University of California
Santa Barbara

Contributions to Modeling of Operational Risk in Banks

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Statistics and Applied Probability

by

Anna S. Tchernobai

Committee in charge:

Professor Svetlozar T. Rachev, Chairperson

Professor Sreenivasa Rao Jammalamadaka, Co-Chair

Professor John Hsu

Professor Yuedong Wang

June 2006

The dissertation of Anna S. Tchernobai is approved:

Professor John Hsu

Professor Yuedong Wang

Professor Sreenivasa Rao Jammalamadaka, Committee Co-Chair

Professor Svetlozar T. Rachev, Committee Chairperson

May 2006

Contributions to Modeling of Operational Risk in Banks

To Makan, Katya, and my parents

Acknowledgments

I would like to begin with expressing my deepest appreciation to my Ph.D. advisor Professor Svetlozar Rachev for his wise direction, patient assistance, and insightful guidance during my doctorate research. It is also to Professor Rachev that I owe the most overwhelming debt of gratitude for assisting me throughout my two prolonged visits to the University of Karlsruhe, Germany, in 2004 and 2005, during which I maintained productive research, made important professional contacts, and got to experience the German culture. Without his generous help, most parts of my Ph.D. work would not have been possible.

Many thanks go to my other dissertation committee members: my deep appreciation goes to Professor Rao Jammalamadaka (a.k.a. "Dr. J"), Professor John Hsu, and Professor Yuedong Wang for their expertise, time, and effort. Special thanks go to the faculty and all the graduate students of the Statistics Department of UCSB for providing a great research environment, holding regular BBQ's on the Goleta beach, and Raya's amazing Thanksgiving cooking! Dr. J's PSTAT 207 A-B-C courses were excellent! Thanks go to Troy for his frequent computer help, and to Gail, Juliana, Denna, and Claudia for the help with miscellaneous issues. I am grateful to Professor Steve LeRoy from the Economics Department for his insightful advices.

Danke schön to my German friends and colleagues: Stefan Trück and Christian Menn. A special thanks to Theda Schmidt for her wonderful organizational skills, selfless concern, and warm friendship while I was on my trips to Karlsruhe.

Arigatoo gozaimasu to my dear Japanese friend Ikumi of over 10 years, thanks to whom I still remember how to write and speak Japanese! Many thanks are due to Mr. Yamada, Mr. Narimoto, and Mr. Aragane from Tokyo.

I wish to thank all my terrific friends who have made my stay in Santa Barbara an enjoyable and unforgettable experience and who have distracted me every so often from my work: Violeta, Markus, Ömür, Gökhan & Bibi, Agil, Özçan, Karine, Pauline, Vanessa, Maurizio, Kambiz, Soo, Gary, Carlo, and many more. I have learned so much from each one of them and will always cherish their friendship. I look forward to repeating those movie nights, coffee breaks, dining out, and Halloween parties. Our trips to San Diego, San Francisco, Bryce and Zion Canyons, and gambling in Las Vegas were truly a blast! My warmest thanks go to my dear friend Payam Naghshtabrizi who has taught me to be open-minded and to respect the beauty and complexity of human nature; I will also miss our work-outs in the RecCen that were frequently followed by a cheeseburger at In-N-Out.

Finally, words alone cannot express the thanks I owe to my loving husband, Makan, for his unconditional love, kindness, support, constant inspiration, and always being there for me. Many special appreciations go to Katya, my twin sister, for her humor and sisterly support. Finally, more than to anyone else, my endless thanks and gratitude go to my beautiful wonderful mother and my great father who have given me all the invaluable opportunities in life and who have always encouraged me to do my best.

Curriculum Vitae of Anna S. Tchernobai

Education

Ph.D. in Statistics and Applied Probability, Emphasis in Mathematical and Empirical Finance. University of California at Santa Barbara, June 2006.

M.A. in Economics, Emphasis in Business Economics. University of California at Santa Barbara, June 2002.

M.Sc. in Economics and Finance. University of Warwick, Warwick Business School, United Kingdom, August 2000.
Research advisor: Professor Archie Pitts.

B.A. in Economics. Jouchi Daigaku (Sophia University), Japan, March 1999. Graduated with Distinction. Language of instruction: Japanese.

Experience

Teaching Assistant. University of California at Santa Barbara, Department of Economics and Department of Statistics and Applied Probability, September 2001-March 2006.

Visiting Researcher. University of Karlsruhe, School of Economics and Business Engineering, Institute for Statistics, Econometrics and Mathematical Finance, Germany, September-December 2004 and April-June 2005.

Awards, Honors, and Recognition

Nominated for campus-wide Outstanding Teaching Assistant Award, Academic Senate, University of California at Santa Barbara, 2005.

Graduate Student Conference Travel Grant, University of California at Santa Barbara, 2004.

President's Study Award, University of California at Santa Barbara, 2003-2004.

Award for Academic Excellence, awarded annually to top 1% of students, Jouchi Daigaku, Japan, 1997.

International Student Fellowship, Jouchi Daigaku, Japan, 1995-1999.

Award Academic Excellence, Moritani Foundation, Japan, 1995-1999.

3rd place JAPANESE LANGUAGE SPEECH CONTEST, Jouchi Daigaku, Japan, 1997.

Professional Activities

Member of SSRN, AFA, ASA.

Participant in STATISTICAL MODELING IN FINANCE conference, Fox School of Business and Management and Statistics Department, Temple University, Philadelphia, March 2006.

Organizer and host of GRADUATE STUDENT SEMINAR, Department of Statistics and Applied Probability, University of California at Santa Barbara, January 2005-present.

Participant in OPERATIONAL RISK MANAGEMENT SUMMIT, ASMI, San Francisco, December 2005.

Participant in 30TH INTERNATIONAL CONFERENCE ON STOCHASTIC PROCESSES AND THEIR APPLICATIONS, University of California at Santa Barbara, June 2005.

Participant in SUMMER SCHOOL IN RISK MEASUREMENT AND CONTROL conference, Swiss Institute, Rome, Italy, June 2005.

Participant in EXTREME EVENTS workshop, Center of Applied Statistics and Economics, Humboldt University, Berlin, Germany, October 2004.

Additional Information

Female. Citizen of Russia.

Language skills: Russian (native), English (fluent), Japanese (fluent), Spanish (good), German (basic).

Selected Publications

Book

A. S. CHERNOBAI, S. T. RACHEV, & F. F. FABOZZI, *Operational Risk: A Guide to Basel II Capital Requirements, Models and Analysis*, John Wiley & Sons, 2006.

Book Chapters

1. A. S. CHERNOBAI, C. MENN, S. TRÜCK, S. T. RACHEV, & M. MOSCADELLI, *Treatment of missing data in the field of operational risk: the impacts on parameter estimates, EL, VaR, and CVaR figures*, in E. Davis (ed.) The Advanced Measurement Approach to Operational Risk, RISK Books,

London, 2006,
http://www.pstat.ucsb.edu/research/papers/Treatment_of_incomplete_data_Risk_book_Mar_2006_.pdf.

2. S. T. Rachev, A. S. Chernobai, & C. Menn, *Empirical examination of operational loss distributions*, in M. Morlock *et al* (eds.) Perspectives on Operational Research, Deutscher Universitäts-Verlag/GWV Fachverlage GmbH, Wiesbaden, Germany, 2006,
http://www.pstat.ucsb.edu/research/papers/festschrift_ed.pdf.

3. A. S. Chernobai & S. T. Rachev, *Stable modeling of operational risk*, in M. G. Cruz (ed.) Operational Risk Modelling and Analysis, Theory and Practice, RISK Books, London, 2004.

Refereed Journals

1. A. S. Chernobai & S. T. Rachev, *Applying robust methods to operational risk modeling*, Journal of Operational Risk: 1(1), 2006,
http://www.pstat.ucsb.edu/research/papers/RobustOprisk_20060305ed.pdf.

2. A. S. Chernobai, K. Burneçki, R. Weron, S. Trück, & S. T. Rachev, *Modelling catastrophe claims with left-truncated severity distributions*, Journal of Computational Statistics: 3, 2006,
http://www.pstat.ucsb.edu/research/papers/Chernobai_etal05_20050910.pdf.

3. A. S. Chernobai, C. Menn, S. Trück, & S. T. Rachev, *A note on the estimation of the frequency and severity distribution of operational losses*, Mathematical Scientist: 30(2), 2006,
http://www.pstat.ucsb.edu/research/papers/or_em_12-20.pdf.

Unrefereed Journals

1. M. Moscadelli, A. S. Chernobai, & S. T. Rachev, *Treatment of missing data in the field of operational risk: the impacts on parameter estimates, EL, VaR, and CVaR figures*, Operational Risk, June 2005,
http://www.pstat.ucsb.edu/research/papers/Treatment%20of%20missing%20data effect%20-revised-%2020050531ed.pdf.

