, and enterprises are also exposed to operational risk events. This discussion, however, will focus on insurance companies and banks because they represent two very influential sectors of national and global economies, and their regulatory frameworks are well developed.

According to the *Basel II* Capital Accord, the formal definition given by the

Basel Committee on Banking Supervision says that operational risk is "the risk of losses resulting from inadequate or failed internal processes, people and systems or from external events" (Bank for International Settlements, 2011). This definition includes legal risk, but excludes strategic risk and reputational risk. Operational risk can be divided into seven types of events: (1) internal fraud; (2) external fraud; (3) employment practices and workplace safety; (4) clients, products, and business practices; (5) damage to physical assets; (6) business disruption and system failures; (7) execution, delivery, and process management. Operational risk is further mapped across eight business lines: (*i*) corporate finance; (*ii*) trading and sales; (*iii*) retail banking; (*iv*) commercial banking; (*v*) payment and settlement; (*vi*) agency services; (*vii*) asset management; (*viii*) retail brokerage. For a detailed description of these event types and business lines, the reader may be referred to Chernobai, Rachev, Fabozzi (2007).

In compliance with the regulatory requirements, financial corporations and insurance companies need to have a reasonable risk management framework of operational risk. Basel II provides a regulatory framework for risk capital allocation for banks, so does *Solvency II* for insurers. Both regulatory frameworks—Basel II and Solvency II—follow a three-pillar structure:

- PILLAR I: minimum capital requirements,
- PILLAR II: supervisory review of the capital adequacy,
- PILLAR III: market discipline and public disclosure.

In order to figure out Pillar I regulatory minimum capital requirements for operational risk, banks and insurance companies may employ three distinct approaches:

- Basic Indicator Approach (BIA),
- Standardized Approach (SA),
- Advanced Measurement Approaches (AMA).

1.2 Operational Risk Modeling

For a bank or an insurance company to determine the risk capital reserve, either for the purpose of an optimal allocation of assets in investment or to meet compliance requirements of regulatory authorities, they need to measure how much to set aside against potential risks and/or losses the company would face. Operational loss data are collected for each combination of event type and business line and then aggregated gradually to an upper level, first over all event types within a particular business line, followed by aggregation of all business lines.

As discussed in Section 1.1, there are three approaches to calculate minimum capital requirements: BIA, SA, and AMA. Using the BIA method, risk capital is reserved for various types of losses subject to a certain ratio of the average annual gross income over the past three years:

$$RC_{\text{BIA}} = \alpha \times GI,$$

where RC_{BIA} denotes risk capital charge under BIA, and GI denotes average annual gross income over the previous three years, with $\alpha = 15\%$.

If the SA technique is used, then there is a specified general indicator that reflects the size of the companie's activities in that area. The capital charge for each business line is calculated by multiplying gross income by a factor assigned to a particular business line, and then aggregated by simple summation. Mathematically speaking, we have the following formula:

$$RC_{\text{SA}} = \sum_{i=1}^{8} \beta_i \times GI_i,$$

where RC_{SA} denotes risk capital charge under SA, and GI_i the past 3-year averaging annual gross income within each business line, with $\beta_i \in [12\%, 18\%]$, $i = 1, 2, \ldots, 8$. In particular, for corporate finance, trading & sales, and payment & settlement, β_i's are equal to 18%; for commercial banking, and agency services, β_i's are 15%; and for retail banking, asset management, and retail brokerage, β_i's are 12%. Note that this technique assumes a perfect positive correlation among the risks, and may result in overestimation of the aggregate risk capital reserve. Moreover, such an approach may not be sensitive to the risk environment that is dynamically changing.

Within the AMA framework, the Loss Distribution Approach (LDA) 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., 0.999 quantile), which is called *value-at-risk* or VaR. That is, if the aggregate loss variable is given as

$$S = \sum_{i=1}^{8} \sum_{k=1}^{7} \sum_{\ell=N_{i,k}^{T}}^{N_{i,k}^{T+1}} X_{i,k}^{\ell},$$

where $X_{i,k}^{\ell}$ denotes the ℓ-th operational loss for business line i, loss type k in the period $[T, T+1]$, and N^T denotes the accrued loss frequency up to time T, then the risk-based capital over the period $[T, T+1]$ is the 99.9% quantile of the distribution function of S. In this work, we will concentrate on the LDA methodology.

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, which is also the focus of this dissertation. 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." Indeed, in practically all situations, banks and insurance companies are able to absorb small and medium-sized losses using regular cashflow accounts. Substantial capital reserves are necessary for large losses that are relatively rare and appear above a high threshold. Consequently, there is shortage of data, and in response to that banks and insurance companies created commercial data consortia, where only observations above a collection threshold are reported. (For example, the most popular consortium of operational risk data is ORX. Some of the largest banks and insurance companies are among its participants. The data reporting threshold is 20,000 Euros.) Using data from such consortia, the LDA type models are constructed. In this context, the main disagreement that still remains among practitioners, regulators, and academic researchers is about the modeling assumptions. Several types of assumptions have been documented in the academic journals and professional monographs, which we discuss next.

1.3 Literature Review

As mentioned above, the risk-based capital is an extreme quantile of the annual aggregate loss variable, which is a combination of loss frequency and loss severity. The latter part of the aggregate model is where the data collection threshold manifests itself. Typical approaches considered for estimation of VaR include: the empirical approach, the "naive" approach, the shifted approach, and the truncated approach (see Chernobai, Rachev, Fabozzi, 2007). Under the empirical approach, if sufficient amount of data is available, then one simply computes the required level sample quantile for loss severity, inserts it in a compound Poisson model and declares that the problem is solved. Clearly, this approach ignores the presence of data collection threshold. The other three approaches are parametric and address the truncation threshold as follows: the naive approach ignores it (i.e., parameters and VaR are estimated assuming there was no threshold); the shifted approach subtracts the threshold value from each observation, estimates model parameters and then computes the corresponding VaR; the truncated approach assumes that the losses are observed conditionally above the threshold. (A more precise description of these approaches is provided in Section 3.2.1.)

Further, Moscadelli, Chernobai, Rachev (2005) considered truncation of both model components—severity and frequency—and found that the ignorance of truncation in frequency has minimal effect on parameter estimates (consequently, VaR), while ignoring severity truncation introduces significant parameter biases. Similar observations have been made by practitioners (Opdyke, 2014): in practice the severity distribution is a primary driver of the capital estimates. In view of this, most sensitivity studies in the literature have focused on stress-testing modeling assumptions involving the loss severity variable. For example, Cavallo, Rosenthal, Wang, Yan (2012), used the lognormal distribution for severity and compared the shifted and truncated approaches. They found that when the true model is a truncated lognormal model, then a (misspecified) shifted lognormal model leads to much smaller standard errors of parameter estimates and produces competitive percentile estimates for certain chosen values of scale parameter, even though the correctly specified model produces more precise percentile estimates in general. On the other hand, when the true model is a shifted lognormal model, the correctly specified model produces substantially more precise extreme percentile estimates than the misspecified truncated model. Moreover, using Vuong's test (it is a likelihood-ratio-type test based on the Kullback-Leibler divergence, which provides valid inference for non-nested models; introduced by Vuong, 1989), they found that these two models are statistically indistinguishable. They concluded that the shifted lognormal model and the truncated lognormal model could be equally valid or invalid choices for modeling operational risk severity.

In addition, a number of research papers and monographs have examined certain aspects of this topic from other perspectives. For instance, for a semi-parametric modeling perspective, Bolance, Guillen, Gustafsson, Nielsen (2012) offer a book-length treatment of the problem. For Bayesian techniques in operational risk, the reader may be referred to Luo, Shevchenko, Donnelly (2007) or a monograph by Shevchenko (2011). In this dissertation, we will concentrate on the empirical and three parametric approaches mentioned above. Our analysis, however, will not be limited to studying the bias of parameter estimators via simulations. It will go further and establish asymptotic normality of model parameter and VaR estimators, and use those results to evaluate the probability of over-estimating VaR measure by a certain percentage. It is anticipated that such estimates will clearly connect modeling assumptions to capital reserves, clarify existing misconceptions, and be fairly easy to explain to practitioners, regulators, and academic researchers.

1.4 Plan of the Dissertation

The remainder of the dissertation is organized as follows. In Chapter 2, we introduce some preliminary tools that are essential for further analysis in later chapters. In particular, we first formulate several asymptotic theorems of mathematical statistics. Then, we present key probabilistic properties of the generalized Pareto distribution, the lognormal distribution, and the Champernowne distribution. (These distributions are commonly used as parametric models for operational risk.)

In Chapter 3, we demonstrate how different assumptions about the data collection threshold in operational risk models lead to substantially different probability models, which in turn leads to *model uncertainty*. The primary objective of this chapter is to thoroughly investigate and understand what effects, if any, model uncertainty has on risk measurements (i.e., VaR estimates). This is done by employing a variety of statistical tools: asymptotic analysis, Monte Carlo simulations, and real-data examples. The numerical examples are based on a data set for legal losses in a business unit (Cruz, 2002). The conclusions of this chapter are somewhat surprising because we find that although standard model validation techniques help reduce the pool of candidate models, in the end they fail to distinguish between two fundamentally different models.

In view of the findings in Chapter 3, in the next chapter we turn to model selection techniques for truncated distributions. We start Chapter 4 with a short review of the well-established model selection tools, such as Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), and then introduce a less-known but quite effective approach based on information complexity (Bozdogan, 1988). In addition, we propose a new model selection method which is based on the asymptotic mean curvature of the model likelihood. The case study of Chapter 3 is revisited to check the effectiveness of these model selection techniques. We find that the new criteria are more effective at distinguishing between the competing models than AIC and BIC.

In Chapter 5, main findings of the dissertation are summarized and future research avenues are discussed. We envision several directions for future studies: theoretical properties of the new information criterion, robust techniques for fitting truncated parametric distributions, and extension of the severity results to various aggregate loss models, including compound Poisson model and others.

Chapter 2

Preliminaries

In this chapter, we provide some theoretical results that are key to further statistical analysis in operational risk modeling. Specifically, in Section 2.1, the asymptotic normality theorems for sample quantiles (equivalently, VaR) and the maximum likelihood estimators of model parameters are presented. The well-known delta method is also provided in this section. In Section 2.2, the generalized Pareto distribution is introduced and a few of its special and limiting cases are discussed. The lognormal distribution and the Champernowne distribution are also introduced in this section.

2.1 Asymptotic Theorems of Mathematical Statistics

Suppose $X_1, \ldots, X_n$ represent a sample of *independent and identically distributed* (*i.i.d.*) continuous random variables with the cumulative distribution function (cdf) G, probability density function (pdf) g, and quantile function (qf) G^{-1}, and let $X_{(1)} \leq \cdots \leq X_{(n)}$ 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 in later sections. Also, throughout the dissertation the notation $\mathcal{AN}$ will be used to denote "asymptotically normal."

Definition 2.1.1. *Let $0 < \beta < 1$. Value-at-risk or VaR is defined as*

$$VaR(\beta) = G^{-1}(\beta)\,, \tag{2.1}$$

where we use the notation $G^{-1}(\beta)$ for a quantile at level β of the distribution G and G^{-1} for the generalized inverse of G.

Since VaR measure is defined as a population quantile, say $G^{-1}(\beta)$ as in (2.1), 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 2.1.2. *[Asymptotic Normality of Sample Quantiles] Let $0 < \beta_1 < \cdots < \beta_k < 1$, with $k > 1$, and suppose that pdf g is continuous, as discussed above. Then the k-variate vector of sample quantiles $\left(X_{(\lceil n\beta_1 \rceil)}, \ldots, X_{(\lceil n\beta_k \rceil)}\right)$ is $\mathcal{AN}$ with the mean vector $\left(G^{-1}(\beta_1), \ldots, G^{-1}(\beta_k)\right)$ and the covariance-variance matrix $\left[\sigma_{ij}^2\right]_{i,j=1}^{k}$ 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*

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

Clearly, in many practical situations the univariate result will suffice, but Theorem 2.1.2 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 that based on maximization of likelihoods (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 2.1.3. *[*Asymptotic Normality of MLEs*] Suppose pdf g is indexed by k unknown parameters,* $(\theta_1, \ldots, \theta_k)$*, and let* $\left(\widehat{\theta}_1, \ldots, \widehat{\theta}_k\right)$ *denote the MLE of those parameters. Then, under the regularity conditions mentioned above,*

$$\left(\widehat{\theta}_1, \ldots, \widehat{\theta}_k\right) \quad \textit{is} \quad \mathcal{AN}\left(\left(\theta_1, \ldots, \theta_k\right),\, \frac{1}{n}\,\mathbf{I}^{-1}\right),$$

where $\mathbf{I} = \left[I_{ij}\right]_{i,j=1}^{k}$ *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 $(k = 1)$*,*

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

Having parameter MLEs, $\left(\widehat{\theta}_1, \ldots, \widehat{\theta}_k\right)$, and knowing their asymptotic distribution is useful. Our ultimate goal, however, is to estimate VaR—a function of $\left(\widehat{\theta}_1, \ldots, \widehat{\theta}_k\right)$—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 2.1.3. For the general theorem and complete technical details, see Serfling (1980, Section 3.3).

Theorem 2.1.4. *[*THE DELTA METHOD*] Suppose that* $\left(\widehat{\theta}_1, \ldots, \widehat{\theta}_k\right)$ *is* $\mathcal{AN}$ *with the parameters specified in Theorem 2.1.3. Let the real-valued functions* $h_1(\theta_1, \ldots, \theta_k)$*,* $\ldots$*,* $h_m(\theta_1, \ldots, \theta_k)$ *represent* m *different risk measures, tail probabilities or other functions of model parameters. Then, under some smoothness conditions on functions* $h_1, \ldots, h_m$*, the vector of MLE-based estimators*

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

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

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

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

2.2 Parametric Distributions for Operational Risk

2.2.1 Generalized Pareto Distribution

The cdf of the three-parameter generalized Pareto distribution GPD is given by

$$F_{\mathrm{GPD}(\mu,\sigma,\gamma)}(x) = \begin{cases} 1 - (1 + \gamma(x-\mu)/\sigma)^{-1/\gamma}, & \gamma \neq 0, \\ 1 - \exp\left(-(x-\mu)/\sigma\right), & \gamma = 0, \end{cases} \tag{2.2}$$

and the pdf by

$$f_{\mathrm{GPD}(\mu,\sigma,\gamma)}(x) = \begin{cases} \sigma^{-1}\,(1 + \gamma(x-\mu)/\sigma)^{-1/\gamma-1}, & \gamma \neq 0, \\ \sigma^{-1}\exp\left(-(x-\mu)/\sigma\right), & \gamma = 0, \end{cases} \tag{2.3}$$

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 $1/\gamma = \alpha$ and $\sigma/\gamma = \mu = \theta$ 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

$$\begin{aligned}
\mathbf{P}\Big\{X \leq x \,\Big|\, X > t\Big\} &= \frac{\mathbf{P}\{t < X \leq x\}}{\mathbf{P}\{X > t\}} \\
&= \frac{F(x) - F(t)}{1 - F(t)} \\
&= \frac{\left(1 + \gamma \dfrac{t-\mu}{\sigma}\right)^{-1/\gamma} - \left(1 + \gamma \dfrac{x-\mu}{\sigma}\right)^{-1/\gamma}}{\left(1 + \gamma \dfrac{t-\mu}{\sigma}\right)^{-1/\gamma}} \\
&= 1 - \left(1 + \gamma \, \frac{x-t}{\sigma + \gamma(t-\mu)}\right)^{-1/\gamma}, \qquad x > t. \qquad (2.4)
\end{aligned}$$

In addition, besides functional simplicity of its cdf and pdf, another attractive feature of the GPD is that its 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 $0 < \beta < 1$, the qf is found by inverting (2.2) and given by

$$F^{-1}_{\text{GPD}(\mu,\,\sigma,\,\gamma)}(\beta) = \begin{cases} \mu + (\sigma/\gamma)\left((1-\beta)^{-\gamma} - 1\right), & \gamma \neq 0, \\ \mu - \sigma \log(1-\beta), & \gamma = 0. \end{cases} \tag{2.5}$$

2.2.2 Lognormal Distribution

The lognormal distribution with parameters μ and σ is defined as the distribution of a random variable X whose logarithm is normally distributed with mean μ and variance σ^2. The two-parameter lognormal distribution has pdf

$$f_{\text{LN}(\mu,\sigma)}(x) = \begin{cases} \dfrac{1}{\sqrt{2\pi}\sigma x}\, e^{-\dfrac{(\log x - \mu)^2}{2\sigma^2}}, & \text{if } x > 0, \\ 0, & \text{if } x \leq 0, \end{cases} \tag{2.6}$$

where $-\infty < \mu < \infty$ is the location parameter and $\sigma > 0$ is the scale parameter.

The two-parameter lognormal distribution has cdf

$$F_{\text{LN}(\mu,\sigma)}(x) = \Phi\left(\frac{\log x - \mu}{\sigma}\right), \quad x > 0, \tag{2.7}$$

by inverting which, the quantile function

$$F^{-1}_{\text{LN}(\mu,\sigma)}(\beta) = e^{\mu + \sigma\Phi^{-1}(\beta)}, \quad 0 < \beta < 1 \tag{2.8}$$

is obtained. Here Φ and Φ^{-1} denote the cdf and qf of the standard normal distribution, respectively.

2.2.3 Champernowne Distribution

The Champernowne distribution is a generalization of the logistic distribution that was introduced by an econometrician D. G. Champernowne (1936, 1952), who developed the distribution to describe the logarithm of income. The Champernowne distribution has pdf

$$f_{\text{CHAMP}(\alpha,M)}(x) = \frac{\alpha M^\alpha x^{\alpha-1}}{(x^\alpha + M^\alpha)^2}, \quad x > 0, \tag{2.9}$$

and cdf

$$F_{\text{CHAMP}(\alpha,M)}(x) = \frac{x^\alpha}{x^\alpha + M^\alpha}, \quad x > 0, \tag{2.10}$$

where parameters $\alpha > 0$ and $M > 0$ represent the shape and median of the distribution, respectively. The Champernowne distribution looks more like a lognormal distribution near 0 when $\alpha > 1$, while converging to a Pareto distribution in the tail (Bolance, Guillen, Gustafsson, Nielsen, 2012). In addition, by inverting (2.10) the quantile function is obtained as

$$F^{-1}_{\text{CHAMP}(\alpha,M)}(\beta) = M \left(\frac{\beta}{1-\beta} \right)^{1/\alpha}, \quad 0 < \beta < 1. \tag{2.11}$$

Chapter 3

Model Uncertainty due to Data Truncation

This chapter is structured as follows. In Section 3.1, we provide motivation for the problem discussed in this chapter. Then, in Section 3.2, we review several typical models used for estimating VaR, describe how model uncertainty due to data truncation emerges, and study effects of that uncertainty on VaR estimates. This is done with parametric examples, where we evaluate the probability of overestimating true VaR for exponential, Lomax, and lognormal distributions, using the theoretical results of Section 2.1 and via Monte Carlo simulations. Then, in Section 3.3, these explorations are further illustrated using a real data set for legal losses in a business unit. The performance of all modeling approaches is illustrated on that real data, including performance of model fitting, model validation, VaR estimates, and model predictive power. Finally, concluding remarks are offered in Section 3.4.

3.1 Introduction

In order to fully understand the problem of model uncertainty in operational risk modeling, in this chapter we will walk the reader through the entire modeling process and investigate how our assumptions affect the end product, which is the estimate of risk-based capital or 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 1.3, 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. In Figure 3.1, we provide graphs of truncated, naive, and shifted probability density functions of two distributions (studied formally in Sections 3.2.2 and 3.2.3): *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 3.5, this task may look straightforward, but its outcomes vary and are heavily influenced by the initial assumptions.

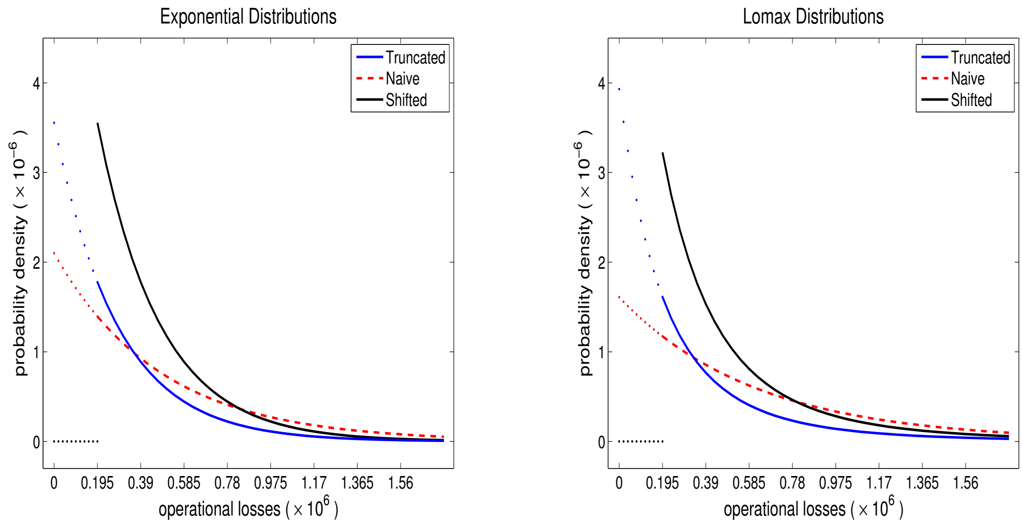

Figure 3.1: Truncated, naive, shifted *Exponential* (σ) and *Lomax* $(\alpha = 3.5, \theta_1)$ probability density functions. Data collection threshold $t = 195,000$, with 50% of data unobserved. Parameters σ and θ_1 are chosen to match those in Tables 3.2 and 3.3 (see Sections 3.2.2 and 3.2.3).

3.2 Capital Reserving

3.2.1 Operational Risk Modeling Approaches

Suppose that $Y_1, \ldots, Y_N$ 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 data collection threshold $t \gg 0$ way above zero. That is, the actually observed variables are X_i's with

$$X_1 \stackrel{d}{=} Y_{i_1} \Big| Y_{i_1} > t, \quad \ldots, \quad X_n \stackrel{d}{=} Y_{i_n} \Big| Y_{i_n} > t, \tag{3.1}$$

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

$$F_*(x) \;=\; \mathbf{P}\Big\{X \leq x \,\Big|\, X > t\Big\} = \frac{\mathbf{P}\{t < X \leq x\}}{\mathbf{P}\{x > t\}} = \frac{F(x) - F(t)}{1 - F(t)}\,, \tag{3.2}$$

$$f_*(x) \;=\; F_*'(x) = \frac{f(x)}{1 - F(t)}\,, \tag{3.3}$$

$$F_*^{-1}(u) \;=\; F^{-1}\Big(u + (1-u)F(t)\Big) \tag{3.4}$$

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 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

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

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. Let us start with an empirical example.

Example 3.2.1. Empirical Approach

As mentioned earlier, the empirical model is restricted to the range of observed data. So it uses data from (3.1), but since the empirical estimator $\widehat{F}(t) = 0$, formulas (3.2)–(3.4) 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 (3.2)–(3.4). In this case, the VaR(β) estimator is simply $\widehat{F}^{-1}(\beta) = X_{(\lceil n\beta \rceil)}$, and as follows from Theorem 2.1.2 (see Section

2.1),

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

We now can evaluate the probability of overestimating true VaR by certain percentage, i.e., we want to study function $H(c) := \mathbf{P}\Big\{X_{(\lceil n\beta \rceil)} > c\,F^{-1}(\beta)\Big\}$ for $c \geq 1$. Using Z to denote the standard normal random variable and Φ for its cdf, and taking into account (3.2)–(3.4), we proceed as follows:

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

where

$$F_*^{-1}(\beta) = F^{-1}(\beta + (1-\beta)F(t)).$$

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 $c \to \infty$, we observe $H(c) \to 0$. Additional numerical illustrations are provided in Table 3.1.

Table 3.1: 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 $n = 100$.)

