Modeling Operational Risk using the Truncation Approach

by

Daniel P. Hadley

B.Sc., Economics, The University of Scranton, 2005

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Science

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Statistics)

The University of British Columbia
(Vancouver)

July 2018

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the thesis entitled:

Modeling Operational Risk using the Truncation Approach

submitted by **Daniel P. Hadley** in partial fulfillment of the requirements for the degree of **Master of Science** in **Statistics**.

Examining Committee:

Natalia Nolde, Statistics
Supervisor

Harry Joe, Statistics
Additional Examiner

Abstract

Banks that use the advanced measurement approach to model operational risk may struggle to develop an internal process that produces stable regulatory capital over time. Large decreases in regulatory capital are scrutinized by regulators while large increases may force banks to set aside more assets than necessary. A major source of this instability arises from the loss severity selection process, especially when the selected distribution families for severity risk categories change year-to-year. In this report, we examine the process of selecting severity distributions from a candidate distribution list within the guidelines of the advanced measurement approach, propose useful tools to aid in selecting an appropriate severity distribution, and analyze the effect of selection criteria on regulatory capital. The log sinh-arcsinh distribution family is added to a list of common candidate severity distributions used by industry. This 4-parameter family solves issues introduced by the 4-parameter g-and-h distribution without sacrificing flexibility and shows promise in outperforming 2-parameter families, reducing the frequency of severity distribution families changing year-to-year. Distribution parameters are estimated using the maximum likelihood approach from loss data truncated at a known minimum reporting threshold. Our severity distribution selection process combines truncation probability estimates with Akaike Information Criterion (AIC), Bayesian Information Criterion, modified Anderson-Darling, QQ-plots, and predictive measures such as the quantile scoring function and out-of-sample AIC, and we discuss some of the challenges associated with this process. We then simulate operational losses and calculate regulatory capital, comparing the effect on regulatory capital of selecting loss severity distributions using AIC versus quantile score. A combination of these two criteria is recommended when selecting loss severity distributions.

Lay Summary

Regulatory capital is the minimum amount of capital that a bank must set aside to cover future operational losses. The advanced measurement approach allows banks to develop internal models to calculate regulatory capital by estimating loss frequency and severity distributions from their internal loss data. A bank's internal model is updated annually, so that regulatory capital reflects the bank's current business environment. Operational loss severity data are often dominated by low probability, high severity events, and the annual selection of loss severity distributions is a major source for year-to-year volatility in the regulatory capital calculation. To mitigate volatility, this thesis analyzes a loss severity distribution selection process by investigating the log-sinh-arcsinh distribution to model loss severities and distribution selection criteria that combine relative measures of overall fit with predictive performance of extreme quantiles.

Preface

This dissertation is original, unpublished work by the author, Daniel Hadley, under the supervision of Professor Natalia Nolde and Professor Harry Joe. The research topic was suggested by the Global Risk Management department at Scotiabank. All simulations and analyses were designed and carried out by the author.

Table of Contents

List of Tables

List of Figures

Glossary

AIC *Akaike's Information Criterion*

AMA *Advanced Measurement Approach*

BASEL II *Basel II framework*

BCBS *Basel Committee on Banking Supervision*

BIC *Bayesian Information Criterion*

BUR *Burr Distribution*

CDF *Cumulative Distribution Function*

EVT *Extreme Value Theory*

GNH *g-and-h Distribution*

GPD *Generalized Pareto Distribution*

LDA *Loss Distribution Approach*

LDCE *Loss Data Collection Exercise*

LGN *Lognormal Distribution*

LGNGPD *Lognormal Body Spliced with Generalized Pareto Tail*

LGNLGN *Lognormal Body Spliced with Lognormal Tail*

LLOG *Loglogistic Distribution*

LSAS *log sinh-arcsinh Distribution*

MLE *Maximum Likelihood Estimation*

OOS *Out-of-Sample*

OR *Operational Risk*

ORX *ORX Global Database*

PDF *Probability Density Function*

PMLE *Penalized Maximum Likelihood Estimation*

QS *Quantile Scoring Function*

RC *Regulatory Capital*

RRC *Regulatory Risk Category*

RV Regularly Varying

SAS *Sinh-arcSinh Distribution*

SRC *Severity Risk Category*

SUBEX Subexponential

SUPEX Superexponential

WBL *Weibull Distribution*

Acknowledgments

I am forever grateful to my supervisor, Professor Natalia Nolde, for her patience, insight, and guidance throughout my research. Natalia is an invaluable resource for not only statistical knowledge, but also for advice on navigating life in graduate school. I am truly humbled by the time she has spent with me, the knowledge she has imparted, and patience with which she has helped me. I hope to one day emulate her qualities as a professor. I also want to thank Professor Harry Joe for being my second reader and whose door is always been open to my ceaseless questions.

Thank you to the Scotiabank Cybersecurity and Risk Analytics Initiative for funding my research, and to the Global Risk Management team at Scotiabank in Toronto for sharing their expertise.

Dedication

To my wife, Jessica, whose unending support and sacrifice allowed me to pursue this dream. She is my biggest advocate and best friend. I am fortunate to have such a loving partner to help me through life's challenges.

To my parents, for setting a good example by being the hardest working people I know.

Chapter 1

Introduction

According to the *Basel II framework* (BASEL II) last updated in 2006, *Operational Risk* (OR) as defined by the *Basel Committee on Banking Supervision* (BCBS) is

> *"the risk of a loss resulting from inadequate or failed internal processes, people and systems, or from external events. This definition includes legal risk but excludes strategic and reputational risk."* [BCBS, 2006]

Thus, OR is the risk arising from execution of a company's business functions [Embrechts and Hofert, 2011].

BASEL II provides recommendations on banking regulations which are issued and updated by BCBS. In this report, we focus on the *Advanced Measurement Approach* (AMA) to calculate the minimum capital requirements for operational risk. BASEL II intentionally grants banks a high degree of flexibility under the AMA to encourage growth in the discipline [Embrechts and Hofert, 2011]. The minimum amount of capital that a bank must set aside for one year to mitigate potential operational losses is called *Regulatory Capital* (RC) and is calculated each year as the 99.9%-quantile of the estimated total operational loss distribution. The total operational loss distribution is commonly estimated using the *Loss Distribution Approach* (LDA), which requires a bank to estimate and select distribution functions for both the loss frequency and severity of each category of operational losses.

A major obstacle facing a bank using the LDA is to develop a procedure that

calculates stable RC from year-to-year. Since RC must be set aside as a reserve, it represents assets that cannot be used freely by the bank to generate revenue. Thus, a bank has a vested interest in minimizing the probability of overestimating operational risk. Since regulators scrutinize large decreases in a bank's RC from one year to the next, a bank may have difficulty correcting an overestimation. Stability in RC is a common problem faced by industry that is not well covered by operational risk research.

A large source of RC's instability is the loss severity distribution selection process, which uses historical data to estimate distributions for a list of candidate distribution families and selects the "best" distribution based on some criteria. This process is repeated annually for each risk category. When a risk category's severity distribution changes from one family to another, the category's contribution to total annual operational risk is likely to change dramatically. Thus, we are interested in finding a flexible distribution family that can outperform other candidates year-after-year.

This flexible distribution family should be able to capture various tail behaviors and model the upper and lower tails separately. Two-parameter distribution families, while popular in industry, are not flexible enough to accomplish both goals. Spliced and mixture distributions can model the two tails separately, but are usually not able to capture various tail behaviors and are thus susceptible to changing distribution families. The four-parameter *g*-and-*h* distribution [Hoaglin, 1985] is a highly flexible distribution commonly used in operational risk, but has many challenges in practice. To maintain the flexibility of a four-parameter family while alleviating problems of the *g*-and-*h* distribution, we investigate the *Sinh-arcSinh Distribution* (SAS) [Jones and Pewsey, 2009] as a candidate model for loss severities.

Besides the list of candidate distribution families, we also analyze criteria for the loss severity distribution selection process. Since RC is the 99.9%-quantile of the total annual operational loss distribution, we are interested in criteria that assess the fit of the extreme right tail. Common estimators such as *Akaike's Information Criterion* (AIC) [Akaike, 1974] and *Bayesian Information Criterion* (BIC) [Schwarz, 1978] use likelihoods that give equal weight to each observation in a dataset and thus may overweight the central portion of the data. We investigate

the *Quantile Scoring Function* (QS) [Gneiting, 2011], which allows us to assess an estimated distribution's performance for specified quantiles, truncation probability estimates, and QQ-plots to aid in the selection of a loss severity distribution.

Banks struggling to calculate stable RC could be a contributing factor to the move away from the AMA. In the next manifestation of the Basel accords, Basel III, the AMA is being removed from the regulatory framework.

> *"The option to use an internal model-based approach for measuring operational risk - the 'Advanced Measurement Approaches' (AMA) - has been removed from the operational risk framework. BCBS believes that modeling of operational risk for regulatory capital purposes is unduly complex and that the AMA has resulted in excessive variability in risk-weighted assets and insufficient levels of capital for some banks."* [BCBS, 2016]

Regardless of regulatory requirements, OR remains a potentially catastrophic threat to a financial institution. In addition to the direct loss resulting from an operational loss event, financial institutions are likely to suffer further damages due to the loss of trust of their customers. These additional losses are considered reputational losses and are specifically excluded from OR. Even as regulation moves away from the AMA, a bank still has plenty of motivation for internal modeling of their OR as evidenced by recent operational losses.

According to the article, "The Final Bill - financial crime" [Economist, 2016], there have been 188 settlements since 2009 for criminal and civil prosecutions against banks costing $219 billion as of August 2016. Eleven firms have paid penalties in excess of 10% of their market capitalization. For example, Bank of America has paid the most in both dollars ($77 billion) and as a percentage of its market capitalization (50%). As a result, banks that saw opportunities for profit by operating in countries where bribery and suspicious transactions are tolerated are now finding the cost of operating in these environments exceed profits. More recently, a March 2018 article in the Wall Street Journal [Strasburg, 2018] reports that Barclays was ordered to pay $2 billion in civil penalties for fraudulently selling mortgage securities that contributed to the 2008 financial crisis. Additionally, two former Barclays executives were considered personally responsible for their role

and ordered to pay $2 million. In April 2018, Wells Fargo was fined $1 billion for the "bank's failures to catch and prevent problems, including improper charges to consumers in its mortgage and auto-lending businesses." [Hayahsi, 2018]. With so much at stake, we believe it is in a bank's best interest to continue to assess its OR exposure using the highly flexible AMA.

This report presents possible modeling procedures within the AMA guidelines and identifies assumptions and methodologies applied to the RC estimation procedure that may contribute to the variability in risk-weighted assets cited by the BCBS. The rest of this report is outlined as follows: Section 2 reviews the LDA under BASEL II and outlines our AMA for estimating the loss severity and frequency distributions for the various severity risk categories. We also address important challenges one may encounter when following our procedure. Section 3 presents a simple, heuristic approach to model estimation and selection of loss severity distributions using simulated data. We select loss severity models from a candidate list based on two different criteria, AIC and QS, and compare the impact of the selection criteria on RC. All simulations and numerical analyses use the R software available at https://www.r-project.org/. Then, we examine the issue of truncation probability estimates. Section 4 summarizes our conclusions and suggests areas for future research.

Chapter 2

Loss Distribution Approach

Under the BASEL II AMA guidelines, bank activities are partitioned into eight business lines and operational losses are categorized into seven event types. We call the occurrence of an operational loss a *loss event*. Loss events are mapped to an intersection of business line and event type, so that each loss event falls into one of 56 business line/event type intersections, called a *Regulatory Risk Category* (RRC).

The eight business lines are corporate finance; trading and sales; retail banking; commercial banking; payment and settlement; agency services; asset management; and retail brokerage. The seven event types are internal fraud; external fraud; employment practices and workplace security; clients, products, and business practices; damage to physical assets; business disruption and system failures; execution, delivery, and process management. As an example, we look at the $2 billion penalty assessed to Barclays [Strasburg, 2018] mentioned in Section 1. The loss amount is $2 billion occurring in year 2018 for the retail banking business line/internal fraud event type. A loss event is limited to one event type, but may affect multiple business lines simultaneously.

The AMA guidelines allow a bank to use another mapping, so long as it is transparent to third parties, approved by the board of directors, and independently reviewed. The number of internal event types and business lines may vary for different banks and usually depends upon the size of the bank and the amount of data available. Regardless of a bank's internal mapping, the bank must be able to map their historical losses to the eight business lines and seven event types outlined

in BASEL II. A loss event has a time stamp, a loss amount, and an associated RRC. The time stamp is usually a fiscal quarter or year and the loss amount is a positive value. The number of loss events that occur in a particular RRC over a given time period is called the *loss frequency*. The loss amount for a loss event is called *loss severity*.

Throughout this report, we exclusively consider an annual loss frequency which has many benefits. First, RC is an annual forecast so it is natural to work with historical data on the same frequency. Secondly, the use of an annual frequency naturally mitigates some of the reporting biases that were evidenced by the 2004 *Loss Data Collection Exercise* (LDCE). According to Dutta and Perry [2006], data collected during the 2004 LDCE show both structural reporting bias and temporal clustering of losses. Structural reporting bias is evidenced by a clear trend in loss events over time, most commonly seen as an increase in loss events as a bank's systems and processes for identifying operational losses improve. However, improved systems and processes may also decrease loss events as risk is identified and mitigated. Structural reporting bias can also affect loss severity, since earlier systems are more likely to catch large losses as opposed to smaller losses. The second type of reporting bias, temporal clustering of losses, commonly manifests as a disproportionate number of losses occurring on the last day of a fiscal quarter or the last fiscal quarter in a year. Thus, an annual frequency alleviates the temporal clustering bias and structural reporting bias for data at higher than annual frequencies, but may not completely address structural reporting bias for annual loss frequencies. We refer to Chavez-Demoulin et al. [2015] for a promising general solution to this problem that uses covariates to model trends in a loss frequency, but we do not incorporate those methods here since our simulated data assume no trend.

Historical loss data are used to estimate the loss frequency distribution and the loss severity distribution by interpreting the historical loss data as the realizations of random variables. Using notation adapted from Embrechts and Hofert [2011], we denote a loss event by

$$\left\{X_{t,n}^{b,l}\right\}, \quad \text{for } t = 1,2,...,T;\ b = 1,2,...,B;\ l = 1,2,...,L;\ n = 1,2,...,N_t^{b,l}; \quad (2.1)$$

where $X_{t,n}^{b,l}$ is a random variable for the loss severity of the n^{th} loss event occurring

in year t for business line b and event type l, and $N_t^{b,l}$ is a random variable for the number of losses occurring in year t for business line b and event type l. Thus, the *total annual operational loss* for next fiscal year can be calculated as

$$S_{T+1} = \sum_{b=1}^{B} \sum_{l=1}^{L} \sum_{n=1}^{N_{T+1}^{b,l}} X_{T+1,n}^{b,l}. \tag{2.2}$$

The goal of the LDA is to estimate the distribution of S_{T+1} and calculate the 99.9% quantile of the distribution of S_{T+1}. This number is the bank's RC for the next year.

The random variable S_{T+1} has two sources of randomness, the loss frequency and the loss severity. The loss frequency, $N_{T+1}^{b,l}$, is a discrete random variable for the number of losses in year $T+1$ in business line b and event type l. The loss severity, $X_{t,n}^{b,l}$, is a non-negative, continuous random variable as defined in (2.1). When calculating RC, we follow two common assumptions in operational risk modeling:

- The loss frequency, $N_{T+1}^{b,l}$, is independent of the loss severity, $X_{t,n}^{b,l}$, for a given (b,l)
- Loss severities are independent and identically distributed within a RRC

The first assumption does not exclude dependence of loss frequencies or severities across business lines/event types. The second assumption treats loss severities as independent and identically distributed through time, so that $X_{t,n}^{b,l}$ is independent of both t and n.

Under an AMA, a bank must have well-documented procedures that justify the ongoing relevance of the historical loss data included in the RC calculation. The historical loss data should reflect all current, material activities, risk exposures, and all relevant losses over a minimum gross loss threshold. For a bank's internal losses, BCBS sets the minimum threshold at \$10,000. The bank's internal loss data must be exclusively used when estimating the loss frequency for each RRC. When estimating loss severity distributions, however, data often contain too few observations to reliably estimate a distribution for each RRC. Therefore, a bank may combine losses across multiple business lines for a given event type to form a *Severity Risk Category* (SRC), so long as the bank's activities across business

lines for a given loss event type are similar enough to justify the assumption of a single loss severity distribution and the mapping of each loss to the standard business line/event type matrix is disclosed. The option for a bank to create SRC's is explicitly stated by BCBS Supervisory Guidelines [BCBS, 2011]:

> *"A bank should determine the optimum balance between granularity of the classes and volume of historical data for each class."*

While the desired number of loss events for a SRC is not given, research by Grooters and Reinink [2013] suggests optimal sample sizes between 500 and 10,000.

Even after forming SRC's, internal data may still be too sparse to estimate loss severity distributions for each SRC. To address this concern, BASEL II allows a banks' internal data to be supplemented by an external database. If the minimum reporting threshold of the data in the database differ from the bank's, the higher threshold is used and applied to all internal and external losses for that SRC. Minimum reporting thresholds and losses measured in different currencies are converted to a common currency using current exchange rates. Some external datasets include losses from banks of various sizes located all over the world, so care should be taken to filter the data so that they are appropriate to a bank's current business activities both in size and scope. For example, the *ORX Global Database* (ORX) contains more than 500,000 loss events whose loss amount exceeds a threshold of €20,000. ORX contains loss events whose loss amount may be multiple times larger than total assets held by a small to medium sized bank and should be excluded for such banks. Also, international banks may face different sources of risk that are not relevant to a local bank. If ORX external data are used to supplement internal SRC data, the higher ORX threshold should be applied to both internal and external data for that SRC. Since some SRC's may need to be supplemented by external data and others may not, a bank may have different minimum reporting thresholds for different SRC's.

For the remainder of this section, we discuss a procedure that uses historical loss data to estimate the loss frequency and severity distributions. When estimating loss frequency and severity distributions, we use SRC's instead of RRC's. To avoid confusion, a brief discussion of the mapping between RRC's and SRC's is presented.

For each loss event type l, a bank may combine losses across one or more

business lines, $b_{i_1}, b_{i_2}, \ldots, b_{i_k}$, and map the associated loss severities $X_{t,n}^{b_{i_j},l}$, for $t = 1,2,\ldots,T; n = 1,2,\ldots,N_t^{b_{i_j},l}; j = 1,2,\ldots,k \leq B$, to one and only one SRC, r. Thus, loss severities are pooled across time and business lines to create the historical loss data in SRC r. These pooled loss severities are treated as independent and identically distributed random variables from some loss severity distribution, F^r.

Loss frequency distributions, however, are estimated separately for each business line/event type, (b,l), by pooling the historical loss frequencies across time: $n_1^{b,l}, n_2^{b,l}, \ldots, n_T^{b,l}$. From this historical dataset, we estimate each loss frequency distribution, $F^{b,l}$. Thus, the loss frequency for SRC r is a random variable,

$$N_t^r = N_t^{b_{i_1},l} + N_t^{b_{i_2},l} + \cdots + N_t^{b_{i_k},l};$$

where $N_t^{b_{i_j},l} \sim F^{b_{i_j},l}$. In some instances, such as if $N_t^{b_{i_j},l} \overset{iid}{\sim} \text{Pois}(\lambda_j)$, the distribution of N_t^r is easily derived as $\text{Pois}(\lambda_1 + \lambda_2 + \cdots + \lambda_k)$.

Using SRC's simplifies our notation so that losses from (2.1) can be rewritten as

$$\left\{X_{t,n}^r\right\}, \quad \text{for } t = 1,2,...,T;\ r = 1,2,...,R;\ n = 1,2,...,N_t^r, \tag{2.3}$$

and the total annual operational loss from equation (2.2) can be rewritten as

$$S_{T+1} = \sum_{r=1}^{R} \sum_{n=1}^{N_{T+1}^r} X_{T+1,n}^r. \tag{2.4}$$

2.1 Loss Severity Distributions

A loss severity distribution, modeled respectively for each SRC, has an associated density function to describe the probability of the loss amount given an operational loss event occurs. For our LDA procedure, we use nine loss severity distribution families to model loss amounts for each SRC. These distributions are listed in Table 2.1. Of these nine candidates, seven are comprised of a single distribution to describe all loss amounts, and two are *piecewise distributions*. A piecewise distribution can potentially handle a situation where loss events in the body and tail

are driven by two unique processes. The *splicing point* of a piecewise distribution is the loss amount that separates the body and tail of the spliced distribution. All losses exceeding the splicing point fall into the tail and all losses below fall into the body. The splicing point is treated as an additional parameter that must be estimated.

The single distribution candidate families are selected to span various tail behaviors and include common distributions employed by industry. In a survey published in 2009, BCBS reported that 33% of surveyed banks use the lognormal distribution and 17% use the Weibull distribution when modeling losses by a single severity distribution. This same survey reported that when modeling the body and tail distributions separately, 14% use lognormal to model the tail while 31% use generalized Pareto [BCBS, 2009].

Since operational loss data only include losses whose loss amounts exceed a minimum threshold, the datasets are incomplete. In particular, since we have no information about the frequency and severity of losses occurring below the threshold, the data are said to be *truncated from below* or *left-truncated*. Note that left-truncated data differ from *left-censored* data, because left-censored data include the number of observations below the minimum threshold.

There are three distinct approaches for handling left-truncated data. The first approach, commonly referred to as the *naive approach*, ignores the threshold and models the dataset as if it were complete. Evidence in the existing literature indicates that this approach may underestimate both the loss frequency and loss severity simultaneously [Baud et al., 2003, Chernobai et al., 2005, Luo et al., 2007]. As a result, the naive approach is likely to underestimate RC. The second approach models the excess loss amount over the minimum threshold. This is referred to as the *shifted approach*, since loss amounts are shifted downward by the amount of the reporting threshold. Luo et al. [2007] states that the shifted approach underestimates loss frequency, but overestimates loss severity. The aggregate effect on RC is thus uncertain. However, Dutta and Perry [2006] use this approach to effectively produce "realistic" RC estimates using the *g*-and-*h* distribution as the loss severity distribution, where realistic RC estimates are estimates that constitute less than 3% of a bank's total assets. Therefore, the shifted approach should not be completely disregarded especially when data and expert opinion for losses below the threshold

Distribution	Notation	Parameters	Tail Behavior
Lognormal	LGN	location: $\mu \in \mathbb{R}$ scale: $\sigma > 0$	SUBEX
Generalized Pareto	GPD	location: $\mu \in \mathbb{R}$ scale: $\theta > 0$ tail: $\xi \in \mathbb{R}$	$\xi > 0 \implies$ RV $\xi = 0 \implies$ Exponential $\xi < 0 \implies$ Bounded above
Burr	BUR	shape 1: $\alpha \in \mathbb{R}$ shape 2: $\gamma > 0$ scale: $\theta > 0$	RV
Weibull	WBL	shape: $a > 0$ scale: $\theta > 0$	$a < 1 \implies$ SUBEX $a > 1 \implies$ SUPEX
Loglogistic	LLOG	shape: $\gamma > 0$ scale: $\theta > 0$	RV
g-and-h	GNH	location: $a \in \mathbb{R}$ scale: $b > 0$ skewness: $g \in \mathbb{R}$ elongation: $h \in \mathbb{R}$	$h > 0 \implies$ RV $h = 0 \implies$ SUBEX $h < 0 \implies$ SUPEX
Log-sinh-arcsinh	LSAS	location: $a \in \mathbb{R}$ scale: $b > 0$ skewness: $\varepsilon \in \mathbb{R}$ elongation: $\delta > 0$	$\delta \leq 0.5 \implies$ RV $0.5 < \delta < 1 \implies$ SUBEX $\delta > 1 \implies$ SUPEX
Lognormal Body Lognormal Tail	LGNLGN	location: $\mu_b \in \mathbb{R}$ scale: $\sigma_b > 0$ splice point: $x_s \in \mathbb{R}$ location: $\mu_u \in \mathbb{R}$ scale: $\sigma_u > 0$	SUBEX
Lognormal Body Pareto Tail	LGNGPD	location: $\mu_b \in \mathbb{R}$ scale: $\sigma_b > 0$ splice point: $x_s \in \mathbb{R}$ tail: $\xi \in \mathbb{R}$	$\xi > 0 \implies$ RV $\xi = 0 \implies$ Exponential $\xi < 0 \implies$ Bounded above

Table 2.1: Candidate distributions: A distribution has a Regularly Varying (RV) right tail if its density decreases to 0 at the rate x^{-b} with $b > 1$, Subexponential (SUBEX) if its density decreases to 0 slower than e^{-x}, but faster than RV, or Superexponential (SUPEX) if its density decreases to 0 faster than e^{-x}. For more theoretical definitions, see Foss et al. [2013].

are unavailable. The third approach, called the *conditional approach* or *truncation approach*, treats operational loss data as left-truncated at the minimum reporting threshold. The truncation approach assumes the following:

- Losses below and above the minimum threshold belong to the same distribution
- For SRC r, loss frequency, N_t^r, and loss severity, $X_{t,n}^r$, can be treated as independent random variables
- For SRC r, all loss severities, $X_{t,n}^r$, are independent and identically distributed random variables from the loss severity distribution, F^r.

While the truncation approach is often favored over the naive approach, see Chapter 9 of Chernobai et al. [2007], a recent study by Yu and Brazauskas [2017] shows that the truncation approach often leads to lower RC estimates than both the naive and shifted approaches. Whether these lower RC estimates are more accurate or not is still up for debate. We believe that the truncation approach introduces uncertainty regarding the proportion of operational losses that occur below the truncation point, which may have contributed to the results of Yu and Brazauskas. We elaborate on this issue in Section 3.4.

The truncation approach estimates the loss severity distribution using only the observed data above the reporting threshold. For n loss events exceeding the minimum threshold in a given SRC, assume loss amounts $X_1, X_2, ..., X_n \overset{iid}{\sim} F^r$, for some loss severity distribution F^r with parameters $\theta \in \Theta$. For the remainder of Section 2.1, we drop the superscript r for loss severity distributions. If we let τ represent the minimum reporting threshold of the given SRC, where τ is always non-random, then the conditional *Cumulative Distribution Function* (CDF) for a loss given that it exceeds τ is defined by

$$\widetilde{F}(x;\Theta,\tau) = \frac{F(x;\Theta) - F(\tau;\Theta)}{1 - F(\tau;\Theta)}. \tag{2.5}$$

The conditional *Probability Density Function* (PDF) for a loss given that it exceeds

τ is

$$\tilde{f}(x;\Theta,\tau) = \frac{d}{dx}\widetilde{F}(x;\Theta,\tau) = \frac{f(x;\Theta)}{1-F(\tau;\Theta)}. \tag{2.6}$$

We can find the estimated unconditional CDF, $F(x)$, and PDF, $f(x)$, by assuming a parametric form for F and performing *Maximum Likelihood Estimation* (MLE) on a sample, $\mathbf{x}$, using the conditional likelihood function,

$$\widetilde{L}(\Theta;\mathbf{x},\tau) = \prod_{i=1}^{n} \tilde{f}(x_i;\Theta,\tau) = \left[1-F(\tau;\Theta)\right]^{-n} \prod_{i=1}^{n} f(x_i;\Theta). \tag{2.7}$$

Maximizing $\widetilde{L}(\Theta;\mathbf{x},\tau)$ over Θ results in distribution parameter estimates $\widehat{\Theta}$.