Technical Reports

1. A. S. Chernobai, C. Menn, S. Trück, & S. T. Rachev, *Estimation of operational Value-at-Risk in the presence of minimum collection thresholds*, University of California at Santa Barbara, 2005,
http://www.pstat.ucsb.edu/research/papers/OpRisk_tecreport.pdf.

2. A. S. Chernobai, S. T. Rachev, & F. F. Fabozzi, *Composite goodness-of-fit tests for left-truncated samples*, University of California at Santa Barbara, 2005,
http://www.pstat.ucsb.edu/research/papers/KSmissing20050604-JFE.pdf.

3. A. S. Chernobai & S. T. Rachev, *Toward effective financial risk management: stable modeling of operational risk*, IFAC conference paper, 2004, http://www.statistik.uni-karlsruhe.de/technical_reports/ifac04.pdf.

4. A. S. Chernobai, *Main banks in Japan: declining role as a corporate governance tool*, M.Sc. thesis, University of Warwick, Warwick Business School, Coventry, United Kingdom, 2000.

Talks

Conference Talks

1. *Modelling catastrophe claims with left-truncated severity distributions*, 30th International Conference on Stochastic Processes and their Applications, University of California at Santa Barbara, June 2006.

2. *Estimation of operational Value-at-Risk in the presence of minimum collection thresholds: a practical approach*, international conference Summer School in Risk Measurement and Control, Swiss Institute, Rome, Italy, June 2006.

Invited Talks

1. *Operational risk: practical solutions*, afternoon talk, Department of Economics and Business Administration, University of Bergamo, Italy, June 2006.

2. *Operational risk: key issues*, morning talk, Department of Economics and Business Administration, University of Bergamo, Italy, June 2006.

3. *Modelling catastrophe claims with left-truncated severity distributions*, Hugo Steinhaus Center for Stochastic Methods, Wrocław University of Technology, Wrocław, Poland, April 2005.

4. *Expectation-Maximization algorithm and some applications*, Graduate Student Seminar, Department of Statistics and Applied Probability, University of California at Santa Barbara, February 2005.

5. *Estimation of operational Value-at-Risk with minimum collection thresholds*, Hugo Steinhaus Center for Stochastic Methods, Wrocław University of Technology, Wrocław, Poland, December 2004.

6. *Operational risk management: recent advances in statistical methodologies*, lecture for Masters students, Institute for Statistics and Mathematical Finance, University of Karlsruhe, Karlsruhe, Germany, October 2004.

Abstract

Contributions to Modeling of Operational Risk in Banks

by

Anna S. Tchernobai

In Part I of the dissertation we review the Basel II Capital Accord and the regulatory capital requirements for operational risk.

Part II of this dissertation addresses statistical and probabilistic assessment of operational risk. Value-at-Risk and Conditional Value-at-Risk are taken as the measure of the buffer capital. Losses are modeled in an actuarial-type compound Cox process framework. A variety of issues is addressed; among the topics discussed are the following. (1) We provide empirical evidence that the intensity factor follows a very specific non-homogeneous form. (2) Internal operational loss databases suffer from reporting bias; practitioners often neglect this issue. We provide theoretical and empirical justification that this bias leads to severe underestimation of the capital charge, and propose methodologies to evaluate the exact information loss and incorporate it into the operational risk model. (3) "Low frequency/ high severity" events are successfully captured by fitting variations of the alpha-Stable distribution to the loss severity data. (4) An innovative EDF-based goodness-of-fit test is designed to evaluate the performance of the loss distributions in the upper quantiles that largely determine the amount of the risk capital. (5) Finally, methodologies for robust modeling of operational risk are addressed.

Contents

Acknowledgments **v**

Curriculum Vitae **vii**

Abstract **xi**

List of Figures **xvii**

1 Introduction **1**

1.1 General Introduction 1

1.2 Main Contributions 2

I Operational Risk Management: Overview **5**

2 What is Operational Risk? **6**

2.1 Introduction 6

2.2 What is Risk? 6

2.3 Definition of Operational Risk 7

2.4 Operational Risk Exposure Indicators 9

2.5 Classification of Operational Risk 9

2.6 Topology of Financial Risks 10

2.7 Appendix: Operational Loss Event Types, Business Lines, and Loss Types . . . 14

3 Operational Risk as Dominant Financial Risk 16

3.1 Introduction . . . 16

3.2 Effects of Globalization and Deregulation: Increased Risk Exposures 16

3.3 Operational Risk is the Dominant Risk in Banks . . . 18

3.4 Examples of High-Magnitude Operational Losses . . . 19

4 Basel II: Regulatory Capital Requirements 21

4.1 Introduction . . . 21

4.2 Capital Allocation for Operational, Market, and Credit Risks . . . 22

4.3 The Basel Capital Accord . . . 23

4.4 Pillar I: Minimum Capital Requirements . . . 24

4.4.1 Three Approaches to Assess Operational Risk Capital Charge 26

4.4.2 Loss Distribution Approach in Detail . . . 26

II Aggregate Stochastic Modeling of Operational Risk 29

5 Exploratory Data Analysis, Value-at-Risk, and Conditional Value-at-Risk 30

5.1 Introduction . . . 30

5.2 Description of the Dataset . . . 31

5.3 Exploratory Data Analysis . . . 32

5.3.1 Time Series of the Loss Process . . . 32

5.3.2 Loss Frequency Process . . . 32

5.3.3 Loss Severity Process . . . 35

5.3.4 Extreme Value Theory for Extreme Losses . . . 42

5.3.5 The Normality Assumption 45
5.4 Compound Poisson Process Model 46
5.5 Value-at-Risk 47
5.6 Conditional Value-at-Risk and Coherent Risk Measures 49

6 α-Stable (Paretian) Distributions 51

6.1 Introduction 51
6.2 Definition of an α-Stable Random Variable 52
6.3 Useful Properties of an α-Stable Random Variable 54
6.4 Estimating Parameters of the α-Stable Distribution 56
6.4.1 Sample Characteristic Function Approach 56
6.4.2 Numerical Approximation of the Density Function Approach 57
6.5 Useful Transformations of α-Stable Random Variables 58
6.5.1 Symmetric α-Stable Random Variable 58
6.5.2 logα-Stable Random Variable 58
6.5.3 Truncated α-Stable Random Variable 59
6.6 Applications to Operational Loss Data 60
6.6.1 Empirical Study with 1980-2002 Public Operational Loss Data 60
6.6.2 Empirical Study of Inter-Arrival Times with 1950-2002 Public Operational Loss Data 66

7 Modeling Operational Frequency with Cox Processes 68

7.1 Introduction 68
7.2 Cox Processes 69
7.2.1 Mixed Poisson Processes 70
7.2.2 Non-Homogeneous Poisson Process with Stochastic Intensity 71

7.3 Renewal Processes 72
7.4 Empirical Studies with Operational Frequency Data 73
7.4.1 Empirical Study with 1980-2002 Public Operational Loss Frequency Data 73

8 Truncated Loss Models 79
8.1 Introduction 79
8.2 Reporting Bias Problem 80
8.3 Truncated Model for Operational Risk 80
8.3.1 Data Specification 81
8.3.2 Parameter Estimation 82
8.4 Simulation Study: Lognormal Example 90
8.4.1 Impact on Parameter Estimates 90
8.4.2 Impact on Expected Aggregate Loss, VaR, and CVaR . . . 92
8.5 Empirical Study with Operational Loss Data 94
8.5.1 Overview of Earlier Studies 94
8.5.2 Empirical Study with 1980-2002 Public Operational Loss Data 95
8.6 Appendix: Results of Empirical Study 108

9 Goodness-of-Fit Tests for Heavy-Tailed Loss Data 123
9.1 Introduction 123
9.2 Overview of Common EDF-Based Goodness-of-Fit Statistics . . . 125
9.3 "Upper Tail" Anderson-Darling Statistic 128
9.3.1 Supremum Class "Upper Tail" Anderson-Darling Statistic 129
9.3.2 Quadratic Class "Upper Tail" Anderson-Darling Statistic 130
9.4 Application to Financial Loss Data 132