		$F(t)=0$			$F(t)=0.5$			$F(t)=0.9$		
c	β	*Light*	*Moderate*	*Heavy*	*Light*	*Moderate*	*Heavy*	*Light*	*Moderate*	*Heavy*
1	0.95	0.500	0.500	0.500	0.944	0.925	0.874	1.000	1.000	0.981
	0.99	0.500	0.500	0.500	0.757	0.736	0.692	0.990	0.955	0.817
1.2	0.95	0.085	0.178	0.331	0.585	0.753	0.824	1.000	1.000	0.978
	0.99	0.177	0.303	0.421	0.409	0.583	0.657	0.918	0.924	0.812
1.5	0.95	0.000	0.010	0.138	0.032	0.326	0.726	0.968	0.996	0.975
	0.99	0.010	0.099	0.309	0.053	0.336	0.600	0.500	0.848	0.804
2	0.95	0.000	0.000	0.015	0.000	0.009	0.523	0.056	0.930	0.968
	0.99	0.000	0.005	0.160	0.000	0.070	0.502	0.010	0.642	0.790

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 $F(t)=0.9$: $\sigma = 84,687$, $\theta_1 = 209,520$, $\theta_2 = 21,667$.

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(0.99) by 20% ($c = 1.2$) for the light-tailed distribution is only 0.177 for $F(t) = 0$, it increases to 0.409 and 0.918 for $F(t) = 0.5$ and 0.9, respectively. If severity follows the heavy-tailed distribution, then $H(1.2)$ is 0.421, 0.657, 0.812 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 VaR(0.99) by 100% seems like a norm (0.642 and 0.790). □

A review of existing methods (see Section 1.3) shows that, besides estimation of

VaR using (3.1)–(3.4) under the truncated framework, there are parametric methods that employ other strategies, such as the naive and shifted approaches. In particular, those two approaches use the data $X_1, \ldots, X_n$ and either ignore t or recognize it in some other way than (3.2)–(3.4). Thus, model uncertainty emerges.

Example 3.2.2. Truncated Approach

The truncated approach uses the observed data $X_1, \ldots, X_n$ and fully recognizes its distributional properties. That is, it takes into account (3.2)–(3.4) and derives MLE values by maximizing the following log-likelihood function:

$$\log \mathcal{L}_T\Big(\theta_1, \ldots, \theta_k \,\Big|\, X_1, \ldots, X_n\Big) \;=\; \sum_{i=1}^{n} \log f_*(X_i) = \sum_{i=1}^{n} \log\left(\frac{f(X_i)}{1 - F(t)}\right), \qquad (3.5)$$

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

Example 3.2.3. Naive Approach

The naive approach uses the observed data $X_1, \ldots, X_n$, but ignores the presence of threshold t. That is, it bypasses (3.2)–(3.4) and derives MLE values by maximizing the following log-likelihood function:

$$\log \mathcal{L}_N\Big(\theta_1, \ldots, \theta_k \,\Big|\, X_1, \ldots, X_n\Big) \;=\; \sum_{i=1}^{n} \log f(X_i). \qquad (3.6)$$

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, VaR(β) estimate is computed by inserting parameter MLEs (the ones found using the naive approach) into $F^{-1}(\beta)$. □

Example 3.2.4. Shifted Approach

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

$$\log \mathcal{L}_S\Big(\theta_1, \ldots, \theta_k \,\Big|\, X_1, \ldots, X_n\Big) \;=\; \sum_{i=1}^{n} \log f(X_i - t). \tag{3.7}$$

By comparing (3.6) and (3.7), 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, $\mathrm{VaR}(\beta)$ is estimated by applying parameter MLEs (the ones found using the shifted approach) to $F^{-1}(\beta) + t$. □

3.2.2 Exponential Models

Suppose $Y_1, \ldots, Y_N$ are *i.i.d.* and follow an exponential distribution, with pdf, cdf, and qf given by (2.3), (2.2), and (2.5), respectively, with $\gamma = 0$ and $\mu = 0$. However, we observe only variable X whose relation to Y is governed by (3.1)–(3.4). Now, by plugging exponential density and/or distribution functions with scale parameter σ into the log-likelihoods (3.5), (3.6), and (3.7), we obtain the corresponding log-likelihoods for exponential models as follows:

$$\begin{aligned} \log \mathcal{L}_T\Big(\sigma \,\Big|\, X_1, \ldots, X_n\Big) &= \sum_{i=1}^{n} \log\left(\frac{f(X_i)}{1 - F(t)}\right) \\ &= \sum_{i=1}^{n} \log\left(\frac{(1/\sigma)e^{-X_i/\sigma}}{e^{-t/\sigma}}\right) \\ &= \sum_{i=1}^{n} \frac{-(X_i - t)}{\sigma} - n \log \sigma, \end{aligned} \tag{3.8}$$

$$\log \mathcal{L}_N\Big(\sigma \,\Big|\, X_1, \ldots, X_n\Big) = \sum_{i=1}^{n} \log f(X_i) = \sum_{i=1}^{n} \log\left(\frac{1}{\sigma} e^{-X_i/\sigma}\right) = \sum_{i=1}^{n} \frac{-X_i}{\sigma} - n \log \sigma, \tag{3.9}$$

$$\log \mathcal{L}_S\Big(\sigma \,\Big|\, X_1, \ldots, X_n\Big) = \sum_{i=1}^{n} \log f(X_i - t) = \sum_{i=1}^{n} \log\left(\frac{1}{\sigma} e^{-(X_i - t)/\sigma}\right) = \sum_{i=1}^{n} \frac{-(X_i - t)}{\sigma} - n \log \sigma. \tag{3.10}$$

Then, by maximizing the log-likelihoods (3.8), (3.9), and (3.10), we get the following MLE formulas for scale parameter σ:

$$\widehat{\sigma}_T = \overline{X} - t, \qquad \widehat{\sigma}_N = \overline{X}, \qquad \widehat{\sigma}_S = \overline{X} - t,$$

where $\overline{X} = n^{-1} \sum_{i=1}^{n} X_i$ and subscripts T, N, S denote "truncated", "naive", "shifted", respectively.

Next, by inserting $\widehat{\sigma}_T$, $\widehat{\sigma}_N$, and $\widehat{\sigma}_S$ into the corresponding qf's as described in Examples 3.2.2 – 3.2.4, we get the following VaR(β) estimators:

$$\widehat{\mathrm{VaR}}_T(\beta) = -\widehat{\sigma}_T \log(1 - \beta),$$

$$\widehat{\mathrm{VaR}}_N(\beta) = -\widehat{\sigma}_N \log(1-\beta),$$

$$\widehat{\mathrm{VaR}}_S(\beta) = -\widehat{\sigma}_S \log(1-\beta) + t.$$

Further, a direct application of Theorem 2.1.3 for $\widehat{\sigma}_T$ (with obvious adjustment for $\widehat{\sigma}_N$), yields that

$$\widehat{\sigma}_T \ \text{ is } \ \mathcal{AN}\left(\sigma, \, \frac{\sigma^2}{n}\right),$$

$$\widehat{\sigma}_N \ \text{ is } \ \mathcal{AN}\left(\sigma + t, \, \frac{\sigma^2}{n}\right),$$

$$\widehat{\sigma}_S \ \text{ is } \ \mathcal{AN}\left(\sigma, \, \frac{\sigma^2}{n}\right),$$

where asymptotic distributions for parameter MLEs of truncated, naive, and shifted models are all derived under the assumption that the true distribution of the observed data is the truncated distribution.

Furthermore, having established $\mathcal{AN}$ for parameter MLEs, we can apply Theorem 2.1.4 and specify asymptotic distributions for VaR estimators. They are as follows:

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

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

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

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 Example 3.2.1, we now define function $H(c) = \mathbf{P}\left\{\widehat{\mathrm{VaR}}(\beta) > c\, F^{-1}(\beta)\right\}$ for $c \geq 1$, the probability of overestimating the target by $(c-1)100\%$, for each parametric VaR estimator and study its behavior:

$$
\begin{aligned}
H_{T}(c) &= \mathbf{P}\Big\{\widehat{\mathrm{VaR}}_{T}(\beta) > c\,F^{-1}(\beta)\Big\}\\
&= \mathbf{P}\left\{\frac{\widehat{\mathrm{VaR}}_{T}(\beta) - \mathbf{E}\left(\widehat{\mathrm{VaR}}_{T}(\beta)\right)}{\sqrt{\mathbf{V}\left(\widehat{\mathrm{VaR}}_{T}(\beta)\right)}} > \frac{c\,F^{-1}(\beta) - \mathbf{E}\left(\widehat{\mathrm{VaR}}_{T}(\beta)\right)}{\sqrt{\mathbf{V}\left(\widehat{\mathrm{VaR}}_{T}(\beta)\right)}}\right\}\\
&\approx\ \mathbf{P}\left\{Z > \frac{-c\sigma\log(1-\beta) + \sigma\log(1-\beta)}{-\sigma\log(1-\beta)/\sqrt{n}}\right\}\\
&= \mathbf{P}\Big\{Z > (c-1)\sqrt{n}\Big\}\\
&= 1 - \Phi\left((c-1)\sqrt{n}\right),
\end{aligned}
$$

$$
\begin{aligned}
H_{N}(c) &= \mathbf{P}\Big\{\widehat{\mathrm{VaR}}_{N}(\beta) > c\,F^{-1}(\beta)\Big\}\\
&= \mathbf{P}\left\{\frac{\widehat{\mathrm{VaR}}_{N}(\beta) - \mathbf{E}\left(\widehat{\mathrm{VaR}}_{N}(\beta)\right)}{\sqrt{\mathbf{V}\left(\widehat{\mathrm{VaR}}_{N}(\beta)\right)}} > \frac{c\,F^{-1}(\beta) - \mathbf{E}\left(\widehat{\mathrm{VaR}}_{N}(\beta)\right)}{\sqrt{\mathbf{V}\left(\widehat{\mathrm{VaR}}_{N}(\beta)\right)}}\right\}\\
&\approx\ \mathbf{P}\left\{Z > \frac{-c\sigma\log(1-\beta) + (\sigma+t)\log(1-\beta)}{-\sigma\log(1-\beta)/\sqrt{n}}\right\}\\
&= \mathbf{P}\Big\{Z > (c-1)\sqrt{n} - \sqrt{n}(t/\sigma)\Big\}\\
&= 1 - \Phi\left((c-1)\sqrt{n} - \sqrt{n}(t/\sigma)\right),
\end{aligned}
$$

$$
\begin{aligned}
H_S(c) &= \mathbf{P}\Big\{\widehat{\mathrm{VaR}}_S(\beta) > c\,F^{-1}(\beta)\Big\} \\
&= \mathbf{P}\left\{\frac{\widehat{\mathrm{VaR}}_S(\beta) - \mathbf{E}\left(\widehat{\mathrm{VaR}}_S(\beta)\right)}{\sqrt{\mathbf{V}\left(\widehat{\mathrm{VaR}}_S(\beta)\right)}} > \frac{c\,F^{-1}(\beta) - \mathbf{E}\left(\widehat{\mathrm{VaR}}_S(\beta)\right)}{\sqrt{\mathbf{V}\left(\widehat{\mathrm{VaR}}_S(\beta)\right)}}\right\} \\
&\approx\ \mathbf{P}\left\{Z > \frac{-c\sigma\log(1-\beta) + \sigma\log(1-\beta) - t}{-\sigma\log(1-\beta)/\sqrt{n}}\right\} \\
&= \mathbf{P}\Big\{Z > (c-1)\sqrt{n} + \sqrt{n}(t/\sigma)\log^{-1}(1-\beta)\Big\} \\
&= 1 - \Phi\left((c-1)\sqrt{n} + \sqrt{n}(t/\sigma)\log^{-1}(1-\beta)\right).
\end{aligned}
$$

Table 3.2 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 3.1. From Table 3.2, 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 3.1) 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 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 3.2: Exponential models: 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 $n = 100$.)

		$F(t) = 0$			$F(t) = 0.5$			$F(t) = 0.9$		
c	β	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
	0.99	0.500	0.500	0.500	0.500	1.000	0.934	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
	0.99	0.023	0.023	0.023	0.023	1.000	0.310	0.023	1.000	0.999
1.5	0.95	0.000	0.000	0.000	0.000	0.973	0.004	0.000	1.000	0.996
	0.99	0.000	0.000	0.000	0.000	0.973	0.000	0.000	1.000	0.500
2	0.95	0.000	0.000	0.000	0.000	0.001	0.000	0.000	1.000	0.010
	0.99	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 $F(t) = 0.9$).

3.2.3 Lomax Models

Suppose $Y_1, \ldots, Y_N$ are *i.i.d.* and follow a Lomax distribution, with pdf, cdf, and qf given by (2.3), (2.2), and (2.5), respectively, with $\alpha = 1/\gamma$, $\theta = \sigma/\gamma$, and $\mu = 0$. However, we observe only variable X whose relation to Y is governed by (3.1)–(3.4). Now, unlike the exponential case, maximization of the log-likelihoods, obtained by plugging Lomax pdf and cdf into (3.5), (3.6), and (3.7),

$$\log \mathcal{L}_T = n \log \alpha + n\alpha \log(\theta + t) - (\alpha + 1) \sum_{i=1}^{n} \log(\theta + X_i), \qquad (3.11)$$

$$\log \mathcal{L}_N = n \log \alpha + n\alpha \log \theta - (\alpha + 1) \sum_{i=1}^{n} \log(\theta + X_i), \qquad (3.12)$$

$$\log \mathcal{L}_S = n \log \alpha + n\alpha \log \theta - (\alpha + 1) \sum_{i=1}^{n} \log(\theta + X_i - t) \qquad (3.13)$$

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 3.1, 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 3.3, where we again observe similar patterns to those of Tables 3.1 and 3.2. 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 $c \to \infty$) 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 3.2. The difference between the T entries in Tables 3.2 and 3.3 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.

Table 3.3: Lomax models: 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 $n = 100$.)

		$F(t)=0$			$F(t)=0.5$			$F(t)=0.9$		
c	β	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
	0.99	0.436	0.436	0.436	0.425	0.926	0.797	0.503	0.888	0.999
1.2	0.95	0.131	0.131	0.131	0.095	0.945	0.791	0.356	0.904	0.999
	0.99	0.219	0.219	0.219	0.154	0.496	0.566	0.228	0.884	0.998
1.5	0.95	0.009	0.009	0.009	0.002	0.626	0.270	0.112	0.879	0.998
	0.99	0.071	0.071	0.071	0.020	0.097	0.265	0.029	0.877	0.957
2	0.95	0.000	0.000	0.000	0.000	0.032	0.010	0.002	0.865	0.984
	0.99	0.011	0.011	0.011	0.001	0.005	0.052	0.000	0.863	0.646

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 $F(t) = 0.9$).

Remark 3.2.5. Asymptotic Justifications

In this Remark, we use asymptotic theorems to double-check some results of the simulation study (Table 3.3). In particular, we can verify the truncated and shifted cases. Let us start with the fact that

$$\left(\widehat{\alpha}_T, \widehat{\theta}_T\right) \quad \text{is} \quad \mathcal{AN}\left((\alpha,\theta), \frac{\mathbf{\Sigma}_T^{Lomax}}{n}\right)$$

and

$$\left(\widehat{\alpha}_S, \widehat{\theta}_S\right) \quad \text{is} \quad \mathcal{AN}\left((\alpha,\theta+t), \frac{\mathbf{\Sigma}_S^{Lomax}}{n}\right),$$

where $\mathbf{\Sigma}_S^{Lomax}$ equals $\mathbf{\Sigma}_T^{Lomax}$ as shown in (A.1) (see Appendix A), since the corresponding parameter estimators have the same variance–covariance structure.

Furthermore, applying Theorem 2.1.4, we have

$$\widehat{\mathrm{VaR}}_T(\beta) \;\text{ is }\; \mathcal{AN}\left(F^{-1}(\beta),\ \mathbf{V}\left(\widehat{\mathrm{VaR}}_T(\beta)\right)\right)$$

and

$$\widehat{\mathrm{VaR}}_S(\beta) \;\text{ is }\; \mathcal{AN}\left(\frac{\theta+t}{\theta}F^{-1}(\beta)+t,\ \mathbf{V}\left(\widehat{\mathrm{VaR}}_S(\beta)\right)\right),$$

where $F^{-1}(\beta) = \theta\left((1-\beta)^{-1/\alpha}-1)\right)$.

We now can evaluate the probability of overestimating true VaR by certain percentage, i.e., we want to study functions

$$\begin{aligned}
H_T(c) &= \mathbf{P}\left\{\widehat{\mathrm{VaR}}_T(\beta) > c\,F^{-1}(\beta)\right\} \\
&= \mathbf{P}\left\{\frac{\widehat{\mathrm{VaR}}_T(\beta) - \mathbf{E}\left(\widehat{\mathrm{VaR}}_T(\beta)\right)}{\sqrt{\mathbf{V}\left(\widehat{\mathrm{VaR}}_T(\beta)\right)}} > \frac{c\,F^{-1}(\beta) - \mathbf{E}\left(\widehat{\mathrm{VaR}}_T(\beta)\right)}{\sqrt{\mathbf{V}\left(\widehat{\mathrm{VaR}}_T(\beta)\right)}}\right\} \\
&\approx \mathbf{P}\left\{Z > \left[c\,F^{-1}(\beta) - F^{-1}(\beta)\right] \times \left[\mathbf{V}\left(\widehat{\mathrm{VaR}}_T(\beta)\right)\right]^{-1/2}\right\} \\
&= 1 - \Phi\left(\left[(c-1)\,F^{-1}(\beta)\right] \times \left[\mathbf{V}\left(\widehat{\mathrm{VaR}}_T(\beta)\right)\right]^{-1/2}\right),
\end{aligned}$$

$$
\begin{aligned}
H_S(c) &= \mathbf{P}\Big\{\widehat{\mathrm{VaR}}_S(\beta) > c\,F^{-1}(\beta)\Big\} \\
&= \mathbf{P}\left\{\frac{\widehat{\mathrm{VaR}}_S(\beta) - \mathbf{E}\left(\widehat{\mathrm{VaR}}_S(\beta)\right)}{\sqrt{\mathbf{V}\left(\widehat{\mathrm{VaR}}_S(\beta)\right)}} > \frac{c\,F^{-1}(\beta) - \mathbf{E}\left(\widehat{\mathrm{VaR}}_S(\beta)\right)}{\sqrt{\mathbf{V}\left(\widehat{\mathrm{VaR}}_S(\beta)\right)}}\right\} \\
&\approx \mathbf{P}\left\{Z > \left[c\,F^{-1}(\beta) - \frac{\theta+t}{\theta}F^{-1}(\beta) - t\right] \times \left[\mathbf{V}\left(\widehat{\mathrm{VaR}}_S(\beta)\right)\right]^{-1/2}\right\} \\
&= 1 - \Phi\left(\left[\left(c - \frac{\theta+t}{\theta}\right) F^{-1}(\beta) - t\right] \times \left[\mathbf{V}\left(\widehat{\mathrm{VaR}}_S(\beta)\right)\right]^{-1/2}\right),
\end{aligned}
$$

where explicit expressions for $\mathbf{V}\left(\widehat{\mathrm{VaR}}_T(\beta)\right)$ and $\mathbf{V}\left(\widehat{\mathrm{VaR}}_S(\beta)\right)$ are provided in (A.1) of Appendix A. The values of functions $H_T(c)$ and $H_S(c)$ in Table 3.4 match closely the corresponding entries in Table 3.3. That justifies the simulation results for the truncated and shifted approaches.

Table 3.4: Lomax models: Theoretical evaluations of functions $H_T(c)$ and $H_S(c)$ for various combinations of c, confidence level β, and proportion of unobserved data $F(t)$. (The sample size is $n = 100$.)

		$F(t) = 0$		$F(t) = 0.5$		$F(t) = 0.9$	
c	β	T	S	T	S	T	S
1	0.95	0.500	0.500	0.500	0.964	0.500	1.000
	0.99	0.500	0.500	0.500	0.803	0.500	0.989
1.2	0.95	0.124	0.124	0.095	0.804	0.332	1.000
	0.99	0.244	0.244	0.166	0.611	0.284	0.973
1.5	0.95	0.002	0.002	0.001	0.286	0.139	1.000
	0.99	0.041	0.041	0.008	0.284	0.077	0.918
2	0.95	0.000	0.000	0.000	0.002	0.015	0.968
	0.99	0.000	0.000	0.000	0.023	0.002	0.688

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 $F(t) = 0.9$).

□

3.2.4 Lognormal Models

Suppose $Y_1, \ldots, Y_N$ are *i.i.d.* and follow a lognormal distribution, with pdf, cdf, and qf given by (2.6), (2.7), and (2.8), respectively. However, we observe only variable X whose relation to Y is governed by (3.1)–(3.4). Numerical maximization of the log-likelihood (3.5) is performed for the truncated lognormal model. We plug in (2.6) and (2.7) into (3.5) to obtain the log-likelihood function for the truncated lognormal model:

$$\log \mathcal{L}_T = C_1 - n \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (\log X_i - \mu)^2 - n \log \left(1 - \Phi\left(\frac{\log t - \mu}{\sigma}\right)\right), \quad (3.14)$$

where $C_1 = -n \log \sqrt{2\pi} - \sum_{i=1}^{n} \log X_i$. Differentiating (3.14) with respect to μ and σ and setting the partial derivatives equal to zero, we can numerically solve for the MLE parameter estimators $\widehat{\mu}_T$ and $\widehat{\sigma}_T$. Explicit formulas for MLEs of the naive and shifted lognormal models are obtained via maximization of the log-likelihoods (3.15) and (3.16), respectively:

$$\log \mathcal{L}_N = C_1 - n \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (\log X_i - \mu)^2, \quad (3.15)$$

$$\log \mathcal{L}_S = C_2 - n \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (\log(X_i - t) - \mu)^2, \quad (3.16)$$

where $C_2 = -n \log \sqrt{2\pi} - \sum_{i=1}^{n} \log(X_i - t)$. The parameter estimators are as follows:

$$\widehat{\mu}_N = \frac{1}{n} \sum_{i=1}^{n} \log (X_i),$$

$$\widehat{\sigma}_N^2 = \frac{1}{n} \sum_{i=1}^{n} [\log (X_i) - \widehat{\mu}_N]^2,$$

$$\widehat{\mu}_S = \frac{1}{n}\sum_{i=1}^{n} \log\left(X_i - t\right),$$

$$\widehat{\sigma}_S^2 = \frac{1}{n}\sum_{i=1}^{n} \left[\log\left(X_i - t\right) - \widehat{\mu}_S\right]^2.$$

In order to evaluate functions $H_T(c)$, $H_N(c)$, $H_S(c)$, we use Monte Carlo simulations and implement the 5-step procedure of Section 3.2.3. Note that now the procedure is modified to generate lognormally distributed data. Simulation parameters are selected as follows:

- Severity distribution Lognormal($\mu, \sigma = 1.6$): $\mu = 10$ (for $F(t) = 0$), $\mu = 12.18$ (for $F(t) = 0.5$), $\mu = 10.13$ (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 3.5, where we observe similar patterns to those of the Lomax model (Table 3.3). Specifically, we see that: The truncated lognormal model produces VaR estimates that are centered around the target; the naive approach does not necessarily overestimate VaR; the shifted approach almost consistently yields VaR overestimation.

Table 3.5: Lognormal models: 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 $n = 100$.)

		$F(t) = 0$			$F(t) = 0.5$			$F(t) = 0.9$		
c	β	T	N	S	T	N	S	T	N	S
1	0.95	0.448	0.473	0.473	0.455	0.850	1.000	0.555	1.000	1.000
	0.99	0.462	0.484	0.484	0.374	0.200	0.982	0.567	0.999	1.000
1.2	0.95	0.181	0.197	0.197	0.203	0.572	0.980	0.499	1.000	1.000
	0.99	0.263	0.278	0.278	0.170	0.065	0.962	0.465	0.981	1.000
1.5	0.95	0.047	0.048	0.048	0.039	0.188	0.866	0.424	1.000	1.000
	0.99	0.083	0.091	0.091	0.035	0.007	0.785	0.302	0.812	1.000
2	0.95	0.000	0.001	0.001	0.000	0.009	0.430	0.285	1.000	1.000
	0.99	0.012	0.014	0.014	0.003	0.001	0.452	0.074	0.338	0.997

NOTE: Threshold t is 0 for $F(t) = 0$ and $195,000$ for $F(t) = 0.5, 0.9$. Lognormal($\mu, \sigma = 1.6$), with $\mu = 10$ (for $F(t) = 0$), $\mu = 12.18$ (for $F(t) = 0.5$), $\mu = 10.13$ (for $F(t) = 0.9$).