We define *truncation probability* as the probability that a loss event's severity is less than or equal to the minimum reporting threshold, given a loss event has occurred. Truncation probability is estimated for each SRC by evaluating the estimated unconditional CDF at the minimum reporting threshold, $F(\tau;\widehat{\Theta})$. When using the truncation approach to estimate loss severity distributions, this estimate is used to correct the downward bias of the historical loss event frequency for the given SRC when simulating losses over the entire SRC distribution. See Section 2.2 for details on estimating the loss frequency.

2.1.1 Estimating Parameters for Single Severity Distribution Candidates

Seven of the nine candidate distributions assume that all loss amounts above and below τ in a given SRC can be estimated by a single distribution. This subsection details the estimation procedure for the unconditional candidate distribution families: lognormal, generalized Pareto, Burr, Weibull, loglogistic, *g*-and-*h*, and log-SaS. See Appendix A for details on the loss severity distributions and their parameterizations.

Under the truncation approach, MLE is performed by minimizing the conditional negative log-likelihood function. Given a sample, $\mathbf{x}$, of n loss amounts from a single SRC, the conditional likelihood function is given by equation (2.7). The

conditional log-likelihood function is

$$\tilde{\ell}(\Theta;\mathbf{x},\tau) = \log\left(\widetilde{L}(\Theta;\mathbf{x},\tau)\right) = -n\,\log\left(1-F(\tau;\Theta)\right) + \sum_{i=1}^{n}\log\left(f(x_i;\Theta)\right). \quad (2.8)$$

The estimated parameters, $\widehat{\Theta}$, are found by maximizing the conditional log-likelihood function, or equivalently, minimizing the negative conditional log-likelihood function, $\widetilde{n\ell}$, defined by

$$\widetilde{n\ell}(\Theta;\mathbf{x},\tau) = -\tilde{\ell}(\Theta;\mathbf{x},\tau).$$

Since we perform MLE as the minimization of the negative conditional log-likelihood function, the maximum likelihood estimates are the solution to the following minimization problem,

$$\widehat{\Theta} = \underset{\Theta}{\operatorname{argmin}}\left(\widetilde{n\ell}(\Theta;\mathbf{x},\tau)\right), \quad (2.9)$$

and thus each candidate distribution, F, has an associated *estimated distribution*, $F(x;\widehat{\Theta})$. From now on, we use likelihood function to refer to the conditional likelihood function from equation (2.7), and we use log-likelihood function to refer to the conditional log-likelihood function from equation (2.8).

2.1.2 Estimating Parameters for Piecewise Severity Distribution Candidates

In addition to the single loss severity distribution candidates, we consider two piecewise distributions, *Lognormal Body Spliced with Lognormal Tail* (LGNLGN) and *Lognormal Body Spliced with Generalized Pareto Tail* (LGNGPD). The spliced distributions assume that the unobservable losses with amounts below τ follow the distribution of the body. Let X be a non-negative random variable generated from the piecewise distribution F, which is comprised of one distribution for the body, F_{body}, and another for the tail, F_{tail}. To derive the conditional PDF of a piecewise distribution, we first derive the conditional CDF and PDF separately for the body and the tail. For the derivation, we treat the splicing point as given, but the splicing point is an additional parameter that must be estimated.

Given a splicing point x_s, minimum reporting threshold τ, where $\tau < x_s$, and parameter vector Θ_b, the conditional CDF for the body can be written as

$$\begin{aligned}
\widetilde{F}_{body}(x;\Theta_b,\tau,x_s) &= P(X \le x | \tau < X \le x_s) \\
&= \frac{P\big(\tau < X \le \min(x,x_s)\big)}{P(\tau < X \le x_s)} \\
&= \frac{F_{body}\big(\min(x,x_s);\Theta_b\big) - F_{body}(\tau;\Theta_b)}{F_{body}(x_s;\Theta_b) - F_{body}(\tau;\Theta_b)} \\
&= \begin{cases} 1 & \text{for } x > x_s \\ \dfrac{F_{body}(x;\Theta_b) - F_{body}(\tau;\Theta_b)}{F_{body}(x_s;\Theta_b) - F_{body}(\tau;\Theta_b)} & \text{for } \tau < x \le x_s \\ 0 & \text{for } x \le \tau, \end{cases}
\end{aligned}$$

where the last equality uses the condition $x \le x_s$ in order for an observation to be in the body. The conditional PDF for the body follows as usual,

$$\begin{aligned}
\tilde{f}_{body}(x;\Theta_b,\tau,x_s) &= \frac{d}{dx}\widetilde{F}_{body}(x;\Theta_b,\tau,x_s) \\
&= \begin{cases} \dfrac{1}{F_{body}(x_s;\Theta_b) - F_{body}(\tau;\Theta_b)} f_{body}(x;\Theta_b) & \text{for } \tau < x \le x_s \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}$$

The conditional CDF and PDF for the tail are derived in the same manner as equations (2.5) and (2.6) for the single severity distributions by treating the splicing point, x_s, as the minimum reporting threshold. Treating x_s as given and Θ_u as the vector of tail distribution parameters, we find the tail CDF and PDF, respectively, to be

$$\widetilde{F}_{tail}(x;\Theta_u,x_s) = \begin{cases} \dfrac{F_{tail}(x;\Theta_u) - F_{tail}(x_s;\Theta_u)}{1 - F_{tail}(x_s;\Theta_u)} & \text{for } x_s < x \\ 0 & \text{otherwise} \end{cases}$$

$$\tilde{f}_{tail}(x;\Theta_u,x_s) = \begin{cases} \dfrac{1}{1 - F_{tail}(x_s;\Theta_u)} f_{tail}(x;\Theta_u) & \text{for } x_s < x \\ 0 & \text{otherwise.} \end{cases}$$

In order to piece the body and tail together, we consider a truncated sample, $\mathbf{x}$, of n observed loss amounts ordered from smallest to largest,

$$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n_b)} \leq x_{(n_b+1)} \leq \cdots \leq x_{(n)},$$

where $n_b := \max\{ j \in 1,2,\ldots,n \mid x_{(j)} \leq x_s \}$. All sample observations less than or equal to x_s are in the body of the sample, and we define the proportion of the sample in the body as $p_b = \frac{n_b}{n}$. The remaining n_u observations, where $n_u = n - n_b$, are in the tail.

To derive the conditional piecewise PDF, $\tilde{f}$, we want to make sure $\tilde{f}$ integrates to 1 over the support of the random loss severities. Let the parameter vector for the piecewise distribution be $\Theta = [\Theta_b \; x_s \; \Theta_u]$. Then, for $a, b > 0$,

$$\int_0^\infty \tilde{f}(x;\Theta,\tau) = a \int_0^\infty \tilde{f}_{body}(x;\Theta_b,\tau,x_s)dx + b \int_0^\infty \tilde{f}_{tail}(x;\Theta_u,x_s)dx = a + b = 1,$$

with constraints

$$\begin{aligned} \int_\tau^{x_s} \tilde{f}(x;\Theta,\tau)dx &= p_b \\ \int_{x_s}^\infty \tilde{f}(x;\Theta,\tau)dx &= 1 - p_b. \end{aligned}$$

Thus, we see that $a = p_b$ and $b = 1 - p_b$. We now have the conditional piecewise PDF

$$\begin{aligned} \tilde{f}(x;\Theta,\tau) &= \begin{cases} p_b \; \tilde{f}_{body}(x;\Theta_b,\tau,x_s) & \text{for } \tau < x \leq x_s \\ (1-p_b) \; \tilde{f}_{tail}(x;\Theta_u,x_s) & \text{for } x_s < x \end{cases} \\ &= \begin{cases} \dfrac{p_b \; f_{body}(x;\Theta_b)}{F_{body}(x_s;\Theta_b) - F_{body}(\tau;\Theta_b)} & \text{for } \tau < x \leq x_s \\ \\ \dfrac{(1-p_b)}{1 - F_{tail}(x_s;\Theta_u)} \; f_{tail}(x;\Theta_u) & \text{for } x_s < x. \end{cases} \end{aligned} \tag{2.10}$$

The conditional body and tail densities are used to find their respective likelihood functions. Then, we perform MLE separately on the body and tail of the sample to find the piecewise distribution that maximizes the sum of the body and tail likelihood functions. An outline of the MLE approach for each of the two piecewise candidate distributions follows.

The piecewise distribution parameter vector, Θ, includes the parameters for the body distribution, Θ_b, the splicing point, x_s, and the parameters for the upper tail distribution Θ_u. To start the MLE procedure, we restrict the set of possible estimates of x_s to the sample percentiles x_p, where $p = 0.3, 0.32, 0.34, ..., 0.96$, so that $\widehat{x}_s \in \{x_{0.3}, x_{0.32}, \ldots, x_{0.96}\}$. For each p, treat x_p as the splicing point. The sample is then split into a body, $\mathbf{x}_{b,p} \in \mathbb{R}^{n_b}$, and a tail $\mathbf{x}_{u,p} \in \mathbb{R}^{n_u}$, where $\mathbf{x}_{b,p} = \{x \in \mathbf{x} \mid x \leq x_p\}$, $\mathbf{x}_{u,p} = \{x \in \mathbf{x} \mid x > x_p\}$, and $n_b + n_u = n$. Since this process is performed for each p, we get 34 possible estimates for the parameter vectors of the piecewise distribution,

$$\widehat{\Theta}_p = \begin{bmatrix} \widehat{\Theta}_{b,p} & x_p & \widehat{\Theta}_{u,p} \end{bmatrix},$$

where $\widehat{\Theta}_{b,p}$ and $\widehat{\Theta}_{u,p}$ are the estimated parameters for the body and tail distributions, respectively, given that x_p is the splicing point. Finally, the piecewise distribution is determined by choosing the value of p that minimizes the negative log-likelihood function. Details are below.

For either the LGNLGN or LGNGPD distribution, the procedure to estimate the distribution of the body is the same. For each x_p, $\mathbf{x}$ is separated into a body $\mathbf{x}_{b,p}$ and a tail $\mathbf{x}_{u,p}$, and the estimated parameters for the distribution of the body, denoted $\widehat{\Theta}_{b,p}$, are found by minimizing the negative log-likelihood function for the body, denoted $\widetilde{n\ell}_{b,p}$. From the conditional piecewise PDF in equation (2.10),

$$\begin{aligned} \widetilde{n\ell}_{b,p}(\Theta_b; \mathbf{x}_{b,p}, \tau, x_p) = n_b \ \log\left(F_{body}(x_p; \Theta_b) - F_{body}(\tau; \Theta_b)\right) - n_b \ \log(p_b) \\ - \sum_{i=1}^{n_b} \log\left(f_{body}(x_i; \Theta_b)\right), \end{aligned}$$

where τ is the minimum reporting threshold determined by the operational loss sample's SRC. For each of the 34 estimates of the splicing point, $\hat{x}_s = x_p$, we estimate the lognormal distribution for the body, $F_{body}(x; \widehat{\Theta}_{b,p})$, by solving the minimization problem

$$\widehat{\Theta}_{b,p} = \underset{\Theta_b}{\operatorname{argmin}} \left(\widetilde{n\ell}_{b,p}(\Theta_b; \mathbf{x}_{b,p}, \tau, x_p) \right).$$

The tail distribution estimation procedure differs for each piecewise candidate. We start with LGNLGN, as the process to estimate the tail is the same as the body.

For each x_p, the estimated parameters for the distribution of the upper tail, denoted $\widehat{\Theta}_{u,p}$, are found by minimizing the negative log-likelihood function for the tail, denoted $\widetilde{n\ell}_{u,p}$. From the conditional piecewise PDF in equation (2.10),

$$\widetilde{n\ell}_{u,p}(\Theta_u; \mathbf{x}_{u,p}, x_p) = n_u \log\big(1 - F_{tail}(x_p; \Theta_u)\big) - n_u \log(1 - p_b) - \sum_{i=1}^{n_u} \log\big(f_{tail}(x_i; \Theta_u)\big).$$

Just as for the body, we end up with 34 estimated lognormal distributions for the tail. For each x_p, $F_{tail}(x; \widehat{\Theta}_{u,p})$ is found by solving the minimization problem

$$\widehat{\Theta}_{u,p} = \underset{\Theta_u}{\operatorname{argmin}} \Big(\widetilde{n\ell}_{u,p}(\Theta_u; \mathbf{x}_{u,p}, x_p) \Big).$$

After estimating the body and tail distributions for each of the 34 splicing point estimates, we have 34 LGNLGN distributions identified by their parameters

$$\widehat{\Theta}_p = \begin{bmatrix} \widehat{\Theta}_{b,p} & x_p & \widehat{\Theta}_{u,p} \end{bmatrix}.$$

The estimated LGNLGN piecewise distribution parameters are found by solving

$$\widehat{\Theta} = \underset{p}{\operatorname{argmin}} \Big(\widetilde{n\ell}_{b,p}(\widehat{\Theta}_{b,p}; \mathbf{x}_{b,p}, \tau, x_p) + \widetilde{n\ell}_{u,p}(\widehat{\Theta}_{u,p}; \mathbf{x}_{u,p}, x_p) \Big),$$

where the p that minimizes the above equation also determines the estimate of the splicing point.

The LGNLGN distribution allows for discontinuity in the PDF at the splicing point. There are arguments both for and against continuity constraints for spliced distributions. For example, imposing continuity constraints on a piecewise density may not lead to a better likelihood measure of fit than allowing for a jump discontinuity. On the other hand, allowing discontinuity seems to favor the fit of the tail as splicing points tend to be chosen in the lower half of possible values. A discussion on the decision to impose continuity constraints on a spliced distribution can be found in Chapter 1 of Peters and Shevchenko [2015].

For the conditional LGNGPD distribution family, we force the body and tail

distributions to be equal at the splicing point by imposing the constraint

$$p_b \; \tilde{f}_{body}(x_s;\Theta_b,\tau,x_s) = (1-p_b) \; \tilde{f}_{tail}(x_s;\Theta_u,x_s)$$

$$\frac{p_b \; f_{body}(x_s;\Theta_b)}{F_{body}(x_s;\Theta_b) - F_{body}(\tau;\Theta_b)} = \frac{1-p_b}{\theta}\Big[1+\xi\frac{x_s-x_s}{\theta}\Big]^{-1-1/\xi}$$

$$\theta = \frac{1-p_b}{p_b} \; \frac{F_{body}(x_s;\Theta_b) - F_{body}(\tau;\Theta_b)}{f_{body}(x_s;\Theta_b)}.$$

This constraint forces the scale parameter of generalized Pareto tail distribution, θ, to be completely determined by the estimated body distribution. As a result, the scale parameter is not treated as a parameter for the LGNGPD distribution. While this continuity constraint prevents jumps in the density, it does not require differentiability of the density at the splicing point. One could impose such a constraint using the derivatives of the conditional body and tail density functions, see Peters and Shevchenko [2015].

Since the scale parameter for the tail distribution is determined by the log-normal body distribution, we only need to estimate the tail parameter, ξ. For the LGNGPD distribution, the losses in the tail of the sample are modeled as the excess losses over the splicing point. This is the "Peaks-Over-Threshold" method in *Extreme Value Theory* (EVT). Given this approach, EVT offers alternative approaches for estimating the splicing point that emphasize the fit of the tail as a generalized Pareto distribution by the use of mean residual life and parameter stability plots. For a practical discussion with examples, see Chapter 4 of Coles [2001].

For each x_p and associated estimate of the body distribution, $F_{body}(x;\widehat{\Theta}_{b,p})$, the estimated scale parameter for the tail distribution is calculated as

$$\hat{\theta}_p = \frac{1-p_b}{p_b} \; \frac{F_{body}(x_s;\widehat{\Theta}_{b,p}) - F_{body}(\tau;\widehat{\Theta}_{b,p})}{f_{body}(x_s;\widehat{\Theta}_{b,p})}.$$

The estimate of the tail parameter, $\hat{\xi}$, is found by minimizing the negative log-

likelihood function for the tail over ξ,

$$\tilde{\ell}_{u,p}(\xi;\mathbf{x}_{u,p},x_p,\hat{\theta}_p) = n_u \log\{1-p_b\} - n_u \log(\hat{\theta}_p) \\ -\Big(\frac{\xi+1}{\xi}\Big)\sum_{i=1}^{n_u}\log\Big(1+\frac{\xi}{\hat{\theta}_p}(x_i-x_p)\Big),$$

and

$$\hat{\xi}_p = \underset{\xi}{\operatorname{argmin}}\Big(\widetilde{n\ell}_{u,p}(\xi;\mathbf{x}_{u,p},x_p,\hat{\theta}_p)\Big).$$

For each x_p, the estimated LGNGPD distribution has four estimated parameters,

$$\widehat{\Theta}_p = \begin{bmatrix}\widehat{\Theta}_{b,p} & x_p & \hat{\xi}_p\end{bmatrix},$$

and the estimated LGNGPD distribution has parameter vector $\widehat{\Theta}$, which is found by solving

$$\widehat{\Theta} = \underset{p}{\operatorname{argmin}}\Big(\widetilde{n\ell}_{b,p}(\widehat{\Theta}_{b,p};\mathbf{x}_{b,p},\tau,x_p) + \widetilde{n\ell}_{u,p}(\hat{\xi}_p;\mathbf{x}_{u,p},x_p,\hat{\theta}_p)\Big),$$

where the p that minimizes the above equation also determines the estimate of the splicing point.

Finally, to derive the unconditional piecewise CDF, PDF, and quantile function, we must normalize the conditional piecewise density to integrate to 1. Let $f(x;\Theta)$ be the unconditional piecewise density with unconditional body density $f_{body}(x;\Theta_b)$ and unconditional tail density $f_{tail}(x;\Theta_u)$. Then, by letting c be the normalizing constant,

$$c\int_{-\infty}^{\infty}\tilde{f}(x;\Theta,\tau)dx = 1,$$

$$ca\int_{-\infty}^{x_s}\tilde{f}_{body}(x;\Theta_b,\tau,x_s)dx + cb\int_{x_s}^{\infty}\tilde{f}_{tail}(x;\Theta_u,x_s)dx = 1,$$

$$c\int_{-\infty}^{x_s}\frac{p_b\, f_{body}(x;\Theta_b)}{F_{body}(x_s;\Theta_b)-F_{body}(\tau;\Theta_b)}dx + c\int_{x_s}^{\infty}\frac{1-p_b}{1-F_{tail}(x_s;\Theta_u)}f_{tail}(x;\Theta_u)dx = 1,$$

$$c\ \frac{p_b\ F_{body}(x_s;\Theta_b)}{F_{body}(x_s;\Theta_b)-F_{body}(\tau;\Theta_b)}+c\ \frac{1-p_b}{1-F_{tail}(x_s;\Theta_u)}\ \Big(1-F_{tail}(x_s;\Theta_u)\Big)=1,$$

$$c\ \frac{p_b}{F_{body}(x_s;\Theta_b)-F_{body}(\tau;\Theta_b)}\ \Big(F_{body}(x_s;\Theta_b)-0\Big)+c\ (1-p_b)=1,$$

$$c\ \frac{F_{body}(x_s;\Theta_b)-(1-p_b)F_{body}(\tau;\Theta_b)}{F_{body}(x_s;\Theta_b)-F_{body}(\tau;\Theta_b)}=1.$$

Finally, we arrive at the unconditional piecewise density function by multiplying the conditional piecewise density by the normalizing constant,

$$f(x;\Theta,\tau)=\begin{cases}\dfrac{p_b\ f_{body}(x;\Theta_b)}{F_{body}(x_s;\Theta_b)-(1-p_b)F_{body}(\tau;\Theta_b)} & \text{for } x\le x_s\\[2ex] c\dfrac{(1-p_b)\ f_{tail}(x;\Theta_u)}{1-F_{tail}(x_s;\Theta_u)} & \text{for } x>x_s.\end{cases}$$

2.1.3 Analysis of Estimated Severity Distributions

To assess the estimation of loss severity candidate distributions, we employ qualitative metrics outlined by Dutta and Perry [2006]:

1. Good Fit - Statistically, how well does the method fit the data?
2. Realistic - If a method fits well in a statistical sense, does it generate a loss distribution with a realistic capital estimate?
3. Well-Specified - Are the characteristics of the fitted data similar to the loss data and logically consistent?
4. Flexible - How well is the method able to reasonably accommodate a wide variety of empirical loss data tail behavior?
5. Simple - Is the method easy to apply in practice, and is it easy to generate random numbers for the purposes of loss simulation?

One measure of model performance is *Akaike's Information Criterion* (AIC) as

developed by Akaike [1974]. The AIC is defined as

$$AIC = -2\,\tilde{\ell}(\widehat{\Theta};\mathbf{x},\tau) + 2\,k, \tag{2.11}$$

where $\widehat{\Theta}$ is the estimated distribution parameter vector as found via MLE and k is the number of estimated parameters in the distribution. Including the number of parameters in the AIC is an attempt to prevent selecting a model that overfits the data. The AIC is a relative performance measure, so while it can pick which model fits the data better than another, it cannot tell if any model is a good fit. Other diagnostics, such as QQ-plots and density plots should be consulted for adequacy.

Another measure that compares model performance is the *Bayesian Information Criterion* (BIC) as developed by Schwarz [1978]. Like AIC, the BIC allows for a comparison across models with different numbers of parameters by incorporating "a mathematical formula for the principle of parsimony in model building". The BIC is defined as

$$BIC = -2\,\tilde{\ell}(\widehat{\Theta};\mathbf{x},\tau) + k\,\log(n), \tag{2.12}$$

where $\widehat{\Theta}$ is the MLE parameter vector, k is the dimensionality of the parameter vector, and n is the number of observations in the sample $\mathbf{x}$. Compared to AIC, BIC favors models with fewer parameters when $n \geq 8$, since model dimensionality is multiplied by $\log(n)$ instead of 2.

The final measure of fit we employ is the modified Anderson-Darling test [Sinclair et al., 1990], which uses the differences between empirical quantiles and estimated quantiles from the candidate distributions while assigning higher weights to higher quantiles than the standard Anderson-Darling test. The idea is to measure goodness-of-fit in the tail from each candidate distribution. The modified Anderson-Darling test statistic is calculated as

$$\widehat{AD} = \frac{n}{2} - 2\,\sum_{i=1}^{n} \widetilde{F}(x_{(i)};\widehat{\Theta},\tau) - \sum_{i=1}^{n} \left[2 - \frac{2i-1}{n}\right] \log\left[1 - \widetilde{F}(x_{(i)};\widehat{\Theta},\tau)\right],$$

where n is the number of observations, $x_{(i)}$ is the i^{th} order statistic such that $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$, and $\widetilde{F}(x;\widehat{\Theta},\tau)$ is the estimated conditional CDF for the candidate

distribution. To perform the modified Anderson-Darling test at the 95% confidence level, we use $\widehat{AD}$ to calculate the p-value,

$$p_{AD} = \left[1 + \exp\left\{2.31 + 1.73\widehat{AD} + \frac{0.275}{\widehat{AD}} - \frac{2}{\sqrt{\widehat{AD}}} - \frac{0.092}{\widehat{AD}^{3/2}}\right\}\right]^{-1}.$$

If $p_{AD} < 0.05$, then the null hypothesis that the data follow the estimated candidate distribution is rejected at the 95% significance level.

We include the results of the modified Anderson-Darling test, but QQ-plots are preferred for their simplicity and interpretability. Especially in the case of model misspecification, the modified Anderson-Darling test may reject a true null hypothesis as we show in Section 3.2.3.

Forecasting ability for the estimated distributions is measured by the *Quantile Scoring Function* (QS) and *Out-of-Sample* (OOS) AIC. Both make OOS predictions to gauge forecasting performance. Following Gneiting [2011], let $\alpha = 0.999$ be the quantile. For a sample $\mathbf{x}$ of n operational losses, let $\mathbf{x}_{(-i)}$, for $i = 1, 2, ..., n$, be the sample with the i-th observation removed. We define $q_{(-i)}$ as the estimated α-quantile when the i^{th} observation is excluded. Then,

$$q_{(-i)} = F^{-1}\left(\alpha; \widehat{\Theta}, \mathbf{x}_{(-i)}\right),$$

where $\widehat{\Theta}$ is found by using the MLE approach outlined in Sections 2.1.1 and 2.1.2 for the sample $\mathbf{x}_{(-i)}$. The quantile scoring function is defined as

$$S(q_{(-i)}, x_i) = \frac{1}{n}\sum_{i=1}^{n}\left(\mathbb{1}(q_{(-i)} \geq x_i) - \alpha\right)\left(q_{(-i)} - x_i\right), \tag{2.13}$$

which is non-negative with values closer to zero indicating better performance.

When $\alpha = 0.999$, the quantile scoring function is asymmetric, penalizing more for severe underestimation than for overestimation. This asymmetric feature should be particularly appealing to regulators, who want to avoid underestimation of risk. The QS should also appeal to financial institutions, who face the possibility of extremely overestimating RC when using candidate distributions capable of modeling heavy tail behavior, see Dutta and Perry [2006]. Since the 99.9% quantile from an

SRC's severity distribution is a good proxy for its contribution to RC, see Section 3.3.2, we employ the QS when selecting a candidate distribution and compare how this affects the RC calculation as opposed to selecting severity distributions by AIC. Since the QS relies heavily on the MLE process, the pitfalls mentioned in Section 2.1.4 become even more important when selecting severity distribution by QS. This is especially true when the MLE algorithm fails to converge due to boundary conditions which can greatly impact the tail behavior.

To see the asymmetry of the QS at $\alpha = 0.999$, we look at a sample of 2500 independent and identically distributed random variables distributed uniformly between 0 and 1000. The true 0.999-quantile is 999, so the QS is minimized when we forecast 999. Forecasts below 999 are penalized more than forecasts that exceed 999 by the same amount. Figure 2.1 plots the quantile scoring function when the 0.999-quantile is forecasted to be 899 thru 1299.

OOS AIC, denoted AIC_{OOS}, is calculated by excluding each year's operational losses, estimating each candidate distribution via MLE on the remaining losses, and calculating the likelihood for the excluded year's data. This process is repeated so that each year's losses have been excluded exactly once. The excluded-year likelihoods are then summed and AIC_{OOS} is calculated as in equation (2.11). These simple forecasting metrics allow us to assess whether the estimated candidate distribution is well-specified and flexible and when combined with QQ-plots, can gauge how realistic the quantile estimates are for each estimated candidate distribution family.