10 Robust Modeling **140**

10.1 Introduction 140

10.2 Outliers in Operational Loss Data 141

10.3 Some Dangers of Using the Classical Approach 143

10.4 Overview of Robust Statistics Methodology 144

10.4.1 Formal Model for Robust Statistics 145

10.4.2 Traditional Methods of Outlier Detection 145

10.4.3 Examples of Non-Robust Estimators 147

10.4.4 Outlier Detection Approach Based On Influence Functions 148

10.4.5 Outlier Rejection Approach and Stress Tests 149

10.5 Literature Review: Applications of Robust Models in Finance . . 150

10.6 Application of Robust Methods to Operational Loss Data 151

10.6.1 Empirical Study with 1980-2002 Operational Loss Data . . 151

10.7 Appendix: Results of Empirical Study 155

11 Conclusions and Future Research Directions **169**

11.1 Conclusions 169

11.2 Future Research Directions 172

Glossary of Acronyms **176**

Bibliography **178**

List of Figures

2.1 Topology of financial risks in banks. 11

4.1 Banks' economic capital allocation for operational, market, and credit risks according to Jorion (2000). 22

4.2 Decomposition of the operational loss distribution into expected (EL), unexpected (UL), and catastrophic operational losses. The distribution represents one-year compound loss distribution based on the one-year frequency and severity distributions. The upper bound for UL is suggested to be estimated at a high confidence level, such as 95-99.9%. The high percentile can be estimated by the Value-at-Risk or Conditional Value-at-Risk measure. 25

5.1 Exploratory data analysis of 1980-2002 public operational loss data: time series. (a) "Relationship," (b) "Human," (c) "Processes," (d) "Technology," and (e) "External." 33

5.2 Exploratory data analysis of 1980-2002 public operational loss data: relative histograms. (a) "Relationship," (b) "Human," (c) "Processes," (d) "Technology," and (e) "External." 38

5.3 Exploratory data analysis of 1980-2002 public operational loss data: log-transformed QQ-plot against Exponential quantiles. (a) "Relationship," (b) "Human," (c) "Processes," (d) "Technology," and (e) "External." 40

5.4 Exploratory data analysis of 1980-2002 public operational loss data: sample mean excess ($e_n(u)$) plot. (a) "Relationship," (b) "Human," (c) "Processes," (d) "Technology," and (e) "External." 41

5.5 Distribution of X (left) and distribution of the excesses over threshold u (right). 43

6.1 α-Stable densities: effects of changing α and β on the form of the α-Stable density. 53

6.2 QQ-plots (logarithmic scale) for logα-Stable distribution fitted to the 1980-2002 public operational loss data. (a) "Relationship," (b) "Human," (c) "Processes," (d) "Technology," and (e) "External." 61

6.3 QQ-plots (logarithmic scale) for symmetric α-Stable distribution fitted to the 1980-2002 public operational loss data. (a) "Relationship," (b) "Human," (c) "Processes," (d) "Technology," and (e) "External." 62

7.1 Annual number of losses (left) and periodogram (right) of the frequency distribution of 1980-2002 public operational loss data. (a) "Relationship," (b) "Human," (c) "Processes," (d) "Technology," and (e) "External." Periodograms for "Processes" and "Technology" losses reveal distinct peaks at frequency 0.32 and 0.41, suggesting a period of 1/0.33=3.03 years and 1/0.41=2.41 years, respectively. 74

7.2 Frequency distributions of 1980-2002 public operational loss data: cumulative number of loss events over time. The plots reveal a non-homogeneous nature of loss occurrence. (a) "Relationship," (b) "Human," (c) "Processes," (d) "Technology," and (e) "External." 76

8.1 Illustration of incomplete loss and frequency data when the loss data are left-truncated. 81

8.2 Illustration of the "naive" approach and the conditional approach. Panel (a) portrays the density estimated under the "naive" approach; panel (b) portrays the conditional density estimated under the conditional approach; panel (c) portrays the unconditional complete-data density estimated under the conditional approach. 83

8.3 Illustration of the bias in the parameters $\hat{\mu}$ and $\hat{\sigma}$ and the fractions of missing data $F_{\hat{\gamma}}(H)$ (denotes by Q), estimated under the "naive" (left) and conditional (right) approaches, under the assumption of Lognormally distributed losses and Poisson frequency, for a range of true values μ_0 and σ_0 and $H = 50$. The figures show the ratio of the estimated parameters to the true parameters of the loss distribution. Bias under the "naive" approach is evident from the discrepancy of the ratios from unity. 91

8.4 Illustration of bias of $\hat{\lambda}$ estimated under the "naive" and conditional approaches for varying fractions of missing data ($F_{\gamma_0}(H)$). 92

8.5 Illustration of the bias in the estimates of one-year EL, 95% VaR, and 95% CVaR under the "naive" (left) and conditional (right) approaches, under the assumption of Lognormal distribution for losses and Poisson frequency, for a range of values for μ and σ, $H = 50$, and $\lambda = 100$. The figures show the ratio of the estimated risk measures to the true risk measures. Bias is evident from the discrepancy of the ratios from unity. 93

9.1 Weights associated with goodness-of-fit statistics assigned to observations in x, for a Lognormal($\mu = 1, \sigma = 1$) example. Left panel: supremum class statistics, right panel: quadratic class statistics. 130

10.1 Exemplary histogram of operational loss data. Extreme events appear as a distinctive tail in the far right of the distribution. . . 142

Chapter 1

Introduction

1.1 General Introduction

Operational risk is defined as *"the risk of loss resulting from inadequate or failed internal processes, people and systems, or from external events."* Financial markets in the last two decades have been highlighted by large-scale financial failures due to incompetence and fraud, such as Barings, Daiwa, Allied Irish Banks, Orange County, Enron, along with man-made and natural disasters, such as "9/11," Hurricanes Andrew and Katrina. As a consequence, operational risk has been acknowledged to overweigh the importance of credit and market risks.

Since 2001, the Basel Committee for the Banking Supervision of the Bank of International Settlements has been requiring banks to set aside regulatory capital amount that would cover potential operational loss. The capital amount must be evaluated on a one-year aggregated basis at a sufficiently high confidence level. Statistical tools are required to accurately assess the frequency and severity distributions.

The presence of so-called "low frequency/ high severity" events poses problems for the modeling of operational risk and calls for models capable of capturing excessive heavy-tailedness in the data.

The goal of this dissertation is to provide some contributions to statistical and probabilistic modeling of operational risk. All theoretical models discussed in this dissertation are supplemented by extensive empirical testing with real operational loss data.

1.2 Main Contributions

Part I

In Part I, we give an extensive account to the development of operational risk awareness by banks in the past two decades. Recent banking failures clearly attributed to losses unrelated to credit and market risk were linked to and justified by drastic developments in the global financial regimes, such as globalization, deregulation, and revolutionary technological innovations. Operational risk is defined and its place among other financial risks is distinctly identified.

Part II

Operational losses occur in an irregular fashion, suggesting that actuarial-type models are relevant for stochastic modeling of operational risk. We develop an optimal compound Cox process model that constitutes two core components: (1) a stochastic intensity function of a non-linear deterministic form, and (2) heavy-tailed claim amounts. An analytic expression for the intensity function is proposed and it is empirically proven to be vastly superior to a homogeneous intensity rate under the simple Poisson assumption. A large class of loss distributions – Exponential, Lognormal, Weibull, Logweibull, Gamma, Pareto, Burr, α-Stable, logα-Stable, and symmetric α-Stable – are applied to real loss data of five types:

“Relationship”, “Human,” “Processes,” “Technology,” and “External,” obtained from a large European public operational loss data provider. We develop a procedure for forecasting 1-year Value-at-Risk and Conditional Value-at-Risk – proxies for the operational risk capital charge. Relevant goodness-of-fit tests backtesting are performed, using bootstrapping and other innovative techniques.

Operational loss data are very heavy-tailed and right-skewed, and require a loss distribution sufficient to capture the “low frequency/ high severity” events. This justifies our extensive use of the *α-Stable (Paretian) distribution* for operational loss modeling. Variations of the α-Stable distribution (logα-Stable, symmetric α-Stable, and truncated α-Stable) are fitted to the loss data; superiority over other loss distributions are evident in the flexibility, better fit in the upper quantiles, and other attractive features. Furthermore, we propose using right-truncated α-Stable distributions that would ensure finite mean, variance, and high moments.

Reporting bias in the operational loss data is addressed: the minimum collection threshold is set at approximately \$10,000 in the internal databases and approximately \$1 million in the external – a reality often neglected by practitioners. We propose a procedure to correctly specify the loss and frequency distributions in order to fully account for the missing data and accurately evaluate Value-at-Risk and Conditional Value-at-Risk. To correctly estimate the parameters of the loss distribution, two methodologies are suggested: the *Expectation-Maximization* algorithm and the *Restricted Maximum Likelihood*. Additionally, we design a procedure for rescaling the intensity parameter of the frequency distribution to account for the exact information loss. Analytic expression of the bias in the loss and frequency distribution parameters and the Value-at-Risk are derived. It is further shown empirically that ignoring the missing data results in severe underestimation of the aggregate mean and the capital charge.

In operational risk, of central concern is the fit of loss distribution to the data in the upper tail, as the upper tail of correctly specified aggregate distribution is the primary determinant of the amount of the capital charge. We develop a novel goodness-of-fit test that puts more weight on the upper quantiles and less weight on the medium and lower quantiles.