3.3 Case Study: Legal Risk

In this section, we illustrate how all the modeling approaches considered so far (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 case study with model-based predictions for quantities below the data collection threshold.

3.3.1 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 (for convenience, the data set is provided in Appendix B). 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 as shown in Figure 3.2, 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.

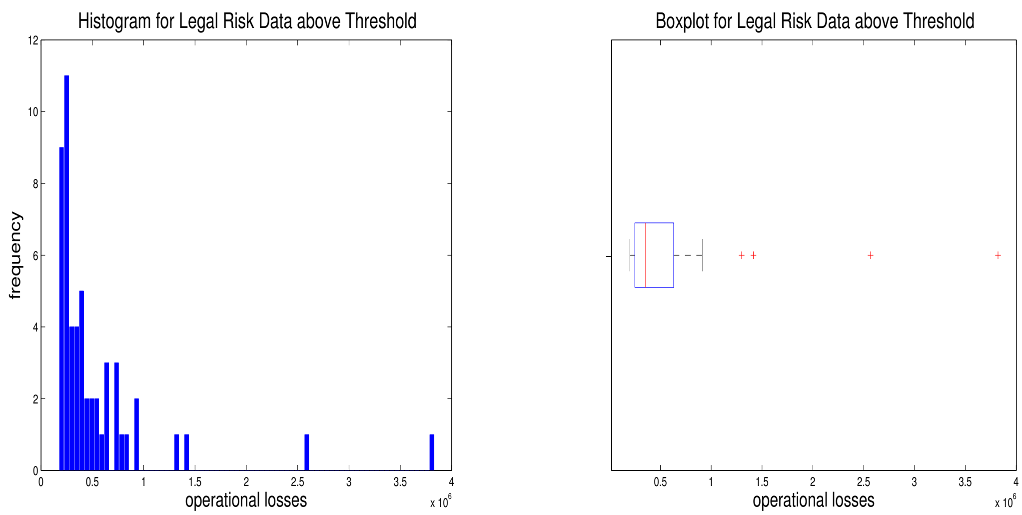

Figure 3.2: Histogram and boxplot for legal risk data above the threshold. Data collection threshold $t = 195,000$.

3.3.2 Model Fitting

We fit three models to the observed data, exponential, Lomax, and lognormal, 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 3.2.2. For the Lomax distribution, we perform numerical maximization of the log-likelihoods (3.11)–(3.13) to estimate model parameters. For the lognormal distribution, we perform numerical maximization of the log-likelihood (3.14) and analytical maximization of the log-likelihoods (3.15) and (3.16) to estimate model parameters. For the data set under consideration, the resulting MLE values are reported in Table 3.6. Also, the corresponding estimates for parameter variances and covariances were computed using Theorem 2.1.3 and the following proposition.

Proposition 3.3.1. *[Asymptotic Distributions of Parameter MLE's.]*

Exponential models:

$$\widehat{\sigma}_{T} \;\; is \;\; \mathcal{AN}\left(\sigma,\, \frac{\sigma^2}{n}\right), \qquad \widehat{\sigma}_{N} \;\; is \;\; \mathcal{AN}\left(\sigma+t,\, \frac{\sigma^2}{n}\right), \qquad \widehat{\sigma}_{S} \;\; is \;\; \mathcal{AN}\left(\sigma,\, \frac{\sigma^2}{n}\right).$$

Lomax models:

$$\left(\widehat{\alpha}_{T}, \widehat{\theta}_{T}\right) \;\; is \;\; \mathcal{AN}\left((\alpha,\theta),\, \frac{1}{n}\mathbf{\Sigma}_{T}^{Lomax}\right),$$

$$\left(\widehat{\alpha}_{N}, \widehat{\theta}_{N}\right) \;\; is \;\; \mathcal{AN}\left(\left(\alpha_{N},\theta_{N}\right),\, \frac{1}{n}\mathbf{\Sigma}_{N}^{Lomax}\right),$$

$$\left(\widehat{\alpha}_{S}, \widehat{\theta}_{S}\right) \;\; is \;\; \mathcal{AN}\left((\alpha,\theta+t),\, \frac{1}{n}\mathbf{\Sigma}_{S}^{Lomax}\right),$$

where entries for $\mathbf{\Sigma}_{T}^{Lomax} = \mathbf{\Sigma}_{S}^{Lomax}$ *and* $\mathbf{\Sigma}_{N}^{Lomax}$ *are provided in (A.1) and (A.2), respectively.*

Lognormal models:

$$\left(\widehat{\mu}_{T}, \widehat{\sigma}_{T}\right) \;\; is \;\; \mathcal{AN}\left((\mu,\sigma),\, \frac{1}{n}\mathbf{\Sigma}_{T}^{LN}\right),$$

$$\left(\widehat{\mu}_{N}, \widehat{\sigma}_{N}\right) \;\; is \;\; \mathcal{AN}\left(\left(\mu+\sigma\kappa, \sigma\sqrt{1+r\kappa-\kappa^2}\right),\, \frac{1}{n}\mathbf{\Sigma}_{N}^{LN}\right),$$

$$\left(\widehat{\mu}_{S}, \widehat{\sigma}_{S}\right) \;\; is \;\; \mathcal{AN}\left(\left(\mu_{S},\sigma_{S}\right),\, \frac{1}{n}\mathbf{\Sigma}_{S}^{LN}\right),$$

where entries for $\mathbf{\Sigma}_{S}^{LN}$, $\mathbf{\Sigma}_{N}^{LN}$, $\mathbf{\Sigma}_{T}^{LN}$ *and other terms are provided in (A.3)–(A.7) and (A.11).*

Proof: See Appendix A. □

Table 3.6: Parameter MLEs (with variance and covariance estimates in parentheses) of the exponential, Lomax and lognormal models, using truncated, naive, and shifted approaches.

Model	*Truncated*	*Naive*	*Shifted*
Exponential	$\widehat{\sigma} = 351,021$ (2.28×10^9)	$\widehat{\sigma} = 546,021$ (5.52×10^9)	$\widehat{\sigma} = 351,021$ (2.28×10^9)
Lomax	$\widehat{\alpha} = 1.91$ (0.569)	$\widehat{\alpha} = 22.51$ $(5,189.86)$	$\widehat{\alpha} = 1.91$ (0.569)
	$\widehat{\theta} = 151,234$ (3.84×10^{10})	$\widehat{\theta} = 11,735,899$ (1.54×10^{15})	$\widehat{\theta} = 346,234$ (3.84×10^{10})
	$(\widehat{cov}(\widehat{\alpha}, \widehat{\theta}) = 138,934)$	$(\widehat{cov}(\widehat{\alpha}, \widehat{\theta}) = 2.82 \times 10^9)$	$(\widehat{cov}(\widehat{\alpha}, \widehat{\theta}) = 138,934)$
Lognormal	$\widehat{\mu} = 10.06$ (2.349)	$\widehat{\mu} = 12.93$ (0.001)	$\widehat{\mu} = 11.81$ (0.042)
	$\widehat{\sigma} = 1.61$ (0.173)	$\widehat{\sigma} = 0.66$ (0.008)	$\widehat{\sigma} = 1.50$ (0.021)
	$(\widehat{cov}(\widehat{\mu}, \widehat{\sigma}) = -0.6008)$	$(\widehat{cov}(\widehat{\mu}, \widehat{\sigma}) = -0.0040)$	$(\widehat{cov}(\widehat{\mu}, \widehat{\sigma}) = -0.0004)$

3.3.3 Model Validation

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

In Figure 3.3, we present plots of the fitted-versus-observed quantiles for the nine models of Section 3.3.2. 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:

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

where $\widehat{G}^{-1}$ is the estimated parametric qf, $X_{(1)} \leq \cdots \leq X_{(54)}$ denote the ordered losses, and $u_i = (i - 0.5)/54$ is the quantile level. For the truncated approach, $\widehat{G}^{-1}(u_i) = \widehat{F}^{-1}\Big(u_i + \widehat{F}(195,000)(1-u_i)\Big)$; 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 3.6.

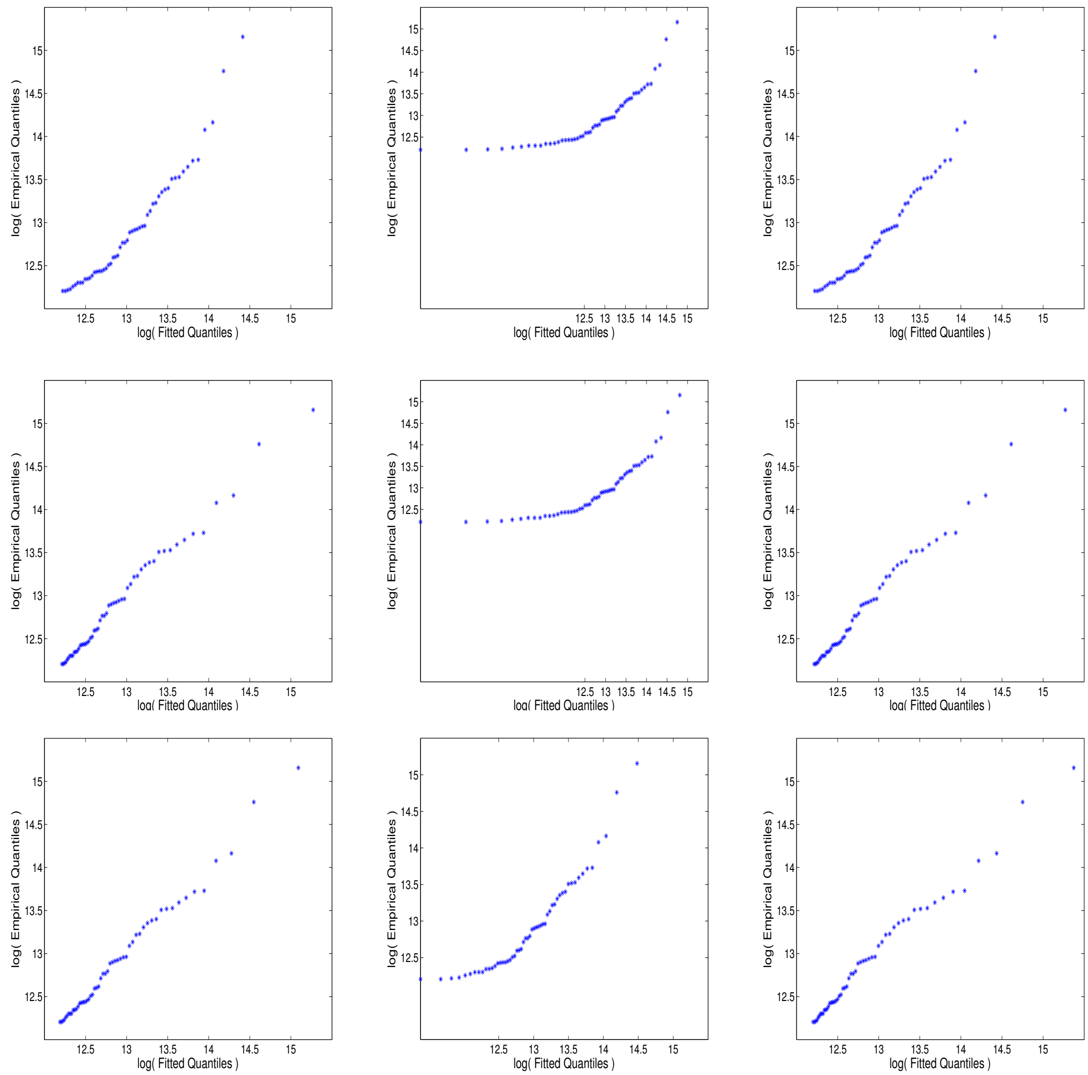

Figure 3.3: Fitted-versus-observed *log*-losses for exponential (top row), Lomax (middle row) and lognormal (bottom row) distributions, using truncated (left), naive (middle), and shifted (right) approaches.

We can see from Figure 3.3 that Lomax and lognormal 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 all distributions, with Lomax and lognormal 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

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

and

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

where $195,000 = X_{(0)} \leq X_{(1)} \leq \cdots \leq X_{(n)} \leq X_{(n+1)} = \infty$ 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 3.6. 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 3.7: Values of KS and AD statistics (with p-values in parentheses) for the fitted models, using truncated, naive, and shifted approaches.

	Kolmogorov-Smirnov			Anderson-Darling		
Model	*Truncated*	*Naive*	*Shifted*	*Truncated*	*Naive*	*Shifted*
Exponential	0.186	0.307	0.186	3.398	4.509	3.398
	(0.004)	(0.000)	(0.004)	(0.000)	(0.000)	(0.000)
Lomax	0.072	0.316	0.072	0.272	4.696	0.272
	(0.632)	(0.000)	(0.631)	(0.671)	(0.000)	(0.678)
Lognormal	0.068	0.136	0.086	0.244	1.614	0.308
	(0.744)	(0.013)	(0.390)	(0.793)	(0.000)	(0.584)

As the results of Table 3.7 suggest, all naive models are rejected by the KS and AD tests (at the 5% significance level), which is consistent with the conclusions based on Q-Q plots. The truncated and shifted exponential models are also rejected, which strengthens our "weak" decisions based on Q-Q plots. Unfortunately, for this data set, neither KS nor the AD test can help us with differentiating between the truncated and shifted Lomax and lognormal models as all of them fit the data very well.

3.3.4 VaR Estimates

Having fitted and validated the models, we now compute several point and interval estimates of VaR(β) for all nine models. The purpose of calculating VaR(β) 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 3.8, where, for completeness, empirical estimates of VaR(β) are also reported. The confidence intervals are derived using Theorem 2.1.4 and based on the variance and covariance estimates from Table 3.6. The following proposition summarizes asymptotic distributions of $\widehat{\mathrm{VaR}}(\beta)$.

Proposition 3.3.2. *[Asymptotic Distributions of VaR Estimators.]*

Exponential models:

$$\widehat{VaR}_{T}^{Exp}(\beta) \;\; is \;\; \mathcal{AN}\left(-\sigma\log(1-\beta),\; \frac{\sigma^2\log^2(1-\beta)}{n}\right),$$

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

$$\widehat{VaR}_{S}^{Exp}(\beta) \;\; is \;\; \mathcal{AN}\left(-\sigma\log(1-\beta)+t,\; \frac{\sigma^2\log^2(1-\beta)}{n}\right).$$

Lomax models:

$$\widehat{VaR}_{T}^{Lomax}(\beta) \;\; is \;\; \mathcal{AN}\left(\theta\left((1-\beta)^{-1/\alpha}-1\right),\; \boldsymbol{V}\left(\widehat{VaR}_{T}^{Lomax}(\beta)\right)\right),$$

$$\widehat{VaR}_{N}^{Lomax}(\beta) \;\; is \;\; \mathcal{AN}\left(\theta_{N}\left((1-\beta)^{-1/\alpha_N}-1\right),\; \boldsymbol{V}\left(\widehat{VaR}_{N}^{Lomax}(\beta)\right)\right),$$

$$\widehat{VaR}_{S}^{Lomax}(\beta) \;\; is \;\; \mathcal{AN}\left((\theta+t)\left((1-\beta)^{-1/\alpha}-1\right),\; \boldsymbol{V}\left(\widehat{VaR}_{S}^{Lomax}(\beta)\right)\right),$$

where variance formulas are provided in (A.12)–(A.14).

Lognormal models:

$$\widehat{VaR}_{T}^{LN}(\beta) \;\; is \;\; \mathcal{AN}\left(e^{\mu+\sigma\Phi^{-1}(\beta)},\; \boldsymbol{V}\left(\widehat{VaR}_{T}^{LN}(\beta)\right)\right),$$

$$\widehat{VaR}_{N}^{LN}(\beta) \;\; is \;\; \mathcal{AN}\left(e^{\mu+\sigma\kappa+\sigma\sqrt{1+r\kappa-\kappa^2}\Phi^{-1}(\beta)},\; \boldsymbol{V}\left(\widehat{VaR}_{N}^{LN}(\beta)\right)\right),$$

$$\widehat{VaR}_{S}^{LN}(\beta) \;\; is \;\; \mathcal{AN}\left(e^{\mu_S+\sigma_S\Phi^{-1}(\beta)},\; \boldsymbol{V}\left(\widehat{VaR}_{S}^{LN}(\beta)\right)\right),$$

where expressions for the variances and other terms are provided in (A.15)–(A.18).

Proof: See Appendix A. □

Table 3.8: Legal Risk: VaR(β) estimates (with 95% confidence intervals in parentheses), measured in millions and based on the fitted models, using truncated, naive, and shifted approaches.

Model	β	*Truncated*	*Naive*	*Shifted*
Exponential	0.95	1.052 (0.771; 1.332)	1.636 (1.199; 2.072)	1.247 (0.966; 1.527)
	0.99	1.617 (1.185; 2.048)	2.515 (1.844; 3.185)	1.812 (1.380; 2.243)
	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 (-0.126; 1.278)	1.670 (1.134; 2.206)	1.514 (0.689; 2.339)
	0.99	1.540 (0.101; 2.979)	2.664 (1.373; 3.954)	3.721 (-0.099; 7.540)
	0.999	5.504 (-0.037; 11.045)	4.214 (0.865; 7.564)	12.797 (-10.925; 36.519)
Lognormal	0.95	0.328 (-0.262; 0.919)	1.220 (0.886; 1.554)	1.768 (0.814; 2.723)
	0.99	0.981 (-0.365; 2.326)	1.912 (1.193; 2.631)	4.559 (1.238; 7.881)
	0.999	3.343 (-0.172; 6.858)	3.163 (1.576; 4.751)	13.889 (0.874; 26.905)

Empirical estimates of VaR(β): 1.416 (for $\beta = 0.95$) and 3.822 (for $\beta = 0.99$ and 0.999).

We see from Table 3.8 that the VaR(β) 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 and lognormal models, 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.

3.3.5 Model Predictions

As the final test of our models, we check their out-of-sample predictive power. Table B.1 (Appendix B) 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.

First of all, we report the estimated probabilities of losses below the data collection threshold, $\widehat{F}(195,000)$. 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). For the lognormal models it is 0.128 (naive) and 0.907 (truncated). Secondly, using these probabilities we can estimate the total, observed and unobserved, number of losses: $\widehat{N} = \dfrac{54}{1-\widehat{F}(195,000)}$. For the exponential models $\widehat{N} = 77.2 \approx 77$ (naive) and $\widehat{N} = 94.1 \approx 94$ (truncated). For the Lomax models $\widehat{N} = 78.3 \approx 78$ (naive) and $\widehat{N} = 262.1 \approx 262$ (truncated). For the lognormal models $\widehat{N} = 61.9 \approx 62$ (naive) and $\widehat{N} = 578.1 \approx 578$ (truncated). Note how different from the rest are the estimates of the truncated Lomax and lognormal models. By the way, one should not forget that these models exhibited the best statistical fit for the observed data.

For predictions that are verifiable, in Table 3.9 we report model-based estimates of the number of losses, the average loss, and the total loss in the interval $\big[150{,}000; 175{,}000\big]$. We also provide the corresponding 95% confidence intervals for the predictions. The intervals were constructed by using the variance and covariance estimates of Table 3.6 in conjunction with Theorem 2.1.4.

Notice that using the data points from Table B.1 it is straightforward to verify that

number of losses: 8
average loss: 156,627
total loss: 1,253,017.

We see from Table 3.9 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 and lognormal models but bad for the other models and/or approaches. As a consequence, 95% confidence intervals based on the truncated Lomax and lognormal models 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, all 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, 2.7, 2.1; total loss: 426,197, 441,155, 344,785).

Table 3.9: Legal Risk: Model-based predictions (with 95% confidence intervals in parentheses) of several statistics for the unobserved losses between \$150,000 and \$175,000.

Model	*Truncated*			*Naive*		
	number of losses	average loss	total loss	number of losses	average loss	total loss
Exponential	4.2	162,352	685,108	2.6	162,405	426,197
	(3.0; 5.5)	(162,312; 162,391)	(452,840; 917,376)	(1.9; 3.4)	(162,379; 162,430)	(141,592; 710,802)
Lomax	9.9	162,017	1,609,649	2.7	162,397	441,155
	(3.3; 16.5)	(161,647; 162,388)	(543,017; 2,676,281)	(1.8; 3.7)	(162,343; 162,451)	(288,324; 593,985)
Lognormal	10.7	161,938	1,736,367	2.1	162,868	344,785
	(6.8; 14.7)	(161,773; 162,103)	(1,095,893; 2,376,842)	(0.6; 3.6)	(162,411; 163,324)	(99,137; 590,432)

3.4 Preliminary Conclusions

In this chapter, 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 question that remains unanswered is: why are standard model validation tools unable to detect flaws of the shifted approach?

Chapter 4

Model Selection for Truncated and Shifted Distributions

The structure of this chapter is as follows. In Section 4.1, we give an introduction to several model selection techniques. After reviewing standard information criteria in Section 4.2, information complexity is explored in Section 4.3, where Fisher information matrices for the truncated and shifted versions of the Lomax, lognormal and Champernowne distributions are used to compute information complexity. Further, in Section 4.4, a new criterion based on the asymptotic mean curvature of the model log-likelihood surface is developed. Finally, in Section 4.5, practical performance of all information criteria presented in this chapter is studied by revisiting the case study of Section 3.3.

4.1 Introduction

Let us recall that the unanswered question from Chapter 3 was why various model validation tools are not able to distinguish between the truncated and shifted approaches, although the latter is fundamentally flawed for estimating operational risk capital. As mentioned in Section 1.3, this same question has been investigated by Cavallo, Rosenthal, Wang, Yan (2012), who applied Vuong's test and found that the truncated lognormal and shifted lognormal models may be statistically indistinguishable. Therefore, here we will revisit this challenging issue but approach it from a different perspective; it will be treated as a model selection problem. In addition to the lognormal distribution, we will also consider Champernowne and Lomax models.

As will be seen in the numerical illustrations of Section 4.5, the well-established model selection tools such as Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are not effective at solving the problem described above. Thus, we have to turn to more refined measures. In particular, we will first consider the information complexity techniques of Bozdogan (1988) and then propose a new criterion based on the asymptotic mean curvature of the log-likelihood surface.

Bozdogan (1988) has introduced so-called *information complexity* (ICOMP) using the inverse Fisher information matrix to penalize the complexity of variance-covariance structure of model's maximum likelihood estimators. His approach is a generalization of the maximal information measure proposed by van Emden (1971). Let $\mathbf{I}_n = n \cdot \mathbf{I}$ denote the Fisher information matrix based on a sample of size n. The ICOMP penalty term is a function of the matrix $\mathbf{I}$ rank, trace and determinant. Its minimum value is zero which is reached when the variances of parameter estimators are equal and the covariances are zeros. This criterion is very effective for regression-type models.

Our numerical experiments using the ICOMP criterion revealed, however, that penalizing the inequality in estimators' variances may not be appropriate for models that involve scale and shape parameters, as those parameters usually take values on quite different scales. A new criterion based on the asymptotic mean curvature (AMC) of the log-likelihood surface is structurally similar to ICOMP, but designed to penalize the flatness of the log-likelihood surface. Flat log-likelihood surfaces is a common technical problem when one deals with truncated distributions (Cope, 2011). Thus, in a two-parameter case, the mean curvature H is the average of two principal curvatures of the log-likelihood surface. When the normal vector is oriented towards above the surface, the mean curvature H is negative. It does approach zero from below for flat surfaces, which prompts us to define the penalty term in the AMC criterion as proportional to $\log(1-1/H)$ (penalty approaches ∞, when the surface becomes flatter).

4.2 Standard Information Criteria

Given a data set, $X_1, \ldots, X_n$, the amount of objective information contained in the data is fixed. However, different models may be fitted to the same data set to extract the information. In general, it is known that when the number of parameters increases, the model likelihood increases as well.