Finally, the truncation probability estimate, $F(\tau;\widehat{\Theta})$, may provide a simple metric to gauge whether an estimated candidate distribution violates the assumption that losses above and below the minimum reporting threshold are generated by the same distribution. In the academic literature, the truncation probability estimate is seldom mentioned, but we feel it is an easily interpretable signal of the appropriateness of one distribution family over another when using the truncation approach. As discussed in Section 2.1.4, large truncation probability estimates (greater than 50%) or extremely small truncation probability estimates (less than 1%) may provide evidence that a candidate distribution family is inappropriate. For example, it may not be realistic to assume more than half of the operational losses in a given SRC, or less than 1% of the losses occur below the threshold. Truncation probabil-

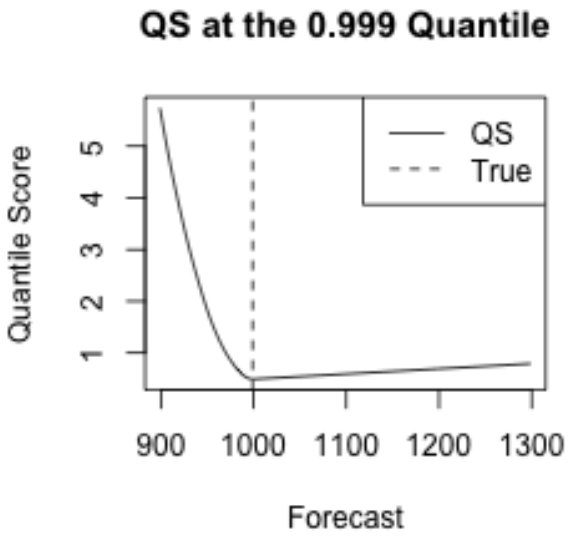

Figure 2.1: Asymmetry of the quantile scoring function at the 0.999 quantile penalizes underestimates more than overestimates: The solid line is the 0.999-quantile score for a sample of 2500 independent Unif$(0, 1000)$ random variables with forecasts of integers 899 to 1299. The true quantile is the dotted vertical line at 999.

ity estimates that are very large provide support for adopting the shifted approach while estimates near zero provide support for using the naive approach. Without more data collection or expert opinion on losses below the threshold, setting conservative bounds on the truncation probability estimate proves quite useful in selecting from estimated candidate distributions.

2.1.4 Challenges with Maximum Likelihood Estimation for Loss Severity Distributions

For particularly flexible distributions such as Burr and log-SaS, their associated negative log-likelihood functions can be numerically unstable. Additionally, the large positive skew that characterizes operational loss severity data often creates a *badly-scaled problem*, where the value of the parameters differ by orders of mag-

nitude. These problems are hard to solve for two reasons. First, different variable magnitudes make it hard to formulate reasonable stopping criteria. Secondly, functions with variables of different magnitudes usually require more iterations to converge. As a result, solving the minimization problem numerically may fail to converge or converge to different local minimums depending upon the parameter starting values.

To increase stability and alleviate the badly-scaled problem, we run our MLE algorithm on the log-losses instead of the raw losses, where appropriate. The log-transform is used for the lognormal, generalized Pareto, log-SaS, lognormal body spliced with lognormal tail, and lognormal body spliced with generalized Pareto tail distributions. Even after using log-losses, the badly-scaled problem still exists for the Burr distribution. This issue can be further reduced through a reparameterization of the log transform of a Burr random variable. See Appendix A.3 for details. To increase our confidence in convergence to a global minimum, a grid of various starting values is used and the parameter values producing the smallest negative log-likelihood are chosen for $\widehat{\Theta}$.

For a sample of loss severities, $\mathbf{x}$, equation (2.9) tells us that $\widehat{\Theta}$ is the value of Θ that minimizes $\widetilde{n\ell}(\Theta;\mathbf{x},\tau)$. Using log-losses $\mathbf{y}=\log(\mathbf{x})$, the same $\widehat{\Theta}$ minimizes $\widetilde{n\ell}\big(\Theta;\mathbf{y},\log(\tau)\big)$. For loss severities and log-loss severities, the minimum values of their negative log-likelihood functions differ by a constant, which is a function of the losses. Thus, the minimum negative log-likelihood of the log-loss data can be easily scaled to enable comparisons between candidate distributions estimated from raw loss severity data and log-loss severity data.

Let X be a random variable for the amount of a loss from a given SRC and let τ be the minimum reporting threshold for that SRC. Then, the conditional CDF and PDF for the loss data are $\widetilde{F}_X(x;\Theta,\tau)$ and $\tilde{f}_X(x;\Theta,\tau)$, respectively. Let $Y=\log(X)$. The conditional CDF and PDF of the log-loss data are:

$$\widetilde{F}_Y\big(y;\Theta,\log(\tau)\big)=\tilde{F}_X\big(e^y;\Theta,\tau\big)$$
$$\tilde{f}_y\big(y;\Theta,\log(\tau)\big)=e^y\,\tilde{f}_X\big(e^y;\Theta,\tau\big).$$

The negative conditional log-likelihood functions for the loss and log-loss data

are respectively

$$\widetilde{n\ell}(\Theta;\mathbf{x},\tau) = -\sum_{i=1}^{n} \log\left(\tilde{f}_X(x_i;\Theta,\tau)\right)$$
$$\widetilde{n\ell}\big(\Theta;\mathbf{y},\log(\tau)\big) = -\sum_{i=1}^{n} \log\left(\tilde{f}_X(x_i;\Theta,\tau)\right) - \sum_{i=1}^{n} y_i.$$

Thus, we must add $\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \log(x_i)$ back to the negative log-likelihood of the log-loss data to make it comparable to the negative log-likelihood of the loss data.

Despite using a grid of starting values and the log-transform of random variables, the MLE algorithm still fails to converge when Gumbel-type distributions are fit to left-truncated data exhibiting regularly varying tail behavior. In these situations, convergence is artificially stopped as one or more of the distribution's parameters approach their boundaries. This situation often manifests itself in extremely large truncation probability estimates (> 0.95). We illustrate this phenomenon, we use asymptotic behavior of order statistics for a specific example.

When a conditional lognormal distribution is estimated by MLE from a sample that exhibits tails heavier than subexponential, the lognormal distribution can mimic the heavier-tailed behavior of an inverse power law by sufficiently increasing the truncation point. This phenomenon is exhibited by distributions in the Gumbel domain of attraction, which includes the lognormal, Weibull (for $0 < a < 1$), and LGNLGN distributions (see Appendix A for parameterizations). We refer to distributions in the Gumbel domain of attraction as *Gumbel-type*.

Using derivations from Section 3.1 of Perline [2005], we first analytically derive an asymptotic approximation to the largest order statistics of Gumbel-type distributions. Secondly, we explicitly show that the conditional lognormal distribution merely mimics an inverse power law and does not obey a true inverse power law. Finally, we graphically show that a sample of independently and identically distributed random variables from a lognormal distribution can mimic an inverse power law in the upper tail at sufficiently high truncation points by comparing the log-log plots of truncated lognormal samples to truncated Pareto samples. All approximations and conclusions are sourced from Perline [2005] unless otherwise specified.

Let $X_{1,1}, X_{1,2}, ..., X_{1,n}$ be n independent observations drawn from a Pareto dis-

tribution with CDF

$$F_X(x;\alpha) = \begin{cases} 1 - \left(\frac{1}{x}\right)^{\alpha} & \text{for } x > 1, \\ 0 & \text{for } x \leq 1. \end{cases}$$

The order statistics of this sample satisfy

$$X_{1,(1)} \geq X_{1,(2)} \geq \cdots \geq X_{1,(n)},$$

and we say a sample satisfies an approximate inverse power law if the order statistics obey

$$X_{1,(j)} \approx \frac{c_n}{j^{\beta}},$$

for $j = 1, 2, ..., n$ and $\beta, c_n > 0$, where c_n depends on n. The plot of the log transform, $\log\left(X_{1,(j)}\right) = Y_{1,(j)}$, against $\log(j)$, should be approximately linear with intercept c_n and slope $-\beta$. This is called a *log-log plot*, and its simplicity motivates our work with the log transform, $Y_{1,(j)}$, which has an exponential distribution with rate parameter α. From pages 69–72 of Beirlant et al. [2004], we know the exponential distribution is Gumbel-type.

If we have a sample of n independent and identically distributed random variables from a Gumbel-type distribution with CDF, $F_Y(y;\Theta)$, then for some fixed j such that $j \ll n$, there exist two sequences of standardizing constants a_n and b_n such that,

$$\lim_{n\to\infty} P\left\{ \frac{Y_{(i)} - a_n}{b_n} \right\} \;=\; \exp(-e^{-y}) \sum_{k=0}^{i-1} \frac{1}{\Gamma(k+1)} e^{-ky},$$

for $1 \leq i \leq j$. The limiting first moment convergence [Polfeldt, 1970] is

$$\lim_{n\to\infty} \mathbb{E}\left[\frac{Y_{(i)} - a_n}{b_n} \right] = \begin{cases} \gamma_{EC} - \sum_{k=1}^{i-1} \frac{1}{k} & \text{for } i > 1, \\ \gamma_{EC} & \text{for } i = 1, \end{cases}$$

for $1 \leq i \leq j$ and where γ_{EC} is Euler's constant, $\lim_{n\to\infty}\left(\sum_{k=1}^{n}\frac{1}{k}-\log n\right)$. Therefore, we can approximate the expected value of the largest order statistics for Gumbel-type distributions where n is large and $i = 1,2,\ldots,j$ such that $j \ll n$, as

$$\mathbb{E}(Y_{(i)}) \approx \begin{cases} a_n + b_n\gamma_{EC} - b_n\sum_{k=1}^{i-1}\frac{1}{k} & \text{for } i > 1, \\ a_n + b_n\gamma_{EC} & \text{for } i = 1. \end{cases}$$

To find a_n and b_n, we use Proposition 1.19 from Resnick [1987] which gives us the equations

$$F(a_n) = 1 - \frac{1}{n}; \qquad\qquad b_n = \frac{1-F(a_n)}{f(a_n)}.$$

For exponential distribution F with rate α, we find the standardizing sequence a_n as

$$\begin{aligned} 1 - e^{-\alpha a_n} &= 1 - \frac{1}{n} \\ a_n &= \frac{1}{\alpha}\log(n). \end{aligned}$$

Standardizing sequence b_n for the exponential distribution is

$$b_n = \frac{1-F\left(\frac{1}{\alpha}\log(n)\right)}{f\left(\frac{1}{\alpha}\log(n)\right)} = \frac{\exp\left\{\frac{\log n}{\alpha}\right\}}{\alpha\exp\left\{\frac{\log n}{\alpha}\right\}} = \frac{1}{\alpha}.$$

Thus for the exponential distribution, we can approximate the first j order statistics with the formula

$$\mathbb{E}(Y_{(i)}) \approx \left(\frac{1}{\alpha}\log(n) + \frac{1}{\alpha}\gamma_{EC}\right) - \frac{1}{\alpha}\sum_{k=1}^{i-1}\frac{1}{k}.$$

Since $\frac{1}{\alpha}\left(\gamma_{EC} - \sum_{k=1}^{i-1}\frac{1}{k}\right) \approx -\log(i)$, the log-log plot of the largest order statistics of $X_{1,(i)}$ is approximately linear with slope $-\frac{1}{\alpha}$.

Now let $X_{2,1}, X_{2,2}, \ldots, X_{2,n}$ be n independent observations drawn from a lognormal distribution with parameterization given in Appendix A.1. Again, we let the

order statistics of this sample satisfy

$$X_{2,(1)} \geq X_{2,(2)} \geq \cdots \geq X_{2,(n)}.$$

The log transform of the j^{th} order statistic, $Y_{2,(j)}$, is normally distributed with location parameter μ and scale parameter σ. From Beirlant et al. [2004], we know the normal distribution is Gumbel-type, so all of the above results hold, except we need to derive the standardizing sequences a_n and b_n for the normal distribution. From Embrechts et al. [1997], we get

$$a_n = \mu + \sigma\sqrt{2\log n} - \sigma\frac{\log\log n + \log 4\pi}{2\sqrt{2\log n}},$$
$$b_n = \frac{\sigma}{\sqrt{2\log n}}.$$

The most important feature is the approximate slope of the log-log plots between the Pareto and lognormal samples. While the Pareto sample has a constant slope in the log-log plot of $-\frac{1}{\alpha}$, the lognormal slope in the log-log plot, $\frac{\sigma}{\sqrt{2\log n}}$ depends on sample size n and is of order $O\big(1/\sqrt{\log n}\big)$. Thus, the largest order statistics of the lognormal only mimic a power law.

To show this graphically, we generate 4 different samples from a Pareto distribution with parameter $\alpha = 1$ of sizes $n_1 = 100, n_2 = 400, n_3 = 10{,}000, n_4 = 100{,}000$. The samples are then truncated at the 10%, 75%, 99%, and 99.9% quantiles. Similarly, we perform the same process for 4 different samples from a lognormal distribution with parameters $\mu = -8$ and $\sigma = 4.5$. The results are presented in Figure 2.2.

The final issue with the MLE approach is specific to the candidate g-and-h distribution family since it is the only candidate distribution family with support on the real numbers. As a result, simulating from the estimated unconditional g-and-h distribution can result in negative loss severities. To prevent negative losses, we instead simulate from the g-and-h distribution truncated at zero using the unconditional MLE parameters. If selecting a g-and-h candidate distribution for a SRC, we must be sure that the probability of a negative observation is very low. Otherwise, the g-and-h distribution truncated at zero will overestimate the probability of an extreme observation.

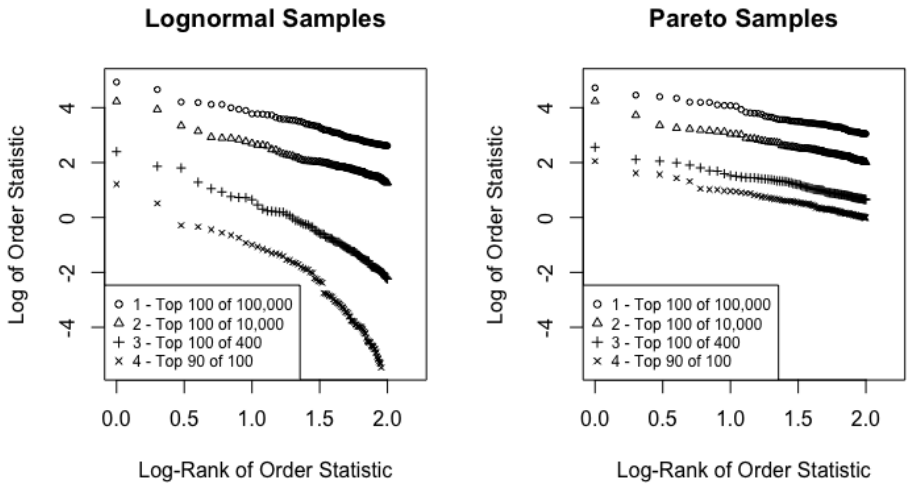

Figure 2.2: Power law mimicking behavior of a lognormal distribution with sufficiently high truncation as seen from a log-log plot: The left-hand plot shows 4 truncated samples of 100, 400, 10,000, and 100,000 independent lognormal random variables with parameters $\mu = -8$ and $\sigma = 4.5$, truncated at the $10\%, 75\%, 99\%$, and 99.9% quantiles, respectively. The right-hand plot shows 4 truncated trunated samples 100, 400, 10,000, and 100,000 independent Pareto random variables with parameter $\alpha = 1$, truncated at the 10%, 75%, 99%, and 99.9% quantiles, respectively. The lognormal samples resemble the constant slope of the Pareto samples when truncation is high enough.

While we did not encounter large probabilities of negative values in our simulations, it is not an unusual situation. In fact, such a scenario is easily created by simulating from a Burr$(\alpha = 0.065, \gamma = 15, \theta = 1.226)$ distribution, truncating the sample at the 2.5% quantile, and fitting a *g*-and-*h* distribution to the truncated sample using MLE. The estimated parameter vector, $\widehat{\Theta}_{mle}$, creates a *g*-and-*h* distribution that is nearly symmetric with extremely fat tails. The problem is that the probability of a negative observation, $F(0;\widehat{\Theta}_{mle})$ is approximately 0.27. Thus, if we use the parameter vector $\widehat{\Theta}_{mle}$ to simulate from a *g*-and-*h* distribution truncated at zero, the density on the positive real numbers is shifted upwards by $\frac{1}{1-F(0;\widehat{\Theta}_{mle})}$,

which over estimates the probability of experiencing an extreme loss.

Since we know that operational loss data are almost always positively skewed, it is reasonable to want a skewness parameter to reflect this. If we simply restrict the skewness parameter space to avoid low positive skew, the MLE algorithm artificially stops when it hits this boundary and thus does not converge. The resulting estimated distribution may not accurately represent the sample due to this premature stopping condition. A much more effective approach is *Penalized Maximum Likelihood Estimation* (PMLE). For example, the PMLE approach that adds 1 to the negative log-likelihood function for each percent of the distribution that falls below zero is a minimization of the form

$$\widehat{\Theta}_{pmle} = \underset{\Theta}{\operatorname{argmin}} \Big(\widetilde{n\ell}(\Theta; \mathbf{x}, \boldsymbol{\tau}) + 100 \cdot F(0; \Theta) \Big),$$

which is modified from equation (2.9). Minimizing this function results in a distribution that is almost identical to the MLE distribution for the right tail, but has a higher probability of experiencing a loss around the mode. This is a desirable result, since simulating from the zero-truncated *g*-and-*h* distribution using the parameter vector $\widehat{\Theta}_{pmle}$ has almost the same probability of an extreme loss as the unconditional *g*-and-*h* distribution with parameters $\widehat{\Theta}_{mle}$. In other words, using PMLE in this situation has a similar effect as simulating from the unconditional *g*-and-*h* distribution with parameter vector $\widehat{\Theta}_{mle}$, but shifting most of the negative probability to the mode and changing the tail very little. Results are shown in Table 2.2 and Figure 2.3.

This is an important point, since the *g*-and-*h* distribution estimated using MLE is essentially useless for our operational risk modeling procedure as we cannot use it for simulations. Using PMLE leads to a more practical distribution while maintaining the salient properties of the data. While we did not conduct specific research into the best penalty term or criteria to penalize, it is worth noting that methods similar to those presented here can be adopted to easily solve a rather frustrating problem.

	MLE	PMLE
Prob of Negative Loss	0.27	0.01
a	1.14	2.11
b	1.51	1.55
g	0.01	1.65
h	1.05	0.32
Negative log-likelihood	3290	3292

Table 2.2: Proportion of distribution below zero, estimated parameters, and the minimized negative log-likelihood when fitting the *g*-and-*h* distribution to a left-truncated sample using MLE and PMLE.

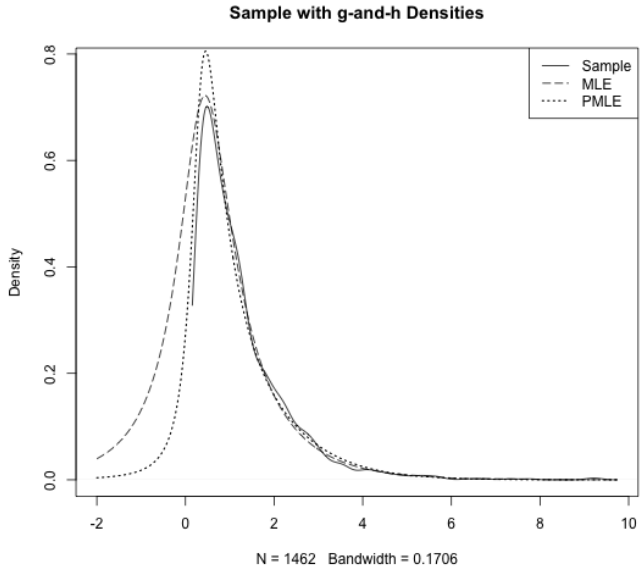

Figure 2.3: Plot of the truncated sample density and the estimated densities under the MLE and PMLE approaches on the log scale. The PMLE parameters produce a right-tail that is almost identical to the MLE distribution, but places a higher probability of a loss occurring around the mode instead of the probability of negative losses estimated by the MLE approach.

2.2 Loss Frequency Distributions

The estimated loss frequency distribution is used to simulate the number of loss events that may occur next year for each business line/event type intersection corresponding to a given SRC. When modeling loss frequency for the LDA, only internal losses are used. Since most databases only include operational loss events whose loss amount exceeds a minimum threshold, loss frequencies based on historical data are biased downwards. Loss events with a loss severity below this threshold occur, but are not reported in the dataset. Failure to acknowledge these losses would underestimate loss frequency but overestimate loss severity, leading to an uncertain impact on RC [Luo et al., 2007].

Under our AMA procedure, we estimate these smaller losses by treating the historical data as a truncated sample, estimating the non-truncated severity distribution, and estimating the truncation probability by evaluating the estimated severity distribution function at the minimum reporting threshold. Thus, we are able to estimate the proportion of operational loss events that go unreported due to the minimum reporting threshold. This proportion is then used to increase the loss event frequency proportionally.

According to BCBS, the two most popular distributions for loss frequency are Poisson followed by negative binomial [BCBS, 2011]. Trends in the annual number of loss events can be modeled via covariates as demonstrated by Chavez-Demoulin et al. [2015] and used for forecasting. Strictly monotone trends can also be detected by the simple Mann-Kendall trend test [Gilbert, 1987, p. 208-217] and exponential smoothing can be used to forecast loss events. If assuming the simplest distribution, Poisson, for the loss frequency, the estimated Poisson rate parameter, $\hat{\lambda}$, is the mean number of observable annual losses. Since this mean number of loss events only includes losses that exceed the minimum reporting threshold, we can derive the estimate of the Poisson rate parameter for loss events both above and below the threshold as

$$\hat{\lambda}^* = \frac{\hat{\lambda}}{1 - F(\tau; \widehat{\Theta})}. \tag{2.14}$$

2.3 Total Annual Loss and Regulatory Capital Estimation

BASEL II defines RC as the 99.9%-quantile of the total annual loss distribution. The total annual loss, S_{T+1}, given by equation (2.4) is restated below,

$$S_{T+1} = \sum_{r=1}^{R} \sum_{n=1}^{N_{T+1}^r} X_{T+1,n}^r,$$

where N_{T+1}^r is the number of loss events in SRC r, and $X_{T+1,n}^r$ is the n^{th} loss severity in SRC r. Equation (2.4) requires forecasts of the number of loss events for SRC r in year $T+1$, N_{T+1}^r, and each loss event's severity, $X_{T+1,n}^r$, for $n = 1, 2, \ldots, N_{T+1}^r$.

Using the relevant historical loss severity data pooled across years for SRC r, loss severity distributions are estimated for each candidate distribution family. One of the nine estimated candidate distributions is chosen as the loss severity distribution for SRC r. Let $F^r(x; \widehat{\Theta})$ be the estimated loss severity distribution for SRC r. If we assume the loss frequency distribution for observable losses in SRC r is Poisson with estimated rate parameter, $\hat{\lambda}_r$, then the loss frequency distribution for all loss events in SRC r is Poisson with estimated rate parameter

$$\hat{\lambda}_r^* = \frac{\hat{\lambda}_r}{1 - F^r(x; \widehat{\Theta})}.$$

To find the loss distribution of S_{T+1}, we use simulation. We can write the total annual operational loss for SRC r as the sum

$$S_{T+1,r} = \sum_{n=1}^{N_{T+1}^r} X_{T+1,n}^r,$$

and we rewrite the total annual loss as

$$S_{T+1} = \sum_{r=1}^{R} S_{T+1,r}.$$

We create N^* simulations of $S_{T+1,r}$ by first simulating a sequence of independent Poisson$\left(\hat{\lambda}_r^*\right)$ random variables $P_1^r, P_2^r, ..., P_{N^*}^r$. For each P_i^r, we generate a

sequence of independent and identically distributed loss severities,

$$X^r_{T+1,1}, X^r_{T+1,2}, ..., X^r_{T+1,P^r_i} \overset{i.i.d.}{\sim} F^r(x; \widehat{\Theta}).$$

Summing these loss severities creates N^* simulations of annual operational losses for SRC r, and we denote the i^{th} simulation of SRC r's annual loss as $S^{(i)}_{T+1,r}$. If we rank each $S^{(i)}_{T+1,r}$ from lowest to highest as $1, 2, ..., N^*$, we can derive the empirical marginal distribution of $S_{T+1,r}$ by dividing the rank of each annual loss by $N^* + 1$. Repeating this process for each SRC, we can derive the R marginal distributions for the annual operational loss from each SRC. We denote the empirical marginal distribution of $S_{T+1,r}$ as F_{S_r}.

To sum across SRC's and arrive at simulations of the total annual operational loss, S_{T+1}, we must account for any dependence of losses across SRC's. This is done by t-copulas as presented by McNeil et al. [2015]. We simulate M^* random vectors of dimensionality R from a multivariate t-distribution. By Sklar's Theorem, we can transform the t-distributed random vectors into uniform random vectors by applying the inverse of the t-distribution function. We derive the inverse of each SRC's marginal distribution, denoted $F^{-1}_{S_r}$, numerically using the `R` function `pchip()` from the `signal` package and apply this inverse to the uniform random vectors. The `pchip()` function performs piecewise cubic Hermite (monotone) interpolation. Finally, by summing the values of each vector we arrive at M^* simulations of S_{T+1}. The 99.9%-quantile estimate of these M^* simulations is our estimate of RC.

Chapter 3

Simulation Studies

In this section, operational loss data both above and below a reporting threshold are simulated for three unique SRC's for the fourteen years 2004 – 2017. Each of the SRC samples are truncated at their true distribution's respective 2.5% quantile, which are treated as known. The frequency and severity distributions are given in Table 3.1.

The loss severity distributions for SRC 1 and SRC 2 are chosen from the best candidate distributions as measured by AIC for actual operational loss data given to the authors. For anonymity, the data are scaled before estimation was performed, so the parameters and selected distributions are for the scaled data, but maintain the salient properties of tail behavior, skewness, and overall shape. Finally, loss severities for SRC 3 are generated from a mixture model to examine the performance of our estimation approach under model misspecification.

For each SRC, we look at the density plots, goodness-of-fit and prediction statistics, the estimated truncation probability, and QQ-plots for each candidate loss severity distribution. Severity distribution parameters are estimated using MLE following the truncation approach. Section 3.2 presents a heuristic procedure using this information to select a loss severity distribution for each SRC. Assuming a Poisson frequency distribution, we then calculate RC for years 2014 - 2018 in Section 3.3.2 and compare the impact on RC when choosing loss severity distributions by AIC versus quantile score. Finally, we briefly revisit the issue of estimated truncation probabilities.