Outliers in the operational loss data are analyzed, and robust modeling is performed. We claim that robust methods can be used as a key diagnostic tool to reveal the influence of outliers on vital statistics and verify empirically that the influence of the highest 5% of loss data accounts for beyond 50% of the total risk exposure.

Part I

Operational Risk Management: Overview

Chapter 2

What is Operational Risk?

2.1 Introduction

Until very recently, it has been believed that banks are exposed to two main risks. In the order of importance they are credit risk (i.e., counterparty failure risk) and market risk (i.e., risk of loss due to changes in market indicators, such as equity prices, interest rates and exchange rates). Operational risk has been regarded as a mere part of "other" risks.

Operational risk is not a new concept for banks: operational losses have been reflected in banks' balance sheets for many decades. They occur in the banking industry every day. Operational risk affects the soundness and operating efficiency of all banking activities and all business units. We begin our discussion with an explanation of the notion of risk.

2.2 What is Risk?

In the financial context, risk is *the* fundamental element that affects financial behavior. There is no unique or uniform definition of risk: different financial institutions may define risk slightly differently, depending on the specifics of their banking structure, operations and investment strategies. The definition of risk also depends on the context.

In the economics literature, generally risk is not necessarily a negative concept, and is understood as *uncertainty about future* or *the dispersion of actual from expected results.* In the context of business investment, risk is *the volatility of expected future cash-flows* (measured, for example, by the standard deviation), and in the context of the Capital Asset Pricing Model (CAPM) is the risk of *asset price volatility due to market-related factors* and is captured by β. Such definitions do not exclude the possibility of positive outcomes. Hence, for the operational risk we need a different definition.[1]

For the purposes of operational risk modeling and analysis, the definitions from insurance are more appropriate, as the notion of risk in insurance has a negative meaning attached to it. Risk is perceived as the *probability and impact of a negative deviation*, the *probability or potential of sustaining a loss*, *"a condition in which there is a possibility of an adverse deviation from a desired outcome that is expected or hoped for"* [159], or *"an expression of the danger that the effective future outcome will deviate from the expected or planned outcome in a negative way"* [78]. As the next step, we need to distinguish operational risk from other categories of financial risk.

2.3 Definition of Operational Risk

Operational risk is largely a firm-specific non-systematic risk: according to the Bank of International Settlements (BIS), *"Unlike market and perhaps credit risk,*

[1] Of course, it is possible that for example an employee error can result in an operational gain rather than loss for the bank, but this possibility is generally ignored for the purpose of operational risk modeling. We do not treat this case.

the [operational] risk factors are largely internal to the bank" [22].

Earlier references on operational risk defined operational risk as *"other risks"*, or *"any risk not categorizes as market and credit risk"*, and *"the risk of loss arising from various types of human or technical errors"* [21]. Other definitions include: risk *"arising from human and technical errors and accidents"* [102], *"a measure of the link between a firm's business activities and the variation in its business results"* [106], and *"the risk associated with operating a business"* [45].

The formal definition that is currently widely accepted was initially proposed by [14] and adopted by the Basel Committee in January 2001 [25]: operational risk was defined as *"the risk of direct or indirect loss resulting from inadequate or failed internal processes, people or systems or from external events."* The industry responded to this definition with criticism regarding the unclarity in the meaning of "direct" and "indirect" losses. A refined definition of operational risk, provided by the September 2001 consultative document, dropped the two terms, hence finalizing the definition of operational risk as [26]:

Definition 1 *(Operational risk) Operational risk is the risk of loss resulting from inadequate or failed internal processes, people or systems or from external events.*

This definition includes legal risk, but excludes strategic and reputational risk. The definition is causal-based, as it provides the breakdown of operational risk by its four major sources:

1. people,
2. processes,
3. systems, and
4. external factors.

2.4 Operational Risk Exposure Indicators

The chance of an operational risk event is increased with a larger number of personnel (due to increased possibility of committing an error) or with a greater transaction volume. A list of operational risk exposure indicators is given in [4, pp. 168-169] [91, pp. 250] [25, Annex 4]. Examples of the operational risk exposure indicators include:

- Gross income;
- Volume of trades or new deals;
- Value of assets under management;
- Value of transactions;
- Number of transactions;
- Number of employees;
- Employees' years of experience;
- Capital structure (debt to equity ratio);
- Historical operational losses;
- Historical insurance claims for operational losses.

2.5 Classification of Operational Risk

The formal definition of operational risk classifies the losses by the sources into four groups: human, process, technology and external losses. Operational risk can be also classified according to a variety of other principles. It can be classified by the nature of the loss (such as internally inflicted or externally inflicted, direct losses or indirect losses), by the degree of expectancy (expected or unexpected), by risk type, event type, business line, or loss type, and by the magnitude (or severity) of loss and frequency of loss. The Basel II Capital Accord classifies operational risk into 7 event type groups, 8 business lines, and identifies 6 operational loss types. Classification of operational risk by event types was discussed by [5]. Detailed description of each of the event types, business lines mapping, and loss types is presented in the Appendix to this chapter.

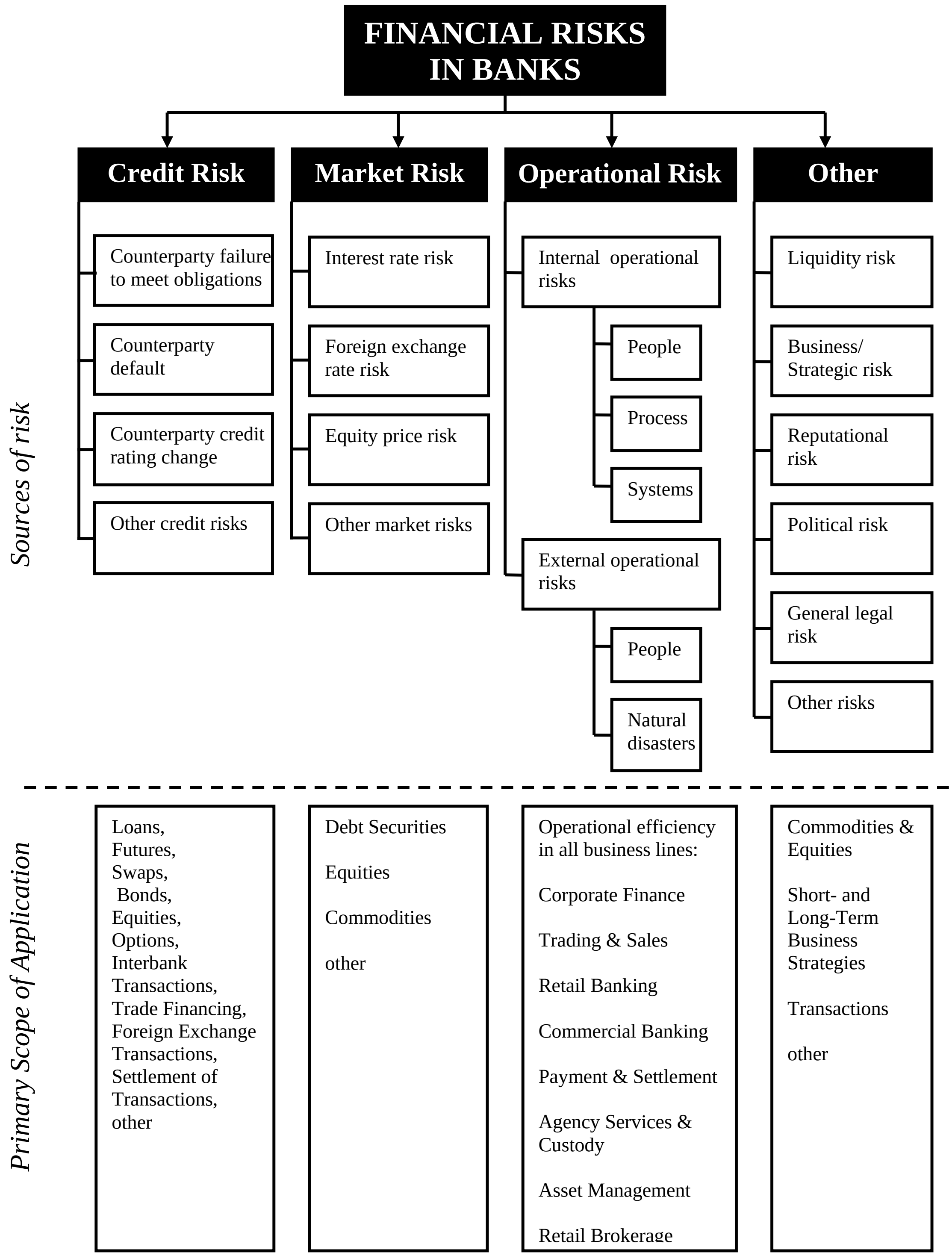
2.6 Topology of Financial Risks

Until recently, credit risk and market risk have been considered as the two largest contributors to banks' risks. Classifications of financial risks were suggested in [45] [158] [156] [76]. In accordance with recent capital requirements and definitions of various financial risks by the Basel Capital Accord (2004), we propose an alternative topology of financial risks, summarized in Figure 2.1 and described below.