Let $\mathcal{L}\left(\theta_1, \ldots, \theta_k \,\middle|\, X_1, \ldots, X_n\right)$ be the likelihood function of a model with k parameters based on a sample of size n, and let $\widehat{\theta}_1, \ldots, \widehat{\theta}_k$ denote the corresponding MLE of those parameters. The Akaike Information Criterion (Akaike, 1973) and Bayesian Information Criterion (Schwarz, 1978) are defined as follows:

$$\mathrm{AIC} = -2\log\mathcal{L}\left(\widehat{\theta}_1, \ldots, \widehat{\theta}_k \,\middle|\, X_1, \ldots, X_n\right) + 2k,$$

$$\mathrm{BIC} = -2\log\mathcal{L}\left(\widehat{\theta}_1, \ldots, \widehat{\theta}_k \,\middle|\, X_1, \ldots, X_n\right) + k\log n.$$

Using these criteria, the preferred model is the one that minimizes AIC and BIC. Both criteria assume that the true model is included in the class of candidate models. Also, we see that there is a competition between the increase in the log-likelihood value and the increase in the number of model parameters in the AIC and BIC formulas. If the increase in the log-likelihood value is not sufficient to compensate the increase in the number of parameters, then it is not worthwhile to have the additional parameters. Note also that the BIC criterion penalizes the model dimensionality more than AIC for $\log n > 2$.

4.3 Information Complexity

The ICOMP criterion (Bozdogan, 1988) penalizes the interdependencies among parameter MLEs and is defined as follows:

$$\text{ICOMP} = -2\log\mathcal{L}\Big(\widehat{\theta}_1,\ldots,\widehat{\theta}_k\,\Big|\,X_1,\ldots,X_n\Big) + 2C_1\left(\mathbf{I}^{-1}(\widehat{\theta}_1,\ldots,\widehat{\theta}_k)\right), \tag{4.1}$$

where

$$C_1\left(\mathbf{I}^{-1}(\widehat{\theta}_1,\ldots,\widehat{\theta}_k)\right) = \frac{s}{2}\log\left(\frac{\text{tr}\left(\mathbf{I}^{-1}(\widehat{\theta}_1,\ldots,\widehat{\theta}_k)\right)}{s}\right) - \frac{1}{2}\log\left(\det\left(\mathbf{I}^{-1}(\widehat{\theta}_1,\ldots,\widehat{\theta}_k)\right)\right),$$

with s, tr, and det denoting the rank, trace, and determinant of $\mathbf{I}^{-1}$ (the inverse of Fisher information matrix), respectively. The log-likelihood function in (4.1) corresponds to either the truncated approach (given by equation (3.5)) or shifted approach (given by equation (3.7)) and can be evaluated for the models of Section 2.2. Likewise, the Fisher information matrix $\mathbf{I}(\theta_1,\ldots,\theta_k)$, defined in Theorem 2.1.3, will be derived using the truncated (3.5) and shifted (3.7) likelihoods for the same parametric distributions. The explicit formulas of ICOMP for the truncated and shifted Lomax, lognormal, and Champernowne distributions are provided in Appendix A. Specifically, ICOMP_T^{Lomax} and ICOMP_S^{Lomax} are given by (A.25) and (A.26), respectively; ICOMP_T^{LN} and ICOMP_S^{LN} are given by (A.27) and (A.28), respectively; ICOMP_T^{Champ} and ICOMP_S^{Champ} are given by (A.29) and (A.30), respectively.

4.4 Asymptotic Mean Curvature

Motivated by the ICOMP criterion, in this section we will derive and analyze the mean curvature of the log-likelihood surface, which has similarities with ICOMP. For the truncated distributions, one often needs to deal with flat log-likelihood surfaces (Cope, 2011), which may cause instability in MLE. The mean curvature is the average of principal curvatures. The curvature of a surface is related to the flatness/steepness of the surface. If the log-likelihood surface is flat, then convergence of numerical algorithms towards the surface maximum is very slow, leading to unstable parameter estimates. A steeper log-likelihood surface will have the log-likelihood value converge faster to its maximum value, leading to more stable parameter estimates. In this sense, it is better to have a steeper log-likelihood surface.

To simplify the derivations (for technical details, see for example, Raussen, 2008, p. 109–132), we will focus on the two-dimensional surfaces (denoted $\ell(\theta_1, \theta_2) = \ell(\boldsymbol{\theta})$), but the general version is not difficult to rederive. To find the mean curvature of the 2-parameter log-likelihood surface, let us start by defining

$$\vec{v} = (\theta_1,\, \theta_2, \ell(\theta_1,\, \theta_2)).$$

The first partial derivatives of $\vec{v}$ with respect to θ_1 and θ_2 are

$$\vec{v}_{\theta_1} = \left(1,\, 0,\, \frac{\partial \ell}{\partial \theta_1}\right), \quad \vec{v}_{\theta_2} = \left(0,\, 1,\, \frac{\partial \ell}{\partial \theta_2}\right),$$

respectively. The second partial derivatives are

$$\vec{v}_{\theta_1\theta_1} = \left(0,\, 0,\, \frac{\partial^2 \ell}{\partial \theta_1^2}\right), \quad \vec{v}_{\theta_1\theta_2} = \left(0,\, 0,\, \frac{\partial^2 \ell}{\partial \theta_1 \theta_2}\right), \quad \vec{v}_{\theta_2\theta_2} = \left(0,\, 0,\, \frac{\partial^2 \ell}{\partial \theta_2^2}\right).$$

Then, the normal vector of $\ell(\theta_1, \theta_2)$ is derived as follows:

$$A = \vec{v}_{\theta_1} \cdot \vec{v}_{\theta_1} = 1 + \left(\frac{\partial \ell}{\partial \theta_1}\right)^2,$$

$$B = \vec{v}_{\theta_1} \cdot \vec{v}_{\theta_2} = \left(\frac{\partial \ell}{\partial \theta_1}\right) \cdot \left(\frac{\partial \ell}{\partial \theta_2}\right),$$

$$C = \vec{v}_{\theta_2} \cdot \vec{v}_{\theta_2} = 1 + \left(\frac{\partial \ell}{\partial \theta_2}\right)^2.$$

$$\vec{v}_{\theta_1} \times \vec{v}_{\theta_2} = \left(-\frac{\partial \ell}{\partial \theta_1}, -\frac{\partial \ell}{\partial \theta_2}, 1\right),$$

$$\vec{n} = \frac{\vec{v}_{\theta_1} \times \vec{v}_{\theta_2}}{|\vec{v}_{\theta_1} \times \vec{v}_{\theta_2}|}$$

$$= \left(\frac{-\frac{\partial \ell}{\partial \theta_1}}{\sqrt{\left(\frac{\partial \ell}{\partial \theta_1}\right)^2 + \left(\frac{\partial \ell}{\partial \theta_2}\right)^2 + 1}}, \frac{-\frac{\partial \ell}{\partial \theta_2}}{\sqrt{\left(\frac{\partial \ell}{\partial \theta_1}\right)^2 + \left(\frac{\partial \ell}{\partial \theta_2}\right)^2 + 1}}, \frac{1}{\sqrt{\left(\frac{\partial \ell}{\partial \theta_1}\right)^2 + \left(\frac{\partial \ell}{\partial \theta_2}\right)^2 + 1}}\right).$$

The mean curvature H follows from these steps:

$$a = \vec{n} \cdot \vec{v}_{\theta_1\theta_1} = \frac{\frac{\partial^2 \ell}{\partial \theta_1^2}}{\sqrt{\left(\frac{\partial \ell}{\partial \theta_1}\right)^2 + \left(\frac{\partial \ell}{\partial \theta_2}\right)^2 + 1}},$$

$$b = \vec{n} \cdot \vec{v}_{\theta_1\theta_2} = \frac{\frac{\partial^2 \ell}{\partial \theta_1 \partial \theta_2}}{\sqrt{\left(\frac{\partial \ell}{\partial \theta_1}\right)^2 + \left(\frac{\partial \ell}{\partial \theta_2}\right)^2 + 1}},$$

$$c = \vec{n} \cdot \vec{v}_{\theta_2\theta_2} = \frac{\frac{\partial^2 \ell}{\partial \theta_2^2}}{\sqrt{\left(\frac{\partial \ell}{\partial \theta_1}\right)^2 + \left(\frac{\partial \ell}{\partial \theta_2}\right)^2 + 1}}.$$

$$H = \left(\frac{1}{2}\right)\left(\frac{aC - 2bB + Ac}{AC - B^2}\right).$$

Thus, the mean curvature of the two-parameter log-likelihood surface is:

$$H = \frac{\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \theta_1^2}\left[1 + \left(\frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_2}\right)^2\right] - 2\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \theta_1 \partial \theta_2}\frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_1}\frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_2} + \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \theta_2^2}\left[1 + \left(\frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_1}\right)^2\right]}{2\sqrt{\left(\left(\frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_1}\right)^2 + \left(\frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_2}\right)^2 + 1\right)^3}}. \tag{4.2}$$

Next, we formulate the following proposition to connect convergence in probability involving the mean curvature of $\ell(\boldsymbol{\theta})$ with an expression in terms of entries of the corresponding Fisher information matrix $\mathbf{I}$.

Proposition 4.4.1. *[Asymptotic Behavior of Mean Curvature of Two-parameter Log-likelihood Surfaces.]*

$$\frac{-H}{\sqrt{n}} - \frac{\frac{1}{n} \cdot tr(\mathbf{I}) + 2 \cdot det(\mathbf{I})}{2\sqrt{(tr(\mathbf{I}) + \frac{1}{n})^3}} \xrightarrow{\mathcal{P}} 0.$$

Proof: See Appendix A. □

Hence, in view of Proposition 4.4.1, we define the new information criterion as

$$\text{AMC} \quad = \quad -2\sum_{i=1}^{n} \log \widetilde{f}(X_i|\hat{\boldsymbol{\theta}}) + 2\log\left(1 - \frac{1}{\mathcal{H}}\right), \tag{4.3}$$

where

$$\mathcal{H} \quad = \quad \frac{n \cdot \text{tr}(\mathbf{I}) + 2n^2 \cdot \det(\mathbf{I})}{-2\sqrt{(n \cdot \text{tr}(\mathbf{I}) + 1)^3}}. \tag{4.4}$$

Note that by Proposition 4.4.1, $\mathcal{H}$ as defined in (4.4) behaves the same way as the mean curvature H in (4.2) asymptotically. The penalty term in (4.3) may be constructed using other transformations instead of the logarithmic transformation. Specific expressions of the AMC criterion for the truncated and shifted Lomax, lognormal and Champernowne distributions are provided in Appendix A (see formulas (A.31)–(A.36)).

4.5 Case Study: Legal Risk - Revisited

Here we revisit the case study of Section 3.3 and include Champernowne models in addition to Lomax and lognormal models. Suppose $Y_1, \ldots, Y_N$ are *i.i.d.* and follow a Champernowne distribution, with pdf, cdf, and qf given by (2.9), (2.10), and (2.11), respectively. Maximization of the log-likelihoods (in terms of the observed data $X_1, \ldots, X_n$) obtained by plugging Champernowne pdf and cdf into (3.5) and

(3.7),

$$\log \mathcal{L}_T = n \log \alpha + (\alpha - 1) \sum_{i=1}^{n} \log(X_i) + n \log(t^\alpha + M^\alpha) - 2 \sum_{i=1}^{n} \log(X_i^\alpha + M^\alpha), \quad (4.5)$$

$$\log \mathcal{L}_S = n \log \alpha + n\alpha \log(M) + (\alpha - 1) \sum_{i=1}^{n} \log(X_i - t) - 2 \sum_{i=1}^{n} \log((X_i - t)^\alpha + M^\alpha) \quad (4.6)$$

does not yield explicit formulas for MLEs, but they can be solved numerically. The resulting MLE values are: $\widehat{\alpha}_T = 1.609$, $\widehat{M}_T = 124,481$ (for truncated) and $\widehat{\alpha}_S = 1.152$, $\widehat{M}_S = 141,105$ (for shifted).

Further, both models pass the visual inspection of Q-Q plots (these will not be presented here) and goodness-of-fit tests. For the truncated model, the p-values are 0.799 (for KS) and 0.415 (for AD). And for the shifted model, they are 0.450 (KS) and 0.407 (AD).

Finally, in Table 4.1 we present the values of all information criteria considered in this chapter for truncated and shifted Champernowne, Lomax, and lognormal models.

Table 4.1: Legal Risk: Information measures for truncated and shifted Champernowne, lognormal, and Lomax models ($t = 195,000$, $n = 54$).

	Champernowne		Lognormal		Lomax	
Criterion	T	S	T	S	T	S
$-2 \log \mathcal{L}$	1,473	1,474	1,472	1,472	1,472	1,472
AIC	1,477	1,478	1,476	1,476	1,476	1,476
BIC	1,481	1,482	1,480	1,480	1,480	1,480
ICOMP	1,498	1,497	1,478	1,472	1,498	1,498
AMC	1,478	1,479	1,478	1,473	1,477	1,477

As we can see from the table, the six models are indistinguishable using traditional information criteria such as AIC and BIC, since the values of each criterion are very close for all models, with the shifted Champernowne model having a little bit larger value though. The use of new information measures (ICOMP and AMC) does not help with Lomax or Champernowne models, but has the ability to distinguish shifted lognormal model from the truncated lognormal model, favoring the shifted one since it has lower values of ICOMP and AMC. Moreover, if we compare across Lomax, lognormal, and Champernowne models, ICOMP values are close for Lomax and Champernowne models, but significantly differ from those of lognormal models. This is due to quite different scales of shape, scale, and location parameters, while ICOMP penalizes the inequality in variances of parameters. Note that this phenomenon disappears in AMC values. Despite being close in terms of information criteria and passing all model validation tests, the models produce very different VaR (0.99) estimates from $980,790$ (truncated lognormal) to $7,812,655$ (shifted Champernowne). The other models produce VaR estimates between $1,539,996$ and $4,704,211$.

Chapter 5

Conclusions and Future Research

5.1 Concluding Remarks

In this dissertation, we have concentrated on statistical aspects of operational risk modeling. Over the last couple decades, operational risk has emerged as a major risk factor for banking and insurance industries, public organizations, and private businesses. Due to a number of bankruptcies by well-established internationally-known institutions that were, in part, caused by operational risk events (e.g., collapse of Barings Bank in 1995 and Lehman Brothers in 2008), this risk is now viewed as a serious threat to global economy. In such environment regulation of banking and insurance industries is crucial, and it has been evolving rapidly. Basel III and Solvency II are the latest editions of international regulatory frameworks for banks and insurers, respectively. Moreover, the so-called AMA (Advanced Measurement Approaches) methodologies are now commonly applied to set regulatory minimum capital reserves. All this progress, however, generates new challenges, including statistical modeling challenges.

Within the AMA framework, the LDA (Loss Distribution Approach) is the most sophisticated tool for estimating operational risk capital which is defined as an extreme quantile of the annual aggregate loss distribution, and called value-at-risk or VaR. The LDA implementation, however, is tricky because the available loss severities used for constructing LDA models are observed above a certain (high) threshold. Such a setup provoked numerous discussions and disagreements on how to treat the data collection threshold in operational risk modeling. Several approaches for estimating VaR are documented in the academic and professional literatures: 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 leading to model uncertainty. It is worth mentioning here that the model uncertainty considered in this dissertation is an epistemic one, not a random uncertainty. As our careful analysis has shown, it can be reduced (but not completely eliminated) by employing sound model validation tools and in some cases 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).

To quantify the effect of model uncertainty on risk measurements, we used asymptotic theorems of mathematical statistics (e.g., asymptotic normality of sample quantiles, large-sample properties of maximum likelihood estimators, and the delta method), Monte Carlo simulations, and real-data examples. This effect has been evaluated by computing the probability of each approach producing conservative capital allocations based on the VaR measure. As specific parametric examples we have employed exponential, Lomax, lognormal and Champernowne distributions. For numerical illustrations, we have relied on a data set for legal losses in a business unit (Cruz, 2002). Similar to other authors who studied some aspects of this problem in the past, we have concluded 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 investigations we have conducted, however, raised new questions: why standard model validation tools are unable to distinguish between the truncated and shifted approaches? And how to choose between several truncated models that pass typical goodness-of-fit tests, but produce markedly different VaR estimates? In the second part of the dissertation, we have proceeded to explore these issues further treating them as a model selection problem. Using standard measures such as AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) we could not get satisfactory answers to the questions formulated above. This has prompted us to search for more refined measures. Therefore, we first have identified a criterion based on information complexity (Bozdogan, 1988) and then proposed a new criterion based on the asymptotic mean curvature of the model log-likelihood. Application of all model selection measures to the legal losses data have shown that these criteria are more effective at distinguishing between the competing models than AIC and BIC.

5.2 Future Research

The research presented in this dissertation 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 popular in practice distributions such as Burr or loggamma. If the chosen model lends itself to analytic investigations, then our approach in Section 3.2.2 is a blueprint for analysis. Otherwise, one may follow Section 3.2.3 for a simulations-based approach.

Another interesting direction would be to study further the newly proposed AMC (Asymptotic Mean Curvature) criterion. One could try to generalize it to arbitrary dimension surfaces and/or replace mean curvature with Gaussian curvature. Also, understanding statistical properties of AMC would help to refine this method. Overall, development of model-selection strategies for truncated, but not necessarily nested, models may be quite challenging due to flatness of the truncated likelihoods, a phenomenon frequently encountered in practice (see Cope, 2011). This raises a question whether it is worthwhile to stay with maximum likelihood procedures for fitting operational risk models or better pursue other parameter estimation methods.

Further, a research direction that has a number of desirable properties and may also help with the flat likelihood problem is robust model fitting. There are several excellent contributions to this topic in the operational risk literature (see Horbenko, Ruckdeschel, Bae, 2011, Opdyke and Cavallo, 2012, and Chau, 2013), but more work could be done. In particular, the method of trimmed moments introduced by Brazauskas, Jones, Zitikis (2009) and Brazauskas (2009) and extended to several loss severity models by Brazauskas and Kleefeld (2009, 2011, 2014) and Kleefeld and Brazauskas (2012) offers various degrees of robustness and is computationally efficient. However, it needs to be redesigned for truncated distributions and hence become applicable to operational risk models.

Finally, the ultimate goal of risk management is to measure and manage all relevant risks simultaneously in a consolidated framework, and to align actual risk and regulatory capital more closely (see, e.g., Tarullo, 2008, p. 158). Thus aggregation of the severity and frequency distributions into the compound process for a particular time frame is necessary. This is typically done using compound Poisson models, but other compound models could be explored as well. Finding operational VaR (a quantile of the aggregate distribution) then is an inversion problem that may involve numerical algorithms such as Monte Carlo approximations, fast Fourier transform, and Panjer recursion.

Bibliography

[1] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *Petrox, B.N., Csaki, F., Editors. Second International Symposium on Information Theory. Budapest. Academiai Kiado.* 267–281

[2] AMA Group (2013). *AMA Quantification Challenges: AMAG Range of Practice and Observations on "The Thorny LDA Topics"*. Risk Management Association.

[3] Arnold, B.C. (2015). *Pareto Distributions*, 2nd edition. Chapman & Hall.

[4] Bank for International Settlements (2011). Operational risk – supervisory guidelines for the advanced measurement approaches. *Bank for International Settlements* **6**, 1–55.

[5] 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.

[6] Bolance, C., Guillen, M., Gustafsson, J., Nielsen, J.P. (2012). *Quantitative Operational Risk Models*. CRC Press.

[7] Bozdogan, H. (1988). ICOMP: A new model-selection criterion. In Hans H. Bock (Ed.), Classification and related methods of data analysis, pp. 599–608. Amsterdam: Elsevier Science (North-Holland).

[8] Brazauskas, V. (2009). Robust and efficient fitting of loss models: diagnostic tools and insights. *North American Actuarial Journal* **13**(3), 356–369.

[9] Brazauskas, V., Jones, B., Zitikis, R. (2009). Robust fitting of claim severity distributions and the method of trimmed moments. *Journal of Statistical Planning and Inference* **139**(6), 2028–2043.

[10] Brazauskas, V., Kleefeld, A. (2009). Robust and efficient fitting of the generalized Pareto distribution with actuarial applications in view. *Insurance: Mathematics and Economics* **45**(3), 424–435.

[11] Brazauskas, V., Kleefeld, A. (2011). Folded- and log-folded-t distributions as models for insurance loss data. *Scandinavian Actuarial Journal* **2011**(1), 59–74.

[12] Brazauskas, V., Kleefeld, A. (2014). Authors' reply to "Letter to the Editor regarding folded models and the paper by Brazauskas and Kleefeld (2011)" by Scollnik. *Scandinavian Actuarial Journal* **2014**(8), 753–757.

[13] 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.

[14] Champernowne, D. G. (1936). The Oxford meeting, September 25–29. *Econometrica* **5**(October 1937).

[15] Champernowne, D. G. (1952). The graduation of income distributions. *Econometrica* **20**, 591–615.

[16] Chau, J. (2013). *Robust Estimation in Operational Risk Modeling*. Master's Thesis, Department of Mathematics, Utrecht University.

[17] Chernobai, A.S., Rachev, S.T., Fabozzi, F.J. (2007). *Operational Risk: A Guide to Basel II Capital Requirements, Models, and Analysis*. Wiley.

[18] Cope, E. (2011). Penalized likelihood estimators for truncated data. *Journal of Statistical Planning and Inference* **141**(1), 345–358.

[19] Cruz, M.G. (2002). *Modeling, Measuring and Hedging Operational Risk*. Wiley.

[20] Horbenko, N., Ruckdeschel, P., Bae, T. (2011). Robust estimation of operational risk. *Journal of Operational Risk* **6**(2), 3–30.

[21] Kleefeld, A., Brazauskas, V. (2012). A statistical application of the quantile mechanics approach: MTM estimators for the parameters of t and gamma distributions. *European Journal of Applied Mathematics* **23**(5), 593–610.

[22] Klugman, S.A., Panjer, H.H., Willmot, G.E. (2012). *Loss Models: From Data to Decisions*, 4th edition. Wiley.

[23] 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.

[24] 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.

[25] Office of the Comptroller of the Currency (2011). Supervisory guidance on model risk management. *SR Letter* **11**(7), 1–21.

[26] Opdyke, J.D. (2014). Estimating operational risk capital with greater accuracy, precision, and robustness. *Journal of Operational Risk* **9**(4), 3–79.

[27] 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.

[28] Raussen, M. (2008). *Elementary Differential Geometry: Curves and Surfaces*. http://people.math.aau.dk/∼raussen/INSB/AD2-11/book.pdf. Accessed on May 2, 2016.

[29] Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics* **6**, 461–464.

[30] Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley.

[31] Shao, J. (1999). *Mathematical Statistics*. Springer, New York.

[32] Shevchenko, P.V. (2011). *Modelling Operational Risk Using Bayesian Inference*. Springer.

[33] Tarullo, D. K. (2008). *Banking on Basel: The Future of International Financial Regulation*. Peterson Institute for International Economics, Washington, DC.

[34] van Emden, M.H. (1971). *An Analysis of Complexity*. In Mathematical Centre Tracts, volume 35. Mathematisch Centrum.

[35] Vuong, Q.H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* **57**(2), 307–333.