SRC	Truncation Point	Frequency Distribution	Frequency Parameters	Severity Distribution	Severity Parameters
1	$\tau = 1.167$	Poisson	$\lambda = 100$	Burr	$\alpha = 0.065$
					$\gamma = 15$
					$\theta = 1.226$
2	$\tau = 3.147$	Poisson	$\lambda = 100$	log-SaS	$a = 1.06$
					$b = 0.37$
					$\varepsilon = 1.65$
					$\delta = 0.97$
3	$\tau = 1.133$	Poisson	$\lambda = 100$	$X \sim \beta X_1 + (1-\beta)X_2$	$\beta = 0.3$
				$X_1 \sim$ LGN	$\mu = 0.4$
					$\sigma = 0.16$
				$X_2 \sim$ Burr	$\alpha = 0.065$
					$\gamma = 15$
					$\theta = 1.226$

Table 3.1: Table of frequency and severity distributions used to simulate operational losses for three SRC's for fourteen years spanning 2004 - 2017. The frequency distribution is the same for each SRC, $Pois(\lambda = 100)$. The loss severity distributions for SRC 1, SRC 2, and SRC 3 are Burr, log-SaS, and a mixture model where component 1 is simulated from lognormal and component 2 is simulated from Burr.

3.1 Exploratory Data Analysis

To create truncation in our data, we truncate each SRC at their known 2.5% quantile and only use the truncated sample for the remainder of this section. Summary statistics for the truncated SRC's are given below in Table 3.2. As is common in operational loss data, the data exhibit extreme right-skewness as evidenced by the mean $>$ median and the maximum value $\gg$ the 75% quantile.

Time series plots of the annual number of observable loss events are presented in Figure 3.1. Since all loss frequencies were simulated from a Pois($\lambda = 100$) distribution and then truncated at the each loss severity's 2.5% quantile, we know that the true Poisson rate for observable losses is 97.5. We assume the loss frequency distribution is Poisson with no trend, so any perceived trend in the loss events is disregarded. Under each plot is the mean and variance of the number of observable

SRC	Sample Size	Minimum	Qtile 0.25	Median	Mean	Qtile 0.75	Maximum
1	1368	1.169	1.69	2.497	10.23	4.97	2491
2	1342	3.155	4.705	7.965	34.8	18.86	2168
3	1365	1.135	1.49	1.898	7.406	3.788	1432

Table 3.2: Summary Statistics for each simulated SRC named SRC 1, SRC 2, and SRC 3, respectively. From left to right, the columns show the name of the SRC, sample size, minimum observable loss, 25^{th} percentile, median, mean, 75^{th} percentile, and maximum loss.

losses from 2004-2017.

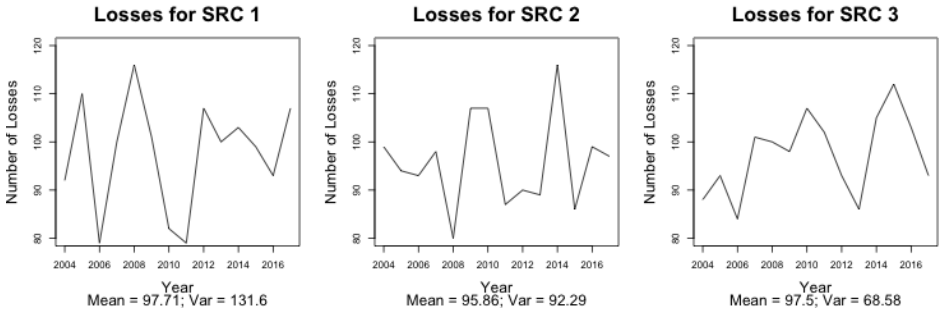

Figure 3.1: Number of observable annual losses from 2004 - 2017 for each SRC with the mean and variance under each plot

Figure 3.2 presents the sample densities for log-losses in the top row and the histogram of smallest 75% of the raw losses. Since the data exhibit extreme skewness, sample densities of the log-losses better highlight the differences between the SRC's. The histograms allow us to see whether there is a clear mode in the raw data or if the mode may fall below the truncation point for an underlying unimodal distribution.

For each SRC, we create 30 evenly spaced buckets ranging from the minimum loss in each SRC to its 75% empirical quantile. Some subjectivity is needed to

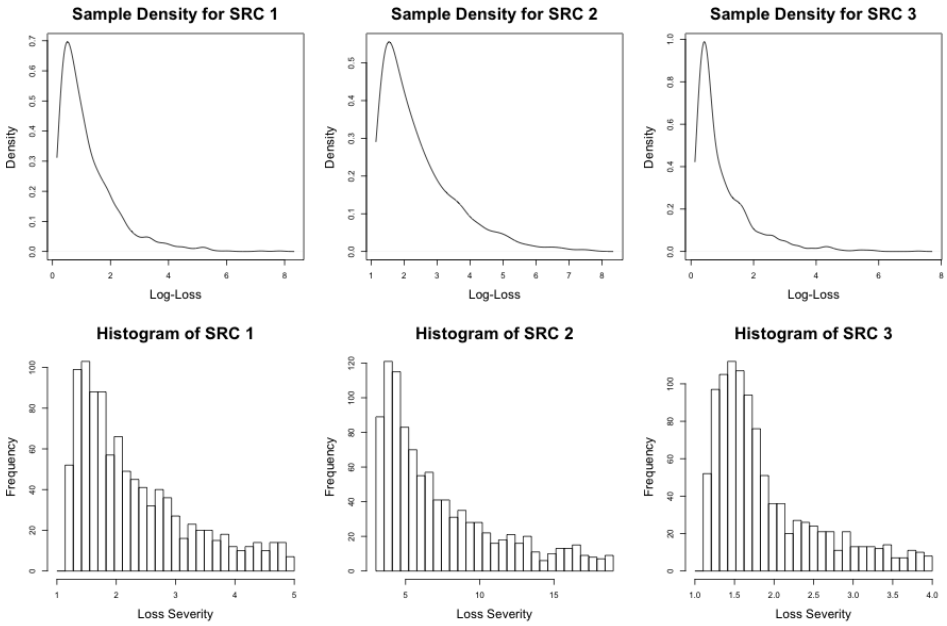

Figure 3.2: Top Row: Sample densities of the log-losses for each SRC; Second Row: Histogram for the smallest 75% of raw losses

interpret the histograms, as modes may arise out of the number of buckets used and not necessarily from the data. Viewing these same histograms with 25 and 20 buckets may appear to yield different stories. One thing that this visualization can tell us is that any turn in the densities for SRC's 1 and 2 must be sharp, since the mode appears in the first few buckets and losses cannot be negative. If we assume the underlying severity distribution is unimodal, we expect to have a very low truncation point between zero and the first bucket. See Section 3.4 for further analysis of the issues with truncation probability estimation and how one may use these histograms in practice.

3.2 Loss Severity Distribution Estimation and Selection

In this section, we use all of the historical truncated loss data to estimate and select distributions for each SRC. This process is somewhat subjective, since one must use their own judgement to gauge whether a distribution seems reasonable.

3.2.1 SRC 1

To generate losses for SRC 1, we first simulate 14 independent and identically distributed Pois($\lambda = 100$) random variables, one for each of the 14 years encompassing 2004 - 2017, inclusive. The loss frequencies for SRC 1 can be represented as a time series $\{N_t^1\}$, where $N_t^1 \overset{iid}{\sim} \text{Pois}(\lambda = 100)$ for $t = 2004, 2005, \ldots, 2017$. For each simulated loss frequency n_t, we generate n_t independent and identically distributed loss severities from a Burr($\alpha = 0.065, \gamma = 15, \theta = 1.226$) distribution and assign them to year t. Finally, all loss severities are then truncated at $\tau = 1.167$, the 2.5% quantile from the true loss severity distribution.

We perform MLE on the truncated sample as outlined in Section 2.1. Figure 3.3 plots the sample density log-losses against the densities for the true loss severity distribution and each estimated candidate distribution on the log scale. The densities shown are conditional densities to emphasize fit to the truncated sample. The plot of the true underlying distribution against the sample gives us a good indication of how representative the sample is of its generating process. Density plots are a good sanity check to make sure the MLE algorithm is producing reasonable results.

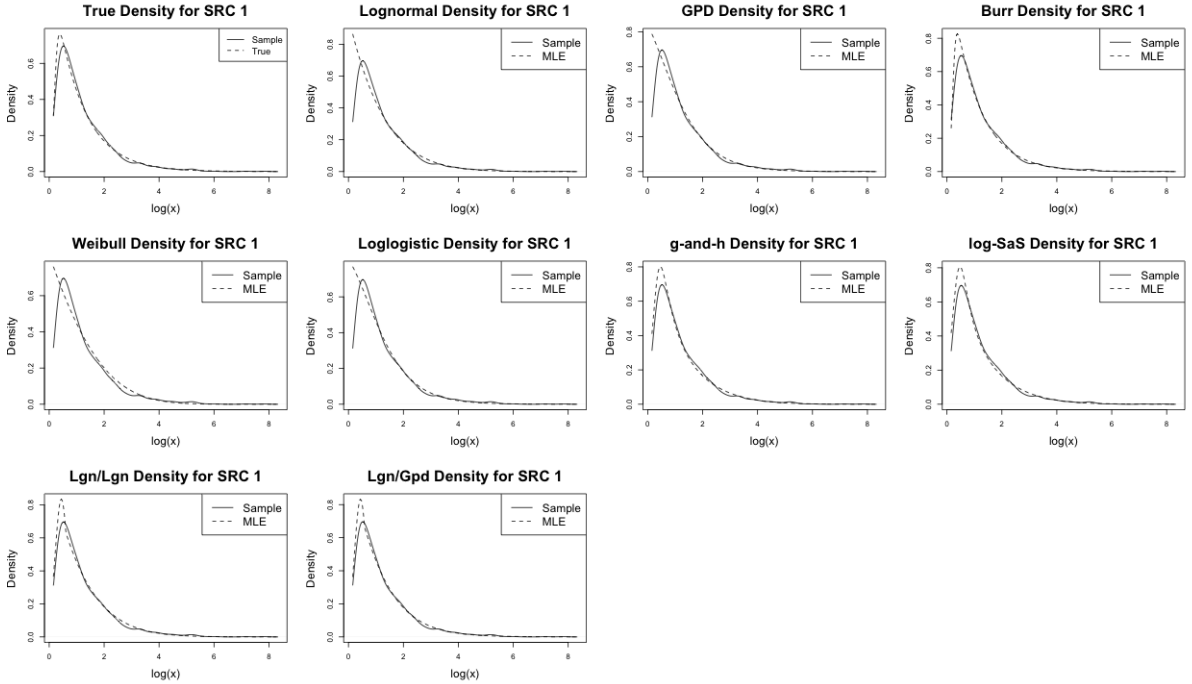

Figure 3.3: The sample density plotted against the true severity model and each estimated candidate distribution for SRC 1 on the log scale. The number of observations in SRC 1 is 1368, and the smoothing bandwidth is 0.1709.

The density plots do not rule out any candidate distributions since all estimated densities seem to fit the data, but the lognormal, generalized Pareto, Weibull and loglogistic distributions are unable to capture the mode of the sample density. These are all 2-parameter distribution families and are typically not able to accommodate lower and upper tail behaviors. This is indicative of a high truncation point. Also, the density plots do not tell us much about the fit in the extreme right tail, a portion of the distribution that is of much interest for OR modeling, and cannot be used to compare fit or predictive ability across models. Table 3.3 gives the truncation probability estimate, BIC, AIC, modified Anderson-Darling test at the 95% confidence level, quantile score, out-of-sample AIC, and estimated 99.9% quantile for each estimated candidate distribution. Values from the underlying true model are given in the first row, and all subsequent rows are sorted from best to worst AIC. The number in parentheses in the other columns shows the rank from best to worst within a given column. The rank for the 99.9% quantile estimate is by distance to the true model's 99.9% quantile. Tables 3.4 and 3.5 are analogous to Table 3.3 for SRC's 2 and 3, respectively.

The Burr distribution has the best fit based on AIC and BIC, passes the modified Anderson-Daring test, a reasonable truncation probability, and the best 99.9% quantile estimate. We consider "reasonable" truncation probabilities to be $0.01 \leq F(\tau;\widehat{\Theta}) \leq 0.5$ and excludes the lognormal, generalized Pareto, Weibull, and loglogistic distributions. The Burr distribution also dominates the predictive measures of the QS and out-of-sample AIC. A logical choice for the loss severity distribution of SRC 1 is Burr.

Finally, QQ-plots give a good visualization of the fit in the right-tail and are presented in Figure 3.4. Like the density plots, we show the QQ-plots for the true model as well as the estimated candidate distributions. The sample quantiles are given on the x-axis while the estimated quantiles are given on the y-axis. Thus, points below the diagonal line indicate quantiles that are underestimated by the candidate distribution. The QQ-plots inform us that all candidate distributions significantly underestimate the two largest observations.

	Trunc Pr	BIC	AIC	A-D	Quantile Score	OOS AIC	0.999 Qtile
True	0.025	NA	NA	NA	NA	NA	1464
Burr	0.017	6170 (1)	6154	Accept	2.31 (1)	6159 (1)	1109 (1)
Lgn/Gpd	0.043	6182 (2)	6161	Accept	2.48 (2)	6176 (4)	950 (2)
g-and-h	0.093	6183 (3)	6162	Accept	2.63 (8)	6171 (2)	530 (7)
log-SaS	0.073	6184 (4)	6164	Accept	2.62 (7)	6172 (3)	634 (5)
Lgn/Lgn	0.044	6195 (5)	6169	Accept	2.57 (5)	6181 (5)	463 (8)
Loglogistic	0.665	6198 (6)	6188	Accept	2.53 (4)	6192 (6)	654 (4)
GPD	0.735	6200 (7)	6189	Accept	2.52 (3)	6193 (7)	685 (3)
Lognormal	0.995	6209 (8)	6199	Accept	2.59 (6)	6202 (8)	552 (6)
Weibull	0.980	6233 (9)	6223	Reject	2.88 (9)	6228 (9)	220 (9)

Table 3.3: SRC 1 selection critera from left to right: truncation probability estimate, BIC, AIC, modified Anderson-Darling test at the 95% confidence level, QS, out-of-sample AIC, and estimated 99.9% quantile. Values from the true model are given in the first row, and subsequent rows are sorted by AIC from best to worst. Ranks from best (1) to worst (9) are presented for other criteria. The rank for the 99.9% quantile estimate is by distance to the true quantile. Lgn/Lgn and Lgn/Gpd refer to the spliced distributions.

3.2.2 SRC 2

To generate losses for SRC 2, we again simulate 14 independent and identically distributed Pois($\lambda = 100$) random variables. For each simulated loss frequency n_t, we generate n_t independent and identically distributed loss severities from a log-SaS($a = 1.06, b = 0.37, \varepsilon = 1.65, \delta = 0.97$) distribution and assign them to year t. Finally, all loss severities are then truncated at $\tau = 3.147$, the 2.5% quantile from the true loss severity distribution.

Figure 3.5 plots the sample density log-losses against the densities for the true loss severity distribution and each estimated candidate distribution on the log scale. All distributions fit the data, but we note that the lognormal, generalized Pareto, Weibull, and loglogistic distributions are not able to capture the mode seen in the sample density. Therefore, their estimated truncation probabilities are going to be high.

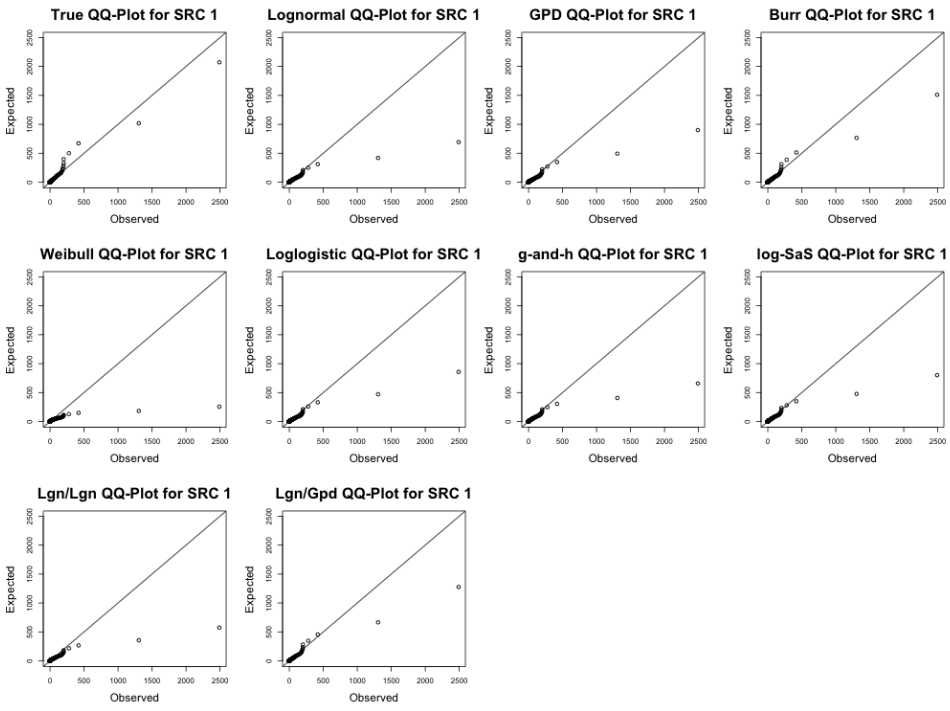

Figure 3.4: QQ-Plots for the true model and each estimated candidate distribution for SRC 1. Estimated quantiles are given on the vertical axis with empirical quantiles along the horizontal. Points below the 45° line indicate underestimates the empirical quantile.

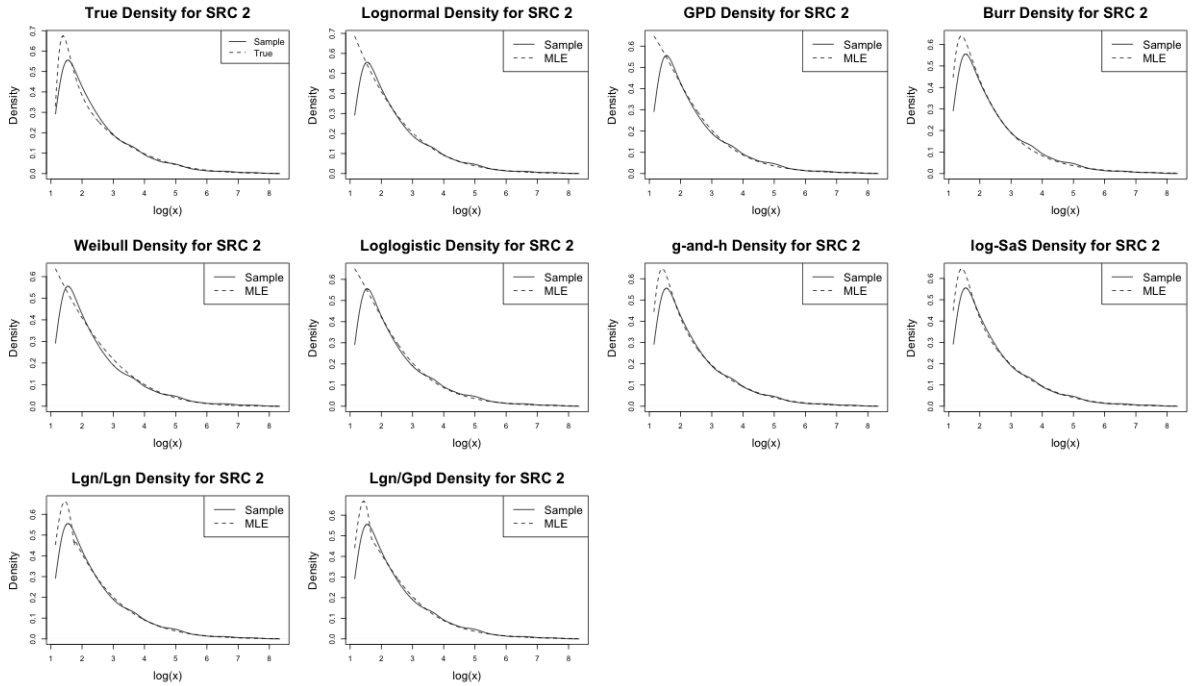

Figure 3.5: The sample density plotted against the true severity model and each estimated candidate distribution for SRC 2 on the log scale. The number of observations in SRC 2 is 1342, and the smoothing bandwidth is 0.2209.

From Table 3.4, Both the *g*-and-*h* and log-SaS distributions are appropriate for modeling SRC 2. We give a slight edge to log-SaS due to its superior quantile score. An important observation is that the Weibull distribution has the best QS even though it has the worst fit based on AIC. The QQ-plots in Figure 3.6 can help explain this. Remember that the quantile score at the 99.9% quantile is very asymmetric and penalizes more for underestimating the quantile than overestimating. From the QQ-plots below, we see that the Weibull distribution can get relatively close to the extreme quantiles, but the distribution is likely to underestimate the extreme right tail. However, all other distributions overestimate the 99.9% quantile by so much that it overcomes the QS's asymmetry. This is an important feature of the quantile score from the bank's perspective, accurately estimating the right-tail without the extreme overestimation seen in literature. Finally, the Weibull distribution would not be selected for its QS due to its extremely high truncation probability. The MLE parameters presented in Appendix B.2 show the Weibull distribution's scale parameter is hitting the boundary condition $\theta > 0$.

	Trunc Pr	BIC	AIC	A-D	Quantile Score	OOS AIC	0.999 Qtile
True	0.025	NA	NA	NA	NA	NA	2586
g-and-h	0.108	9837 (4)	9816	Accept	2.95 (3)	9825 (2)	2989 (2)
log-SaS	0.079	9837 (5)	9816	Accept	2.86 (2)	9825 (1)	2891 (1)
Lgn/Lgn	0.099	9845 (9)	9819	Accept	3.8 (5)	9831 (7)	3836 (5)
Lgn/Gpd	0.087	9842 (8)	9821	Accept	7.43 (6)	9832 (8)	7777 (8)
Lognormal	0.968	9834 (1)	9824	Accept	3.42 (4)	9828 (4)	3454 (3)
Burr	0.075	9840 (6)	9824	Accept	14.02 (9)	9828 (3)	14050 (9)
GPD	0.671	9836 (3)	9826	Accept	7.72 (8)	9830 (5)	7757 (7)
Loglogistic	0.697	9836 (2)	9826	Accept	7.5 (7)	9830 (6)	7535 (6)
Weibull	0.978	9842 (7)	9831	Accept	2.45 (1)	9836 (9)	1544 (4)

Table 3.4: SRC 2 selection critera from left to right: truncation probability estimate, BIC, AIC, modified Anderson-Darling test at the 95% confidence level, QS, out-of-sample AIC, and estimated 99.9% quantile. Values from the true model are given in the first row, and subsequent rows are sorted by AIC from best to worst. Ranks from best (1) to worst (9) are presented for other criteria. The rank for the 99.9% quantile estimate is by distance to the true quantile. Lgn/Lgn and Lgn/Gpd refer to the spliced distributions.

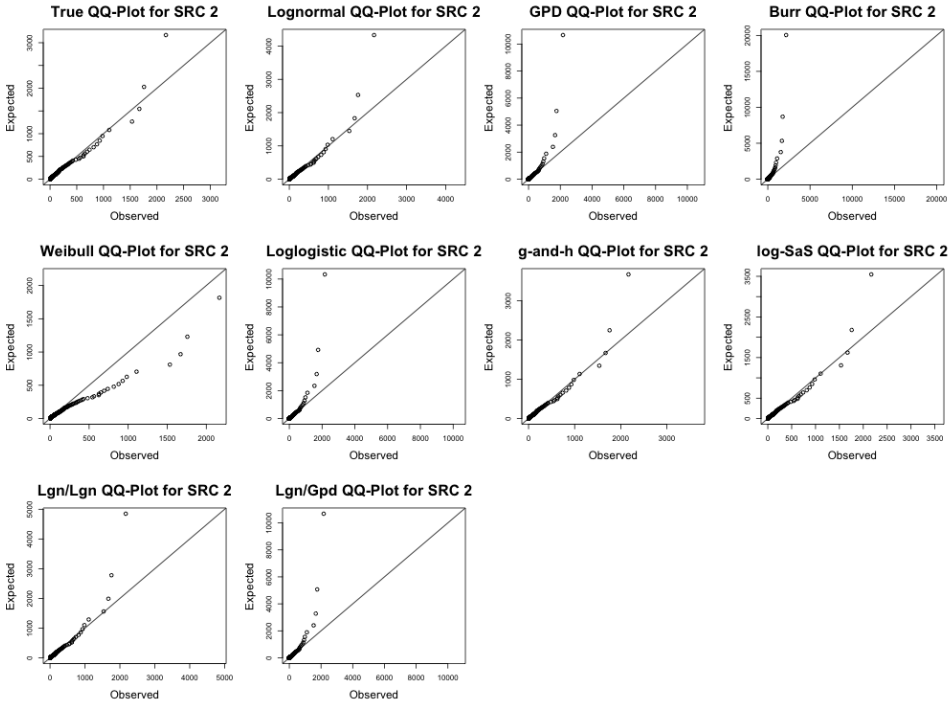

Figure 3.6: QQ-Plots for the true model and each estimated candidate distribution for SRC 2. Estimated quantiles are given on the vertical axis with empirical quantiles along the horizontal. Points below the 45° line indicate underestimates the empirical quantile.

3.2.3 SRC 3

To generate losses for SRC 3, we again simulate 14 independent and identically distributed Pois($\lambda = 100$) random variables. For each simulated loss frequency n_t, we generate n_t independent and identically distributed loss severities from a mixture distribution, $F(x;\Theta) = \beta F_1(x;\Theta_1) + (1-\beta)F_2(x;\Theta_2)$; where $\beta = 0.3$, $F_1(x;\Theta_1)$ is LN($\mu = 0.16, \sigma = 0.4$), and $F_2(x;\Theta_2)$ is Burr($\alpha = 0.065, \gamma = 15, \theta = 1.226$). The loss severities are assigned to year t. Finally, all loss severities are then truncated at $\tau = 1.133$, the 2.5% quantile from the true loss severity distribution.

Figure 3.7 plots the sample density log-losses against the densities for the true loss severity distribution and each estimated candidate distribution on the log scale.

All distributions fit the data, but we note that the lognormal, generalized Pareto, Weibull, and loglogistic distributions are not able to capture the mode seen in the sample density.