1. *Credit Risk*: the potential that a bank borrower or counterparty will fail to meet its obligations in accordance with agreed terms (BIS definition).

2. *Market Risk*: the risk of losses (in on- and off-balancesheet positions) arising from movements in market prices, including interest rates, exchange rates and equity values (BIS definition). It is the risk of the potential change in the value of a portfolio of financial instruments resulting from the movement of market rates, underlying prices and volatilities. The major components of market risk are the interest rate risk, equity position risk, foreign exchange risk, and commodity risk.

Figure 2.1: Topology of financial risks in banks.



3. *Operational Risk*: the risk of loss resulting from inadequate or failed internal processes, people or systems or from external events (BIS definition). As already mentioned, operational risk includes legal risks, which includes, but is not limited to, exposure to fines, penalties, or punitive damages resulting from supervisory actions, as well as private settlements (BIS definition).

4. *Liquidity Risk*: the risk of inability to fund increases in assets and meet obligations as they come due (BIS definition), such as inability to raise money in the long-term or short-term debt capital markets, or an inability to access the repurchase and securities lending markets. An alternative definition by [45] says that liquidity risk is the risk that the institution will not be able to execute a transaction at the prevailing market price because there is, temporarily, no appetite for the deal on the "other side" of the market. Liquidity risk is often considered part of market risk.

5. *Business and Strategic Risk*: the risk that a bank would have to modify the line of behavior and activity in order to cope with changes in the economic and financial environment in which it operates. For example, a new competitor can change the business paradigm, or new strategic initiatives (such as development of a new business line or re-engineering an existing business line, for example, e-banking) can expose bank to strategic risk [45]. Many strategic risks are involved with the timing issue [24].

6. *Reputational Risk*: the risks mainly associated with the customer, i.e. the risk of failure to meet customers' expectations. Banks with a large private banking sector and e-banking activities are especially vulnerable to reputational risk. Reputational risk includes risks related to customer data and privacy protection, e-banking and e-mail services, timely and proper infor-

mation disclosure, etc. [24].

7. *Political Risk*: the risk of an adverse impact on bank's activities due to changes in country and/or regional political or economical pressures, such as monetary controls. Changes in political policies may adversely affect the ability of clients or counterparties located in that country or region to obtain foreign exchange or credit and, therefore, to perform their obligations to the bank.

8. *General Legal Risk*: the risk that a bank would have to modify its activities due to changes in the country's legal system or law enforcements. Examples include a potential impact of a change in tax codes.

2.7 Appendix: Operational Loss Event Types, Business Lines, and Loss Types

Table 2.1: Business line mapping according to the Basel II Capital Accord.

Business Unit	**Business Line**
Investment Banking	Corporate Finance
	Trading and Sales
Banking	Retail Banking
	Commercial Banking
	Payment and Settlement
	Agency Services
Others	Asset Management
	Retail Brokerage

Table 2.2: Loss types and definitions according to the Basel II Capital Accord.

Loss Type	**Contents**
Write-downs	Direct reduction in value of assets due to theft, fraud, unauthorized activity or market and credit losses arising as a result of operational events
Loss of recourse	Payments or disbursements made to incorrect parties and not recovered
Restitution	Payments to clients of principal and/or interest by way of restitution, or the cost of any other form of compensation paid to clients
Legal liability	Judgements, settlements and other legal costs
Regulatory and compliance	Taxation penalties, fines, or the direct cost of any other penalties, such as license revocations
Loss of or damage to assets	Direct reduction in value of physical assets, including certificates, due to some kind of accident (e.g., neglect, accidents, fires, earthquakes)

Table 2.3: Event types and descriptions according to the Basel II Capital Accord.

Event Type	Definition and Categories
Internal fraud	Acts intended to defraud, misappropriate property or circumvent regulations, the law or company policy, which involves at least one internal party. *Categories:* unauthorized activity and theft and fraud.
External fraud	Acts of a type intended to defraud, misappropriate property or circumvent the law, by a third party. *Categories:* (1) theft and fraud and (2) systems security.
Employment Practices and Workplace Safety	Acts inconsistent with employment, health or safety laws or agreements, from payment of personal injury claims, or from diversity/discrimination events. *Categories:* (1) employee relations, (2) safe environment, and (3) diversity and discrimination.
Clients, Products and Business Practices	Unintentional or negligent failure to meet a professional obligation to specific clients (including fiduciary and suitability requirements), or from the nature or design of a product. *Categories:* (1) suitability, disclosure and fiduciary, (2) improper business or market practices, (3) product flaws, (4) selection, sponsorship and exposure, and (5) advisory activities.
Damage to Physical Assets	Loss or damage to physical assets from natural disaster or other events. *Categories:* Disasters and other events.
Business Disruption and System Failures	Disruption of business or system failures. *Categories:* systems.
Execution, Delivery and Process Management	Failed transaction processing or process management, from relations with trade counterparties and vendors. *Categories:* (1) transaction capture, execution and maintenance, (2) monitoring and reporting, (3) customer intake and documentation, (4) customer/client account management, (5) trade counterparties, and (6) vendors and suppliers.

Chapter 3

Operational Risk as Dominant Financial Risk

3.1 Introduction

Until recently, it has been generally considered that credit risk and market risk are the major sources of risk for a financial institution, while the importance of other risks such as operational risk has been largely underestimated. In this chapter we focus on the significance of operational risk in the financial industry.

3.2 Effects of Globalization and Deregulation: Increased Risk Exposures

In the course of the last couple of decades, the global financial industry has been highlighted by several pronounced trends, which have been in response to increased investors' appetites:

- globalization and deregulation,

- accelerated technological innovation,
- revolutionary advances in the information network, and
- increase in the scope of financial services and products.

Examples of some global financial changes include the following:

- "Big Bang" reform in the London Stock Exchange, October 1986: introduction of automated screen-based trading, eliminated fixed commissions on security trades and allowed securities firms to act as brokers and dealers.
- "Big Bang" financial deregulation reform, Japan, 1998: liberalization of banking, insurance, and stock exchange markets and boosting the competition of the Japanese financial market relative to the European and American markets.
- The Financial Services Act of 1999, United States, 1999: the bill repealed the 1933 Glass-Steagall Act's restrictions on bank and securities firm affiliations and allowed affiliations among financial service companies, including banks, registered investment companies, securities firms, and insurance companies – formerly prohibited under the Bank Holdings Act of 1956; it also called for the expansion of the range of financial services allowed by banks.
- Formation of the European Union and adoption of a single currency, Euro, 1990s: integration on the cultural, economic, and political levels.
- Collapse of the Soviet Regime, early 1990s: a result was creation of a massive new market for capital flows.

As a side-effect of these global financial trends and policies, outsourcing, expansion of the scope of financial services, and large-scale mergers and acquisitions

(M&A) have become more frequent around the globe. These, in turn, inevitably result in an elevated exposure of the financial institutions to various sources of risk. As a simple example, increased use of computer-based banking services is vulnerable to viruses and computer failures, and credit card fraud. When business units expand, this requires additional employees – this may increase the number of errors committed and increase the hazard of fraudulent activities.

3.3 Operational Risk is the Dominant Risk in Banks

As result of global shifts in the financial industry, discussed in §3.2, previously non-existent or insignificant risk factors have become a large (or larger) part of the complex risk profiles of financial institutions. Up until recently, cash flow fluctuations of a larger scale, that are more likely to be incurred by an institution/bank's operation practices rather than market or credit risk related factors, have not been well-managed [106].

Without exaggeration, operational risk is the most striking of all, and has been the subject of heated discussions among risk managers, regulators, and academics in the last several years. In 1999, the Basel Committee for Banking Supervision (BCBS) confirmed this by the following statement [23]: *"…an informal survey [22] that highlights the growing realisation of the significance of* **risks other than credit and market risks, such as operational risk,** *which have been at the heart of some important banking problems in recent years…"* As Roger W. Ferguson, Vice Chairman of the Board of Governors of the Federal Reserve System, stated, *"In an increasingly technologically driven banking system, operational risks have become an even larger share of total risk. Frankly, at some banks, they are* *probably the dominant risk."*[1] Major banks share the same view. As an example, a report by the HSBC Group (2004) states that *"...regulators are increasingly focusing on operational risk ... This extends to operational risk the principle of supporting credit and market risk with capital, since arguably it is operational risk that potentially poses the greatest risk."*[2]

3.4 Examples of High-Magnitude Operational Losses

The world financial system has been shaken by a number of banking failures over the last 20 years, and the risks – that particularly internationally active banks have had to deal with – have become more complex and challenging. More than 100 operational losses exceeding $100 million in value each, and a number of losses exceeding $1 billion, have impacted financial firms globally since the end of 1980s.[3] There is no question that the cause is unrelated to market or credit risks. Such large-scale losses have resulted in bankruptcies, mergers, or substantial equity price declines of a large number of highly recognized financial institutions. Several examples of such losses that occurred in the last two decades are briefly presented in Table 3.4.