Appendix A: Proofs and Derivations

Proof of Proposition 3.3.1:

Exponential models: Recall (3.8):

$$\ell_T = \log \mathcal{L}_T = \sum_{i=1}^{n} \frac{-(X_i - t)}{\sigma} - n \log \sigma,$$

$$\frac{\partial \ell_T}{\partial \sigma} = \sum_{i=1}^{n} \frac{X_i - t}{\sigma^2} - \frac{n}{\sigma}.$$

Put $n = 1$ and use generic X:

$$\frac{\partial^2 \ell_T}{\partial \sigma^2} = \frac{-2(X - t)}{\sigma^3} + \frac{1}{\sigma^2}.$$

$$J = -\mathbf{E}\left[\frac{\partial^2 \ell_T}{\partial \sigma^2}\right] = \frac{2}{\sigma^3}\mathbf{E}\left[X - t\right] - \frac{1}{\sigma^2} = \frac{2\sigma}{\sigma^3} - \frac{1}{\sigma^2} = \frac{1}{\sigma^2},$$

$$\mathbf{E}\left(\overline{X}\right) = \mathbf{E}(X) = \sigma + t,$$

$$\widehat{\sigma}_T = \overline{X} - t, \quad \widehat{\sigma}_N = \overline{X}, \quad \widehat{\sigma}_S = \overline{X} - t,$$

$$\mathbf{E}\left(\widehat{\sigma}_T\right) = \sigma, \quad \mathbf{E}\left(\widehat{\sigma}_N\right) = \sigma + t, \quad \mathbf{E}\left(\widehat{\sigma}_S\right) = \sigma.$$

$$\mathbf{V}\left(\widehat{\sigma}_T\right) = \mathbf{V}\left(\widehat{\sigma}_N\right) = \mathbf{V}\left(\widehat{\sigma}_S\right) = \frac{1}{n} \cdot J^{-1} = \frac{\sigma^2}{n}.$$

Lomax models: Put $1/\gamma = \alpha$, $\mu = 0$, $\sigma/\gamma = \theta$ in (2.2) and (2.3) to get Lomax cdf and pdf:

$$F(x) = 1 - \left(\frac{\theta}{\theta + x}\right)^{\alpha}, \quad f(x) = \frac{\alpha\theta^{\alpha}}{(\theta + x)^{\alpha+1}}, \quad x > 0.$$

Truncated Lomax pdf:

$$f_*(x) = \frac{f(x)}{1 - F(t)} = \frac{\alpha(\theta + t)^{\alpha}}{(\theta + x)^{\alpha+1}}, \quad x > t.$$

Recall (3.11):

$$\ell_T = \log \mathcal{L}_T = n \log \alpha + n\alpha \log(\theta + t) - (\alpha + 1) \sum_{i=1}^{n} \log(\theta + X_i),$$

$$\begin{cases} \dfrac{\partial \ell_T}{\partial \alpha} & = \dfrac{n}{\alpha} + n \log(\theta + t) - \displaystyle\sum_{i=1}^{n} \log(\theta + X_i) = 0, \\ \dfrac{\partial \ell_T}{\partial \theta} & = \dfrac{n\alpha}{\theta + t} - (\alpha + 1) \displaystyle\sum_{i=1}^{n} \dfrac{1}{\theta + X_i} = 0, \end{cases}$$

eliminating α to obtain

$$n - \left[1 + \frac{1}{n} \sum_{i=1}^{n} \log\left(\frac{\theta + X_i}{\theta + t}\right)\right] \sum_{i=1}^{n} \frac{\theta + t}{\theta + X_i} = 0,$$

which is solved numerically for $\widehat{\theta}_T$, and

$$\widehat{\alpha}_T = \frac{1}{\displaystyle\frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{\widehat{\theta}_T + X_i}{\widehat{\theta}_T + t}\right)}.$$

Similarly,

$$\widehat{\alpha}_N = \frac{1}{\displaystyle\frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{\widehat{\theta}_N + X_i}{\widehat{\theta}_N}\right)},$$

$$\widehat{\alpha}_S = \frac{1}{\displaystyle\frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{\widehat{\theta}_S + X_i - t}{\widehat{\theta}_S}\right)}.$$

Note that

$$\widehat{\theta}_S = \widehat{\theta}_T + t, \quad \widehat{\alpha}_S = \widehat{\alpha}_T,$$

since

$$\mathbf{P}\Big\{X \le x \,\Big|\, X > t\Big\} \;=\; 1 - \left(\frac{\theta + t}{\theta + t + x - t}\right)^{\alpha}, \qquad x > t,$$

as shown in (2.4), which suggests that maximizing

$$\ell_T = \log \mathcal{L}_T = \sum_{i=1}^{n}\log\left(\frac{f(X_i)}{1 - F(t)}\right)$$

is equivalent to maximizing

$$\ell_S = \log \mathcal{L}_S = \sum_{i=1}^{n}\log f(X_i - t).$$

$$\mathbf{E}\left(\widehat{\alpha}_T\right) = \alpha, \quad \mathbf{E}\left(\widehat{\alpha}_N\right) = \alpha_N, \quad \mathbf{E}\left(\widehat{\alpha}_S\right) = \alpha,$$

$$\mathbf{E}\left(\widehat{\theta}_T\right) = \theta, \quad \mathbf{E}\left(\widehat{\theta}_N\right) = \theta_N, \quad \mathbf{E}\left(\widehat{\theta}_S\right) = \theta + t.$$

$$\boldsymbol{\Sigma} = \mathbf{I}^{-1}, \mathbf{I} = \begin{bmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{bmatrix}.$$

$$\mathbf{V}\left(\widehat{\alpha}\right) = \frac{1}{n} \cdot \frac{I_{22}}{I_{11}I_{22} - I_{12}^2},$$

$$\mathbf{V}\left(\widehat{\theta}\right) = \frac{1}{n} \cdot \frac{I_{11}}{I_{11}I_{22} - I_{12}^2},$$

$$\mathbf{COV}\left(\widehat{\alpha}, \widehat{\theta}\right) = \frac{1}{n} \cdot \frac{-I_{12}}{I_{11}I_{22} - I_{12}^2}.$$

To derive $\mathbf{I}$ for the truncated Lomax model, set $n = 1$ and use generic X.

$$I_{11} = -\mathbf{E}\left[\frac{\partial^2 \ell_T}{\partial \alpha^2}\right] = -\mathbf{E}\left[-\frac{1}{\alpha^2}\right] = \frac{1}{\alpha^2},$$

$$I_{12} = I_{21} = -\mathbf{E}\left[\frac{\partial^2 \ell_T}{\partial \alpha \partial \theta}\right] = -\mathbf{E}\left[\frac{1}{\theta + t} - \frac{1}{X + \theta}\right]$$

$$= -\frac{1}{\theta + t} + \int_t^\infty (x+\theta)^{-1} \frac{\alpha(\theta + t)^\alpha}{(\theta + x)^{\alpha+1}}\, dx$$

$$= -\frac{1}{\theta + t} + \frac{\alpha}{(t+\theta)^{-\alpha}} \cdot \frac{(t+\theta)^{-\alpha-1}}{\alpha + 1} = -\frac{1}{\theta + t} \cdot \frac{1}{\alpha + 1},$$

$$I_{22} = -\mathbf{E}\left[\frac{\partial^2 \ell_T}{\partial \theta^2}\right] = -\mathbf{E}\left[-\frac{\alpha}{(\theta + t)^2} + \frac{\alpha + 1}{(X + \theta)^2}\right]$$

$$= \frac{\alpha}{(\theta + t)^2} - (\alpha + 1)\int_t^\infty (x + \theta)^{-2} \cdot \frac{\alpha(\theta + t)^\alpha}{(\theta + x)^{\alpha+1}}\, dx$$

$$= \frac{\alpha}{(\theta + t)^{-\alpha}} \cdot \frac{(t+\theta)^{-\alpha-2}}{\alpha + 2} = \frac{\alpha}{(\alpha + 2)(t + \theta)^2},$$

$$\frac{1}{n}\mathbf{\Sigma}_{S}^{Lomax} = \begin{bmatrix} \mathbf{V}\left(\widehat{\alpha}_{S}\right) & \mathbf{COV}\left(\widehat{\alpha}_{S}, \widehat{\theta}_{S}\right) \\ \\ \mathbf{COV}\left(\widehat{\alpha}_{S}, \widehat{\theta}_{S}\right) & \mathbf{V}\left(\widehat{\theta}_{S}\right) \end{bmatrix}$$

$$= \frac{1}{n}\mathbf{\Sigma}_{T}^{Lomax} = \begin{bmatrix} \mathbf{V}\left(\widehat{\alpha}_{T}\right) & \mathbf{COV}\left(\widehat{\alpha}_{T}, \widehat{\theta}_{T}\right) \\ \\ \mathbf{COV}\left(\widehat{\alpha}_{T}, \widehat{\theta}_{T}\right) & \mathbf{V}\left(\widehat{\theta}_{T}\right) \end{bmatrix}$$

$$= \begin{bmatrix} \dfrac{\alpha^2(\alpha+1)^2}{n} & \dfrac{\alpha(\alpha+1)(\alpha+2)(t+\theta)}{n} \\ \\ \dfrac{\alpha(\alpha+1)(\alpha+2)(t+\theta)}{n} & \dfrac{(\alpha+1)^2(\alpha+2)(t+\theta)^2}{n\alpha} \end{bmatrix}, \qquad \text{(A.1)}$$

and similarly,

$$\frac{1}{n}\mathbf{\Sigma}_{N}^{Lomax} = \begin{bmatrix} \mathbf{V}\left(\widehat{\alpha}_{N}\right) & \mathbf{COV}\left(\widehat{\alpha}_{N}, \widehat{\theta}_{N}\right) \\ \\ \mathbf{COV}\left(\widehat{\alpha}_{N}, \widehat{\theta}_{N}\right) & \mathbf{V}\left(\widehat{\theta}_{N}\right) \end{bmatrix}$$

$$= \begin{bmatrix} \dfrac{\alpha^2(\alpha+1)^2}{n} & \dfrac{\alpha(\alpha+1)(\alpha+2)\theta}{n} \\ \\ \dfrac{\alpha(\alpha+1)(\alpha+2)\theta}{n} & \dfrac{(\alpha+1)^2(\alpha+2)\theta^2}{n\alpha} \end{bmatrix}. \qquad \text{(A.2)}$$

Lognormal models:

$$\mathbf{E}[\widehat{\mu}_{S}] = \mathbf{E}[\log(X-t)] = \mu_{S},$$

$$\mathbf{E}[\widehat{\sigma}_S^2] = \mathbf{V}[\log(X - t)] = \sigma_S^2.$$

$$\frac{1}{n}\mathbf{\Sigma}_S^{LN} = \begin{bmatrix} \mathbf{V}\left(\widehat{\mu}_S\right) & \mathbf{COV}\left(\widehat{\mu}_S,\, \widehat{\sigma}_S\right) \\ \\ \mathbf{COV}\left(\widehat{\mu}_S,\, \widehat{\sigma}_S\right) & \mathbf{V}\left(\widehat{\sigma}_S\right) \end{bmatrix}$$

$$= \begin{bmatrix} \dfrac{\sigma_S^2}{n} & 0 \\ \\ 0 & \dfrac{\sigma_S^2}{2n} \end{bmatrix}. \tag{A.3}$$

$$\begin{aligned} \mathbf{E}[\widehat{\mu}_N] &= \mathbf{E}[\log(X)] = \mathbf{E}[\log(Y) \,\big|\, Y > t] \\ &= \mathbf{E}[\mu + \sigma Z \,\big|\, Z > r] = \mu + \sigma\kappa, \end{aligned} \tag{A.4}$$

where

$$\kappa = \frac{\phi(r)}{1 - \Phi(r)}, \quad r = \frac{\log(t) - \mu}{\sigma}. \tag{A.5}$$

$$\begin{aligned} \mathbf{E}\left[\log^2 X\right] &= \mathbf{E}[\log^2(Y) \,\big|\, Y > t] = \mathbf{E}[(\mu + \sigma Z)^2 \,\big|\, Z > r] \\ &= \mu^2 + 2\mu\sigma\ \mathbf{E}[Z \,\big|\, Z > r] + \sigma^2\ \mathbf{E}[Z^2 \,\big|\, Z > r] \\ &= \mu^2 + 2\mu\sigma\kappa + \sigma^2\ \frac{r\phi(r) + 1 - \Phi(r)}{1 - \Phi(r)} \\ &= \mu^2 + 2\mu\sigma\kappa + \sigma^2(r\kappa + 1). \end{aligned}$$

$$\begin{aligned}\mathbf{E}[\widehat{\sigma}^2_{\scriptscriptstyle N}] &= \mathbf{V}[\log(X)] = \mathbf{E}[\log^2(X)] - \mathbf{E}[\log(X)]^2 \\ &= \mu^2 + 2\mu\sigma\kappa + \sigma^2(r\kappa+1) - (\mu+\sigma\kappa)^2 \\ &= \sigma^2(1 + r\kappa - \kappa^2). \end{aligned} \tag{A.6}$$

$$\begin{aligned}\frac{1}{n}\mathbf{\Sigma}^{\scriptscriptstyle LN}_{\scriptscriptstyle N} &= \begin{bmatrix} \mathbf{V}\left(\widehat{\mu}_{\scriptscriptstyle N}\right) & \mathbf{COV}\left(\widehat{\mu}_{\scriptscriptstyle N},\,\widehat{\sigma}_{\scriptscriptstyle N}\right) \\ & \\ \mathbf{COV}\left(\widehat{\mu}_{\scriptscriptstyle N},\,\widehat{\sigma}_{\scriptscriptstyle N}\right) & \mathbf{V}\left(\widehat{\sigma}_{\scriptscriptstyle N}\right) \end{bmatrix} \\ &= \frac{\sigma^2}{n}\cdot\begin{bmatrix} \dfrac{2+3r\kappa}{2+3r\kappa-4\kappa^2} & \dfrac{-2\kappa}{2+3r\kappa-4\kappa^2} \\ & \\ \dfrac{-2\kappa}{2+3r\kappa-4\kappa^2} & \dfrac{1}{2+3r\kappa-4\kappa^2} \end{bmatrix}. \end{aligned} \tag{A.7}$$

Recall (3.14):

$$\ell_{\scriptscriptstyle T} = \log\mathcal{L}_{\scriptscriptstyle T} = C_1 - n\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(\log X_i - \mu)^2 - n\log\left(1 - \Phi\left(\frac{\log t - \mu}{\sigma}\right)\right),$$

$$\begin{cases} \dfrac{\partial \ell_{\scriptscriptstyle T}}{\partial\mu} &= \dfrac{1}{\sigma^2}\displaystyle\sum_{i=1}^{n}(\log(X_i) - \mu) - \dfrac{n}{\sigma}\cdot\kappa, \\ \dfrac{\partial \ell_{\scriptscriptstyle T}}{\partial\sigma} &= \dfrac{-n}{\sigma} + \dfrac{1}{\sigma^3}\displaystyle\sum_{i=1}^{n}(\log(X_i) - \mu)^2 - \dfrac{n}{\sigma}\cdot r\kappa, \end{cases}$$

Put $n = 1$ and use generic X to derive $\mathbf{I}$:

$$\frac{\partial^2 \ell_T}{\partial \mu^2} = \frac{\partial}{\partial \mu}\left(\frac{\log X - \mu}{\sigma^2} - \frac{1}{\sigma} \cdot \frac{\phi\left(\frac{\log t - \mu}{\sigma}\right)}{1 - \Phi\left(\frac{\log t - \mu}{\sigma}\right)}\right)$$

$$= \frac{-1}{\sigma^2} + \frac{1}{\sigma^2} \cdot \frac{\frac{\log t - \mu}{\sigma^2} \cdot \phi\left(\frac{\log t - \mu}{\sigma}\right) \cdot (-\sigma)\left(1 - \Phi(r)\right) + \phi^2(r)}{\left(1 - \Phi(r)\right)^2}$$

$$= \frac{-1}{\sigma^2}\left(1 + r\kappa - \kappa^2\right),$$

$$\frac{\partial^2 \ell_T}{\partial \mu \partial \sigma} = \frac{\partial}{\partial \sigma}\left(\frac{\log X - \mu}{\sigma^2} - \frac{1}{\sigma} \cdot \frac{\phi\left(\frac{\log t - \mu}{\sigma}\right)}{1 - \Phi\left(\frac{\log t - \mu}{\sigma}\right)}\right)$$

$$= \frac{-2(\log X - \mu)}{\sigma^3} + \frac{\kappa}{\sigma^2}$$

$$- \frac{1}{\sigma} \cdot \frac{\frac{(\log t - \mu)^2}{\sigma^3} \cdot \phi\left(\frac{\log t - \mu}{\sigma}\right)\left(1 - \Phi(r)\right) + \phi(r)\left(-\frac{\log t - \mu}{\sigma^2}\right)\phi(r)}{\left(1 - \Phi(r)\right)^2}$$

$$= \frac{-2(\log X - \mu)}{\sigma^3} + \frac{\kappa}{\sigma^2} - \frac{r^2\kappa}{\sigma^2} + \frac{r\kappa^2}{\sigma^2},$$

$$\frac{\partial^2 \ell_T}{\partial \sigma^2} = \frac{\partial}{\partial \sigma}\left(\frac{-1}{\sigma} + \frac{(\log X - \mu)^2}{\sigma^3} - \frac{\log t - \mu}{\sigma^2} \cdot \frac{\phi\left(\frac{\log t - \mu}{\sigma}\right)}{1 - \Phi\left(\frac{\log t - \mu}{\sigma}\right)}\right)$$

$$= \frac{1}{\sigma^2} - \frac{3(\log X - \mu)^2}{\sigma^4} + \frac{2(\log t - \mu)}{\sigma^3} \cdot \kappa$$

$$- \frac{r}{\sigma} \cdot \frac{\frac{(\log t - \mu)^2}{\sigma^3} \cdot \phi\left(\frac{\log t - \mu}{\sigma}\right)\left(1 - \Phi(r)\right) + \phi(r)\left(-\frac{\log t - \mu}{\sigma^2}\right)\phi(r)}{\left(1 - \Phi(r)\right)^2}$$

$$= \frac{1}{\sigma^2} - \frac{3(\log X - \mu)^2}{\sigma^4} + \frac{2r\kappa}{\sigma^2} - \frac{r^3\kappa}{\sigma^2} + \frac{r^2\kappa^2}{\sigma^2},$$

$$I_{11} = -\mathbf{E}\left[\frac{\partial^2 \ell_T}{\partial \mu^2}\right] = \frac{1}{\sigma^2}\left(1 + r\kappa - \kappa^2\right), \tag{A.8}$$

$$I_{12} = I_{21} = -\mathbf{E}\left[\frac{\partial^2 \ell_{T}}{\partial\mu\partial\sigma}\right] = \frac{2}{\sigma^3}\cdot\mathbf{E}\left[\log X\right] - \frac{2\mu}{\sigma^3} - \frac{\kappa}{\sigma^2} + \frac{r^2\kappa}{\sigma^2} - \frac{r\kappa^2}{\sigma^2}$$

$$= \frac{2}{\sigma^3}(\mu+\sigma\kappa) - \frac{2\mu}{\sigma^3} - \frac{\kappa}{\sigma^2} + \frac{r^2\kappa}{\sigma^2} - \frac{r\kappa^2}{\sigma^2}$$

$$= \frac{\kappa(1+r^2-r\kappa)}{\sigma^2}, \tag{A.9}$$

$$I_{22} = -\mathbf{E}\left[\frac{\partial^2 \ell_{T}}{\partial\sigma^2}\right] = \frac{3\cdot\mathbf{E}\left[\log^2 X\right]}{\sigma^4} - \frac{6\mu\cdot\mathbf{E}\left[\log X\right]}{\sigma^4} + \frac{3\mu^2}{\sigma^4} + \frac{-1-2r\kappa+r^3\kappa-r^2\kappa^2}{\sigma^2}$$

$$= \frac{3(\mu^2+2\mu\sigma\kappa+\sigma^2(r\kappa+1))}{\sigma^4} - \frac{6\mu(\mu+\sigma\kappa)}{\sigma^4} + \frac{3\mu^2}{\sigma^4} + \frac{-1-2r\kappa+r^3\kappa-r^2\kappa^2}{\sigma^2}$$

$$= \frac{2+r\kappa+r^3\kappa-r^2\kappa^2}{\sigma^2}, \tag{A.10}$$

$$I_{11}I_{22} - I_{12}^2 = \frac{2+3r\kappa-2\kappa^2+r^3\kappa-r\kappa^3+r^4\kappa^2-2r^3\kappa^3+r^2\kappa^4}{\sigma^4}.$$

$$\frac{1}{n}\mathbf{\Sigma}_{T}^{LN} = \begin{bmatrix} \mathbf{V}\left(\widehat{\mu}_{T}\right) & \mathbf{COV}\left(\widehat{\mu}_{T},\,\widehat{\sigma}_{T}\right) \\ \mathbf{COV}\left(\widehat{\mu}_{T},\,\widehat{\sigma}_{T}\right) & \mathbf{V}\left(\widehat{\sigma}_{T}\right) \end{bmatrix}$$

$$= \frac{\sigma^2}{n}\cdot\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \tag{A.11}$$

where

$$a_{11} = \frac{2+r\kappa+r^3\kappa-r^2\kappa^2}{2+3r\kappa-2\kappa^2+r^3\kappa-r\kappa^3+r^4\kappa^2-2r^3\kappa^3+r^2\kappa^4},$$

$$a_{12} = a_{21} = \frac{-\kappa(1+r^2-r\kappa)}{2+3r\kappa-2\kappa^2+r^3\kappa-r\kappa^3+r^4\kappa^2-2r^3\kappa^3+r^2\kappa^4},$$

$$a_{22} = \frac{1 + r\kappa - \kappa^2}{2 + 3r\kappa - 2\kappa^2 + r^3\kappa - r\kappa^3 + r^4\kappa^2 - 2r^3\kappa^3 + r^2\kappa^4}.$$

□

Proof of Proposition 3.3.2:

Exponential models:

$$\mathbf{E}\left(\widehat{\sigma}_T\right) = \sigma, \quad \mathbf{V}\left(\widehat{\sigma}_T\right) = \frac{\sigma^2}{n}.$$

$$\mathbf{E}\left(\widehat{\sigma}_N\right) = \sigma + t, \quad \mathbf{V}\left(\widehat{\sigma}_N\right) = \frac{\sigma^2}{n}.$$

$$\mathbf{E}\left(\widehat{\sigma}_S\right) = \sigma, \quad \mathbf{V}\left(\widehat{\sigma}_S\right) = \frac{\sigma^2}{n}.$$

$$\widehat{\mathrm{VaR}}_T(\beta) = -\widehat{\sigma}_T \log(1-\beta),$$

$$\widehat{\mathrm{VaR}}_N(\beta) = -\widehat{\sigma}_N \log(1-\beta),$$

$$\widehat{\mathrm{VaR}}_S(\beta) = -\widehat{\sigma}_S \log(1-\beta) + t.$$

$$\mathbf{E}\left(\widehat{\mathrm{VaR}}_T(\beta)\right) = -\sigma \log(1-\beta),$$

$$\mathbf{E}\left(\widehat{\mathrm{VaR}}_N(\beta)\right) = -(\sigma + t) \log(1-\beta),$$

$$\mathbf{E}\left(\widehat{\mathrm{VaR}}_S(\beta)\right) = -\sigma \log(1-\beta) + t.$$

Differentiating $F^{-1}(\beta) = -\sigma \log(1-\beta)$, we get

$$\frac{\partial F^{-1}(\beta)}{\partial \sigma} = -\log(1-\beta).$$

Applying Theorem 2.1.4,

$$\mathbf{V}\left(\widehat{\mathrm{VaR}}_{T}(\beta)\right) = \mathbf{V}\left(\widehat{\mathrm{VaR}}_{N}(\beta)\right) = \mathbf{V}\left(\widehat{\mathrm{VaR}}_{T}(\beta)\right) = \frac{\sigma^2 \log^2(1-\beta)}{n}.$$

Lomax models:

$$\mathbf{E}\left(\widehat{\alpha}_{T}\right) = \alpha, \quad \mathbf{E}\left(\widehat{\theta}_{T}\right) = \theta, \quad \mathbf{E}\left(\widehat{\mathrm{VaR}}_{T}^{Lomax}(\beta)\right) = \theta\left((1-\beta)^{-1/\alpha} - 1\right).$$