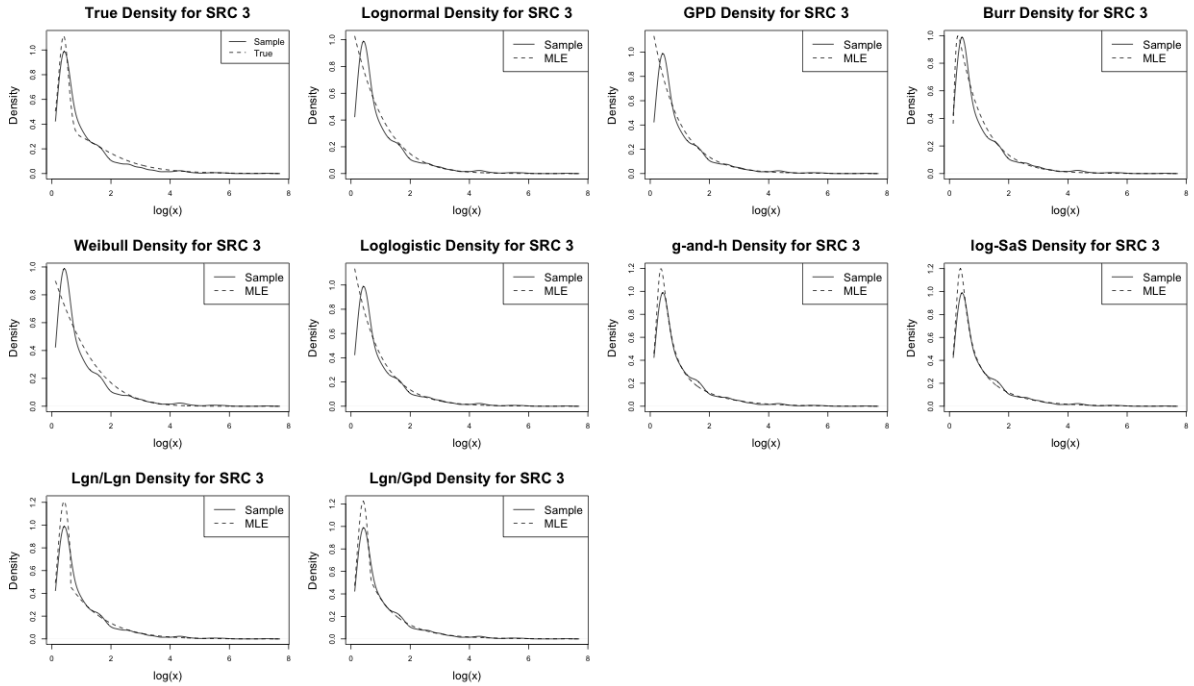

Figure 3.7: The sample density plotted against the true severity model and each estimated candidate distribution for SRC 3 on the log scale. The number of observations in SRC 3 is 1365, and the smoothing bandwidth is 0.1479.

Table 3.5 gives the truncation probability estimate, BIC, AIC, modified Anderson-Darling test at the 95% confidence level, QS, out-of-sample AIC, and estimated 99.9% quantile for each estimated candidate distribution. Values from the underlying true model are given in the first row, and all subsequent rows are sorted from best to worst AIC.

	Trunc Pr	BIC	AIC	A-D	Quantile Score	OOS AIC	0.999 Qtile
True	0.025	NA	NA	NA	NA	NA	1015
Lgn/Gpd	0.050	5070 (1)	5050	Accept	1.33 (6)	5067 (1)	961 (1)
Lgn/Lgn	0.056	5082 (2)	5056	Accept	1.18 (2)	5070 (2)	389 (6)
log-SaS	0.052	5082 (3)	5061	Accept	1.39 (7)	5072 (3)	926 (2)
g-and-h	0.082	5083 (4)	5062	Accept	1.31 (5)	5073 (4)	743 (3)
Burr	0.018	5122 (5)	5107	Reject	1.15 (1)	5112 (5)	377 (7)
GPD	1.000	5153 (7)	5142	Reject	1.18 (4)	5142 (6)	501 (4)
Loglogistic	1.000	5153 (6)	5142	Reject	1.18 (3)	5143 (7)	498 (5)
Lognormal	0.995	5179 (8)	5168	Reject	1.49 (8)	5183 (8)	201 (8)
Weibull	0.987	5244 (9)	5233	Reject	2.05 (9)	5255 (9)	111 (9)

Table 3.5: SRC 3 selection critera from left to right: truncation probability estimate, BIC, AIC, modified Anderson-Darling test at the 95% confidence level, QS, out-of-sample AIC, and estimated 99.9% quantile. Values from the true model are given in the first row, and subsequent rows are sorted by AIC from best to worst. Ranks from best (1) to worst (9) are presented for other criteria. The rank for the 99.9% quantile estimate is by distance to the true quantile. Lgn/Lgn and Lgn/Gpd refer to the spliced distributions.

The results of LGNLGN (denoted Lgn/Lgn in Table 3.5) distribution in Table 3.5 should be scrutinized. Due to the high truncation probability estimate for the single lognormal distribution, it is likely the case that the lognormal body spliced with lognormal tail distribution is unable to converge for the tail distribution's MLE. A quick investigation into the estimated parameters presented in Appendix B.3 answers this question for us. The estimated splicing point for the Lgn/Lgn distribution is 1.8977. Remember from Section 2.1.3 that the lognormal tail is treated as a truncated distribution with a truncation point equal to the splicing point. The lognormal upper tail parameters are $\mu_u = -7.477$ and $\sigma_u = 3.1516$, and the truncation

probability of the lognormal tail is 0.995. This is a boundary condition in the MLE algorithm and indicates that the Lgn/Lgn distribution results cannot be used.

Additionally, one may wonder why the modified Anderson-Darling test is rejecting the Burr distribution when Burr has the best QS, and we know the true tail behavior is Burr. Since losses for SRC 3 are generated from a mixture distribution with distinct components, the estimated Burr parameters are affected by the body of the losses generated by the lognormal component resulting in an underestimated 99.9% quantile. Thus, the estimated Burr distribution may have a tail that behaves differently from the mixture distribution's Burr component. Another issue is caused by the weights used in calculating the test statistic. Since 30% of the distribution is generated by a lognormal distribution, the weighting function should assign more weight to the tail than our calculation. More importantly, this shows the effectiveness of using the QS and QQ-plots for assessing a distribution's ability to model tail behavior over the modified Anderson-Darling test.

Finally, we observe the effect of using AIC and BIC that focus on the central portion of the distribution when trying to estimate the 99.9% quantile. The Burr distribution, which is the true tail behavior for SRC 3, is unable to capture the tail behavior due to the mode of the losses occurring in the body of the distribution most of which are generated from lognormal. This scenario of model misspecification, which is likely to occur when working from a candidate distribution list, demonstrates the importance of combining criteria that focus on overall fit, such as AIC, with performance at extreme quantiles, such as QS.

3.3 Loss Severity Selection Criteria and Regulatory Capital

In this section, we calculate RC for the years $2014, 2015, \ldots, 2018$ under real-world conditions faced by a bank using AMA. To calculate RC for year $T+1$, for $T+1 = 2014, 2015, \ldots, 2018$, we remove all of the operational losses for the years $T+1, T+2, \ldots, 2018$. This recreates the challenge of estimating a forward-looking RC. Since we know the true value of RC, we are able to compare each year's calculation to the true value. All estimated parameters are available in Appendices A.2, A.3, and A.4 for SRC 1, 2, and 3, respectively.

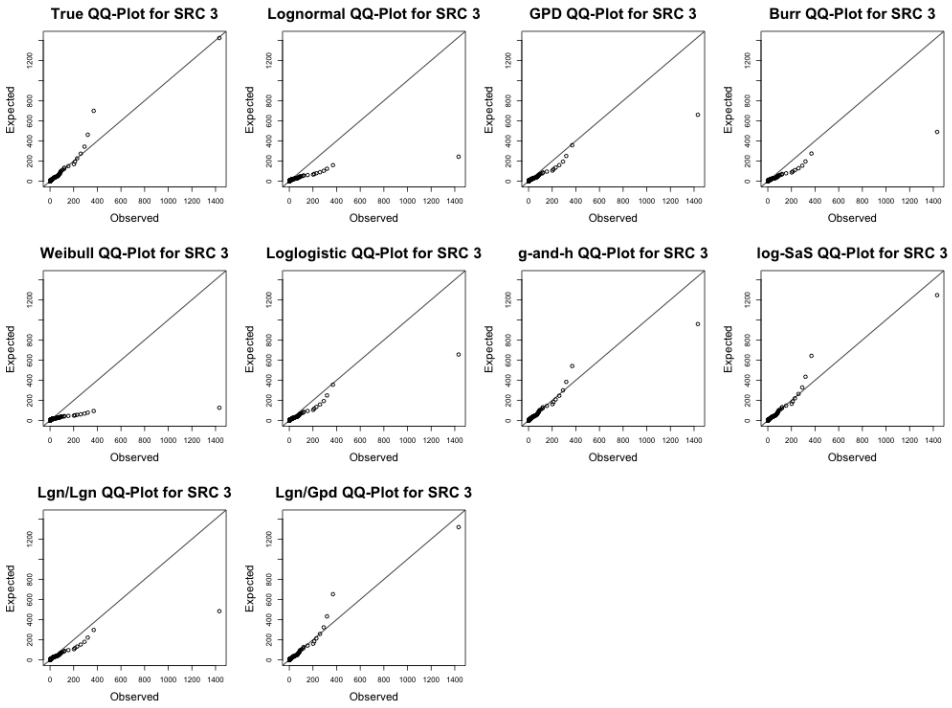

Figure 3.8: QQ-Plots for the true model and each estimated candidate distribution for SRC 3. Estimated quantiles are given on the vertical axis with empirical quantiles along the horizontal. Points below the 45° line indicate underestimates the empirical quantile.

All historical losses for the years $2004, 2005, \ldots, T$ are used to estimate parameters for the candidate severity distributions using the MLE approach. Loss severity distribution selection for each SRC is performed using two separate objective criteria. First, we use lowest AIC and truncation probability estimate less than 0.5. The selected loss severity distributions are then used to derive the Poisson rate parameter for all losses. Then, we can empirically derive the marginal distributions for total annual operational loss from each SRC, denoted F_{S_r}, see Section 2.3. Finally, we use t-copulas with 10 degrees of freedom and correlation parameters 0 and 0.1, respectively, to find the distribution for the total annual operational loss. From this distribution, we calculate RC as the 99.9% quantile. The process is then repeated using quantile score instead of AIC.

3.3.1 Marginal Distributions

For each SRC, Figures 3.9, 3.10, and 3.11 show side-by-side plots of the 99.9% quantile from the best loss severity distribution each year based on AIC and quantile score and the 99.9% quantile of the SRC's marginal distribution for total annual loss. An important observations from these plots is that the loss severity quantile can be used as a proxy for the total annual loss quantile.

For SRC 1, the plots are given in Figure 3.9. The Burr distribution wins every year based on AIC and all but one year when using QS. In 2015, the *g*-and-*h* distribution had the best QS. We observe that changes in the winning distribution can cause drastic changes in the extreme quantiles. The underestimation of the *g*-and-*h* distribution in 2015 is due to the QS selecting a distribution based only on the 99.9% quantle. Thus, we see that the rest of the distribution may contribute a lot to the total annual risk for SRC 1, and measures of overall fit should also be considered when selecting a distribution.

For SRC 2, the plots are given in Figure 3.10. When selecting the best distribution by AIC, log-SaS wins in 2014 and 2015, but is then beat by *g*-and-*h* for the remaining years. When selecting the loss severity distribution based on QS, log-SaS wins every year.

For SRC 3, the plots are given in Figure 3.11. When selecting the best distribution by AIC, the lognormal body spliced with generalized Pareto tail wins every year. When using QS, the Burr distribution wins. SRC 3 creates the model misspecification scenario and yields the most interesting results.

Using AIC alone may lead to selecting a severity distribution that drastically overestimates risk, which is a deterrent from the bank's perspective. We see this demonstrated in Figure 3.11 for the earlier years in the simulation. The QS selection criteria, however, chooses distributions that consistently underestimate the 99.9% quantile of the distribution of SRC 3. This scenario is also seen when analyzing SRC 1. Selecting severity distributions by the best QS can severely underestimate risk. This signals that the other 99% of the distribution has considerable impact on the annual losses for SRC 3.

If we believe that model misspecification is likely to occur, then these results promote the use of AIC and BIC together with the QS, truncation probability esti-

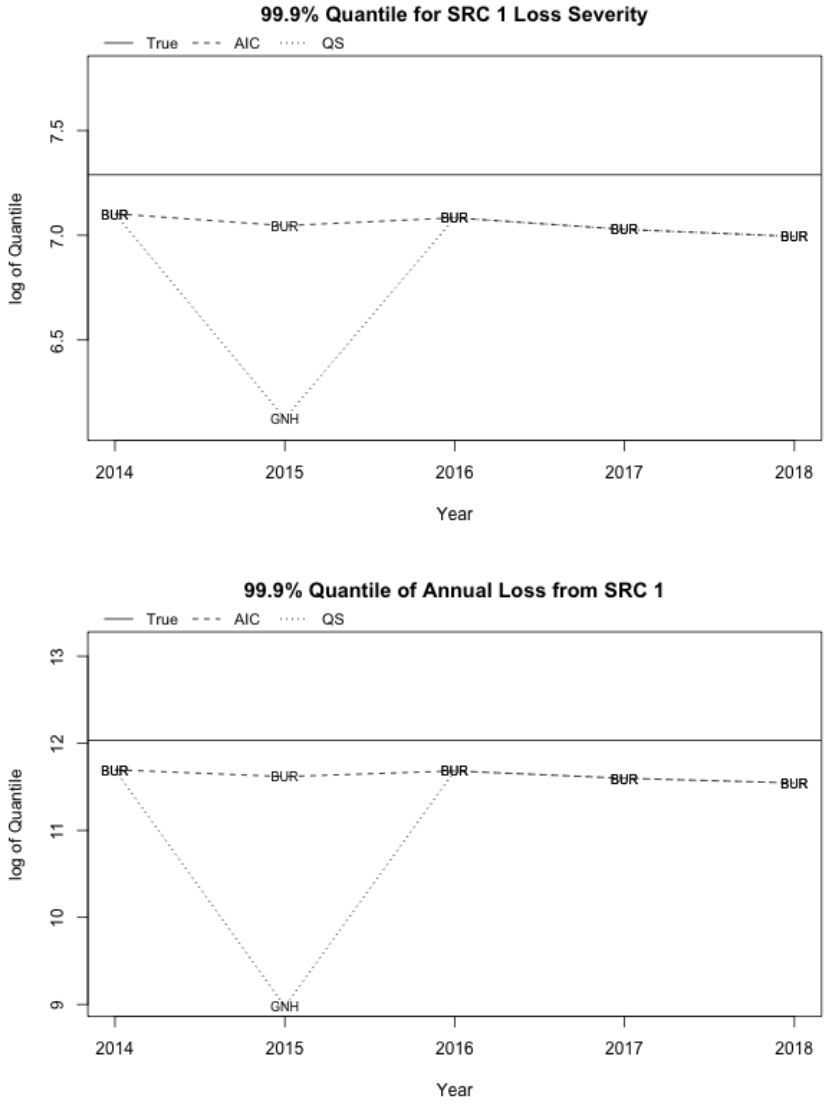

Figure 3.9: The top plot shows the log of the 99.9% quantile from SRC 1's loss severity distribution as selected by AIC and QS. The shorthand name for loss severity distribution shows each year's selection. The bottom plot shows the log of the 99.9% quantile for SRC 1's total annual loss marginal distribution.

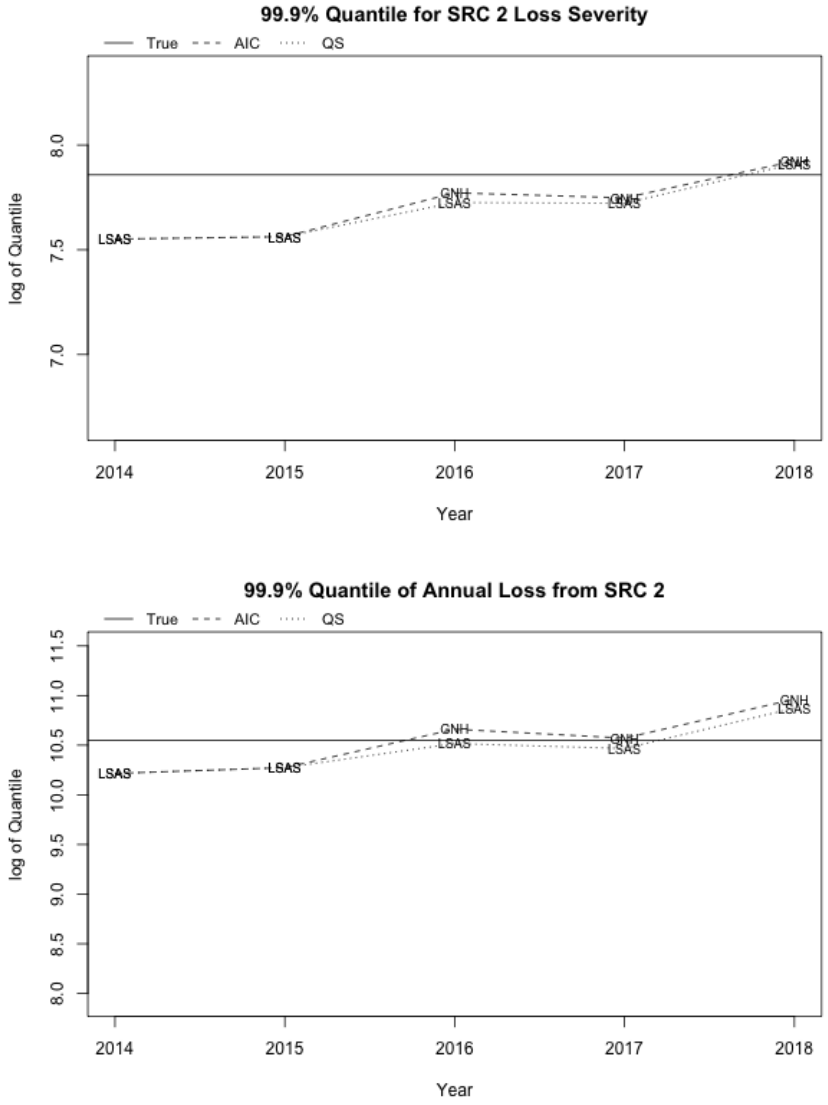

Figure 3.10: The top plot shows the log of the 99.9% quantile from SRC 2's loss severity distribution as selected by AIC and QS. The shorthand name for loss severity distribution shows each year's selection. The bottom plot shows the log of the 99.9% quantile for SRC 2's total annual loss marginal distribution.

mates, and graphical tools like QQ-plots and density plots to get a full picture of the loss severity data.

3.3.2 Regulatory Capital

We use the loss severity distributions from the previous section to compare RC when selecting severity distribution using AIC versus QS. Since there are no data to assess the dependence across SRC's, we use t-copulas with 10 degrees of freedom and correlation parameters 0 and 0.1 when combining losses for total annual operational loss. Using these two separate correlation parameters, we see that RC barely changes, as presented in Figure 3.12, suggesting little sensitivity to changes in the dependence between SRC's. This analysis is limited, however, and using various degrees of freedom with different correlation parameters can provide a better picture of sensitivity.

We see a similar story as told by the marginal distribution quantile plots, namely that the quantile score calculates RC consistently below that of the true value and RC calculated by choosing loss severity distributions by AIC. This seems to suggest that accurate RC calculations should consider more than just the 99.9% quantile.

3.4 Challenges of Truncation Probabilities

Without any data collection or expert opinion of the operational loss severities below an SRC's minimum reporting threshold, it is difficult to set a general interval of reasonable truncation probability estimates. While we can set conservatively wide ranges, such as $(0.01, 0.5)$, it may be possible that the actual proportion of an SRC's losses below the threshold fall outside this interval. This problem of truncation probabilities exists throughout the industry with no consensus on how to handle it [AMA Group, 2013]. Using either the naive approach or the shifted approach avoids truncation probabilities altogether, but they have their own detractions as discussed in Section 2.1.

In addition to the challenges directly associated with estimating truncation probability, a study by Yu and Brazauskas [2017] analyzed the differences in RC

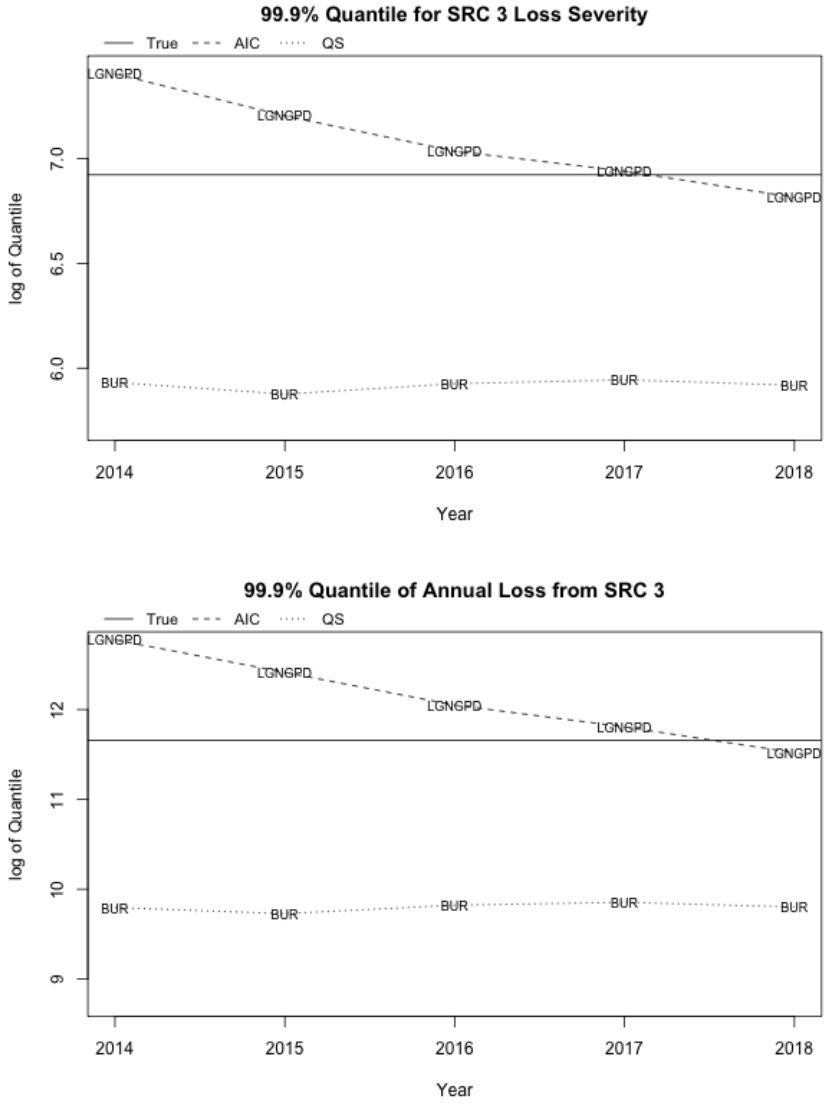

Figure 3.11: The top plot shows the log of the 99.9% quantile from SRC 3's loss severity distribution as selected by AIC and quantile score. The shorthand name for loss severity distribution shows each year's selection. The bottom plot shows the log of the 99.9% quantile for SRC 3's total annual loss marginal distribution.

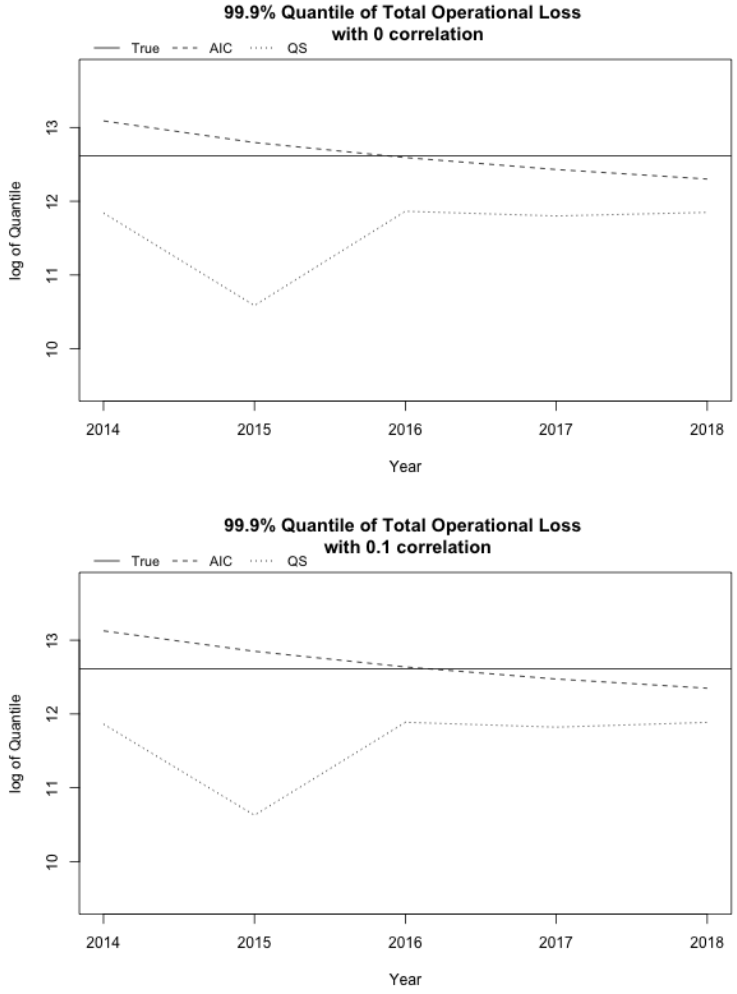

Figure 3.12: The top plot shows the log of RC as the 99.9% quantile for total annual operational loss when using a t-copula with 10 degrees of freedom and correlation parameter of 0 to combine losses across SRC's. The bottom plot shows the log of RC with a correlation parameter of 0.1. In both plots, the true RC is the solid line, the dashed line is RC using estimated distributions selected by AIC, and the dotted line shows RC using estimated distributions selected by QS.

estimates between the three approaches and concluded that the truncation approach provides the lowest RC (called VaR by Yu and Brazauskas) estimates when using LDA:

> *"We demonstrate that for a fixed probability distribution, the choice of the truncation approach yields lowest VaR estimates, which may be viewed as beneficial to the bank, whilst the naive and shifted approaches lead to higher estimates of VaR."*

The uncertainty surrounding the truncation probability estimate may be an explanation for the phenomenon of RC estimates being systematically lower when using the truncation approach. If the truncation probability estimate is too high, then too many of the simulated losses may be too small.

To investigate this idea further, we simulate losses in a scenario where we have more information than in reality. We assume a known distribution and 1000 independent samples each of size 2500 from the same Burr distribution used to generate SRC 1. We then truncate each sample at the actual 2.5%, 5%, 10%, 20%, and 40% quantiles and estimate the truncation probability assuming the Burr distribution. Figure 3.13 shows the distribution of truncation probability estimates under the known model for different truncation points.

As expected, when the truncation point is small, the distribution of truncation probability estimates is very positively skewed. As the truncation point is increased, we see larger standard errors. With a truncation point at the 20% quantile, we have a bimodal distribution with many estimates far exceeding the actual value. Without any knowledge of the losses below the truncation point, we acknowledge that truncation probability estimates introduce more uncertainty in calculating RC.