References on the discussion of the Orange County collapse include [101] [103] [99], and personal accounts of the case by Nick Leeson include his monographs [112] [113]; references on the Barings Bank case include [108] [109] [72] [11] [10] [42]; Daiwa Bank's case was analyzed by [111] [73]. Other well-known examples from

[1] From the 108th session on The New Basel Capital Accord Proposal, Hearing before the Committee on Banking, Housing and Urban Affairs, United States Senate, 2003.

[2] HSBC Operational Risk Consultancy group was founded in 1990, and is a division of HSBC Insurance Brokers.

[3] Large internationally active banks typically experience between 50 and 80 losses exceeding $1 million per year [49].

Table 3.1: Examples of high-magnitude operational losses from the global financial industry in the last two decades.

Year	Name	Impact	Description
1994	Orange County	>$1.7 bln, bankruptcy	Incompetence (Robert Citron, treasurer), lack of expert risk oversight, and poor internal surveillance and control
1995	Barings Bank	>$1 bln, bankruptcy	Internal fraud (Nick Leeson, trader), unauthorized trading, and poor internal surveillance and control
1995	Daiwa Bank	>$1.1 bln, S&P downgrading from A to BBB	Internal fraud and illegal trading (Toshihide Iguchi, trader), and poor internal surveillance and control
2001	"9/11" Terrorist Attack	Civilian and property loss, business disruptions	Terrorism externally inflicted
2002	Allied Irish Banks	>$0.7 bln	Fraudulent activities (John Rusnak, trader) and poor interla surveillance and control

the financial industry include losses incurred by Bank of Credit and Commerce International (1991, fraud) [15] [1], Bankers Trust (1994, fraud), NatWest Markets (1997, error, incompetence) [148], Nomura Securities (1997, fraud), and the Enron collapse (2001) [117] [155] [74] [32]. Some individual case studies are discussed in [41] [46] [45] [66].

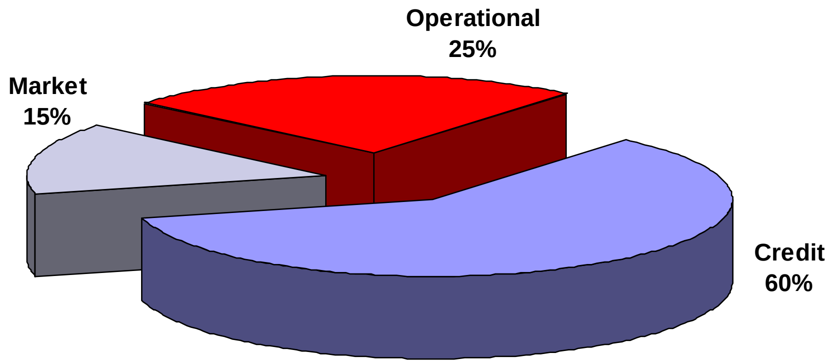
Chapter 4

Basel II: Regulatory Capital Requirements

4.1 Introduction

While one possibility for banks to manage operational risk would be to make efforts to prevent it, another option would be to try to protect themselves against potential consequences. Capital requirements for operational risk were proposed by the Basel Committee for Banking Supervision (BCBS) of the Bank of International Settlements (BIS) in 1999, and the first guidelines were released in 2001. This chapter reviews these capital requirements and makes an important link with subsequent chapters related to modeling issues.

Figure 4.1: Banks' economic capital allocation for operational, market, and credit risks according to Jorion (2000).



4.2 Capital Allocation for Operational, Market, and Credit Risks

The primary role of capital charge is to serve as a buffer to protect against the damage resulting from risk. It can also be seen as a self-insurance tool. There are two types of risk capital, and clear distinction should be made.

- *Economic capital* is defined as the amount of capital market forces dictate for risk in a bank.

- *Regulatory capital* is the amount of capital necessary to provide adequate coverage of banks' exposures to financial risks. A one year *Minimum Regulatory Capital* (MRC) is calculated as 8% of reported risk-weighted assets for the year in question.

Large internationally-active banks allocate roughly $2 billion to $7 billion to operational risk [49]. Current estimates suggest that the allocation of total financial risk of a bank is roughly 60% to credit risk, 15% to market risk and liquidity risk, and 25% to operational risk [102]; see Figure 4.1. [46] suggests 50%, 15%, and 35%, respectively, and [45] suggest 70%, 10%, and 20%, respectively. The portion of economic capital allocated to operational risk has been reported to range between 15% and 25%, [25]. This implies an average of 20%.

4.3 The Basel Capital Accord

The central goal of the Basel Capital Accord is to set guidelines for the estimation of the regulatory capital charge for operational, credit, and market risks, and set necessary managerial principles for the risks.

In July 1988, the Basel Committee released the Capital Accord, now commonly referred to as "Basel I". The primary objective was to establish minimum capital standards designed to protect against credit risk.[1] In April 1993 including market risk into the scope of risks subject to capital requirements was discussed and the capital accord was broadened in 1996 [20] [21]. After two years, reflecting the developments in the financial industry in the preceding years, the Basel Committee decided to undertake a comprehensive amendment of the Basel I and account for the diversity of risks taken by banks. The new capital accord of 1998 is now known as "Basel II". The document "Operational Risk Management" was released in 1998 and discussed the importance of operational risk as a substantial financial risk factor [21]. No discussion regarding the requirement of a capital charge against operational risk had been made until the breaking point in January 2001 when the consultative document "Operational Risk" was released [25].

The Basel II has undergone a number of amendments and was finalized in June 2004.[2] Under the capital accord, operational risk regulatory capital, esti-

[1] Although Basel I discussed the capital requirements for only credit risk, it used a "broad-brush" approach, as it was constructed in such a way as to also implicitly cover other risks. See [19] for the description.

[2] A number of amendments to the 2001 proposal have been released by the Basel Committee since 2001. See the BIS official website `http://www.bis.org` for a full list of downloadable

mated separately by every bank, is designed to reflect the exposure to operational risk. The accord defines and sets detailed instructions on the capital assessment of operational risk and proposes several approaches that banks may consider to estimate the operational capital charge, as well as outlines necessary managerial and disclosure requirements.

The deadline for implementation of the Capital Accord has been provisionally set to year-end 2007, with some transitional adjustments. The scope of application is mainly internationally active banks and their subsidiaries including securities companies. The organization of the Basel II uses a three pillar structure and addresses three types of risk: credit risk, market risk, and operational risk:

- *Pillar I:* Minimum capital requirements;
- *Pillar II:* Supervisory review of an institution's capital adequacy and internal assessment process;
- *Pillar III:* Market discipline through public disclosure of various financial and risk indicators.

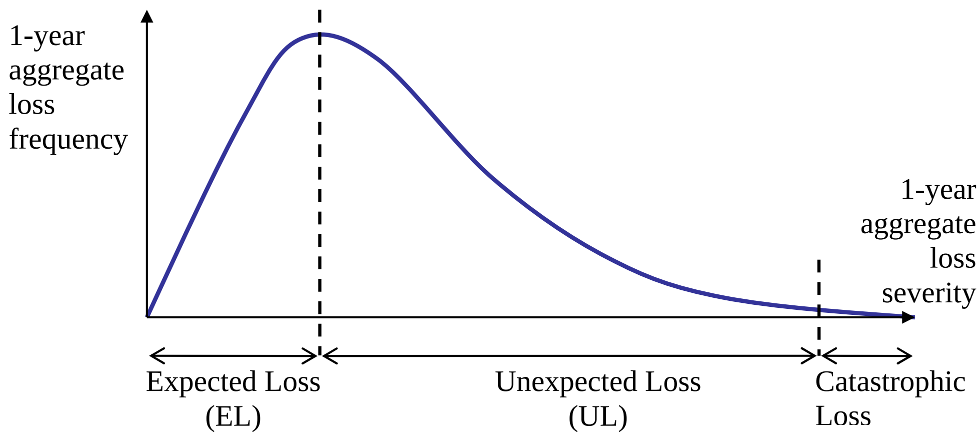
The main focus of this dissertation is to provide a necessary statistical model to be used for the estimation of the minimum capital under the Pillar I.

4.4 Pillar I: Minimum Capital Requirements

In 2001 BIS suggested that the capital charge for operational risk should cover unexpected losses (UL) due to operational risk, and that provisions should cover expected losses (EL); see Figure 4.2. This is due to the fact that for many banking activities with a highly likely incidence of expected regular operational risk losses

publications.