Differentiating $F^{-1}(\beta) = \theta\left((1-\beta)^{-1/\alpha} - 1\right)$, we obtain

$$\frac{\partial F^{-1}(\beta)}{\partial \alpha} = \frac{\theta(1-\beta)^{-1/\alpha}\log(1-\beta)}{\alpha^2},$$

$$\frac{\partial F^{-1}(\beta)}{\partial \theta} = (1-\beta)^{-1/\alpha} - 1.$$

Since $\widehat{\mathrm{VaR}}_{T}^{Lomax}(\beta)$ is obtained by plugging unbiased MLE parameter estimates into $F^{-1}(\beta)$, applying Theorem 2.1.4,

$$\begin{aligned}\mathbf{V}\left(\widehat{\mathrm{VaR}}_{T}^{Lomax}(\beta)\right) = {} & \left(\frac{\partial F^{-1}(\beta)}{\partial \alpha}\right)^2 \mathbf{V}\left(\widehat{\alpha}_{T}\right) + \left(\frac{\partial F^{-1}(\beta)}{\partial \theta}\right)^2 \mathbf{V}\left(\widehat{\theta}_{T}\right) \\ & + 2\frac{\partial F^{-1}(\beta)}{\partial \alpha}\frac{\partial F^{-1}(\beta)}{\partial \theta}\mathbf{COV}\left(\widehat{\alpha}_{T}, \widehat{\theta}_{T}\right),\end{aligned} \tag{A.12}$$

where the variance-covariance matrix is given in (A.1).

$$\mathbf{E}\left(\widehat{\alpha}_{N}\right) = \alpha_{N}, \quad \mathbf{E}\left(\widehat{\theta}_{N}\right) = \theta_{N}, \quad \mathbf{E}\left(\widehat{\mathrm{VaR}}_{N}^{Lomax}(\beta)\right) = \theta_{N}\left((1-\beta)^{-1/\alpha_N} - 1\right).$$

$\widehat{\text{VaR}}_N^{Lomax}(\beta)$ is obtained by plugging biased MLE estimates of α and θ into $F^{-1}(\beta)$, and applying Theorem 2.1.4,

$$\begin{aligned}\mathbf{V}\left(\widehat{\text{VaR}}_N^{Lomax}(\beta)\right) = &\left(\frac{\theta_N(1-\beta)^{-1/\alpha_N}\log(1-\beta)}{\alpha_N^2}\right)^2 \mathbf{V}\left(\widehat{\alpha}_N\right) \\ &+ \left((1-\beta)^{-1/\alpha_N}-1\right)^2 \mathbf{V}\left(\widehat{\theta}_N\right) \\ &+ 2\left(\frac{\theta_N(1-\beta)^{-1/\alpha_N}\log(1-\beta)}{\alpha_N^2}\right)\left((1-\beta)^{-1/\alpha_N}-1\right) \\ &\times \mathbf{COV}\left(\widehat{\alpha}_N, \widehat{\theta}_N\right), \qquad (A.13)\end{aligned}$$

where the variance-covariance matrix is given in (A.2).

$$\mathbf{E}\left(\widehat{\alpha}_S\right) = \alpha, \quad \mathbf{E}\left(\widehat{\theta}_S\right) = \theta + t, \quad \mathbf{E}\left(\widehat{\text{VaR}}_S^{Lomax}(\beta)\right) = (\theta + t)\left((1-\beta)^{-1/\alpha} - 1\right).$$

$\widehat{\text{VaR}}_S^{Lomax}(\beta)$ is obtained by plugging unbiased MLE estimate of α and biased MLE estimate of θ into $F^{-1}(\beta) + t$, and applying Theorem 2.1.4,

$$\begin{aligned}\mathbf{V}\left(\widehat{\text{VaR}}_S^{Lomax}(\beta)\right) = &\left(\frac{\partial F^{-1}(\beta)}{\partial \alpha} + \frac{t(1-\beta)^{-1/\alpha}\log(1-\beta)}{\alpha^2}\right)^2 \mathbf{V}\left(\widehat{\alpha}_S\right) \\ &+ \left(\frac{\partial F^{-1}(\beta)}{\partial \theta}\right)^2 \mathbf{V}\left(\widehat{\theta}_S\right) \\ &+ 2\left(\frac{\partial F^{-1}(\beta)}{\partial \alpha} + \frac{t(1-\beta)^{-1/\alpha}\log(1-\beta)}{\alpha^2}\right)\frac{\partial F^{-1}(\beta)}{\partial \theta} \\ &\times \mathbf{COV}\left(\widehat{\alpha}_S, \widehat{\theta}_S\right), \qquad (A.14)\end{aligned}$$

where the variance-covariance matrix is given in (A.1).

Lognormal models: Differentiating $F^{-1}(\beta) = e^{\mu + \sigma\Phi^{-1}(\beta)}$, we obtain

$$\frac{\partial F^{-1}(\beta)}{\partial \mu} = e^{\mu + \sigma\Phi^{-1}(\beta)},$$

$$\frac{\partial F^{-1}(\beta)}{\partial \sigma} = \Phi^{-1}(\beta) \cdot e^{\mu + \sigma\Phi^{-1}(\beta)}.$$

$$\mathbf{E}\left(\widehat{\mu}_T\right) = \mu, \quad \mathbf{E}\left(\widehat{\sigma}_T\right) = \sigma, \quad \mathbf{E}\left(\widehat{\mathrm{VaR}}_T^{LN}(\beta)\right) = e^{\mu + \sigma\Phi^{-1}(\beta)}.$$

Applying Theorem 2.1.4,

$$\begin{aligned} \mathbf{V}\left(\widehat{\mathrm{VaR}}_T^{LN}(\beta)\right) &= \left(\frac{\partial F^{-1}(\beta)}{\partial \mu}\right)^2 \mathbf{V}\left(\widehat{\mu}_T\right) + \left(\frac{\partial F^{-1}(\beta)}{\partial \sigma}\right)^2 \mathbf{V}\left(\widehat{\sigma}_T\right) \\ &\quad + 2\frac{\partial F^{-1}(\beta)}{\partial \mu}\frac{\partial F^{-1}(\beta)}{\partial \sigma}\mathbf{COV}\left(\widehat{\mu}_T,\, \widehat{\sigma}_T\right), \end{aligned} \tag{A.15}$$

where the variance-covariance matrix is given in (A.11).

$$\mathbf{E}\left(\widehat{\mu}_N\right) = \mu + \sigma\kappa, \quad \mathbf{E}\left(\widehat{\sigma}_N\right) = \sigma\sqrt{1 + r\kappa - \kappa^2},$$

$$\mathbf{E}\left(\widehat{\mathrm{VaR}}_N^{LN}(\beta)\right) = e^{\mu + \sigma\kappa + \sigma\sqrt{1 + r\kappa - \kappa^2}\Phi^{-1}(\beta)}, \tag{A.16}$$

with r and κ as defined in (A.5). Applying Theorem 2.1.4,

$$\mathbf{V}\left(\widehat{\mathrm{VaR}}_{N}^{LN}(\beta)\right) = \left(e^{\mu + \sigma\kappa + \sigma\sqrt{1+r\kappa-\kappa^2}\Phi^{-1}(\beta)}\right)^2 \mathbf{V}\left(\widehat{\mu}_N\right)$$

$$+ \left(\left(\kappa + \sqrt{1+r\kappa-\kappa^2}\Phi^{-1}(\beta)\right) e^{\mu + \sigma\kappa + \sigma\sqrt{1+r\kappa-\kappa^2}\Phi^{-1}(\beta)}\right)^2$$

$$\times \mathbf{V}\left(\widehat{\sigma}_N\right)$$

$$+ 2\left(\kappa + \sqrt{1+r\kappa-\kappa^2}\Phi^{-1}(\beta)\right)\left(e^{\mu + \sigma\kappa + \sigma\sqrt{1+r\kappa-\kappa^2}\Phi^{-1}(\beta)}\right)^2$$

$$\times \mathbf{COV}\left(\widehat{\mu}_N, \widehat{\sigma}_N\right), \qquad (A.17)$$

where the variance-covariance matrix is given in (A.7).

$$\mathbf{E}\left(\widehat{\mu}_S\right) = \mu_S, \quad \mathbf{E}\left(\widehat{\sigma}_S\right) = \sigma_S, \quad \mathbf{E}\left(\widehat{\mathrm{VaR}}_{S}^{LN}(\beta)\right) = e^{\mu_S + \sigma_S\Phi^{-1}(\beta)}.$$

Applying Theorem 2.1.4,

$$\mathbf{V}\left(\widehat{\mathrm{VaR}}_{S}^{LN}(\beta)\right) = \left(e^{\mu_S + \sigma_S\Phi^{-1}(\beta)}\right)^2 \mathbf{V}\left(\widehat{\mu}_S\right)$$

$$+ \left(\Phi^{-1}(\beta)e^{\mu_S + \sigma_S\Phi^{-1}(\beta)}\right)^2 \mathbf{V}\left(\widehat{\sigma}_S\right)$$

$$+ 2\Phi^{-1}(\beta)\left(e^{\mu_S + \sigma_S\Phi^{-1}(\beta)}\right)^2 \mathbf{COV}\left(\widehat{\mu}_S, \widehat{\sigma}_S\right), \qquad (A.18)$$

where the variance-covariance matrix is given in (A.3). □

Derivation of I for Champernowne models:

Truncated Champernowne pdf and cdf:

$$f_*(x) = \frac{f(x)}{1 - F(t)} = \frac{\alpha x^{\alpha-1}(t^\alpha + M^\alpha)}{(x^\alpha + M^\alpha)^2}, \quad x > t,$$

$$F_*(x) = \frac{F(x)}{1 - F(t)} = \frac{x^\alpha(t^\alpha + M^\alpha)}{M^\alpha(x^\alpha + M^\alpha)}, \quad x > t.$$

Recall (4.5):

$$\ell_T = n \log \alpha + (\alpha - 1) \sum_{i=1}^{n} \log(X_i) + n \log(t^\alpha + M^\alpha) - 2 \sum_{i=1}^{n} \log(X_i^\alpha + M^\alpha),$$

$$\begin{cases} \dfrac{\partial \ell_T}{\partial \alpha} & = \dfrac{n}{\alpha} + \displaystyle\sum_{i=1}^{n} \log X_i + n \cdot \dfrac{t^\alpha \log t + M^\alpha \log M}{t^\alpha + M^\alpha} - 2 \displaystyle\sum_{i=1}^{n} \dfrac{X_i^\alpha \log X_i + M^\alpha \log M}{X_i^\alpha + M^\alpha}, \\ \\ \dfrac{\partial \ell_T}{\partial M} & = n \cdot \dfrac{\alpha M^{\alpha-1}}{t^\alpha + M^\alpha} - 2 \displaystyle\sum_{i=1}^{n} \dfrac{\alpha M^{\alpha-1}}{X_i^\alpha + M^\alpha}, \end{cases}$$

Put $n = 1$ and use generic X:

$$\frac{\partial^2 \ell_T}{\partial \alpha^2} = \frac{-1}{\alpha^2} + \frac{t^\alpha M^\alpha \left(\log t - \log M\right)^2}{\left(t^\alpha + M^\alpha\right)^2} - 2 \cdot \frac{X^\alpha M^\alpha \left(\log X - \log M\right)^2}{\left(X^\alpha + M^\alpha\right)^2},$$

$$\frac{\partial^2 \ell_T}{\partial M \partial \alpha} = \frac{-2M^{\alpha-1}}{X^\alpha + M^\alpha} + \frac{2\alpha M^{\alpha-1} X^\alpha \left(\log X - \log M\right)}{\left(X^\alpha + M^\alpha\right)^2}$$

$$+ \frac{M^{\alpha-1}}{t^\alpha + M^\alpha} - \frac{\alpha M^{\alpha-1} t^\alpha \left(\log t - \log M\right)}{\left(t^\alpha + M^\alpha\right)^2},$$

$$\frac{\partial^2 \ell_T}{\partial M^2} = -2 \cdot \frac{\alpha(\alpha-1)M^{\alpha-2}}{X^\alpha + M^\alpha} + 2 \cdot \frac{(\alpha M^{\alpha-1})^2}{\left(X^\alpha + M^\alpha\right)^2}$$

$$+ \frac{\alpha(\alpha-1)M^{\alpha-2}}{t^\alpha + M^\alpha} - \frac{(\alpha M^{\alpha-1})^2}{\left(t^\alpha + M^\alpha\right)^2},$$

$$\mathbf{E}\left[\frac{X^\alpha M^\alpha \left(\log X - \log M\right)^2}{\left(X^\alpha + M^\alpha\right)^2}\right] = \int_t^\infty \frac{x^\alpha M^\alpha \left(\log x - \log M\right)^2}{\left(x^\alpha + M^\alpha\right)^2} \; d\frac{x^\alpha (t^\alpha + M^\alpha)}{M^\alpha (x^\alpha + M^\alpha)}$$

$$= \frac{t^\alpha + M^\alpha}{M^\alpha} \int_{(t/M)^\alpha}^\infty \frac{w \log^2 w}{\alpha^2 \left(w+1\right)^2} \; d\frac{w}{w+1}$$

$$= \frac{t^\alpha + M^\alpha}{6\alpha^2 M^\alpha} \int_{(t/M)^\alpha}^\infty \log^2 w \cdot \frac{6w}{\left(w+1\right)^4} \; dw,$$

$$I_{11} = -\mathbf{E}\left[\frac{\partial^2 \ell_{\scriptscriptstyle T}}{\partial \alpha^2}\right] = \frac{1}{\alpha^2} - \frac{t^\alpha M^\alpha \left(\log t - \log M\right)^2}{\left(t^\alpha + M^\alpha\right)^2} + 2\cdot\mathbf{E}\left[\frac{X^\alpha M^\alpha \left(\log X - \log M\right)^2}{\left(X^\alpha + M^\alpha\right)^2}\right]$$

$$= \frac{1}{\alpha^2} - \frac{t^\alpha M^\alpha \left(\log t - \log M\right)^2}{\left(t^\alpha + M^\alpha\right)^2}$$

$$+ \frac{t^\alpha + M^\alpha}{3\alpha^2 M^\alpha}\int_{(t/M)^\alpha}^{\infty} \log^2 w \cdot \frac{6w}{\left(w+1\right)^4}\, dw, \tag{A.19}$$

$$\mathbf{E}\left[\frac{X^\alpha M^{\alpha-1}\left(\log X - \log M\right)}{\left(X^\alpha + M^\alpha\right)^2}\right] = \int_t^\infty \frac{x^\alpha M^{\alpha-1}\left(\log x - \log M\right)}{\left(x^\alpha + M^\alpha\right)^2}\, d\frac{x^\alpha(t^\alpha + M^\alpha)}{M^\alpha(x^\alpha + M^\alpha)}$$

$$= \frac{t^\alpha + M^\alpha}{M^{\alpha+1}}\int_{(t/M)^\alpha}^{\infty} \frac{w\log w}{\alpha\left(w+1\right)^2}\, d\frac{w}{w+1}$$

$$= \frac{t^\alpha + M^\alpha}{6\alpha M^{\alpha+1}}\int_{(t/M)^\alpha}^{\infty} \log w \cdot \frac{6w}{\left(w+1\right)^4}\, dw,$$

$$\mathbf{E}\left[\frac{1}{X^\alpha + M^\alpha}\right] = \int_t^\infty \frac{1}{x^\alpha + M^\alpha}\, d\frac{x^\alpha(t^\alpha + M^\alpha)}{M^\alpha(x^\alpha + M^\alpha)}$$

$$= \frac{t^\alpha + M^\alpha}{M^{2\alpha}}\int_{(t/M)^\alpha}^{\infty} \frac{1}{\left(w+1\right)^3}\, dw$$

$$= \frac{t^\alpha + M^\alpha}{2M^{2\alpha}((t/M)^\alpha + 1)^2} = \frac{1}{2(t^\alpha + M^\alpha)},$$

$$I_{12} = I_{21} = -\mathbf{E}\left[\frac{\partial^2 \ell_T}{\partial M \partial \alpha}\right]$$

$$= 2M^{\alpha-1} \cdot \mathbf{E}\left[\frac{1}{X^\alpha + M^\alpha}\right] - 2\alpha \cdot \mathbf{E}\left[\frac{X^\alpha M^{\alpha-1}\left(\log X - \log M\right)}{\left(X^\alpha + M^\alpha\right)^2}\right]$$

$$- \frac{M^{\alpha-1}}{t^\alpha + M^\alpha} + \frac{\alpha M^{\alpha-1} t^\alpha \left(\log t - \log M\right)}{\left(t^\alpha + M^\alpha\right)^2}$$

$$= -\frac{t^\alpha + M^\alpha}{3M^{\alpha+1}} \int_{(t/M)^\alpha}^{\infty} \log w \cdot \frac{6w}{\left(w+1\right)^4}\; dw$$

$$+ \frac{\alpha M^{\alpha-1} t^\alpha \left(\log t - \log M\right)}{\left(t^\alpha + M^\alpha\right)^2}, \tag{A.20}$$

$$\mathbf{E}\left[\frac{1}{(X^\alpha + M^\alpha})^2\right] = \int_t^\infty \frac{1}{(x^\alpha + M^\alpha)^2}\; d\frac{x^\alpha (t^\alpha + M^\alpha)}{M^\alpha (x^\alpha + M^\alpha)}$$

$$= \frac{t^\alpha + M^\alpha}{M^{3\alpha}} \int_{(t/M)^\alpha}^{\infty} \frac{1}{\left(w+1\right)^4}\; dw$$

$$= \frac{t^\alpha + M^\alpha}{3M^{3\alpha}((t/M)^\alpha + 1)^3} = \frac{1}{3(t^\alpha + M^\alpha)^2},$$

$$
\begin{aligned}
I_{22} &= -\mathbf{E}\left[\frac{\partial^2 \ell_{T}}{\partial M^2}\right] \\
&= 2\alpha(\alpha-1)M^{\alpha-2}\cdot \mathbf{E}\left[\frac{1}{X^\alpha + M^\alpha}\right] - 2(\alpha M^{\alpha-1})^2 \cdot \mathbf{E}\left[\frac{1}{(X^\alpha + M^\alpha}\right)^2\Bigg] \\
&- \frac{\alpha(\alpha-1)M^{\alpha-2}}{t^\alpha + M^\alpha} + \frac{(\alpha M^{\alpha-1})^2}{(t^\alpha + M^\alpha)^2} \\
&= 2\alpha(\alpha-1)M^{\alpha-2}\cdot \frac{1}{2(t^\alpha + M^\alpha)} - 2(\alpha M^{\alpha-1})^2 \cdot \frac{1}{3(t^\alpha + M^\alpha)^2} \\
&- \frac{\alpha(\alpha-1)M^{\alpha-2}}{t^\alpha + M^\alpha} + \frac{(\alpha M^{\alpha-1})^2}{(t^\alpha + M^\alpha)^2} \\
&= \frac{(\alpha M^{\alpha-1})^2}{3\left(t^\alpha + M^\alpha\right)^2}. \qquad \text{(A.21)}
\end{aligned}
$$

Shifted Champernowne pdf:

$$
f(x-t) = \frac{\alpha M^\alpha (x-t)^{\alpha-1}}{((x-t)^\alpha + M^\alpha)^2}, \quad x > t
$$

Recall (4.6):

$$
\ell_{S} = n\log\alpha + n\alpha\log M + (\alpha-1)\sum_{i=1}^{n}\log\left(X_i - t\right) - 2\sum_{i=1}^{n}\log((X_i - t)^\alpha + M^\alpha),
$$

$$\begin{cases} \dfrac{\partial \ell_S}{\partial \alpha} & = \dfrac{n}{\alpha} + n \log M + \displaystyle\sum_{i=1}^{n} \log(X_i - t) - 2 \sum_{i=1}^{n} \frac{(X_i - t)^\alpha \log(X_i - t) + M^\alpha \log M}{(X_i - t)^\alpha + M^\alpha}, \\ \\ \dfrac{\partial \ell_S}{\partial M} & = \dfrac{n\alpha}{M} - 2 \displaystyle\sum_{i=1}^{n} \frac{\alpha M^{\alpha-1}}{(X_i - t)^\alpha + M^\alpha}, \end{cases}$$

Put $n = 1$ and use generic X:

$$\frac{\partial^2 \ell_S}{\partial \alpha^2} = \frac{-1}{\alpha^2} - 2 \cdot \frac{(X - t)^\alpha \log^2(X - t) + M^\alpha \log^2 M}{(X - t)^\alpha + M^\alpha}$$

$$+ 2 \cdot \frac{\left((X - t)^\alpha \log(X - t) + M^\alpha \log M\right)^2}{\left((X - t)^\alpha + M^\alpha\right)^2},$$

$$\frac{\partial^2 \ell_S}{\partial M \partial \alpha} = \frac{1}{M} - 2 \cdot \frac{M^{\alpha-1} + \alpha M^{\alpha-1} \log M}{(X - t)^\alpha + M^\alpha}$$

$$+ 2 \cdot \frac{\alpha M^{\alpha-1} \left((X - t)^\alpha \log(X - t) + M^\alpha \log M\right)}{\left((X - t)^\alpha + M^\alpha\right)^2},$$

$$\frac{\partial^2 \ell_S}{\partial M^2} = \frac{-\alpha}{M^2} - 2 \cdot \frac{\alpha(\alpha - 1) M^{\alpha-2}}{(X - t)^\alpha + M^\alpha} + 2 \cdot \frac{(\alpha M^{\alpha-1})^2}{\left((X - t)^\alpha + M^\alpha\right)^2},$$

$$I_{11} = -\mathbf{E}\left[\frac{\partial^2 \ell_S}{\partial \alpha^2}\right] = \frac{1}{\alpha^2} + 2\mathbf{E}\left[\frac{(X-t)^\alpha \log^2(X-t) + M^\alpha \log^2 M}{(X-t)^\alpha + M^\alpha}\right]$$

$$- 2\mathbf{E}\left[\frac{\left((X-t)^\alpha \log(X-t) + M^\alpha \log M\right)^2}{\left((X-t)^\alpha + M^\alpha\right)^2}\right]$$

$$= \frac{1}{\alpha^2} + \frac{\log^2 M}{3} + 2\int_0^\infty \frac{\alpha M^\alpha y^{2\alpha-1} \log^2 y}{(y^\alpha + M^\alpha)^3}\, dy$$

$$- 2\int_0^\infty \frac{\alpha M^\alpha y^{3\alpha-1} \log^2 y}{(y^\alpha + M^\alpha)^4}\, dy - 4\int_0^\infty \frac{\alpha M^{2\alpha} \log M y^{2\alpha-1} \log y}{(y^\alpha + M^\alpha)^4}\, dy, \tag{A.22}$$

$$I_{12} = I_{21} = -\mathbf{E}\left[\frac{\partial^2 \ell_S}{\partial M \partial \alpha}\right]$$

$$= \frac{-1}{M} + 2\mathbf{E}\left[\frac{M^{\alpha-1} + \alpha M^{\alpha-1} \log M}{(X-t)^\alpha + M^\alpha}\right]$$

$$- 2\mathbf{E}\left[\frac{\alpha M^{\alpha-1}\left((X-t)^\alpha \log(X-t) + M^\alpha \log M\right)}{\left((X-t)^\alpha + M^\alpha\right)^2}\right]$$

$$= \frac{\alpha \log M}{3M} - 2\int_0^\infty \frac{\alpha^2 M^{2\alpha-1} y^{2\alpha-1} \log y}{(y^\alpha + M^\alpha)^4}\, dy, \tag{A.23}$$

$$
\begin{aligned}
I_{22} &= -\mathbf{E}\left[\frac{\partial^2 \ell_S}{\partial M^2}\right] \\
&= \frac{\alpha}{M^2} + 2\mathbf{E}\left[\frac{\alpha(\alpha-1)M^{\alpha-2}}{(X-t)^\alpha + M^\alpha}\right] - 2\mathbf{E}\left[\frac{(\alpha M^{\alpha-1})^2}{\left((X-t)^\alpha + M^\alpha\right)^2}\right] \\
&= \frac{\alpha^2}{3M^2}. \qquad \text{(A.24)}
\end{aligned}
$$

Formulas for ICOMP and AMC:

Rewriting (4.1) as

$$
\text{ICOMP} = -2\log\mathcal{L} + 2\log(I_{11} + I_{22}) - \log(I_{11}I_{22} - I_{12}^2) - 2\log(2),
$$

$$
\begin{aligned}
\text{ICOMP}_T^{Lomax} &= -2\left(n\log\alpha + n\alpha\log(\theta+t) - (\alpha+1)\sum_{i=1}^{n}\log(\theta+X_i)\right) \\
&+ 2\log\left(\frac{1}{\alpha^2} + \frac{\alpha}{(\alpha+2)(t+\theta)^2}\right) \\
&- \log\left(\frac{1}{\alpha(\alpha+1)^2(\alpha+2)(t+\theta)^2}\right) - 2\log(2), \qquad \text{(A.25)}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{ICOMP}_S^{Lomax} = -2 &\left(n \log \alpha + n\alpha \log \theta_S - (\alpha+1) \sum_{i=1}^{n} \log\left(\theta_S + X_i - t\right) \right) \\
&+ 2 \log \left(\frac{1}{\alpha^2} + \frac{\alpha}{(\alpha+2)(\theta_S)^2} \right) \\
&- \log \left(\frac{1}{\alpha(\alpha+1)^2(\alpha+2)(\theta_S)^2} \right) - 2\log(2), \qquad \text{(A.26)}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{ICOMP}_T^{LN} = -2 &\Bigg(-n \log \sqrt{2\pi} - \sum_{i=1}^{n} \log X_i - n \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(\log X_i - \mu\right)^2 \\
&- n \log \left(1 - \Phi \left(\frac{\log t - \mu}{\sigma} \right) \right) \Bigg) + 2 \log \left(\frac{3 + 2r\kappa - \kappa^2 + r^3\kappa - r^2\kappa^2}{\sigma^2} \right) \\
&- \log \left(\frac{2 + 3r\kappa - 2\kappa^2 + r^3\kappa - r\kappa^3 + r^4\kappa^2 - 2r^3\kappa^3 + r^2\kappa^4}{\sigma^4} \right) \\
&- 2\log(2), \qquad \text{(A.27)}
\end{aligned}
$$

with r and κ as defined in (A.5).

$$
\begin{aligned}
\mathrm{ICOMP}_S^{LN} = -2 &\Bigg(-n \log \sqrt{2\pi} - \sum_{i=1}^{n} \log(X_i - t) - n \log \sigma_S \\
&- \frac{1}{2\sigma_S^2} \sum_{i=1}^{n} \left(\log(X_i - t) - \mu_S\right)^2 \Bigg) + 2 \log \left(\frac{3}{\sigma_S^2} \right) - \log \left(\frac{2}{\sigma_S^4} \right) \\
&- 2\log(2), \qquad \text{(A.28)}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{ICOMP}_T^{Champ} = &-2\Bigg(n\log\alpha + (\alpha-1)\sum_{i=1}^{n}\log(X_i) + n\log(t^\alpha + M^\alpha) \\
&- 2\sum_{i=1}^{n}\log(X_i^\alpha + M^\alpha)\Bigg) \\
&+ 2\log\Bigg(\frac{1}{\alpha^2} - \frac{t^\alpha M^\alpha \left(\log t - \log M\right)^2}{\left(t^\alpha + M^\alpha\right)^2} \\
&+ \frac{t^\alpha + M^\alpha}{3\alpha^2 M^\alpha}\int_{(t/M)^\alpha}^{\infty} \log^2 w \cdot \frac{6w}{\left(w+1\right)^4}\, dw + \frac{(\alpha M^{\alpha-1})^2}{3\left(t^\alpha + M^\alpha\right)^2}\Bigg) \\
&- \log\Bigg(\bigg(\frac{1}{\alpha^2} - \frac{t^\alpha M^\alpha \left(\log t - \log M\right)^2}{\left(t^\alpha + M^\alpha\right)^2} \\
&+ \frac{t^\alpha + M^\alpha}{3\alpha^2 M^\alpha}\int_{(t/M)^\alpha}^{\infty} \log^2 w \cdot \frac{6w}{\left(w+1\right)^4}\, dw\bigg)\cdot\bigg(\frac{(\alpha M^{\alpha-1})^2}{3\left(t^\alpha + M^\alpha\right)^2}\bigg) \\
&- \bigg(-\frac{t^\alpha + M^\alpha}{3M^{\alpha+1}}\int_{(t/M)^\alpha}^{\infty} \log w \cdot \frac{6w}{\left(w+1\right)^4}\, dw \\
&+ \frac{\alpha M^{\alpha-1} t^\alpha \left(\log t - \log M\right)}{\left(t^\alpha + M^\alpha\right)^2}\bigg)^2\Bigg) \\
&- 2\log(2), \qquad \text{(A.29)}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{ICOMP}_{S}^{Champ} &= -2\Bigg(n\log\alpha + n\alpha\log(M) + (\alpha-1)\sum_{i=1}^{n}\log(X_i - t) \\
&- 2\sum_{i=1}^{n}\log((X_i-t)^{\alpha} + M^{\alpha})\Bigg) \\
&+ 2\log\Bigg(\frac{1}{\alpha^2} + \frac{\log^2 M}{3} + 2\int_0^{\infty} \frac{\alpha M^{\alpha} y^{2\alpha-1}\log^2 y}{(y^{\alpha}+M^{\alpha})^3}\, dy \\
&- 2\int_0^{\infty} \frac{\alpha M^{\alpha} y^{3\alpha-1}\log^2 y}{(y^{\alpha}+M^{\alpha})^4}\, dy - 4\int_0^{\infty} \frac{\alpha M^{2\alpha}\log M y^{2\alpha-1}\log y}{(y^{\alpha}+M^{\alpha})^4}\, dy \\
&+ \frac{\alpha^2}{3M^2}\Bigg) \\
&- \log\Bigg(\Bigg(\frac{1}{\alpha^2} + \frac{\log^2 M}{3} + 2\int_0^{\infty} \frac{\alpha M^{\alpha} y^{2\alpha-1}\log^2 y}{(y^{\alpha}+M^{\alpha})^3}\, dy \\
&- 2\int_0^{\infty} \frac{\alpha M^{\alpha} y^{3\alpha-1}\log^2 y}{(y^{\alpha}+M^{\alpha})^4}\, dy - 4\int_0^{\infty} \frac{\alpha M^{2\alpha}\log M y^{2\alpha-1}\log y}{(y^{\alpha}+M^{\alpha})^4}\, dy\Bigg) \\
&\times \left(\frac{\alpha^2}{3M^2}\right) \\
&- \left(\frac{\alpha\log M}{3M} - 2\int_0^{\infty} \frac{\alpha^2 M^{2\alpha-1} y^{2\alpha-1}\log y}{(y^{\alpha}+M^{\alpha})^4}\, dy\right)^2\Bigg) \\
&- 2\log(2). \qquad \text{(A.30)}
\end{aligned}
$$

Rewriting (4.3) as

$$\text{AMC} = -2\log\mathcal{L} + 2\log\left(1 + \frac{2\sqrt{(nI_{11}+nI_{22}+1)^3}}{nI_{11}+nI_{22}+2n^2(I_{11}I_{22}-I_{12}^2)}\right),$$

$$\text{AMC}_T^{Lomax} = -2\left(n\log\alpha + n\alpha\log(\theta+t) - (\alpha+1)\sum_{i=1}^{n}\log(\theta+X_i)\right)$$

$$+ 2\log\left(1 + \frac{2\sqrt{\left(\frac{n}{\alpha^2}+\frac{n\alpha}{(\alpha+2)(t+\theta)^2}+1\right)^3}}{\frac{n}{\alpha^2}+\frac{n\alpha}{(\alpha+2)(t+\theta)^2}+\frac{2n^2}{\alpha(\alpha+1)^2(\alpha+2)(t+\theta)^2}}\right), \quad \text{(A.31)}$$

$$\text{AMC}_S^{Lomax} = -2\left(n\log\alpha + n\alpha\log\theta_S - (\alpha+1)\sum_{i=1}^{n}\log\left(\theta_S+X_i-t\right)\right)$$

$$+ 2\log\left(1 + \frac{2\sqrt{\left(\frac{n}{\alpha^2}+\frac{n\alpha}{(\alpha+2)(\theta_S)^2}+1\right)^3}}{\frac{n}{\alpha^2}+\frac{n\alpha}{(\alpha+2)(\theta_S)^2}+\frac{2n^2}{\alpha(\alpha+1)^2(\alpha+2)(\theta_S)^2}}\right), \quad \text{(A.32)}$$

$$\text{AMC}_T^{LN} = -2\Bigg(-n\log\sqrt{2\pi} - \sum_{i=1}^{n}\log X_i - n\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(\log X_i-\mu)^2$$

$$- n\log\left(1-\Phi\left(\frac{\log t-\mu}{\sigma}\right)\right)\Bigg)$$

$$+ 2\log\left(1 + \frac{2\sqrt{(nI_{11}^{lnT}+nI_{22}^{lnT}+1)^3}}{nI_{11}^{lnT}+nI_{22}^{lnT}+2n^2(I_{11}^{lnT}I_{22}^{lnT}-(I_{12}^{lnT})^2)}\right), \quad \text{(A.33)}$$

with I_{11}^{lnT}, I_{12}^{lnT}, I_{22}^{lnT} as derived in (A.8)–(A.10), respectively.

$$\begin{aligned}\mathrm{AMC}_S^{LN} = -2\Bigg(&-n\log\sqrt{2\pi} - \sum_{i=1}^{n}\log(X_i - t) - n\log\sigma_S \\ &- \frac{1}{2\sigma_S^2}\sum_{i=1}^{n}\left(\log(X_i - t) - \mu_S\right)^2\Bigg) \\ &+ 2\log\left(1 + \frac{2\sqrt{\left(n\cdot\frac{1}{\sigma_S^2} + n\cdot\frac{2}{\sigma_S^2} + 1\right)^3}}{n\cdot\frac{1}{\sigma_S^2} + n\cdot\frac{2}{\sigma_S^2} + 2n^2\left(\frac{2}{\sigma_S^4}\right)}\right),\end{aligned} \tag{A.34}$$

$$\begin{aligned}\mathrm{AMC}_T^{Champ} = -2\Bigg(&n\log\alpha + (\alpha - 1)\sum_{i=1}^{n}\log(X_i) + n\log(t^\alpha + M^\alpha) \\ &- 2\sum_{i=1}^{n}\log(X_i^\alpha + M^\alpha)\Bigg) \\ &+ 2\log\left(1 + \frac{2\sqrt{(nI_{11}^{champT} + nI_{22}^{champT} + 1)^3}}{nI_{11}^{champT} + nI_{22}^{champT} + 2n^2(I_{11}^{champT}I_{22}^{champT} - (I_{12}^{champT})^2)}\right),\end{aligned} \tag{A.35}$$

with I_{11}^{champT}, I_{12}^{champT}, I_{22}^{champT} as derived in (A.19)–(A.21), respectively.

$$\begin{aligned}\text{AMC}_S^{Champ} = -2\Bigg(& n\log\alpha + n\alpha\log(M) + (\alpha-1)\sum_{i=1}^{n}\log(X_i - t) \\ & - 2\sum_{i=1}^{n}\log((X_i - t)^{\alpha} + M^{\alpha})\Bigg) \\ & + 2\log\left(1 + \frac{2\sqrt{(nI_{11}^{champS} + nI_{22}^{champS} + 1)^3}}{nI_{11}^{champS} + nI_{22}^{champS} + 2n^2(I_{11}^{champS}I_{22}^{champS} - (I_{12}^{champS})^2)}\right),\end{aligned} \tag{A.36}$$

with I_{11}^{champS}, I_{12}^{champS}, I_{22}^{champS} as derived in (A.22)–(A.24), respectively.

Proof of Proposition 4.4.1: By the weak law of large numbers (WLLN), we have the following convergence in probability as $n \to \infty$,

$$-\frac{1}{n}\frac{\partial^2 \ell(\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T} = -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 \log\widetilde{f}(X_i|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\Bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \xrightarrow{\mathcal{P}} \mathbf{I}(\boldsymbol{\theta}),$$

where $\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n}\log\widetilde{f}(X_i|\boldsymbol{\theta})$ and $\widetilde{f}$ denotes either truncated or shifted pdf. In particular, we have

$$-\frac{1}{n}\frac{\partial^2 \ell(\hat{\boldsymbol{\theta}})}{\partial\theta_1^2} \xrightarrow{\mathcal{P}} I_{11}(\boldsymbol{\theta}), \quad -\frac{1}{n}\frac{\partial^2 \ell(\hat{\boldsymbol{\theta}})}{\partial\theta_1\partial\theta_2} \xrightarrow{\mathcal{P}} I_{12}(\boldsymbol{\theta}), \quad \text{and} \quad -\frac{1}{n}\frac{\partial^2 \ell(\hat{\boldsymbol{\theta}})}{\partial\theta_2^2} \xrightarrow{\mathcal{P}} I_{22}(\boldsymbol{\theta}).$$

Similarly, the following convergence in probability result can be proven using WLLN when $n \to \infty$:

$$\frac{1}{n}\frac{\partial\ell(\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}}\frac{\partial\ell(\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}^T} \xrightarrow{\mathcal{P}} \mathbf{I}(\boldsymbol{\theta}),$$

since for $i.i.d.$ data $X_1, \ldots, X_n$, the following steps are easily justified:

$$\begin{aligned}
\mathbf{E}\left[\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\right] &= \mathbf{V}\left[\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right] + \mathbf{E}\left[\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]^2 \\
&= \mathbf{V}\left[\sum_{i=1}^{n} \frac{\partial \log \widetilde{f}(X_i|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right] \\
&= n\left(\mathbf{V}\left[\frac{\partial \log \widetilde{f}(X_i|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]\right) \\
&= n\mathbf{I}(\boldsymbol{\theta}),
\end{aligned}$$

while

$$\mathbf{E}\left[\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right] = 0.$$

And in particular, we have

$$\frac{1}{n}\left(\frac{\partial \ell(\hat{\boldsymbol{\theta}})}{\partial \theta_1}\right)^2 \xrightarrow{\mathcal{P}} I_{11}(\boldsymbol{\theta}), \quad \frac{1}{n}\frac{\partial \ell(\hat{\boldsymbol{\theta}})}{\partial \theta_1}\frac{\partial \ell(\hat{\boldsymbol{\theta}})}{\partial \theta_2} \xrightarrow{\mathcal{P}} I_{12}(\boldsymbol{\theta}), \quad \text{and} \quad \frac{1}{n}\left(\frac{\partial \ell(\hat{\boldsymbol{\theta}})}{\partial \theta_2}\right)^2 \xrightarrow{\mathcal{P}} I_{22}(\boldsymbol{\theta}).$$

Since convergence in probability is preserved under addition and multiplication, and since both the square root function and the reciprocal function are continuous, by the continuous mapping theorem (see Shao, 1999, p. 42), the result follows by applying above convergence in probability results to (4.2). □

Appendix B: Data Set

Table B.1: Unobserved and observed costs of legal events (below and above $195,000, respectively).

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
200,000.00	220,357.00	252,000.00	332,000.00	416,562.38	630,200.00	907,077.00
200,000.00	229,368.50	255,414.00	350,000.00	423,319.62	650,000.00	917,000.00
202,077.38	230,000.00	260,000.00	350,000.00	426,000.00	660,000.00	1,299,345.00
204,450.00	232,500.00	270,341.11	360,000.00	483,711.60	734,900.00	1,415,988.00
210,536.56	239,102.93	274,509.80	394,672.11	505,947.00	742,651.56	2,567,921.00
214,634.95	248,341.96	294,835.23	400,203.01	550,000.00	750,000.00	3,821,987.00
220,000.00	250,000.00	297,035.48	406,001.47	556,000.00	800,000.00	
220,070.00	251,489.59	301,527.50	410,060.72	600,000.34	845,000.00	

SOURCE: Cruz (2002), page 57.