In light of the challenges in estimating truncation probability, we propose a simple alternative. The purpose of the truncation approach is to estimate an entire distribution over which we can simulate losses for each SRC when using the LDA. Using the histograms presented in Figure 3.2, we can make reasonable assumptions about the number of losses below the truncation point. For example, the histograms for SRC 1 and SRC 3 seem to be indicating a mode for the losses around the 2nd or 3rd bucket. Under the assumptions of the truncation approach combined with our list of candidate distributions, it would be reasonable to assume that the number

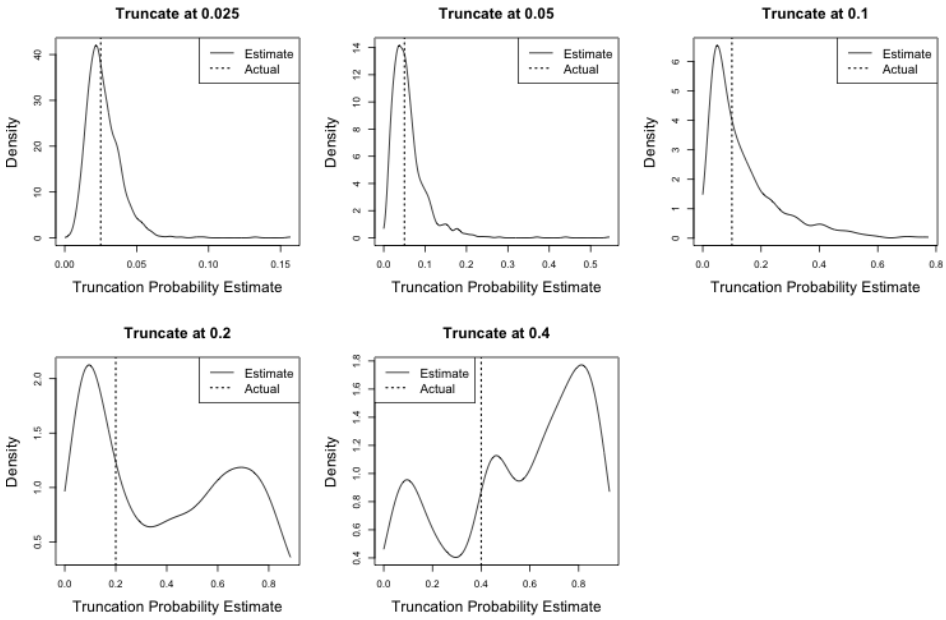

Figure 3.13: Distribution of truncation probability estimates for 1000 truncated samples of size 2500 at the 2.5%, 5%, 10%, 20%, and 40% quantiles. Maximum likelihood estimation is performed using the Burr distribution. Even under these optimal conditions, there is a lot of uncertainty in the truncation probability estimate.

of losses for each bucket below the truncation point can be bounded above by the mode of the observable losses. Under this assumption, we could eliminate candidate distributions that do not have a truncation probability below this "worst case" scenario. Unfortunately, this would not help us with losses exhibited by an SRC that does not show a clear mode in the histogram and whose mode might lie below the truncation point. Any data collection for losses below the threshold, even in aggregate, improves the truncation approach by setting bounds on the truncation probability estimates.

As mentioned in Section 2.1.4, the truncation probability is a useful tool for eliminating inappropriate distributions that seek to mimic tail behavior by increasing the truncation probability. Using the same 1000 Burr samples used to generate

Figure 3.13, estimating the truncation probably using the lognormal distribution produces no estimates below 0.9. These high estimates would lead to extremely high Poisson rates for the loss frequencies when simulating losses. Therefore, the speed and efficiency of using the truncation approach to select a loss severity distribution is increased by considering the truncation probability estimate.

Chapter 4

Conclusion

While both regulators and financial institutions share the main objective to measure OR as accurately as possible, they have diametrically opposing priorities. Regulators want to minimize a bank's exposure to financial ruin from an operational loss and thus want operational risk modeled as accurately as possible while minimizing underestimation. Since a bank must set aside assets equal to RC to cover potential losses, the bank wants to accurately model operational risk while minimizing overestimation. While we do not pose a solution to this problem, the research presented supports the quantile score as a function that considers the priorities of both stakeholders.

Since the 99.9% quantile of a loss severity distribution is a good proxy of a SRC's total annual loss, selecting distributions based on the quantile score is intuitive and easily interpretable. The asymmetry of the quantile score at the 99.9% quantile penalizes underestimation more than overestimation which aligns with the priorities of regulators. However, the quantile score also penalizes overestimation, and severe overestimation is a concern for banks that is also seen in operational risk research [Dutta and Perry, 2006]. Thus, the quantile score accounts for both the regulator's and the bank's priorities.

A concern when using the quantile score at the 99.9% quantile to select a severity distribution is that it ignores the rest of the distribution, which can make up a significant portion of a SRC's total annual loss. One way to address this issue is to calculate the quantile score at multiple quantiles. Since we are most concerned with

the right-tail of the loss severity distributions, we suggest using the "letter values" presented in Hoaglin [1985] which are comprised of the quantiles $\left(1-(1/2)^n\right)$ for $n = 2,3,\ldots,8$. If the focus is on achieving the best quantile score in the right-tail, however, then the maximum likelihood approach may not provide parameters that minimize the quantile score objective function, equation (2.13). Minimizing equation (2.13) is a computationally taxing process, so due diligence to eliminate inappropriate distributions such as those with unrealistically high truncation probabilities would be prudent.

The quantile score should be complemented with estimators such as AIC and BIC, which consider the entire distribution with a focus around the mode, and some combination of these estimators should provide better distribution selection than only using one. We noticed in the SRC 3 simulation that AIC and quantile score picked very different distributions, and this may be a signal that loss severities are generated from separate processes in the body and the tail, so combining these measures is especially important when there is model misspecification. A single distribution may dominate the quantile score, but is unable to capture the true loss generating process if it is a mixture model or some other combination of distributions. The opposite is also true. A distribution that dominates in AIC may fail to accurately capture the extreme right tail of the loss generating process and lead to drastic over or underestimations.

The log-SaS distribution is a highly flexible distribution that performs well when modeling loss severities generated from different processes when using maximum likelihood estimation. Log-SaS solves the problems of the *g*-and-*h* distribution by having a support on the positive real numbers, a monotonic transformation over the entire parameter space, and an analytical inverse. The log-SaS distribution offers these advantages while maintaining the flexibility to capture various tail behaviors and shapes. Another benefit of the log-SaS severity distribution is that the parameters can be estimated using a log-transform of the loss severity data, aiding in the convergence of numerical algorithms and mitigating the badly-scaled problem.

Issues with maximum likelihood estimation and truncation probability estimates should be included in the loss severity distribution selection process, as they can quickly eliminate inappropriate distributions. When using the truncation

approach, extremely high truncation probabilities are often the result of a non-converging maximum likelihood algorithm due to a lighter-tailed distribution's effort to mimic heavier-tailed behavior. Even if the algorithm is converging, high truncation probabilities may simply be unrealistic and more data collection and/or expert opinion is desired for losses below the minimum reporting threshold. For the *g*-and-*h* distribution specifically, care must be taken to analyze the results of maximum likelihood to avoid situations where too much of the distribution lies below zero.

This report has laid the groundwork for future work in loss severity distribution selection that incorporates both regulators' and banks' preferences to operational risk modeling. Inclusion of the quantile score shows promise in capturing the right-tail of loss severity data, and future work may provide optimal quantiles to incorporate and the effectiveness of estimating distribution parameters using the quantile score in equation (2.13) as the objective function. We also present a new perspective to maximum likelihood estimation for the *g*-and-*h* distribution that includes penalties to the portion of the distribution below zero. This penalized likelihood approach can be improved by assessing the optimal penalty term and the weight given to the penalty and may also be applied to other distributions to eliminate power tail mimicking of Gumbel-type distributions. We also stress that the truncation approach can benefit from additional data collection for losses below the reporting threshold and encourage banks to start this collection process.

Bibliography

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, AC-19(6):716–723.

AMA Group (2013). Ama quantification challenges: Amag range of practice and observations on 'the thorny lda topics'. Industry position paper, The Risk Management Association.

Baud, N., Frachot, A., and Roncalli, T. (2003). How to avoid over-estimating capital charge for operational risk? *Opertaional Risk - Risk's Newsletter*.

BCBS (2006). International convergence of capital measurement and capital standards. Technical report, Bank for International Settlements.

BCBS (2009). Observed range of practice in key elements of advanced measurement approaches (ama). Survey, Bank for International Settlements.

BCBS (2011). Operational risk - supervisory guidelines for the advanced measurement approaches. Technical report, Bank for International Settlements.

BCBS (2016). Standardised measurement approach for operational risk. Consultative document, Bank for International Settlements.

Beirlant, J., Goegebeur, Y., Teugels, J., and Segers, J. (2004). *Statistics of Extremes: Theory and Applications*. John Wiley & Sons, Ltd.

Chavez-Demoulin, V., Embrechts, P., and Hofert, M. (2015). An extreme value approach for modeling operational risk losses depending on covariates. *The Journal of Risk and Insurance*, 83(3):735–776.

Chernobai, A., Menn, C., Truck, S., and Rachev, S. (2005). A note on the estimation of the frequency and severity distribution of operational losses. *The Mathematical Scientist*, 30(2):1–10.

Chernobai, A., Rachev, S., and Fabozzi, F. (2007). *Operational Risk: A Guide to Basel II Capital Requirements, Models, and Analysis*. John Wiley & Sons, Inc.

Coles, S. (2001). *An Introduction to Statistical Modelling of Extreme Values*. Springer Series in Statistics.

Degen, M., Embrechts, P., and Lambrigger, D. (2007). The quantitative modeling of operational risk: Between g-and-h and evt. *Astin Bulletin*, 37(2):265–291.

Dutta, K. and Perry, J. (2006). A tale of tails: An empirical analysis of loss distribution models for estimating operational risk capital. Working Papers 06-13, Federal Reserve Bank of Boston.

Economist (2016). The final bill - financial crime. *The Economist*, 11.

Embrechts, P. and Hofert, M. (2011). Practices and issues in operational risk modeling under basel ii. *Lithuanian Mathematical Journal*, 51(2):180–193.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer.

Foss, S., Korshunov, D., and Zachary, S. (2013). *An Introduction to Heavy-Tailed and Subexponential Distributions*. Springer Science + Business Media New York, 2nd edition.

Gilbert, R. (1987). *Statistical Methods for Environmental Pollution Monitoring*. Van Nostrand Reinhold Company, Inc.

Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494):746–762.

Grooters, D. and Reinink, B. (2013). Data set size for operational risk modelling. Technical report, Black Cat's Walk.

Hayahsi, Y. (2018). Wells fargo to pay $1 billion to settle risk management claims. *Wall Street Journal*, 20.

Hoaglin, D. (1985). Summarizing shape numerically: The g-and-h distributions. In Hoaglin, D., Mosteller, F., and Tukey, J., editors, *Exploring Data Tables, Trends, and Shapes*, chapter 11, pages 461–513. John Wiley & Sons, Inc.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R*. Springer Science + Business Media New York, 6th edition.

Jones, M. and Pewsey, A. (2009). Sinh-arcsinh distributions. *Biometrika*, 96(4):761–780.

Lee, S. Y. and Kim, J. H. (2018). Exponentiated generalized pareto distribution: Properties and applications towards extreme value theory. *Communications in Statistics - Theory and Methods*, page DOI: 10.1080/03610926.2018.1441418.

Luo, X., Shevchenko, P., and Donnelly, J. (2007). Addressing the impact of data truncation and parameter uncertainty on operational risk estimates. *The Journal of Operational Risk*, 2(4):3–26.

McNeil, A. J., Frey, R., and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques, Tools*. Princeton University Press, Princeton.

Panjer, H. (2006). *Operational Risk: Modeling Analytics*. John Wiley & Sons, Inc.

Perline, R. (2005). Strong, weak and false inverse power laws. *Statistical Science*, 20(1):68–88.

Peters, G. and Shevchenko, P. (2015). *Advances in Heavy Tailed Risk Modeling: A Handbook of Operational Risk*. John Wiley & Sons, Inc.

Polfeldt, T. (1970). Asymptotic results in non-regular estimation. *Scandinavian Actuarial Journal*, Supplemental Volumes 1-2.

Resnick, S. (1987). *Extreme Values, Regular Variation, and Point Processes*. Springer.

Ross, S. (2010). *A First Course in Probability*. Pearson Education, Inc., 8th edition.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464.

Sinclair, C., Spurr, B., and Ahmad, M. (1990). Modified anderson-darling test. *Communications in Statistics - Theory and Methods*, 19(10):3677–3686.

Strasburg, J. (2018). Barclays to pay $2 billion to resolve mortgage-securities claims. *Wall Street Journal*, 29.

Yu, D. and Brazauskas, V. (2017). Model uncertainty in operational risk modeling due to data truncation: A single risk case. *Risks*, 5(49).

Appendix A

Loss Severity Distributions

A.1 Lognormal Distribution

Let Y be a normally distributed random variable with location parameter μ and scale parameter σ. Since $Y \sim N(\mu, \sigma^2)$, then $X = e^Y$ is lognormally distributed with location parameter μ and scale parameter σ, denoted $\text{LN}(\mu, \sigma^2)$. To derive the lognormal distribution, let $Z \sim N(0,1)$. Then $Y = \mu + \sigma Z$ and

$$F_X(x) = P(X \leq x) = P\big(e^Y \leq x\big) = P\big(\mu + \sigma Z \leq \log(x)\big) = P\left(Z \leq \frac{\log(x) - \mu}{\sigma}\right)$$
$$= \Phi\left(\frac{\log(x) - \mu}{\sigma}\right),$$

where Φ is the standard normal CDF. Let ϕ be the standard normal density function. Then for $\Theta = [\mu \ \sigma]$, where $\mu \in \mathbb{R}$ and $\sigma > 0$,

$$F_X(x;\Theta) = \begin{cases} \Phi\left(\frac{\log(x)-\mu}{\sigma}\right) & \text{for } x > 0, \\ 0 & \text{for } x \leq 0; \end{cases}$$
$$f_X(x;\Theta) = \begin{cases} \frac{1}{\sigma x}\phi\left(\frac{\log(x)-\mu}{\sigma}\right) & \text{for } x > 0, \\ 0 & \text{for } x \leq 0; \end{cases}$$

$$F_X^{-1}(p;\Theta) = \exp\left\{\mu + \sigma\Phi^{-1}(p)\right\} \quad \text{for } 0 < p < 1.$$

The conditional CDF and PDF for a lognormally distributed random variable with minimum threshold τ are derived from equations (2.5) and (2.6), respectively. The quantile function is found by setting $\widetilde{F}(x;\Theta,\tau) = p$, for $0 < p < 1$, and solving for x:

$$\widetilde{F}(x;\Theta,\tau) = \begin{cases} \frac{\Phi\left[\left(\log(x)-\mu\right)/\sigma\right] - \Phi\left[\left(\log(\tau)-\mu\right)/\sigma\right]}{1-\Phi\left[\left(\log(\tau)-\mu\right)/\sigma\right]} & \text{for } x > \tau, \\ 0 & \text{for } x \leq \tau; \end{cases}$$

$$\widetilde{f}(x;\Theta,\tau) = \begin{cases} \frac{\phi\left[\left(\log(x)-\mu\right)/\sigma\right]}{\sigma x\left[1-\Phi\left[\left(\log(\tau)-\mu\right)/\sigma\right]\right]} & \text{for } x > \tau, \\ 0 & \text{for } x \leq \tau; \end{cases}$$

$$\widetilde{F}^{-1}(p;\Theta,\tau) = \exp\left\{\mu + \sigma\Phi^{-1}\left[(1-p)\Phi\left(\frac{\log(\tau)-\mu}{\sigma}\right) + p\right]\right\} \quad \text{for } 0 < p < 1.$$

From equation (2.7), we derive the conditional likelihood function as

$$\widetilde{L}(\Theta;\mathbf{x},\tau) = \left[\sigma - \sigma\,\Phi\left(\frac{\log(\tau)-\mu}{\sigma}\right)\right]^{-n} \prod_{i=1}^{n} \frac{1}{x_i}\,\phi\left(\frac{\log(x_i)-\mu}{\sigma}\right).$$

From equation (2.8), the conditional log-likelihood function is

$$\begin{aligned}\tilde{\ell}(\Theta;\mathbf{x},\tau) = &-n\log(\sigma) - n\log\left[1-\Phi\left(\frac{\log(\tau)-\mu}{\sigma}\right)\right] \\ &+ \sum_{i=1}^{n}\log\left[\phi\left(\frac{\log(x_i)-\mu}{\sigma}\right)\right] - \sum_{i=1}^{n}\log(x_i).\end{aligned}$$

A.2 Generalized Pareto Distribution

As presented in Coles [2001], for a sequence of random variables, $X_1, X_2, \ldots, X_n \overset{iid}{\sim} H$, let

$$M_n = \max\{X_1, ..., X_n\}.$$

Suppose that there exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that

$$H(b_n + a_n z) = P\{(M_n - b_n)/a_n \leq z\} \to G(z) \quad \text{as } n \to \infty$$

for a non-degenerate distribution function G. Then, G is a member of the generalized extreme value (GEV) family,

$$G(z) = \exp\Big\{ -\Big[1 + \xi\Big(1 + \frac{z-\mu}{\sigma}\Big)\Big]^{-1/\xi}\Big\},$$

defined on $\big\{z : 1 + \xi(z-\mu)/\sigma > 0\big\}$, where $\mu, \sigma > 0$ and $\mu, \xi \in \mathbb{R}$. For large n,

$$\begin{aligned} P\{(M_n - b_n)/a_n \leq z\} &\approx G(z) \\ P\{M_n \leq z\} &\approx G(b_n + a_n z) \\ &= G^*(z), \end{aligned}$$

where G^* is another distribution from the GEV family.

Finally, for large enough $u > \mu$ and $\theta = \sigma + \xi(u - \mu)$, the distribution function of $(X - u)$ conditioned on $X > u$ is approximately

$$F(x) = \begin{cases} 1 - \Big[1 + \frac{\xi}{\theta}(x-u)\Big]^{-1/\xi} & \text{for } \xi \neq 0, \\ 1 - \exp\big\{ -\frac{x-u}{\theta}\big\} & \text{for } \xi = 0, \end{cases}$$

defined for $x > u$ when $\xi > 0$ and $u < x < u - \theta/\xi$ when $\xi < 0$, and $\xi = 0$ interpreted as $\xi \to 0$ leading to an exponential distribution for the excess $x - u$ with parameter $1/\theta$. This is called the **GPD!** (**GPD!**) family.

For our purposes, we work exclusively with the **GPD!** where $\xi > 0$, so all functions for the remainder of Appendix A.2 assume tail parameter $\xi > 0$. The CDF, PDF, and quantile function for the **GPD!** family with parameter vector $\Theta = [\,\xi \;\; \theta\,]$, where $\xi > 0$ is the tail parameter, $\theta > 0$ the scale parameter, and u is the treshold, are

$$F(x;\Theta,u) = \begin{cases} 1-\left(1+\frac{\xi}{\theta}(x-u)\right)^{-1/\xi} & \text{for } x>u, \\ 0 & \text{for } x\le u; \end{cases}$$

$$f(x;\Theta,u) = \begin{cases} \frac{1}{\theta}\left(1+\frac{\xi}{\theta}(x-u)\right)^{-\frac{1+\xi}{\xi}} & \text{for } x>u, \\ 0 & \text{for } x\le u; \end{cases}$$

$$F^{-1}(p;\Theta,u) = u+\frac{\theta}{\xi}\left((1-p)^{-\xi}-1\right) \qquad \text{for } 0<p<1.$$

For single severity candidate distributions, the threshold $u = 0$ is not a parameter and thus is not included in the parameter vector. When estimating the parameters for the **GPD!** family assuming a single severity distribution, operational loss amounts are treated as excesses over 0. Since our data are from a truncated sample with minimum reporting threshold τ, we calculate the conditional CDF and PDF for a generalized Pareto distributed random variable using equations (2.5) and (2.6), respectively. The quantile function is found by setting $\widetilde{F}(x;\Theta,\tau) = p$ and solving for x. Assuming $\tau > u$, the conditional CDF, PDF, and quantile functions are

$$\widetilde{F}(x;\Theta,u,\tau) = \begin{cases} 1-\left(\frac{\theta+\xi(x-u)}{\theta+\xi(\tau-u)}\right)^{-1/\xi} & \text{for } x>\tau, \\ 0 & \text{for } x\le\tau; \end{cases}$$

$$\widetilde{f}(x;\Theta,u,\tau) = \begin{cases} \frac{1}{\theta+\xi(\tau-u)}\left(\frac{\theta+\xi(x-u)}{\theta+\xi(\tau-u)}\right)^{-\frac{1+\xi}{\xi}} & \text{for } x>\tau, \\ 0 & \text{for } x\le\tau; \end{cases}$$

$$\widetilde{F}^{-1}(p;\Theta,u,\tau) = u+\frac{(1-p)^{-\xi}\left(\theta+\xi(\tau-u)\right)-\theta}{\xi} \quad \text{for } 0<p<1.$$

For the LGNGPD distribution, the tail distribution is the unconditional **GPD!** with the threshold equal to the splicing point, $u = x_s$.

Using equation (2.7), the conditional likelihood function for the single severity generalized Pareto candidate distribution is

$$\widetilde{L}(\Theta;\mathbf{x},u,\tau) = \left[\theta+\xi(\tau-u)\right]^{n/\xi}\prod_{i=1}^{n}\left[\theta+\xi(x_i-u)\right]^{-\frac{1+\xi}{\xi}},$$

and from equation (2.8), the conditional log-likelihood function for the single severity generalized Pareto candidate distribution is

$$\tilde{\ell}(\Theta;\mathbf{x},u,\tau) = \tfrac{n}{\xi}\log\left\{\theta+\xi(\tau-u)\right\} - \left(\tfrac{1+\xi}{\xi}\right)\sum_{i=1}^{n}\log\left\{\theta+\xi(x_i-u)\right\}.$$

To aid in the convergence of the MLE algorithm as mentioned in Section 2.1.4, we fit the log-transform of a generalized Pareto random variable. Let $X \sim \text{GPD}(\xi,\theta)$ with threshold u, and let $Y = \log(X)$. The CDF and PDF for Y are derived, respectively, as

$$F_Y\big(y;\Theta,\log(u)\big) = F(e^y;\Theta,u) \tag{A.1}$$

$$f_Y\big(y;\Theta,\log(u)\big) = e^y f(e^y;\Theta,u) \tag{A.2}$$

Then Y is called the exponentiated generalized Pareto distribution [Lee and Kim, 2018], and the CDF and PDF for Y, using equations (A.1) and (A.2) are

$$F_Y\big(y;\Theta,\log(u)\big) = \begin{cases} 1-\left(1+\frac{\xi}{\theta}\left(e^y-u\right)\right)^{-1/\xi} & \text{for } y>\log(u),\\ 0 & \text{for } y\le\log(u);\end{cases}$$

$$f_Y\big(y;\Theta,\log(u)\big) = \begin{cases} \frac{e^y}{\theta}\left(1+\frac{\xi}{\theta}\left(e^y-u\right)\right)^{-\frac{1+\xi}{\xi}} & \text{for } y>\log(u),\\ 0 & \text{for } y\le\log(u).\end{cases}$$

Assuming $\tau > u \ge 0$, the conditional CDF and PDF are

$$\widetilde{F}_Y\big(y;\Theta,\log(u),\log(\tau)\big) = \begin{cases} 1-\left(\frac{\theta+\xi(e^y-u)}{\theta+\xi(\tau-u)}\right)^{-1/\xi} & \text{for } y>\log(\tau),\\ 0 & \text{for } y\le\log(\tau);\end{cases}$$

$$\widetilde{f}_Y\big(y;\Theta,\log(u),\log(\tau)\big) = \begin{cases} \frac{e^y}{\theta+\xi(\tau-u)}\left(\frac{\theta+\xi(e^y-u)}{\theta+\xi(\tau-u)}\right)^{-\frac{1+\xi}{\xi}} & \text{for } y>\log(\tau),\\ 0 & \text{for } y\le\log(\tau).\end{cases}$$

By letting $\mathbf{y} = \log(\mathbf{x})$ and using equation (2.7), the conditional likelihood function for the log-losses is

$$\widetilde{L}(\Theta; \mathbf{y}, u, \tau) = \left[\theta + \xi(\tau - u)\right]^{n/\xi} \exp\left\{\sum_{i=1}^{n} y_i\right\} \prod_{i=1}^{n} \left[\theta + \xi(e^{y_i} - u)\right]^{-\frac{1+\xi}{\xi}},$$

and from equation (2.8), the conditional log-likelihood function for the log-losses is

$$\tilde{\ell}(\Theta; \mathbf{y}, u, \tau) = \tfrac{n}{\xi} \log\left\{\theta + \xi(\tau - u)\right\} + \sum_{i=1}^{n} y_i - \left(\tfrac{1+\xi}{\xi}\right) \sum_{i=1}^{n} \log\left\{\theta + \xi(e^{y_i} - u)\right\}.$$

A.3 Burr Distribution

The parameterization of the candidate Burr Distribution is a generalized three-parameter version of the Pareto distribution derived from Chapter 6 of Chernobai et al. [2007]. This parameterization allows for greater flexibility due to an additional shape parameter and is derived as

$$\begin{aligned} F(x) &= 1 - \left(\frac{\beta}{\beta + x^{\gamma}}\right)^{\alpha} = 1 - \left(\frac{\beta + x^{\gamma}}{\beta}\right)^{-\alpha} = 1 - \left(1 + \frac{x^{\gamma}}{\beta}\right)^{-\alpha} \\ &= 1 - \left(1 + \left(\frac{x}{\theta}\right)^{\gamma}\right)^{-\alpha}, \end{aligned}$$

where $\theta = \beta^{1/\gamma}$. This parameterization of the Burr distribution has shape parameters $\alpha,\ \gamma > 0$ and scale parameter $\theta > 0$. The flexibility of this parameterization is in its ability to capture both Pareto and loglogistic distributions. When $\gamma = 1$, the Burr distribution reduces to the Pareto distrbution with power tail $x^{-\alpha}$. When $\alpha \leq 1$, we have a very heavy-tailed distribution with infinite mean and variance. When $\alpha = 1$, the Burr distribution reduces to the loglogistic distribution.