Figure 4.2: Decomposition of the operational loss distribution into expected (EL), unexpected (UL), and catastrophic operational losses. The distribution represents one-year compound loss distribution based on the one-year frequency and severity distributions. The upper bound for UL is suggested to be estimated at a high confidence level, such as 95-99.9%. The high percentile can be estimated by the Value-at-Risk or Conditional Value-at-Risk measure.



(such as fraud losses in credit card books), they are deducted from reported income in the particular year. Therefore, in 2001 BIS proposed to calibrate the capital charge for operational risk based on both EL and UL, but to deduct the amount due to provisioning and loss deduction (rather than EL) from the minimum capital requirement [25].

However, accounting rules in many countries do not provide a robust and clear approach to setting provisions, for example allowing for provisions set only for future obligations related to events that have already occurred. In this sense, they may not accurately reflect the true scope of EL. Therefore, in the 2004 final version of the accord it was proposed to estimate the capital charge as a sum of EL and UL and to allow to subtract the EL portion in those cases when the bank is able to demonstrate its ability to capture the EL by its internal business practices [28].

Exhibit 4.2 illustrates the dimensions of operational risk. *Catastrophic loss* is the loss in excess of the upper boundary of the estimated UL, such as 99.9% Value-at-Risk.[3] It requires no capital coverage; however, insurance coverage may be considered. Catastrophic loss is often called *stress loss*.

4.4.1 Three Approaches to Assess Operational Risk Capital Charge

Three approaches have been finalized for assessing the operational risk capital charge [25] [26]:

1. *the Basic Indicator Approach* (BIA),
2. *the Standardized Approach* (TSA),
3. *the Advanced Measurement Approaches* (AMA).

The BIA and TSA are often referred to as the *top-down approaches* in the sense that the capital charge is allocated according to a fixed proportion of income (bank's gross income under BIA and separately for each business line under TSA), and the AMA are called the *bottom-up approaches* in the sense that the capital charge is estimated from the actual internal loss data. We focus on the AMA that is relevant for our subsequent discussion.

4.4.2 Loss Distribution Approach in Detail

Loss Distribution Approach is the most sophisticated within AMA. Under the LDA, it is suggested that bank's activities are classified into a matrix of "business lines/ event type" combinations (certainly, the actual number of business lines

[3]The notion of Value-at-Risk will be discussed in Chapter 5.

and event types depends on the complexity of the bank structure). For a general case of 8 business lines and 7 event types, bank deals with a 56-cell matrix of possible pairs. For each pair, the key task is to estimate the loss severity and loss frequency distributions. Based on these two estimated distributions, the bank then computes the probability distribution function of the cumulative operational loss.

The operational capital charge is computed as the simple sum of the one-year Value-at-Risk (VaR) measure (with confidence level such as 99.9%) for each "business line/ risk type" pair. The 99.9th percentile means that the capital charge is sufficient to cover losses in all but the worst 0.1% of adverse operational risk events. That is, there is a 0.1% chance that banks will not be able to cover adverse operational losses.

Under simplifying assumptions, for the general case – 8 business lines ($j = 1, 2, \ldots, 8$) and 7 loss event types ($k = 1, 2, \ldots, 7$) – the capital charge can be expressed as:

$$\mathrm{K}_{\mathrm{LDA}} = \sum_{j=1}^{8} \sum_{k=1}^{7} \mathrm{VaR}_{jk}.$$

The advantages of the LDA are as follows:

1. Highly risk sensitive: makes direct use of bank's internal loss data;
2. No assumptions are made about relationship between expected and unexpected losses;
3. Applicable to banks with solid databases;
4. Provided that an estimation methodology is correct, LDA provides an accurate capital charge.

A common criticism of LDA is that estimating the capital charge by a simple sum of the Value-at-Risk measures implies a *perfect correlation* among the "business line/ event type" combinations. A modified version of the LDA approach would take into consideration such correlation effects. Other drawbacks include:

1. Loss distributions may be complicated to estimate. Therefore, the approach can create *model risk* (i.e., wrong estimates due to the misspecification of the model);

2. VaR confidence level is currently not agreed upon, and whether 99.9% or higher/lower percentile is considered makes significant difference on the capital charge;

3. Extensive internal data sets (at least 5 years) are required;

4. The approach lacks the forward-looking component, because the risk assessment is based primarily on the past loss history.

Part II

Aggregate Stochastic Modeling of Operational Risk

Chapter 5

Exploratory Data Analysis, Value-at-Risk, and Conditional Value-at-Risk

5.1 Introduction

Operational risk possesses unique characteristics that distinguish it from other sources of financial risk. The nature of operational risk is very different from that of market risk and credit risk. In fact, operational losses share many similarities with insurance claims, suggesting that most actuarial models can be a natural choice of the model for operational risk, and models well developed by the insurance industry can be almost exactly applied to operational risk. In this chapter we will introduce our dataset on which all our empirical studies are based, and will begin our discussion of statistical modeling of operational risk.

5.2 Description of the Dataset

We here introduce the dataset which lies in the basis of our subsequent empirical studies in this dissertation. We work with operational loss data for the period 1980 to 2002 that comprises publicly announced operational loss events throughout the world, obtained from a major European data provider. The original loss data covered losses in the period 1950-2002; however, we excluded observations prior to 1980 because of relatively few data points available that is most likely explained by poor data recording practices. A few recorded data points appeared below \$1 million in nominal value, so we excluded them from the dataset, to make it more consistent with the conventional threshold for external databases of \$1 million. The loss amounts have been adjusted for inflation using the Consumer Price Index from the U.S. Department of Labor. Loss data are classified into five types:

- "Relationship:" events related to legal issues, negligence, and salesrelated fraud,
- "Human:" events related to employee errors, physical injury, and internal fraud,
- "Processes:" events related to business errors, supervision, security, and transactions,
- "Technology:" events related to technology and computer failure and telecommunications, and
- "External:" events related to natural and man-made disasters and external fraud.

For each operational loss, the date of occurrence (more precisely, the date on which the state of affairs of the event was considered "closed" or "assumed closed") and the loss amount in U.S. dollars was available. For some data points the exact date was not available, and only the year was known. Information on the year suffices when a Poisson process model is considered, with one year as a unit of time. However, the exact date becomes important when one wishes to investigate the distribution of the inter-arrival times.

5.3 Exploratory Data Analysis

In this section, we investigate some properties of the operational loss data.