Appendix C: Computer Code

```
% MATLAB code
%
%
% Table 3.3
% Compute how often Lomax VaR estimates
% exceed multiple of target VaR by simulation
clear
count_trunc = 0;count_naive = 0;count_shift = 0;

% Record times of MLE nonconvergent and convergent
nonconvcount = 0;
convcount = 0;

% Run the following procedure 1000 times
iter_max = 1000;

% for iteration = 1: iter_max
while convcount < iter_max
% Lomax models:
% F(t) = 0.5
alpha = 3.5; theta = 890355; t = 195000; N = 100;
xi = 1/alpha;
sigma = theta/alpha;
% generate Lomax data
x = gprnd(xi,sigma,0,N,1);
beta = .95; % VaR level
VaR_target = sigma*((1-beta)^-xi-1)/xi;
% multiple of target VaR
c = 1;
VaR_target = c*VaR_target;
% observed data above the threshold
x = x(x>t);

% fit naive model to select MLE convergence
% produce parameter and VaR estimates
param_naive = lomaxfitNLL(x);
% criterion to select MLE convergence
errNonConv = getGlobalerrNonConv;
if (errNonConv == 0)
    nonconvcount = nonconvcount + 1;
    continue % next iteration
% elseif (errNonConv == 1)
```

```
%      convcount = convcount + 1;
end
     xi_hat_nMLE = 1/param_naive(1);
     sigma_hat_nMLE = param_naive(2)/param_naive(1);
if ((1-beta)^-xi_hat_nMLE-1 > 0)
     VaR_naive = sigma_hat_nMLE*((1-beta)^-xi_hat_nMLE-1)/xi_hat_nMLE;
else
    % (1-beta)^-xi_hat_nMLE-1 == 0
    % convergent to exponential
     VaR_naive = -sigma_hat_nMLE*log(1-beta);
end

if VaR_naive > VaR_target
     count_naive = count_naive + 1;
end

% fit shifted model to select MLE convergence,
% produce parameter and VaR estimates
param = lomaxfitNLL(x-t+realmin);
% criterion to select MLE convergence
errNonConv = getGlobalerrNonConv;
if (errNonConv == 0)
     nonconvcount = nonconvcount + 1;
     continue % next iteration
elseif (errNonConv == 1)
     convcount = convcount + 1;
else (errNonConv == -1)
     error(message('stats:gpfit:NoSolution'));
end
     xi_hat_sMLE = 1/param(1);
     sigma_hat_sMLE = param(2)/param(1);
if ((1-beta)^-xi_hat_sMLE-1 > 0)
     VaR_shift = sigma_hat_sMLE*((1-beta)^-xi_hat_sMLE-1)/xi_hat_sMLE+t;
else
    % (1-beta)^-xi_hat_sMLE-1 == 0
    % convergent to exponential
     VaR_shift = -sigma_hat_sMLE*log(1-beta)+t;
end

if VaR_shift > VaR_target
     count_shift = count_shift + 1;
end

% fit truncated model, parameter and VaR estimates
     xi_hat_tMLE = xi_hat_sMLE;
     sigma_hat_tMLE = sigma_hat_sMLE - xi_hat_sMLE*t;
if ((1-beta)^-xi_hat_tMLE-1 > 0)
     VaR_trunc = sigma_hat_tMLE*((1-beta)^-xi_hat_tMLE-1)/xi_hat_tMLE;
else
    % (1-beta)^-xi_hat_tMLE-1 == 0
    % convergent to exponential
     VaR_trunc = -sigma_hat_tMLE*log(1-beta);
end

if VaR_trunc > VaR_target
     count_trunc = count_trunc + 1;
end

end
overest_p_trunc = count_trunc/iter_max
```

```
overest_p_naive = count_naive/iter_max
overest_p_shift = count_shift/iter_max
% end of Table 3.3
%
%
% Table 3.5
% Compute how often lognormal VaR estimates
% exceed multiple of target VaR by simulation:
% Similar to what have been done in Table 3.3
% just generate lognormal data and
% fit truncated, naive, shifted lognormal models
% end of Table 3.5
%
%
% Table 3.6
% Fit exponential/Lomax/lognormal models
% to Cruz data by MLE
clear
Data = textread('cruz-data.txt');
% reshape data
Data = reshape(Data, 1, []);
N = length(Data);
% threshold
t = 195000;
% observed data above the threshold
x = Data(Data>=t);
x = sort(x);
x_reverse = sort(x, 'descend');
% number of observations
n = numel(x);
% VaR level
p = 0.99;

% truncated exponential parameter estimate
theta_hat_tMLE = exp_trunc_mle(x, t);

% naive exponential parameter estimate
theta_hat_nMLE = mean(x);

% shifted exponential parameter estimate
theta_hat_sMLE = mean(x-t);

% fit naive Lomax model, parameter estimates
param_naive = lomaxfitFPD(x);
    alpha_hat_nMLE = param_naive(1);
    theta_hat_nMLE = param_naive(2);
    xi_hat_nMLE = 1/param_naive(1);
    sigma_hat_nMLE = param_naive(2)/param_naive(1);

% fit shifted Lomax model, parameter estimates
param_shift = lomaxfitNLL(x-t+realmin);
     alpha_hat_sMLE = param_shift(1);
     theta_hat_sMLE = param_shift(2);
     xi_hat_sMLE = 1/param_shift(1);
     sigma_hat_sMLE = param_shift(2)/param_shift(1);

% fit truncated Lomax model, parameter estimates
    alpha_hat_tMLE = alpha_hat_sMLE;
    theta_hat_tMLE = theta_hat_sMLE - t;
    xi_hat_tMLE = xi_hat_sMLE;
```

```
    sigma_hat_tMLE = sigma_hat_sMLE - xi_hat_sMLE*t;

% fit truncated lognormal model
[mu_hat_tMLE, sigma_hat_tMLE] = logn_trunc_mle(x, t);
% Underlying lognormal distribution function with
% truncated MLE for the observed data
pd2_trunc = makedist('logn', mu_hat_tMLE,sigma_hat_tMLE);

% Method of moments (naive MM) fitting lognormal model
% Parameter esimates by naive MM for lognormal model,
% which is equivalent to naive MLE for lognormal model
mu_hat_nMLE = mean(log(x));
sigma_hat_nMLE = std(log(x));

% Parameter estimates of shifted lognormal model
mu_hat_sMLE = sum(log(x - t))/n;
% Notice here divided by (n-1)
sigma_hat_sMLE=sqrt(sum((log(x-t)-mu_hat_sMLE).^2)/(n-1));
% end of Table 3.6
%
%
% Table 3.8
% VaR level
p = 0.99;
% truncated VaR estimate for exponential model
VaR_p_tMLE = theta_hat_tMLE*-log(1-p);
VaR_p_tMLE_std = theta_hat_tMLE*-log(1-p)/sqrt(n);
% 95% Confidence Interval
VaR_p_tMLE_CI = [VaR_p_tMLE - norminv(0.975)*VaR_p_tMLE_std, ...
    VaR_p_tMLE + norminv(0.975)*VaR_p_tMLE_std];

% naive VaR estimate for exponential model
VaR_p_nMLE = theta_hat_nMLE*-log(1-p);
VaR_p_nMLE_std = theta_hat_nMLE*-log(1-p)/sqrt(n);
% 95% Confidence Interval
VaR_p_nMLE_CI = [VaR_p_nMLE - norminv(0.975)*VaR_p_nMLE_std, ...
    VaR_p_nMLE + norminv(0.975)*VaR_p_nMLE_std];

% shifted VaR estimate for exponential model
VaR_p_sMLE = theta_hat_sMLE*-log(1-p)+t;
VaR_p_sMLE_std = theta_hat_sMLE*-log(1-p)/sqrt(n);
% 95% Confidence Interval
VaR_p_sMLE_CI = [VaR_p_sMLE - norminv(0.975)*VaR_p_sMLE_std, ...
    VaR_p_sMLE + norminv(0.975)*VaR_p_sMLE_std];
%
%
% naive VaR for Lomax model
VaR_p_nMLE = sigma_hat_nMLE*((1-p)^-xi_hat_nMLE-1)/xi_hat_nMLE;

% shifted VaR for Lomax model
VaR_p_sMLE = sigma_hat_sMLE*((1-p)^-xi_hat_sMLE-1)/xi_hat_sMLE+t;

% truncated VaR for Lomax model
VaR_p_tMLE = sigma_hat_tMLE*((1-p)^-xi_hat_tMLE-1)/xi_hat_tMLE;
%
%
% Compute VaR for truncated lognormal model
VaR_p_tMLE = [];
for p = 0.9:.01:0.99
    % 90 percentile through 99 percentile
```

```
    VaR_p_tMLE = [VaR_p_tMLE, exp(mu_hat_tMLE+sigma_hat_tMLE*norminv(p))];
end

% Compute VaR for shifted lognormal model
VaR_p_sMLE = [];
for p = 0.9:.01:0.99
    % 90 through 99 percentile
    VaR_p_sMLE = [VaR_p_sMLE, exp(mu_hat_sMLE+sigma_hat_sMLE*norminv(p))+t];
end

% Compute VaR for naive lognormal model
VaR_p_nMLE = [];
for p = 0.9:.01:0.99
    % 90 through 99 percentile
    VaR_p_nMLE = [VaR_p_nMLE, exp(mu_hat_nMLE+sigma_hat_nMLE*norminv(p))];
end
% end of Table 3.8
%
%
% Table 3.9
% Verify delta method for number of loss between a and b
% for truncated exponential model
a = 150000;
b = 175000;
% run dataExpMLE first to obtain theta_hat_tMLE
sigma = theta_hat_tMLE;

% partial derivative of number of loss between a and b with respect to sigma
part_sigma = n*(exp((t-b)/sigma)*(t-b)/sigma^2-exp((t-a)/sigma)*(t-a)/sigma^2);

% variance of parameter estimator
var_sigma = sigma^2/n;
% variance of number of loss between a and b based on covariance from delta method
fprintf(' variance of number of loss between a and b \n ')
var_numLoss = part_sigma^2*var_sigma

% confidence level
beta = 0.95;
% number of loss between a and b
numLoss = n*(-exp((t-b)/sigma)+exp((t-a)/sigma));
fprintf(' CI of number of loss between a and b \n ')
% confidence interval for number of loss between a and b
% for truncated exponential model
[numLoss-norminv(beta+(1-beta)/2)*sqrt(var_numLoss)...
    numLoss+norminv(beta+(1-beta)/2)*sqrt(var_numLoss)]
%
%
% Verify delta method for average loss between a and b
% for truncated exponential model
%
% partial derivative of average loss between a and b with respect to sigma
part_sigma = 1-(1/sigma^2)*(exp(-(a+b)/sigma)*(a-b)^2 ...
    /(exp(-a/sigma)-exp(-b/sigma))^2);

% variance of average loss between a and b based on covariance from delta method
fprintf(' variance of average loss between a and b \n ')
var_aveLoss = part_sigma^2*var_sigma

% confidence level
beta = 0.95;
```

```
% average loss between a and b
aveLoss = sigma+(a*exp(-a/sigma)-b*exp(-b/sigma))/(exp(-a/sigma)-exp(-b/sigma));
fprintf(' CI of average loss between a and b \n ')
% confidence interval for average loss between a and b
% for truncated exponential model
[aveLoss-norminv(beta+(1-beta)/2)*sqrt(var_aveLoss)...
    aveLoss+norminv(beta+(1-beta)/2)*sqrt(var_aveLoss)]
%
%
% Verify delta method for total loss between a and b
% for truncated exponential model
%
% partial derivative of total loss between a and b with respect to sigma
part_sigma = n*((1+(a/sigma+1/sigma)*(a/sigma-t/sigma))*exp(-(a-t)/sigma) ...
    -(1+(b/sigma+1/sigma)*(b/sigma-t/sigma))*exp(-(b-t)/sigma));
% variance of total loss between a and b based on covariance from delta method
fprintf(' variance of total loss between a and b \n ')
var_totLoss = part_sigma^2*var_sigma

% confidence level
beta = 0.95;
% total loss between a and b
totLoss = n*((a+sigma)*exp(-(a-t)/sigma)-(b+sigma)*exp(-(b-t)/sigma));
fprintf(' CI of total loss between a and b \n ')
% confidence interval for total loss between a and b
% for truncated exponential model
[totLoss-norminv(beta+(1-beta)/2)*sqrt(var_totLoss)...
    totLoss+norminv(beta+(1-beta)/2)*sqrt(var_totLoss)]
%
%
% Verify delta method for number of loss between a and b
% for truncated Lomax model
a = 150000;
b = 175000;
% run dataLomaxMLE first
alpha = 1/xi_hat_tMLE;
theta = sigma_hat_tMLE/xi_hat_tMLE;

% partial derivatives of number of loss between a and b
% with respect to alpha and theta
part_alpha = n*((1+(t-a)/(a+theta))^alpha*log(1+(t-a)/(a+theta)) ...
    - (1+(t-b)/(b+theta))^alpha*log(1+(t-b)/(b+theta)));
part_theta = n*(alpha*(1+(t-a)/(a+theta))^(alpha-1)*(a-t)/(a+theta)^2 ...
    - alpha*(1+(t-b)/(b+theta))^(alpha-1)*(b-t)/(b+theta)^2);

% variance-covariance of parameter estimators
var_alpha = (1/xi_hat_tMLE)^2*(1/xi_hat_tMLE+1)^2/n;
var_theta = (d+sigma_hat_tMLE/xi_hat_tMLE)^2*(1/xi_hat_tMLE+2) ...
    *(1/xi_hat_tMLE+1)^2/n/(1/xi_hat_tMLE);
cov_alpha_theta = (1/xi_hat_tMLE) * (1/xi_hat_tMLE+1) * (1/xi_hat_tMLE+2) ...
    * (d+sigma_hat_tMLE/xi_hat_tMLE) / n;
% variance of number of loss between a and b based on covariance from delta method
fprintf(' variance of number of loss between a and b \n ')
var_numLoss = part_alpha^2*var_alpha + 2*part_alpha*part_theta*cov_alpha_theta ...
    + part_theta^2*var_theta

% confidence level
beta = 0.95;
% number of loss between a and b
numLoss = n*((1+(t-a)/(a+theta))^alpha-(1+(t-b)/(b+theta))^alpha);
```

```
fprintf(' CI of number of loss between a and b \n ')
% confidence interval for number of loss between a and b
% for truncated Lomax model
[numLoss-norminv(beta+(1-beta)/2)*sqrt(var_numLoss)...
    numLoss+norminv(beta+(1-beta)/2)*sqrt(var_numLoss)]
%
%
% Verify delta method for average loss between a and b
% for truncated Lomax model
R = (b+theta)/(a+theta);

% partial derivatives of average loss between a and b
% with respect to alpha and theta
part_alpha = (a*R^alpha-b+alpha*a*R^alpha*log(b+theta)-alpha*b*log(a+theta)) ...
    /(alpha-1)/(R^alpha-1)-alpha*(a*R^alpha-b) ...
    *(R^alpha-1+(alpha-1)*R^alpha*log(b+theta)-(alpha-1)*log(a+theta)) ...
    /(alpha-1)^2/(R^alpha-1)^2-theta/(alpha-1)^2;
part_theta = (alpha/(alpha-1))*((a*alpha*R^alpha/(b+theta)-b*alpha/(a+theta)) ...
    *(R^alpha-1)-(a*R^alpha-b)*(alpha*R^alpha/(b+theta)-alpha/(a+theta))) ...
    /(R^alpha-1)^2 + 1/(alpha-1);

% variance of average loss between a and b based on covariance from delta method
fprintf(' variance of average loss between a and b \n ')
var_aveLoss = part_alpha^2*var_alpha + 2*part_alpha*part_theta*cov_alpha_theta ...
    + part_theta^2*var_theta

% confidence level
beta = 0.95;
% average loss between a and b
aveLoss = ((-a*alpha-theta)*R^alpha+b*alpha+theta)/(1-alpha)/(R^alpha-1);
fprintf(' CI of average loss between a and b \n ')
% confidence interval for average loss between a and b
% for truncated Lomax model
[aveLoss-norminv(beta+(1-beta)/2)*sqrt(var_aveLoss)...
    aveLoss+norminv(beta+(1-beta)/2)*sqrt(var_aveLoss)]
%
%
% Verify delta method for total loss between a and b
% for truncated Lomax model
%
% total loss between a and b
totLoss = n*((t+theta)/(a+theta))^alpha*((-a*alpha-theta)/(1-alpha) ...
    +(b*alpha+theta)/(1-alpha)/((b+theta)/(a+theta))^alpha);

% partial derivatives of total loss between a and b
% with respect to alpha and theta
part_alpha = totLoss*log(t+theta) + n*((t+theta)/(a+theta))^alpha ...
    *(-a*(1-alpha)+(a*alpha+theta)*(-1+(1-alpha)*log(a+theta)))/(1-alpha)^2 ...
    + n*((t+theta)/(b+theta))^alpha*(b*(1-alpha)-(b*alpha+theta) ...
    *(-1+(1-alpha)*log(b+theta)))/(1-alpha)^2;
part_theta = totLoss*alpha/(t+theta) + n*((t+theta)/(a+theta))^alpha ...
    *(-1+(a*alpha+theta)*alpha/(a+theta))/(1-alpha) ...
    + n*((t+theta)/(b+theta))^alpha*(1-(b*alpha+theta)*alpha/(b+theta))/(1-alpha);

% variance of total loss between a and b based on covariance from delta method
fprintf(' variance of total loss between a and b \n ')
var_totLoss = part_alpha^2*var_alpha + 2*part_alpha*part_theta*cov_alpha_theta ...
    + part_theta^2*var_theta

% confidence level
```

```
beta = 0.95;
fprintf(' CI of total loss between a and b \n ')
% confidence interval for total loss between a and b
% for truncated Lomax model
[totLoss-norminv(beta+(1-beta)/2)*sqrt(var_totLoss)...
    totLoss+norminv(beta+(1-beta)/2)*sqrt(var_totLoss)]
%
%
% Verify delta method for number of loss between a and b
% for truncated lognormal model
a = 150000;
b = 175000;
% run dataLognormalMLE first
mu = mu_hat_tMLE;
sigma = sigma_hat_tMLE;
r = (log(t)-mu)/sigma;
ra = (log(a)-mu)/sigma;
rb = (log(b)-mu)/sigma;
K = normpdf(r)/(1-normcdf(r));
% partial derivatives of number of loss between a and b
% with respect to mu and sigma
part_mu = n*((1-normcdf(r))*(normpdf(rb)-normpdf(ra)) ...
    +(normcdf(rb)-normcdf(ra))*normpdf(r))/((1-normcdf(r))^2)/(-sigma);
part_sigma = n*((1-normcdf(r))*(normpdf(rb)*(log(b)-mu)-normpdf(ra) ...
    *(log(a)-mu))+(normcdf(rb)-normcdf(ra))*normpdf(r)*(log(t)-mu)) ...
    /((1-normcdf(r))^2)/(-sigma^2);

% Fisher information matrix entries
I11 = (1-r*K/sigma-K^2)/sigma^2;
I12 = K/sigma^2 - r^2*K/sigma^3 - r*K^2;
I22 = (2+r*K+r^3*K-r^2*K^2)/sigma^2;

% variance-covariance of parameter estimators
var_mu = I22/(I11*I22-I12^2)/n;
var_sigma = I11/(I11*I22-I12^2)/n;
cov_mu_sigma = -I12/(I11*I22-I12^2)/n;
% variance of number of loss between a and b based on covariance from delta method
fprintf(' variance of number of loss between a and b \n ')
var_numLoss = part_mu^2*var_mu + 2*part_mu*part_sigma*cov_mu_sigma ...
    + part_sigma^2*var_sigma

% confidence level
beta = 0.95;
% number of loss between a and b
numLoss = n*(normcdf(rb)-normcdf(ra))/(1-normcdf(r));
fprintf(' CI of number of loss between a and b \n ')
% confidence interval for number of loss between a and b
% for truncated lognormal model
[numLoss-norminv(beta+(1-beta)/2)*sqrt(var_numLoss)...
    numLoss+norminv(beta+(1-beta)/2)*sqrt(var_numLoss)]
%
%
% Verify delta method for average loss between a and b
% for truncated lognormal model
%
% lognpdf_funx is lognpdf_fun * x as integrand
lognpdf_funx = @(x) exp(-(log(x)-mu).^2/2/sigma^2)/sigma/sqrt(2*pi);
% lognpdf_funx_dmu is lognpdf_fun * x differentiated w.r.t. mu as integrand
lognpdf_funx_dmu = @(x) exp(-(log(x)-mu).^2/2/sigma^2) ...
    .*(log(x)-mu)/sigma^3/sqrt(2*pi);
```

```
% lognpdf_funx_dsigma is lognpdf_fun * x differentiated w.r.t. sigma as integrand
lognpdf_funx_dsigma = @(x) exp(-(log(x)-mu).^2/2/sigma^2) ...
    .*(-1/sigma+(log(x)-mu).^2/sigma^3)/sigma/sqrt(2*pi);

% partial derivatives of average loss between a and b
% with respect to mu and sigma
part_mu = ((normcdf(rb)-normcdf(ra))*integral(lognpdf_funx_dmu,a,b) ...
    - integral(lognpdf_funx,a,b)*(normpdf(rb)-normpdf(ra))/-sigma) ...
    /(normcdf(rb)-normcdf(ra))^2;
part_sigma = ((normcdf(rb)-normcdf(ra))*integral(lognpdf_funx_dsigma,a,b) ...
    - integral(lognpdf_funx,a,b)*(normpdf(rb)*rb-normpdf(ra)*ra)/-sigma) ...
    /(normcdf(rb)-normcdf(ra))^2;

% variance of average loss between a and b based on covariance from delta method
fprintf(' variance of number of loss between a and b \n ')
var_aveLoss = part_mu^2*var_mu + 2*part_mu*part_sigma*cov_mu_sigma ...
    + part_sigma^2*var_sigma

% confidence level
beta = 0.95;
% average loss between a and b
aveLoss = integral(lognpdf_funx,a,b)/(normcdf(rb)-normcdf(ra));
fprintf(' CI of average loss between a and b \n ')
% confidence interval for average loss between a and b
% for truncated lognormal model
[aveLoss-norminv(beta+(1-beta)/2)*sqrt(var_aveLoss)...
    aveLoss+norminv(beta+(1-beta)/2)*sqrt(var_aveLoss)]
%
%
% Verify delta method for total loss between a and b
% for truncated lognormal model
%
% partial derivatives of total loss between a and b
% with respect to mu and sigma
part_mu = n*((1-normcdf(r))*integral(lognpdf_funx_dmu,a,b) ...
    - integral(lognpdf_funx,a,b)*normpdf(r)/sigma)/(1-normcdf(r))^2;
part_sigma = n*((1-normcdf(r))*integral(lognpdf_funx_dsigma,a,b) ...
    - integral(lognpdf_funx,a,b)*normpdf(r)*r/sigma)/(1-normcdf(r))^2;

% variance of total loss between a and b based on covariance from delta method
fprintf(' variance of number of loss between a and b \n ')
var_totLoss = part_mu^2*var_mu + 2*part_mu*part_sigma*cov_mu_sigma ...
    + part_sigma^2*var_sigma

% confidence level
beta = 0.95;
% total loss between a and b
totLoss = n*integral(lognpdf_funx,a,b)/(1-normcdf(r));
fprintf(' CI of total loss between a and b \n ')
% confidence interval for total loss between a and b
% for truncated lognormal model
[totLoss-norminv(beta+(1-beta)/2)*sqrt(var_totLoss)...
    totLoss+norminv(beta+(1-beta)/2)*sqrt(var_totLoss)]
%
% end of Table 3.9
%
%
% Table 4.1
%
%
```

```
% Run dataLomaxMLE first
% Compute ICOMP AMC for truncated Lomax model
xi = xi_hat_tMLE;
% alpha is shape parameter of Lomax distribution
alpha = 1/xi;
sigma = sigma_hat_tMLE;
% theta is scale parameter of Lomax distribution
theta = sigma/xi;
% t is the data collection threshold
t;
% I_11 is the negative expectation of the second partial
% derivative of the log-likelihood function with respect
% to alpha
I_11 = ( 1/alpha^2 );
% I_12 is the negative expectation of the cross partial
% derivative of the log-likelihood function with respect
% to alpha & theta
I_12 = ( -1/((alpha+1)*(theta+t)) );
% I_22 is the negative expectation of the second partial
% derivative of the log-likelihood function with respect
% to theta
I_22 = ( alpha/((alpha+2)*(theta+t)^2) );
% H is the mean curvature, i.e. the average of the two
% principal curvatures of the log-likelihood surface
H = (n*(I_11+I_22)+2*n^2*(I_11*I_22-I_12^2)) ...
/ (-2*sqrt(n*I_11+n*I_22+1)^3);
format long
% ICOMP is information complexity using inverse
% of Fisher information matrix
ICOMP = 2*neg_loglik_trunc + 2*log(I_11+I_22) ...
- log(I_11*I_22-I_12^2) - 2*log(2)
% AMC is new model selection criteria: Information
% criteria incorporating both negative log-likelihood
% and the asymptotic mean curvature of the log-likelihood surface
AMC = 2*neg_loglik_trunc + 2*log(1-1/H)
%
%
% Run dataLomaxMLE first
% Compute ICOMP and AMC for shifted Lomax model
xi = xi_hat_sMLE;
% alpha is shape parameter of Lomax distribution
alpha = 1/xi;
sigma = sigma_hat_sMLE;
theta_shift = sigma/xi;
% theta is scale parameter of Lomax distribution
theta = theta_shift;

% Fisher information entries
I_11 = ( 1/alpha^2 );
I_12 = ( -1/((alpha+1)*(theta)) );
I_22 = ( alpha/((alpha+2)*(theta)^2) );
% mean curvature
H = (n*(I_11+I_22)+2*n^2*(I_11*I_22-I_12^2)) ...
/ (-2*sqrt(n*I_11+n*I_22+1)^3);
format long
% information complexity criterion
ICOMP = 2*neg_loglik_shift + 2*log(I_11+I_22) ...
- log(I_11*I_22-I_12^2) - 2*log(2)
% asymptotic mean curvature criterion
AMC = 2*neg_loglik_shift + 2*log(1-1/H)
```

```
%
%
% Run dataLognormalMLE first
% Compute ICOMP and AMC for truncated lognormal model
% sigma is scale parameter of Lognormal distribution
sigma = sigma_hat_tMLE;
% mu is location parameter of Lognormal distribution
mu = mu_hat_tMLE;
% t is the data collection threshold
r = (log(t)-mu)/sigma;
K = normpdf(r)/(1-normcdf(r));
% I_11 is the negative expectation of the second partial
% derivative of the log-likelihood function with respect
% to mu
I_11 = (1+r*K-K^2)/sigma^2;
% I_12 is the negative expectation of the cross partial
% derivative of the log-likelihood function with respect
% to mu & sigma
I_12 = ( -K/sigma^2+r^2*K/sigma^2-r*K^2/sigma^2) ...
+ 2*(Exp_star_logX-mu)/sigma^3;
% I_22 is the negative expectation of the second partial
% derivative of the log-likelihood function with respect
% to sigma
I_22 = ( -(1+2*r*K)/sigma^2+r^3*K/sigma^2-r^2*K^2/sigma^2 ) ...
+ 3*Exp_star/sigma^4;
% Same expression to compute mean curvature,
% information complexity criterion,
% asymptotic mean curvature criterion
% as for truncated Lomax model
%
%
% Run dataLognormalMLE first
% Compute ICOMP and AMC for shifted lognormal model
% sigma is scale parameter of Lognormal distribution
sigma = sigma_hat_sMLE;
% mu is location parameter of Lognormal distribution
mu = mu_hat_sMLE;
% Fisher information entry I_11
I_11 = 1/sigma^2;
% Define function handle for integrand log(x-t)f(x)
fh1 = @(x) log(x-t) ...
.*normpdf((log(x-t)-mu)./sigma)./((x-t).*sigma);
% Expectation of log(X) w.r.t. the conditional distribution
% where the support of X is from t to infinity
Exp_logXminust = integral(fh1,t,inf);
% I_12
I_12 = 2*(Exp_logXminust-mu)/sigma^3;
% Define function handle for integrand (log(x)-mu)^2*f(x)
fh2 = @(x) (log(x-t)-mu).^2 ...
.*normpdf((log(x-t)-mu)./sigma)./((x-t).*sigma);
% Expectation of (log(X)-mu)^2 w.r.t. the conditional distribution
% where the support of X is from t to infinity
Exp = integral(fh2,t,inf);
% I_22
I_22 = 3*Exp/sigma^4 - 1/sigma^2;
% Same expression to compute mean curvature,
% information complexity criterion,
% asymptotic mean curvature criterion
% as for shifted Lomax model
%
```

```
%
% Run dataChampernowneMLE first
% Compute ICOMP and AMC for truncated Champernowne model
% alpha is shape parameter of Champernowne distribution
alpha = tao_hat_tMLE;
% M is scale parameter of Champernowne distribution
M = theta_hat_tMLE;
% define the function handle for the integrand of
% (log(w))^2 times density of beta prime (2, 2)
fh2 = @(x) log(x).^2.*6.*x./(x+1).^4;
% I_11 is the negative expectation of the second partial
% derivative of the log-likelihood function with respect
% to alpha
I_11 = ( 1/alpha^2 - t^alpha*M^alpha*(log(t)-log(M))^2/(t^alpha+M^alpha)^2 ) ...
+ integral(fh2,(t/M)^alpha,inf)/(3*alpha^2) ...
* (1+t^alpha/M^alpha);
% define the function handle for the integrand of
% (log(w))^2 times density of beta prime (2, 2)
fh1 = @(x) log(x).*6.*x./(x+1).^4;
% I_12 is the negative expectation of the cross partial
% derivative of the log-likelihood function with respect
% to alpha & M
I_12 = ( 1/M/(1+(t/M)^alpha)^2 - integral(fh1,(t/M)^alpha,inf)/3/M ) ...
* (1+t^alpha/M^alpha) ...
- M^(alpha-1)/(t^alpha+M^alpha) ...
+ alpha*t^alpha*M^(alpha-1)*(log(t)-log(M))/(t^alpha+M^alpha)^2;
% I_22 is the negative expectation of the second partial
% derivative of the log-likelihood function with respect
% to M
I_22 = ( alpha*(alpha-1)/M^2/(1+(t/M)^alpha)^2 ...
- 2*alpha^2/(3*M^2)/(1+(t/M)^alpha)^3 ) ...
* (1+t^alpha/M^alpha) ...
- alpha*(alpha-1)*M^(alpha-2)/(t^alpha+M^alpha) ...
+ alpha^2*M^(2*alpha-2)/(t^alpha+M^alpha)^2;
% Same expression to compute mean curvature,
% information complexity criterion,
% asymptotic mean curvature criterion
% as for truncated Lomax model
%
%
% Run dataChampernowneMLE first
% Compute ICOMP AMC for shifted Champernowne model
% alpha is shape parameter of Champernowne distribution
alpha = tao_hat_sMLE;
% M is scale parameter of Champernowne distribution
M = theta_hat_sMLE;
% define the function handles for the integrand
f1 = @(y) alpha*M^alpha*y.^(2*alpha-1).*log(y).^2 ...
    ./(y.^alpha+M^alpha).^3;
f2 = @(y) alpha*M^alpha*y.^(3*alpha-1).*log(y).^2 ...
    ./(y.^alpha+M^alpha).^4;
f3 = @(y) alpha*M^(2*alpha)*log(M)*y.^(2*alpha-1).*log(y) ...
    ./(y.^alpha+M^alpha).^4;
% I_11
I_11 = 1/alpha^2 + 2*integral(f1,0,inf) + log(M)^2 ...
    - 2*integral(f2,0,inf) - 4*integral(f3,0,inf) ...
    -2*log(M)^2/3;
% define the function handle for the integrand
f4 = @(y) alpha^2*M^(2*alpha-1)*y.^(2*alpha-1).*log(y) ...
    ./(y.^alpha+M^alpha).^4;
```

```
% I_12
I_12 = alpha*log(M)/(3*M) - 2*integral(f4,0,inf);
% I_22
I_22 = alpha^2/(3*M^2);
% Same expression to compute mean curvature,
% information complexity criterion,
% asymptotic mean curvature criterion
% as for shifted Lomax model
% end of Table 4.1
%
%
% end of MATLAB code
```

Curriculum Vitae

Daoping Yu

2006	B.S., Information Management and Information Systems South China Normal University Guangzhou, China
2010	M.S., Mathematics and Statistics East Tennessee State University Johnson City, Tennessee
2016	Ph.D., Mathematics University of Wisconsin-Milwaukee Milwaukee, Wisconsin