The CDF, PDF, and quantile function for the Burr distribution with parameter vector $\Theta = [\,\alpha\ \gamma\ \theta\,]$ are

$$F(x; \Theta) = \begin{cases} 1 - \left[1 + \left(\frac{x}{\theta}\right)^{\gamma}\right]^{-\alpha} & \text{for } x > 0, \\ 0 & \text{for } x \leq 0; \end{cases}$$

$$f(x;\Theta) = \begin{cases} \frac{\alpha\gamma x^{-1}\left(\frac{x}{\theta}\right)^{\gamma}}{\left(1+\left(\frac{x}{\theta}\right)^{\gamma}\right)^{1+\alpha}} & \text{for } x > 0, \\ 0 & \text{for } x \leq 0; \end{cases}$$

$$F^{-1}(p;\Theta) = \theta\left((1-p)^{-1/\alpha} - 1\right)^{1/\gamma} \quad \text{for } 0 < p < 1.$$

The conditional CDF and PDF for a Burr distributed random variable with minimum threshold τ are derived from equations (2.5) and (2.6), respectively. The quantile function is found by setting $\widetilde{F}(x;\tau) = p$ and solving for x:

$$\widetilde{F}(x;\Theta,\tau) = \begin{cases} 1 - \left(\frac{\theta^{\gamma}+\tau^{\gamma}}{\theta^{\gamma}+x^{\gamma}}\right)^{\alpha} & \text{for } x > \tau, \\ 0 & \text{for } x \leq \tau; \end{cases}$$

$$\widetilde{f}(x;\Theta,\tau) = \begin{cases} \alpha\gamma\, x^{\gamma-1}\left[\theta^{\gamma}+\tau^{\gamma}\right]^{\alpha}\left[\theta^{\gamma}+x^{\gamma}\right]^{-\alpha-1} & \text{for } x > \tau, \\ 0 & \text{for } x \leq \tau; \end{cases}$$

$$\widetilde{F}^{-1}(p;\Theta,\tau) = \left[(1-p)^{-1/\alpha}\left[\theta^{\gamma}+\tau^{\gamma}\right] - \theta^{\gamma}\right]^{1/\gamma} \quad \text{for } 0 < p < 1.$$

Using equation (2.7), the conditional likelihood function for the Burr candidate distribution is

$$\widetilde{L}(\Theta;\mathbf{x},\tau) = (\alpha\gamma)^{n}\left[\theta^{\gamma}+\tau^{\gamma}\right]^{n\alpha}\prod_{i=1}^{n} x_i^{\gamma-1}\left[\theta^{\gamma}+x_i^{\gamma}\right]^{-\alpha-1}.$$

From equation (2.8), the conditional log-likelihood function for the Burr candidate distribution is

$$\widetilde{\ell}(\Theta;\mathbf{x},\tau) = n\log(\alpha\gamma) + n\alpha\log\left[\theta^{\gamma}+\tau^{\gamma}\right] + (\gamma-1)\sum_{i=1}^{n}\log(x_i) - (\alpha+1)\sum_{i=1}^{n}\log\left[\theta^{\gamma}+x_i^{\gamma}\right].$$

MLE for the Burr distribution is performed on the log-loss data. Unfortunately, the log-losses do not eliminate the badly-scaled problem discussed in Section 2.1.4. To address this, we again reparameterize the Burr distribution. Making the follow-

ing substitutions,

$$\alpha^* = \alpha\gamma, \qquad \zeta = \gamma, \qquad \eta = \theta,$$

the Burr CDF becomes

$$F(x;\Theta) = 1 - \left(1 + \left(\frac{x}{\eta}\right)^{\zeta}\right)^{-\alpha^*/\zeta},$$

where * is used to differentiate the new α^* parameter from the original parameterization. In this parameterization, α^* is a tail parameter, and ζ is a shape parameter that can affect the lower tail. The Burr distribution is unimodal when $\zeta > 1$.

Now, let $X \sim \text{Burr}(\alpha, \gamma, \theta)$ and let $Y = \log(X)$. Y is said to have a *generalized logistic* distribution, since the parameterization can be used to model both logistic and exponentiated Pareto random variables. Using equations (A.1) and (A.2), the CDF, PDF, and quantile function for the generalized logistic distribution with parameter vector $\Theta = [\,\alpha^* \;\zeta\;\eta\,]$ where $\alpha^* \in \mathbb{R}$ and $\zeta, \eta > 0$ are

$$F_Y(y;\Theta) = 1 - \left[1 + \left(\frac{e^y}{\eta}\right)^{\zeta}\right]^{-\alpha^*/\zeta} \qquad \text{for } y \in \mathbb{R};$$

$$f_Y(y;\Theta) = \frac{\alpha^* \left(e^y/\eta\right)^{\zeta}}{\left[1 + \left(e^y/\eta\right)^{\zeta}\right]^{\alpha^*/\zeta+1}} \qquad \text{for } y \in \mathbb{R};$$

$$F_Y^{-1}(p;\Theta) = \log(\eta) + \frac{1}{\zeta}\,\log\left\{\left(1-p\right)^{-\zeta/\alpha^*} - 1\right\} \quad \text{for } 0 < p < 1.$$

The conditional CDF and PDF for a generalized logistic distributed random variable with minimum threshold $\log(\tau)$ are derived from equations (2.5) and (2.6). The quantile function is found by setting $\widetilde{F}_Y\left(y;\Theta,\log(\tau)\right) = p$ and solving for y:

$$\widetilde{F}_Y\left(y;\Theta,\log(\tau)\right) = \begin{cases} 1 - \left[\left(\eta^{\zeta} + e^{\zeta\tau}\right)/\left(\eta^{\zeta} + e^{\zeta y}\right)\right]^{\alpha^*/\zeta} & \text{for } y > \log(\tau), \\ 0 & \text{for } y \le \log(\tau); \end{cases}$$

$$\widetilde{f}_Y\big(y;\Theta,\log(\tau)\big) = \begin{cases} \big(\alpha^* e^{\zeta y}\big)\big(\eta^\zeta + e^{\zeta\tau}\big)/\big(\eta^\zeta + e^{\zeta y}\big)^{\alpha^*/\zeta+1} & \text{for } y > \log(\tau), \\ 0 & \text{for } y \le \log(\tau); \end{cases}$$

$$F_Y^{-1}\big(p;\Theta,\log(\tau)\big) = \log(\eta) + \frac{1}{\zeta}\,\log\left\{\frac{1+\Big(e^\tau/\eta\Big)^{\zeta^*}}{(1-p)^{\zeta/\alpha^*}} - 1\right\} \quad \text{for } 0 < p < 1.$$

All estimated parameters are converted back to the original Burr(α,γ,θ) distribution throughout the report.

By letting $\mathbf{y} = \log(\mathbf{x})$ and using equation (2.7), the conditional likelihood function for the generalized logistic candidate distribution is

$$\widetilde{L}\big(\Theta;\mathbf{y},\log(\tau)\big) = \alpha^{*n}\left(\eta^\zeta + \exp\big\{\zeta\tau\big\}\right)^{n\alpha^*/\zeta}\prod_{i=1}^{n}\frac{\exp\big\{\zeta y_i\big\}}{\Big[\eta^\zeta + \exp\big\{\zeta y_i\big\}\Big]^{\alpha^*/\zeta+1}}.$$

From equation (2.8), the conditional log-likelihood function for the generalized logistic candidate distribution is

$$\begin{aligned}\tilde{\ell}\big(\Theta;\mathbf{y},\log(\tau)\big) =& n\,\log(\alpha^*) + \frac{n\alpha^*}{\zeta}\log\big(\eta^\zeta + e^{\zeta\tau}\big) + \zeta\sum_{i=1}^{n} y_i \\ &- \Big(\frac{\alpha^*}{\zeta}+1\Big)\sum_{i=1}^{n}\log\big(\eta^\zeta + e^{\zeta y_i}\big).\end{aligned}$$

A.4 Weibull Distribution

The Weibull Distribution is one of three limiting distributions in Extreme Value Theory (EVT), see Coles [2001]. In EVT, the Weibull distribution is the GEV distribution with negative shape parameter $\xi < 0$. The Weibull distribution is similar to the shape of a lognormal distribution, but with a thinner tail. This distribution is widely used as the distribution of the lifetime of some object, particularly when the "weakest link" model is appropriate [Ross, 2010]. For parameter vector $\Theta = [\,a\ \theta\,]$, with shape parameter $a > 0$ and scale parameter $\theta > 0$, the CDF, PDF, and quantile function for the Weibull distribution are

$$F(x;\Theta) = \begin{cases} 1-\exp\Big\{-\big[\frac{x}{\theta}\big]^a\Big\} & \text{for } x>0, \\ 0 & \text{for } x\le 0; \end{cases}$$

$$f(x;\Theta) = \begin{cases} \frac{a}{\theta}\big[\frac{x}{\theta}\big]^{a-1}\exp\Big\{-\big[\frac{x}{\theta}\big]^a\Big\} & \text{for } x>0, \\ 0 & \text{for } x\le 0; \end{cases}$$

$$F^{-1}(p;\Theta) = \theta\big[-\log(1-p)\big]^{1/a} \qquad \text{for } p\in(0,1).$$

The conditional CDF and PDF for a Weibull distributed random variable with minimum threshold $\tau > 0$ are derived from equations (2.5) and (2.6), respectively. The quantile function is found by setting $\widetilde{F}(x;\Theta,\tau) = p$ and solving for x.

$$\widetilde{F}(x;\Theta,\tau) = \begin{cases} 1-\exp\Big\{\big(\frac{\tau}{\theta}\big)^a-\big(\frac{x}{\theta}\big)^a\Big\} & \text{for } x>\tau, \\ 0 & \text{for } x\le\tau; \end{cases}$$

$$\widetilde{f}(x;\Theta,\tau) = \begin{cases} \frac{a}{\theta}\big[\frac{x}{\theta}\big]^{a-1}\exp\Big\{\big(\frac{\tau}{\theta}\big)^a-\big(\frac{x}{\theta}\big)^a\Big\} & \text{for } x>\tau, \\ 0 & \text{for } x\le\tau; \end{cases}$$

$$\widetilde{F}^{-1}(p;\Theta,\tau) = \Big\{(1-p)^{-1/\alpha}\ \big[\theta^\gamma+\tau^\gamma\big]-\theta^\gamma\Big\}^{1/\gamma} \qquad \text{for } 0<p<1.$$

Using equation (2.7), the conditional likelihood function for the Weibull candidate distribution given a sample $\mathbf{x}\in\mathbb{R}^n$ is

$$\widetilde{L}(\Theta;\mathbf{x},\tau) = a^n\ \theta^{-na}\ \exp\Big\{n(\tau/\theta)^a\Big\}\ \prod_{i=1}^n x_i^{a-1}\ \exp\big\{-\big(x_i/\theta\big)^a\big\},$$

and from equation (2.8), the conditional log-likelihood function for the Weibull candidate distribution is

$$\tilde{\ell}(\Theta;\mathbf{x},\tau) = n\log(a)-na\log(\theta)+n\Big(\frac{\tau}{\theta}\Big)^a+(a-1)\sum_{i=1}^n\log(x_i)-\sum_{i=1}^n\Big(\frac{x_i}{\theta}\Big)^a.$$

A.5 Loglogistic Distribution

Like Weibull distribution, the loglogistic distribution is similar in shape to the lognormal distribution, but with a heavier right-tail than both the lognormal and Weibull. From [Panjer, 2006, p. 62], the loglogistic distribution has shape parameter $\gamma > 0$ and scale parameter $\theta > 0$. The CDF, PDF, and quantile function for the loglogistic distribution are

$$F(x;\Theta) = \begin{cases} \left[1+\left(x/\theta\right)^{-\gamma}\right]^{-1} & \text{for } x > 0, \\ 0 & \text{for } x \leq 0; \end{cases}$$

$$f(x;\Theta) = \begin{cases} \dfrac{\gamma\left(x/\theta\right)^{\gamma}}{x\left[1+\left(x/\theta\right)^{\gamma}\right]^{2}} & \text{for } x > 0, \\ 0 & \text{for } x \leq 0; \end{cases}$$

$$F^{-1}(p;\Theta) = \theta\left(\frac{1-p}{p}\right)^{-1/\gamma} \quad \text{for } 0 < p < 1.$$

The conditional CDF and PDF for a loglogistic distributed random variable with minimum threshold $\tau > 0$ are derived from equations (2.5) and (2.6), respectively. The quantile function is found by setting $\widetilde{F}(x;\Theta,\tau) = p$ and solving for x.

$$\widetilde{F}(x;\Theta,\tau) = \begin{cases} \left[x^{\gamma}-\tau^{\gamma}\right]/\left[\theta^{\gamma}+x^{\gamma}\right] & \text{for } x > \tau, \\ 0 & \text{for } x \leq \tau; \end{cases}$$

$$\widetilde{f}(x;\Theta,\tau) = \begin{cases} \gamma\, x^{\gamma-1}\ [\theta^{\gamma}+\tau^{\gamma}]/[\theta^{\gamma}+x^{\gamma}]^{2} & \text{for } x > \tau, \\ 0 & \text{for } x \leq \tau; \end{cases}$$

$$\widetilde{F}^{-1}(p;\Theta,\tau) = \left[\frac{p\theta^{\gamma}+\tau^{\gamma}}{1-p}\right]^{-1/\gamma} \quad \text{for } 0 < p < 1.$$

Using equation (2.7), the conditional likelihood function for the loglogistic

candidate distribution given a sample $\mathbf{x} \in \mathbb{R}^n$ is

$$\widetilde{L}(\Theta; \mathbf{x}, \tau) = \gamma^n \, [\theta^\gamma + \tau^\gamma]^n \prod_{i=1}^{n} \left\{ x_i^{\gamma-1} / [\theta^\gamma + x_i^\gamma]^2 \right\},$$

and from equation (2.8), the conditional log-likelihood function for the loglogistic candidate distribution is

$$\tilde{\ell}(\Theta; \mathbf{x}, \tau) = n\log(\gamma) + n\log[\theta^\gamma + \tau^\gamma] + (\gamma - 1)\textstyle\sum_{i=1}^{n} \log(x_i) - 2\sum_{i=1}^{n} \log[\theta^\gamma + x_i^\gamma].$$

A.6 *g*-and-*h* Distribution

The *g*-and-*h* distribution, as presented by Hoaglin [1985], is a four-parameter generalization of the lognormal distribution resulting from a transformation of a standard normal random variable. The transformational nature of the *g*-and-*h* distribution allows one to use the quantiles of sample data relative to the quantiles of the standard normal distribution to estimate the parameters. This method, called the *percentile method*, allows for the skewness and elongation parameters, g and h respectively, to be estimated as functions of the standard normal percentiles, $g(z_p)$ and $h(z_p)$ for percentile p, leading to an extremely flexible distribution. This flexibility enables the *g*-and-*h* distribution to model fatter or thinner tails and positive or negative skewness when compared to the lognormal distribution. Full details are available in Hoaglin [1985].

Let $Z \sim N(0,1)$. Then the random variable X^* such that

$$X^* = A_{g,h}(Z) = \left(\frac{e^{gZ} - 1}{g} \right) e^{hZ^2/2},$$

is said to have a standard *g*-and-*h* distribution with skewness parameter g and elongation parameter h. We refer to X^* as "standard" since the *g*-and-*h* location and scale parameters are 0 and 1, respectively. When $g = 0$, there is no skewness. This can be seen by writing the Taylor expansion of e^{gz},

$$\begin{aligned} e^{gz} &= 1 + gz + \frac{(gz)^2}{2!} + \frac{(gz)^3}{3!} + \ldots \\ \frac{e^{gz} - 1}{g} &= z + g\frac{z^2}{2!} + g^2\frac{z^3}{3!} + \ldots, \end{aligned}$$

and letting $g = 0$ in the right-hand side of the last equation. To further generalize, let $a \in \mathbb{R}$ be a location parameter and $b > 0$ be a scale parameter. Then random variable X, where

$$X = a + b \cdot A_{g,h}(Z) = a + b \left(\frac{e^{gZ} - 1}{g} \right) e^{hZ^2/2},$$

is said to have a g-and-h distribution.

For our purposes, estimating parameters using the percentile method has major drawbacks. First, the percentile method allows the elongation and skewness parameters to change unsystematically, which can cause $A_{g,h}(Z)$ to be non-monotonic in Z with potentially multiple turning points. Also, h may take on negative values which also causes non-monotonicity in $A_{g,h}(Z)$. While non-monotonicity is not a problem, per se, multiple turning points cause the g-and-h PDF to become unwieldy when calculating likelihoods for MLE, AIC, and BIC. More importantly, however, non-monotonicity in $A_{g,h}(Z)$ may lead to undefined regions, where some observations from the sample do not have a defined inverse. When this occurs, there is no way to calculate a likelihood for these observations. The undefined regions tend to include the largest observations, which we are most concerned with when modeling OR. Thus, the observations with the highest impact on operational losses are the ones that we are least able to model.

The second drawback of the percentile method occurs as the number of statistically significant parameters increase. When g and h are allowed to be functions, the percentile method can easily result in 5 or more significant parameters, leading to the phenomenon known as *overfitting* the data [James et al., 2013]. As a result, the model suffers prediction accuracy which is the goal of the AMA. Also, while $g(z_p)$ and $h(z_p)$ are assumed to be polynomials, there is no clear consensus on how to estimate the degree of the polynomials and their coefficients.

Finally, the selection of which percentiles to use when estimating parameters g and h by the percentile method is subjective and can greatly affect the estimates. There is little research done in picking the optimal percentiles for various scenarios. Hoaglin suggests using the "letter values", the $\frac{3}{4}, \frac{7}{8}, \frac{15}{16}, \frac{31}{32}, \frac{63}{64}, \frac{127}{128}, \frac{255}{256}$ percentiles. This allows the upper tail of the distribution to be measured with some precision relative to other areas of the distribution. Once the percentiles have been selected, the second drawback still exists to estimate $g(z_p)$ and $h(z_p)$. Hoaglin assumes a

linear form, but selects the model visually. For these reasons, the percentile method is better used as an exploratory data analysis technique to gauge the systematic and unsystematic properties of the sample data than as an estimation procedure.

Given the drawbacks associated with the percentile method and to compare the estimated g-and-h distribution to the other candidate distributions, the maximum likelihood approach is used when estimating the g-and-h distribution for loss severity. The maximum likelihood approach requires that $A_{g,h}(Z)$ be monotonic, so we restrict both g and h to be constant, with $h > 0$. As shown by Degen et al. [2007], this restriction forces the g-and-h distribution to have regularly varying tails with index $-1/h$.

The skewness parameter, g, signifies both the direction and magnitude of skewness. Positive g signifies positive skewness and larger g signifies more skewness. When $g = 0$ and $h > 0$, the distribution is symmetric with fatter tails than the normal distribution. If $g = 1$ and $h = 0$, we have a shifted lognormal distribution. In the context of positively skewed distributions, the lognormal distribution is assumed to have "neutral elongation". As a result, restricting the elongation parameter to positive values is not unreasonable when the tail is fatter than lognormal. The limitation of a constant g may not be optimal, but is the same restriction imposed on the other candidate distributions used to model operational loss severities.

Let $A_{g,h}^{-1}\left(\frac{X-a}{b}\right)$ be the inverse standard g-and-h transformation which does not have an analytical form, and let $A_{g,h}^{'}(Z)$ be the derivative of the standard g-and-h transformation. Also, let $\Phi(z)$ and $\phi(z)$ be the standard normal CDF and PDF, respectively. Let $\Phi^{-1}(p)$ be the standard normal quantile function for $0 < p < 1$. Then, for parameter vector $\Theta = \left[\begin{array}{cccc} a & b & g & h \end{array}\right]$ with $a, b, g, h \in \mathbb{R}$ and $b, h > 0$,

$$F(x;\Theta) = \Phi\left[A_{g,h}^{-1}\left(\frac{x-a}{b}\right)\right] \quad \text{for } x \in \mathbb{R};$$

$$f(x;\Theta) = \frac{\phi\left[A_{g,h}^{-1}\left(\frac{x-a}{b}\right)\right]}{b\, A_{g,h}^{'}\left[A_{g,h}^{-1}\left(\frac{x-a}{b}\right)\right]} \qquad \text{for } x \in \mathbb{R};$$

$$F^{-1}(p;\Theta) = a + b\, A_{g,h}\left(\Phi^{-1}(p)\right) \qquad \text{for } 0 < p < 1.$$

The g-and-h distribution has support on the real number line, so the g-and-h loss severity distribution can lead to negative losses when performing simulations. To avoid this, simulations are performed from the conditional g-and-h distribution with minimum threshold $\tau = 0$. Using the conditional distribution for simulations is reasonable when the estimated truncation probability is sufficiently small. See Section 2.1.4 for details.

The conditional CDF and PDF for a g-and-h distributed random variable with minimum threshold τ are derived from equations (2.5) and (2.6), respectively. The quantile function is found by setting $\widetilde{F}(x;\Theta,\tau) = p$ and solving for x.

$$\widetilde{F}(x;\Theta,\tau) = \begin{cases} \frac{\Phi\left[A_{g,h}^{-1}\left((x-a)/b\right)\right] - \Phi\left[A_{g,h}^{-1}\left((\tau-a)/b\right)\right]}{1-\Phi\left[A_{g,h}^{-1}\left((\tau-a)/b\right)\right]} & \text{for } x > \tau, \\ 0 & \text{for } x \leq \tau; \end{cases}$$

$$\widetilde{f}(x;\Theta,\tau) = \begin{cases} \frac{1}{1-\Phi\left[A_{g,h}^{-1}\left((\tau-a)/b\right)\right]} \frac{\phi\left[A_{g,h}^{-1}\left((x-a)/b\right)\right]}{b\, A_{g,h}'\left[A_{g,h}^{-1}\left((x-a)/b\right)\right]} & \text{for } x > \tau, \\ 0 & \text{for } x \leq \tau; \end{cases}$$

$$\widetilde{F}^{-1}(p;\Theta,\tau) = a + bA_{g,h}\left\{\Phi^{-1}\left[(1-p)\,\Phi\left[A_{g,h}^{-1}\left(\frac{\tau-a}{b}\right)\right] + p\right]\right\} \quad \text{for } 0 < p < 1.$$

The `R` function `pchip()` from the `signal` package is used to numerically derive the inverse function, $A_{g,h}^{-1}\left(\frac{x-a}{b}\right)$. The `pchip()` function performs piecewise cubic Hermite (monotone) interpolation.

Using equation (2.7), the conditional likelihood function for the single severity g-and-h candidate distribution given a sample $\mathbf{x} \in \mathbb{R}^n$ is

$$\widetilde{L}(\Theta;\mathbf{x},\tau) = b^{-n}\left[1-\Phi\left(A_{g,h}^{-1}\left((\tau-a)/b\right)\right)\right]^{-n} \prod_{i=1}^{n} \frac{\phi\left[A_{g,h}^{-1}\left((x_i-a)/b\right)\right]}{A_{g,h}'\left[A_{g,h}^{-1}\left((x_i-a)/b\right)\right]},$$

and from equation (2.8), the conditional log-likelihood function for the single

severity g-and-h candidate distribution is

$$\tilde{\ell}(\Theta;\mathbf{x},\tau) = -n\ \log(b) - n\ \log\left\{1-\Phi\left[A_{g,h}^{-1}\big((\tau-a)/b\big)\right]\right\} + \sum_{i=1}^{n}\log\left\{\phi\left[A_{g,h}^{-1}\big((x_i-a)/b\big)\right]\right\} - \sum_{i=1}^{n}\log\left\{A_{g,h}^{'}\left[A_{g,h}^{-1}\big((x_i-a)/b\big)\right]\right\}.$$

A.7 log-SaS Distribution

The *Sinh-arcSinh Distribution* (SAS) distribution introduced by Jones and Pewsey [2009] is a four parameter distribution resulting from a transformation of a standard normal random variable, similar to the g-and-h distribution. The SAS distribution uses a monotonic transformation with an analytical inverse giving it two major advantages over g-and-h when calculating likelihoods.

Let $Z \sim N(0,1)$ and $\varepsilon, \delta \in \mathbb{R}$ with $\delta > 0$. Also, let $\sinh(\cdot)$ be the hyperbolic sine function and $\sinh^{-1}(\cdot)$ be the inverse hyperbolic sine function (arcsinh). Then the random variable Y^*, such that

$$Y^* = A_{\varepsilon,\delta}(Z) = \sinh\left\{\frac{\sinh^{-1}(Z)+\varepsilon}{\delta}\right\},$$

is said to have a standard SaS distribution with skewness parameter ε and tailweight parameter δ. We consider Y^* a "standard" SaS random variable since it assumes a location parameter of 0 and scale parameter of 1. To further generalize, let $a, b \in \mathbb{R}$ be the location and scale parameters, respectively, with $b > 0$. Then,

$$Y = a + b\cdot A_{\varepsilon,\delta}(Z) = a + b\ \sinh\left\{\frac{\sinh^{-1}(Z)+\varepsilon}{\delta}\right\},$$

where $Y \sim \mathrm{SaS}(a,b,\varepsilon,\delta)$.

We note that the generalized SAS transformation, $a + b\,A_{\varepsilon,\delta}(\cdot)$, has a closed-form inverse given by

$$A_{\varepsilon,\delta}^{-1}\left(\frac{y-a}{b}\right) = \sinh\left[\delta\cdot\sinh^{-1}\left(\frac{y-a}{b}\right) - \varepsilon\right].$$

To calculate the density function of a SAS random variable, we need the derivative

of the inverse given by

$$\frac{d}{dy}A_{\varepsilon,\delta}^{-1}\left(\frac{y-a}{b}\right) = \cosh\left[\delta \cdot \sinh^{-1}\left(\frac{y-a}{b}\right) - \varepsilon\right] \frac{\delta}{b\sqrt{(\frac{y-a}{b})^2+1}} dy,$$

where $\cosh(\cdot)$ is the hyperbolic cosine function.

As pointed out by Jones and Pewsey [2009], ε gives both the magnitude and direction of skewness. A positively skewed distribution will have parameter value $\varepsilon > 0$, with larger ε indicating more skewness. The tailweight parameter, δ, has a negative relation to the thickness of the tails. As the value of δ approaches zero from the right, the tail behavior becomes heavier. For example, distributions with heavier tails than the normal distribution have a tailweight parameter $0 < \delta < 1$. A tailweight parameter greater than 1 indicates thinner tails than the normal distribution.

Let $\Phi(\cdot)$ represent the standard normal CDF and $\Phi^{-1}(\cdot)$ be the standard normal quantile function. Then, for parameter vector $\Theta = [\,a\ b\ \varepsilon\ \delta\,]$ with $a, b, \varepsilon, \delta \in \mathbb{R}$ and $b, \delta > 0$,

$$F_Y(y;\Theta) = \Phi\left[A_{\varepsilon,\delta}^{-1}\left(\frac{y-a}{b}\right)\right] \quad \text{for } y \in \mathbb{R};$$

$$f_Y(y;\Theta) = \phi\left[A_{\varepsilon,\delta}^{-1}\left(\frac{y-a}{b}\right)\right] \left(\frac{d}{dy}A_{\varepsilon,\delta}^{-1}\left(\frac{y-a}{b}\right)\right) \quad \text{for } y \in \mathbb{R};$$

$$F_Y^{-1}(p;\Theta) = a + b\, A_{\varepsilon,\delta}\left(\Phi^{-1}(p)\right) \quad \text{for } 0 < p < 1.$$

The conditional CDF and PDF for a SaS distributed random variable with minimum threshold τ^* are derived from equations (2.5) and (2.6), respectively. The quantile function is found by setting $\widetilde{F}_Y(y;\Theta,\tau^*) = p$ and solving for y.

$$\widetilde{F}_Y(y;\Theta,\tau^*) = \begin{cases} \frac{\Phi\left[A_{\varepsilon,\delta}^{-1}\left((y-a)/b\right)\right] - \Phi\left[A_{\varepsilon,\delta}^{-1}\left((\tau^*-a)/b\right)\right]}{1-\Phi\left[A_{\varepsilon,\delta}^{-1}\left((\tau^*-a)/b\right)\right]} & \text{for } y > \tau^*, \\ 0 & \text{for } y \le \tau^*; \end{cases}$$

$$\widetilde{f}_y(x;\Theta,\tau^*) = \begin{cases} \frac{1}{1-\Phi\left[A_{\varepsilon,\delta}^{-1}\left(\frac{\tau^*-a}{b}\right)\right]}\ \phi\left[A_{\varepsilon,\delta}^{-1}\left(\frac{y-a}{b}\right)\right]\ \left(\frac{d}{dy}A_{\varepsilon,\delta}^{-1}\left(\frac{y-a}{b}\right)\right) & \text{for } y > \tau^*, \\ 0 & \text{for } y \leq \tau^*; \end{cases}$$

$$\widetilde{F}_Y^{-1}(p;\Theta,\tau^*) = a + bA_{\varepsilon,\delta}\left\{\Phi^{-1}\left[(1-p)\ \Phi\left[A_{\varepsilon,\delta}^{-1}\left(\frac{\tau^*-a}{b}\right)\right] + p\right]\right\},$$

for $0 < p < 1$.

Using equation (2.7), the conditional likelihood function for the SaS candidate distribution is

$$\widetilde{L}(\Theta;\mathbf{x},\tau^*) = \frac{1}{\left\{1-\Phi\left[A_{\varepsilon,\delta}^{-1}\left(\frac{\tau^*-a}{b}\right)\right]\right\}^n}\ \prod_{i=1}^{n}\left\{\phi\left[A_{\varepsilon,\delta}^{-1}\left(\frac{y_i-a}{b}\right)\right]\frac{d}{dy}A_{\varepsilon,\delta}^{-1}\left(\frac{y_i-a}{b}\right)\right\}.$$

From equation (2.8), the conditional log-likelihood function for the single severity SaS candidate distribution is

$$\begin{aligned}\tilde{\ell}(\Theta;\mathbf{x},\tau^*) = {} & n\log(\delta) - n\log(b) - n\log\left\{1-\Phi\left[A_{\varepsilon,\delta}^{-1}\left(\frac{\tau^*-a}{b}\right)\right]\right\} \\ & + \sum_{i=1}^{n}\log\left\{\cosh\left[\delta\cdot\sinh^{-1}\left(\frac{y_i-a}{b}\right)-\varepsilon\right]\right\} \\ & + \sum_{i=1}^{n}\log\left\{\phi\left[A_{\varepsilon,\delta}^{-1}\left(\frac{y_i-a}{b}\right)\right]\right\} - \frac{1}{2}\sum_{i=1}^{n}\log\left\{\left(\frac{y_i-a}{b}\right)^2+1\right\}.\end{aligned}$$

When attempting to fit the SAS distribution to actual operational loss data, we find that MLE suffers from the badly-scaled problem from Section 2.1.4. To alleviate this issue, we treat losses as *log sinh-arcsinh Distribution* (LSAS) random variables.

If $Y \sim \text{SaS}(a,b,\varepsilon,\delta)$, then $X = e^Y$ has a LSAS distribution with parameter vector $\Theta = [\,a\ b\ \varepsilon\ \delta\,]$, denoted by $X \sim \text{lsas}(a,b,\varepsilon,\delta)$. Using the log-SaS distribution solves another problem of the *g*-and-*h* distribution since the support for the log-SaS distribution is only the positive real numbers. Thus, the log-SaS distribution solves the problems associated with the *g*-and-*h* distribution without sacrificing the flexibility of a four parameter distribution and proves to be an incredibly useful

distribution for modeling loss severities.

Let $X = e^Y$. We can find the CDF and PDF of X using the formulas

$$F(x;\Theta) = F_Y\big(\log(x);\Theta\big);$$
$$f(x;\Theta) = \frac{1}{x}\, f_Y\big(\log(x);\Theta\big).$$

Then, the CDF, PDF, and quantile function for the LSAS distribution are

$$F(x;\Theta) = \begin{cases} \Phi\left[A_{\varepsilon,\delta}^{-1}\left(\frac{\log(x)-a}{b}\right)\right] & \text{for } x > 0, \\ 0 & \text{for } x \le 0; \end{cases}$$

$$f(x;\Theta) = \begin{cases} \frac{1}{x}\,\phi\left[A_{\varepsilon,\delta}^{-1}\left(\frac{\log(x)-a}{b}\right)\right]\left(\frac{d}{dx}A_{\varepsilon,\delta}^{-1}\left(\frac{\log(x)-a}{b}\right)\right) & \text{for } x > 0, \\ 0 & \text{for } x \le 0; \end{cases}$$

$$F^{-1}(p;\Theta) = \exp\left\{a + b\, A_{\varepsilon,\delta}\Big(\Phi^{-1}(p)\Big)\right\} \qquad \text{for } 0 < p < 1.$$

To derive the conditional CDF, PDF, and quantile function, let the loss data have minimum reporting threshold $\tau = \exp\{\tau^*\}$. The functions are written in terms of the functions for $Y = \log(X)$ for brevity.

$$\widetilde{F}(x;\Theta,\tau) = \widetilde{F}_Y\big(\log(x);\Theta,\tau^*\big) \qquad \text{for } x > 0;$$
$$\widetilde{f}(x;\Theta,\tau) = \widetilde{f}_y\big(\log(x);\Theta,\tau^*\big) \qquad \text{for } x > 0;$$
$$\widetilde{F}^{-1}(p;\Theta) = \exp\left\{\widetilde{F}_Y^{-1}(p;\Theta,\tau^*)\right\} \qquad \text{for } 0 < p < 1.$$

Since we use the log-loss data to estimate the distribution parameters, the conditional likelihood and log-likelihood function for the SAS distribution are used on the log transform of the log-loss data.

A.8 Lognormal Spliced Lognormal Distribution

Using the derivation from Section 2.1.2, we can calculate the unconditional CDF, PDF, and quantile function for the piecewise LGNLGN distribution. For ease of notation, let

$$D_1(\Theta, \tau) = F_{body}(x_s; \Theta_b) - (1 - p_b) F_{body}(\tau; \Theta_b);$$

$$D_2(\Theta, \tau) = \frac{1 - p_b}{D_1(\Theta, \tau)} \frac{F_{body}(x_s; \Theta_b) - F_{body}(\tau; \Theta_b)}{1 - F_{tail}(x_s; \Theta_u)},$$

where

$$F_{body}(x; \Theta_b) = \Phi\left[\frac{\log(x) - \mu_b}{\sigma_b}\right]; \qquad F_{tail}(x; \Theta_u) = \Phi\left[\frac{\log(x) - \mu_u}{\sigma_u}\right],$$

and $\Theta = \begin{bmatrix} \Theta_b & x_s & \Theta_u \end{bmatrix}$, $\Theta_b = \begin{bmatrix} \mu_b & \sigma_b \end{bmatrix}$ and $\Theta_u = \begin{bmatrix} \mu_u & \sigma_u \end{bmatrix}$. Also, let

$$f_{body}(x; \Theta_b) = \frac{1}{\sigma_b x} \phi\left[\frac{\log(x) - \mu_b}{\sigma_b}\right]; \quad f_{tail}(x; \Theta_u) = \frac{1}{\sigma_u x} \phi\left[\frac{\log(x) - \mu_u}{\sigma_u}\right],$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the standard normal CDF and PDF, respectively.

Then, the CDF, PDF, and quantile function for the LGNLGN distribution are

$$F(x; \Theta) = \begin{cases} \dfrac{p_b}{D_1(\Theta, \tau)} F_{body}(x; \Theta_b) & \text{for } 0 < x \leq x_s, \\ \dfrac{p_b}{D_1(\Theta, \tau)} F_{body}(x_s; \Theta_b) & \\ \quad + D_2(\Theta, \tau)\left[F_{tail}(x; \Theta_u) - F_{tail}(x_s; \Theta_u)\right] & \text{for } x > x_s, \\ 0 & \text{otherwise;} \end{cases}$$

$$f(x;\Theta) = \begin{cases} \dfrac{p_b}{D_1(\Theta,\tau)} f_{body}(x;\Theta_b) & \text{for } 0 < x \le x_s, \\ D_2(\Theta,\tau)\ f_{tail}(x;\Theta_u) & \text{for } x > x_s, \\ 0 & \text{otherwise;} \end{cases}$$

$$F^{-1}(p;\Theta) = \begin{cases} \exp\left\{ \mu_b + \sigma_b\ \Phi^{-1}\left[\dfrac{p}{p_b}\ D_1(\Theta,\tau) \right] \right\} & \text{for } 0 < p \le p_s, \\ \exp\left\{ \mu_u + \sigma_u\ \Phi^{-1}\left[F_{tail}(x_s;\Theta_u) + \dfrac{p - p_s}{D_2(\Theta,\tau)} \right] \right\} & \text{for } p_s < p < 1, \end{cases}$$

where $p_s = \dfrac{p_b}{D_1(\Theta,\tau)}\ F_{body}(x_s;\Theta_b)$, and $\Phi^{-1}(\cdot)$ is the standard normal quantile function.

The conditional CDF, PDF, and quantile function follow directly from the derivation in Section 2.1.2.

$$\widetilde{F}(x;\Theta,\tau) = \begin{cases} p_b\ \dfrac{F_{body}(x;\Theta_b) - F_{body}(\tau;\Theta_b)}{F_{body}(x_s;\Theta_b) - F_{body}(\tau;\Theta_b)} & \text{for } \tau < x \le x_s, \\ p_b + (1 - p_b)\ \dfrac{F_{tail}(x;\Theta_u) - F_{tail}(x_s;\Theta_u)}{1 - F_{tail}(x_s;\Theta_u)} & \text{for } x > x_s, \\ 0 & \text{for } x \le \tau; \end{cases}$$

$$\widetilde{f}(x;\Theta) = \begin{cases} \dfrac{p_b\, f_{body}(x;\Theta_b)}{F_{body}(x_s;\Theta_b) - F_{body}(\tau;\Theta_b)} & \text{for } \tau < x \leq x_s, \\ (1-p_b)\, \dfrac{f_{tail}(x;\Theta_u)}{1-F_{tail}(x_s;\Theta_u)} & \text{for } x > x_s, \\ 0 & \text{otherwise;} \end{cases}$$

$$\widetilde{F}^{-1}(p;\Theta) = \begin{cases} \exp\left\{ \mu_b + \sigma_b\, \Phi^{-1}\left[\dfrac{p}{p_b} F_{body}(x_s;\Theta_b) \right.\right. \\ \qquad \left.\left. + \dfrac{p_b - p}{p_b} F_{body}(\tau;\Theta_b) \right] \right\} & \text{for } 0 < p \leq p_b, \\ \exp\left\{ \mu_u + \sigma_u\, \Phi^{-1}\left[F_{tail}(x_s;\Theta_u) + \dfrac{p - p_b}{1-p_b} \right.\right. \\ \qquad \left.\left. - \dfrac{p - p_b}{1-p_b} F_{tail}(x_s;\Theta_u) \right] \right\} & \text{for } p_b < p < 1. \end{cases}$$

For the conditional log-likelihood functions of the body and tail, let n_b be the number of observations in the body of the sample and n_u be the number of the observations in the tail of the sample. The conditional log-likelihood for the body is

$$\tilde{\ell}_b(\Theta_b, \mathbf{x}_b, \tau, x_s) = n_b \log(p_b) - n_b \log(\sigma_b) - \sum_{i=1}^{n_b} x_i + \sum_{i=1}^{n_b} \log\left\{ \phi\left(\frac{\log(x_i) - \mu_b}{\sigma_b} \right) \right\} - n_b \log\left\{ \Phi\left(\frac{\log(x_s) - \mu_b}{\sigma_b} \right) - \Phi\left(\frac{\log(\tau) - \mu_b}{\sigma_b} \right) \right\}.$$

Likewise, the conditional log-likelihood function for the upper tail is

$$\tilde{\ell}_u(\Theta_u, \mathbf{x}_u, x_s) = n_u \log\{1-p_b\} - n_u \log(\sigma_u) - \sum_{i=1}^{n_u} x_i + \sum_{i=1}^{n_u} \log\left\{\phi\left(\frac{\log(x_i)-\mu_u}{\sigma_u}\right)\right\} - n_u \log\left\{1-\Phi\left(\frac{\log(x_s)-\mu_u}{\sigma_u}\right)\right\}.$$

A.9 Lognormal Spliced Generalized Pareto Distribution

Using the derivation from Section 2.1.2, we can calculate the unconditional CDF, PDF, and quantile function for the piecewise lgn/gpd distribution. For ease of notation, let

$$D_1(\Theta, \tau) = F_{body}(x_s; \Theta_b) - (1-p_b) F_{body}(\tau; \Theta_b);$$

$$D_2(\Theta, \tau) = \frac{1-p_b}{D_1(\Theta, \tau)} \frac{F_{body}(x_s; \Theta_b) - F_{body}(\tau; \Theta_b)}{1 - F_{tail}(x_s; \Theta_u)},$$

where

$$F_{body}(x; \Theta_b) = \Phi\left(\frac{\log(x)-\mu_b}{\sigma_b}\right);$$

$$F_{tail}(x; \xi, \hat{\theta}, x_s) = 1 - \left(1 + \frac{\xi}{\hat{\theta}}(x - x_s)\right)^{-1/\xi},$$

and $\Theta = [\,\Theta_b \;\; x_s \;\; \xi\,]$, $\Theta_b = [\,\mu_b \;\; \sigma_b\,]$, and $\hat{\theta} = \frac{1-p_b}{p_b} \frac{F_{body}(x_s;\Theta_b) - F_{body}(\tau;\Theta_b)}{f_{body}(x_s;\Theta_b)}$. Also, let

$$f_{body}(x; \Theta_b) = \frac{1}{\sigma_b x} \phi\left[\frac{\log(x)-\mu_b}{\sigma_b}\right],$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the standard normal CDF and PDF, respectively.

Then, the CDF, PDF, and quantile function for the LGNGPD distribution are

$$F(x;\Theta) = \begin{cases} \dfrac{p_b}{D_1(\Theta_b,\tau)} F_{body}(x,\Theta_b) & \text{for } 0 < x \le x_s, \\ \dfrac{p_b}{D_1(\Theta_b,\tau)} F_{body}(x_s,\Theta_b) & \\ \quad + D_2(\Theta,\tau)\left[1 - \left\{1 + \frac{\xi}{\hat{\theta}}(x - x_s)\right\}\right] & \text{for } x > x_s, \\ 0 & \text{otherwise;} \end{cases}$$

$$f(x;\Theta) = \begin{cases} \dfrac{p_b}{D_1(\Theta,\tau)} f_{body}(x,\Theta_b) & \text{for } 0 < x \le x_s, \\ \dfrac{D_2(\Theta,\tau)}{\hat{\theta}} \left(1 + \dfrac{\xi}{\hat{\theta}}(x - x_s)\right)^{-1-1/\xi} & \text{for } x > x_s, \\ 0 & \text{otherwise;} \end{cases}$$

$$F^{-1}(p;\Theta) = \begin{cases} \exp\left\{\mu_b + \sigma_b\, \Phi^{-1}\left[\dfrac{D_1(\Theta,\tau)}{p_b}\, p\right]\right\} & \text{for } 0 < p \le p_s, \\ x_s + \dfrac{\hat{\theta}}{\xi}\left[\left(1 - \dfrac{p - p_s}{D_2(\Theta,\tau)}\right)^{-\xi} - 1\right] & \text{for } p_s < p < 1, \end{cases}$$

where $p_s = \dfrac{p_b}{D_1(\Theta,\tau)}\, \Phi\left[\frac{\log(x_s) - \mu_b}{\sigma_b}\right]$.

The conditional CDF, PDF, and quantile function follow directly from the

derivation in Section 2.1.2.

$$\widetilde{F}(x;\Theta,\tau) = \begin{cases} p_b \, \dfrac{F_{body}(x,\Theta_b) - F_{body}(\tau,\Theta_b)}{F_{body}(x_s,\Theta_b) - F_{body}(\tau,\Theta_b)} & \text{for } \tau < x \leq x_s, \\ p_b + (1-p_b)\left[1 - \left(1 + \dfrac{\xi}{\hat{\theta}}(x - x_s)\right)^{-1/\xi}\right] & \text{for } x > x_s, \\ 0 & \text{for } x \leq \tau; \end{cases}$$

$$\widetilde{f}(x;\Theta) = \begin{cases} \dfrac{p_b \, F_{body}(x,\Theta_b)}{F_{body}(x_s,\Theta_b) - F_{body}(\tau,\Theta_b)} & \text{for } \tau < x \leq x_s, \\ \dfrac{1-p_b}{\hat{\theta}}\left(1 + \dfrac{\xi}{\hat{\theta}}(x - x_s)\right)^{-1-1/\xi} & \text{for } x > x_s, \\ 0 & \text{otherwise;} \end{cases}$$

$$\widetilde{F}^{-1}(p;\Theta) = \begin{cases} \exp\left\{\mu_b + \sigma_b \, \Phi^{-1}\left[\dfrac{p}{p_b} F_{body}(x_s;\Theta_b) \right.\right. \\ \qquad \left.\left. + \dfrac{p_b - p}{p_b} F_{body}(\tau;\Theta_b)\right]\right\} & \text{for } 0 < p \leq p_b, \\ x_s + \dfrac{\hat{\theta}}{\xi}\left[\left(\dfrac{1-p_b}{1-p}\right)^{\xi} - 1\right] & \text{for } p_b < p < 1. \end{cases}$$

For the conditional log-likelihood functions of the body and tail, let n_b be the number of observations in the body of the sample and n_u be the number of the observations in the tail of the sample. The conditional log-likelihood for the body

is

$$\tilde{\ell}_b(\Theta_b, \mathbf{x}_b, \tau, x_s) = n_b \log(p_b) - n_b \log(\sigma_b) - \sum_{i=1}^{n_b} x_i + \sum_{i=1}^{n_b} \log\left\{\phi\left(\frac{\log(x_i) - \mu_b}{\sigma_b}\right)\right\}$$
$$- n_b \log\left\{\Phi\left(\frac{\log(x_s) - \mu_b}{\sigma_b}\right) - \Phi\left(\frac{\log(\tau) - \mu_b}{\sigma_b}\right)\right\}.$$

Likewise, the conditional log-likelihood function for the upper tail is

$$\tilde{\ell}_u(\xi; \mathbf{x}_u \hat{\theta}, x_s) = n_u \log\{1 - p_b\} - n_u \log(\hat{\theta}) - \left(\frac{\xi + 1}{\xi}\right) \sum_{i=1}^{n_u} \log\left(1 + \frac{\xi}{\theta}(x_i - x_s)\right).$$

Appendix B

MLE Results

B.1 MLE Results for SRC 1

Lognormal MLE Parameters

	μ	σ
2014	-7.99	3.29
2015	-8.17	3.31
2016	-8.56	3.38
2017	-8.5	3.36
2018	-8.46	3.35

GPD MLE Parameters

	ξ	θ
2014	0.87	0.48
2015	0.87	0.48
2016	0.88	0.45
2017	0.87	0.46
2018	0.87	0.47

Burr MLE Parameters

	α	γ	θ
2014	0.04	22.66	1.24
2015	0.05	20.62	1.25
2016	0.05	21.1	1.24
2017	0.06	16.95	1.24
2018	0.06	17.27	1.24

Weibull MLE Parameters

	a	θ
2014	0.19	0
2015	0.19	0
2016	0.19	0
2017	0.19	0
2018	0.19	0

Loglogistic MLE Parameters

	γ	θ
2014	1.15	0.65
2015	1.15	0.66
2016	1.14	0.61
2017	1.15	0.62
2018	1.16	0.64

g-and-h MLE Parameters

	a	b	g	h
2014	2.34	1.82	1.74	0.15
2015	2.32	1.78	1.73	0.16
2016	2.28	1.75	1.72	0.19
2017	2.21	1.67	1.65	0.23
2018	2.25	1.69	1.68	0.21

log-SaS MLE Parameters

	a	b	ε	δ
2014	0.39	0.37	0.8	0.76
2015	0.4	0.36	0.78	0.75
2016	0.4	0.36	0.75	0.74
2017	0.43	0.39	0.63	0.72
2018	0.42	0.38	0.68	0.73

Lgn/Lgn MLE Parameters

	μ_b	σ_b	x_s	μ_t	σ_t	p_b
2014	0.51	0.31	2.24	-7.39	3.18	0.44
2015	0.44	0.22	1.86	-7.51	3.16	0.32
2016	0.43	0.21	1.86	-7.54	3.17	0.32
2017	0.49	0.31	2.24	-7.33	3.16	0.44
2018	0.44	0.22	1.85	-7.47	3.14	0.32

Lgn/Gpd MLE Parameters

	μ_b	σ_b	x_s	ξ	θ	p_b
2014	0.49	0.28	2.03	0.96	2.15	0.38
2015	0.44	0.22	1.81	0.94	1.93	0.3
2016	0.43	0.21	1.81	0.94	2	0.3
2017	0.44	0.23	1.8	0.94	1.91	0.3
2018	0.44	0.22	1.8	0.94	1.9	0.3

Figure B.1: SRC 1 MLE parameters for each candidate distribution when including all loss data before each row's designated year. The last row uses all simulated data. Lgn/Lgn and Lgn/Gpd refer to LGNLGN and LGNGPD, respectively.

B.2 MLE Results for SRC 2

Lognormal MLE Parameters

	μ	σ
2014	-3.38	2.89
2015	-2.91	2.8
2016	-3.63	2.99
2017	-3.84	3.04
2018	-4.89	3.26

GPD MLE Parameters

	ξ	θ
2014	1.02	1.66
2015	1.01	1.81
2016	1.05	1.67
2017	1.05	1.62
2018	1.08	1.46

Burr MLE Parameters

	α	γ	θ
2014	0.09	9.81	3.24
2015	0.11	7.98	3.22
2016	0.11	7.98	3.15
2017	0.08	9.82	3.2
2018	0.1	8.18	3.09

Weibull MLE Parameters

	a	θ
2014	0.17	0
2015	0.17	0
2016	0.17	0
2017	0.17	0
2018	0.17	0

Loglogistic MLE Parameters

	γ	θ
2014	0.99	1.62
2015	0.99	1.78
2016	0.96	1.55
2017	0.95	1.49
2018	0.93	1.29

g-and-h MLE Parameters

	a	b	g	h
2014	7.2	7.05	1.87	0.1
2015	7.19	7.06	1.84	0.12
2016	7.11	7.14	1.88	0.13
2017	7.25	7.35	1.9	0.11
2018	6.86	6.9	1.9	0.16

log-SaS MLE Parameters

	a	b	ε	δ
2014	1.22	0.47	1.18	0.92
2015	1.25	0.5	1.07	0.91
2016	1.24	0.51	1.05	0.9
2017	1.21	0.49	1.16	0.91
2018	1.26	0.52	0.96	0.87

Lgn/Lgn MLE Parameters

	μ_b	σ_b	x_s	μ_t	σ_t	p_b
2014	1.46	0.28	5.56	-4.82	3.17	0.34
2015	1.47	0.31	5.85	-4.38	3.1	0.36
2016	1.46	0.32	5.9	-5.39	3.32	0.36
2017	1.45	0.29	5.91	-5.84	3.4	0.36
2018	1.45	0.34	5.83	-6.04	3.47	0.36

Lgn/Gpd MLE Parameters

	μ_b	σ_b	x_s	ξ	θ	p_b
2014	1.47	0.31	5.75	1.02	7.82	0.36
2015	1.53	0.51	7.8	0.97	10.51	0.5
2016	1.46	0.34	5.62	1.06	7.54	0.34
2017	1.45	0.31	5.61	1.04	8	0.34
2018	1.43	0.32	5.4	1.07	7.56	0.32

Figure B.2: SRC 2 MLE parameters for each candidate distribution when including all loss data before each row's designated year. The last row uses all simulated data. Lgn/Lgn and Lgn/Gpd refer to LGNLGN and LGNGPD, respectively.

B.3 MLE Results for SRC 3

Lognormal MLE Parameters

	μ	σ
2014	-7.25	2.86
2015	-7.13	2.82
2016	-7.19	2.84
2017	-7.3	2.88
2018	-7.12	2.81

GPD MLE Parameters

	ξ	θ
2014	0.88	0
2015	0.87	0
2016	0.88	0
2017	0.88	0
2018	0.88	0

Burr MLE Parameters

	α	γ	θ
2014	0.05	25.45	1.18
2015	0.04	26.99	1.17
2016	0.05	25.72	1.17
2017	0.05	24.99	1.17
2018	0.05	23.75	1.17

Weibull MLE Parameters

	a	θ
2014	0.21	0
2015	0.21	0
2016	0.21	0
2017	0.21	0
2018	0.21	0

Loglogistic MLE Parameters

	γ	θ
2014	1.13	0
2015	1.14	0
2016	1.14	0
2017	1.13	0
2018	1.14	0

g-and-h MLE Parameters

	a	b	g	h
2014	1.8	0.96	1.85	0.35
2015	1.79	0.94	1.76	0.39
2016	1.81	0.99	1.82	0.34
2017	1.82	1.01	1.81	0.33
2018	1.83	1.02	1.82	0.31

log-SaS MLE Parameters

	a	b	ε	δ
2014	0.35	0.19	0.66	0.59
2015	0.36	0.2	0.61	0.59
2016	0.34	0.2	0.67	0.6
2017	0.34	0.21	0.68	0.61
2018	0.34	0.21	0.71	0.62

Lgn/Lgn MLE Parameters

	μ_b	σ_b	x_s	μ_t	σ_t	p_b
2014	0.42	0.23	2.16	-7.8	3.33	0.58
2015	0.42	0.24	2.17	-7.62	3.26	0.58
2016	0.42	0.23	2.22	-7.47	3.21	0.58
2017	0.42	0.23	1.9	-7.56	3.19	0.5
2018	0.42	0.22	1.9	-7.48	3.15	0.5

Lgn/Gpd MLE Parameters

	μ_b	σ_b	x_s	ξ	θ	p_b
2014	0.41	0.21	1.94	1.16	1.55	0.52
2015	0.41	0.21	1.95	1.12	1.58	0.52
2016	0.41	0.21	1.96	1.07	1.73	0.52
2017	0.41	0.21	1.98	1.05	1.79	0.52
2018	0.41	0.21	1.98	1.02	1.84	0.52

Figure B.3: SRC 3 MLE parameters for each candidate distribution when including all loss data before each row's designated year. The last row uses all simulated data. Lgn/Lgn and Lgn/Gpd refer to LGNLGN and LGNGPD, respectively.