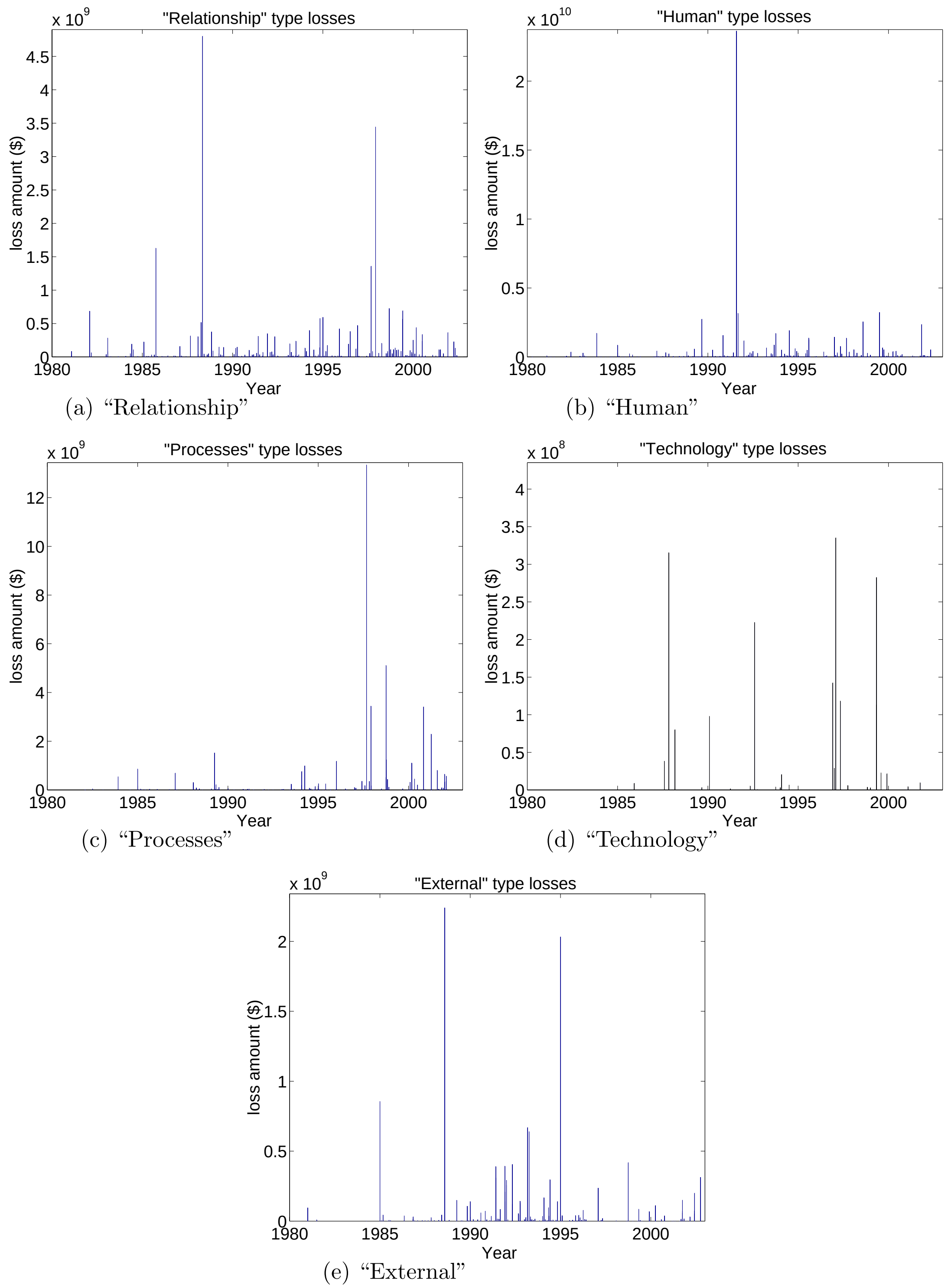
5.3.1 Time Series of the Loss Process

Figure 5.3.1 exhibits the time series of the 1980-2002 public operational loss data for which the exact date of occurrence was available. The time series reveal irregular nature of operational loss arrival process as is seen from varying time intervals between events. Clustering of events also suggests that the frequency of operational losses are on a non-homogeneous nature. Magnitudes of losses indicate high variability in loss amounts.

5.3.2 Loss Frequency Process

One of the difficulties that arise with modeling operational losses has to do with the irregular nature of the event arrival process. In market risk models, market positions are recorded on a frequent basis, many times daily depending on the entity, by marking to market. Price quotes are available daily or for those securities that are infrequently traded, model-based prices are available for marking a position to market. As for credit risk, credit ratings by rating agencies are available. In addition, rating agencies provide credit watches to identify credits that

Figure 5.1: Exploratory data analysis of 1980-2002 public operational loss data: time series. (a) “Relationship,” (b) “Human,” (c) “Processes,” (d) “Technology,” and (e) “External.”



(a) “Relationship”

(b) “Human”

(c) “Processes”

(d) “Technology”

(e) “External”

are candidates for downgrades. In contrast, operational losses occur at irregular time intervals suggesting a process of a *discrete* nature. This makes it similar to the reduced-form models for credit risk, in which the frequency of default (i.e., failure to meet a credit agreement) is of non-trivial concern. Hence, while in market risk it is needed to model only the return distribution in order to obtain VaR, in operational risk both loss severity and frequency distributions are important.

Another problem is related to timing and data recording issue. In market and credit risk models, the impact of a relevant event is almost immediately reflected in the market and credit returns. In an ideal scenario, banks would know how much of operational loss would be borne by the bank from an event at the very moment the event takes place, and would record the loss at this moment. However, from the practical point of view, this appears nearly impossible to implement, because it takes time for the losses to accumulate after an event took place. Therefore, it may take days, months, or even years for the full impact of a particular loss event to be evaluated. Hence, there is the problem of discrepancy (i.e. a time lag) between the occurrence of an event and the time at which the incurred loss is being recorded.

This problem directly affects the method in which banks choose to record their operational loss data. When bank record their operational loss data, they record (i) the amount of loss, and (ii) the corresponding date. We can identify three potential scenarios for the types of date banks might use:

1. *Date of occurrence:* the date on which the event that has led to operational losses actually took place.

2. *Date on which the existence of event has been identified:* the date when bank authorities realize that an event that has led to operational losses has taken or is continuing to take place. Recording a loss at this date may be relevant

in cases when the true date of occurrence is impossible or hard to track.

3. *Accounting date:* the date on which the total amount of operational losses due to a past event are realized and fully measured, and the state of affairs of the event is closed or assumed closed.

Depending on which of the three date types is used, the models for operational risk and conclusions drawn from them may be considerably different. For example, in the third case of accounting dates, we are likely to observe cyclicality/seasonal effects in the time series of the loss data (for example, many loss events would be recorded around the end of December), while in the first and second cases such effects are much less likely to be present in the data. Fortunately, however, selection of the frequency distribution does not have a serious impact on the resulting capital charge [33]. We will treat the operational risk frequency distributions in detail in Chapter 7.

5.3.3 Loss Severity Process

Specifics of Operational Loss Severity Data

The first problem related to the loss severity data deals with the sign of the data. Depending on the movements in the interest or exchange rates, the oscillations in the market returns and indicators can take either positive or negative sign. This is different in the credit and operational risk models - usually, only losses (i.e. negative cash-flows) are assumed to take place.[1] Hence, in modeling operational loss magnitudes, one should either consider fitting the loss distributions that are defined only on positive values, or should use distributions that are defined on negative and positive values, truncated at zero.

[1] Certainly, it is possible that a human error can incur unexpected profits for a bank, but usually this possibility is not considered.

The second problem deals with the high degree of dispersion of loss data. Historical observations suggest that the movements in the market indicators are generally of relatively low magnitude. Bigger losses are usually attributed to credit risk. Finally, although most of the operational losses occur on a daily basis and hence are small in magnitude, the excessive losses of financial institutions are in general due to the operational losses, rather than credit or market risk-related losses. We provided in Chapter 3 examples of high-magnitude operational losses from the financial industry. Empirical evidence indicates that there is a very high degree of dispersion of the operational loss data, with a range from near-zero to billions of dollars. In general, dispersion is measured by variance or standard deviation.[2]

The third problem concerns the shape of the loss distribution. The shape of the data for operational risk is very different from that of market or credit risk. In market risk models, the distribution of the market returns is (often assumed to be) nearly symmetric around zero. Asymmetric cases refer to the data which is either *left-skewed* or *right-skewed* and/or has two or more peaks of different height. Operational losses are highly asymmetric, and empirical evidence on operational risk indicates that the losses are highly skewed to the right, i.e. the right tail of the loss severity distribution is very long.

Empirical evidence on operational losses also indicate a very large number of observations close to zero, and a number of observations of a very high magnitude. The first phenomenon refers to a high kurtosis (i.e. peak) of the data, and the second one indicates heavy tails (or fat tails). Distributions of such data are often described as *leptokurtic*.

[2]Some very heavy-tailed distributions, such as the heavy-tailed Weibull, Pareto, or α-Stable, can have an infinite variance.

Table 5.1: Exploratory data analysis of 1980-2002 public operational loss data: sample descriptive statistics of "Relationship," "Human," "Processes," "Technology," and "External" loss types.

	"Relation."	"Human"	"Proces."	"Techn."	"Exter."
n	849	813	325	67	233
min ($\$ \times 10^6$)	1.07	1.10	1.10	1.13	1.1
max ($\$ \times 10^6$)	6,480	23,630	13,334	830	6,384
mean ($\$ \times 10^6$)	89.86	138.47	285.55	77.43	103.35
median ($\$ \times 10^6$)	14.63	12.32	39.98	11.60	12.89
st. dev. ($\$ \times 10^6$)	360.45	901.51	955.52	136.65	470.24
skewness	11.6429	22.2416	9.1070	3.1761	11.0320
kurtosis	169.9732	570.1188	112.5151	15.7230	140.8799

Sample Descriptive Statistics and Histograms

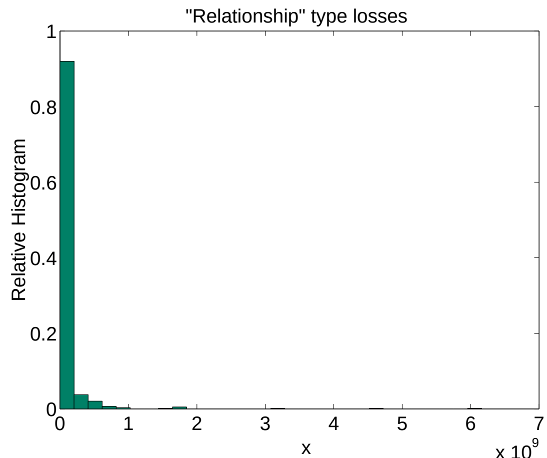
Table 5.1 presents sample descriptive statistics for the 1980-2002 public operational loss data. High dispersion in the data is evident from the high standard deviation figures. Skewness and kurtosis of the data are captured by the sample skewness and kurtosis coefficients calculated by:

$$\mathrm{sk} = \frac{\sum_{j=1}^{n}(x_j - \bar{x})^3}{(n-1)s^3}, \tag{5.1}$$

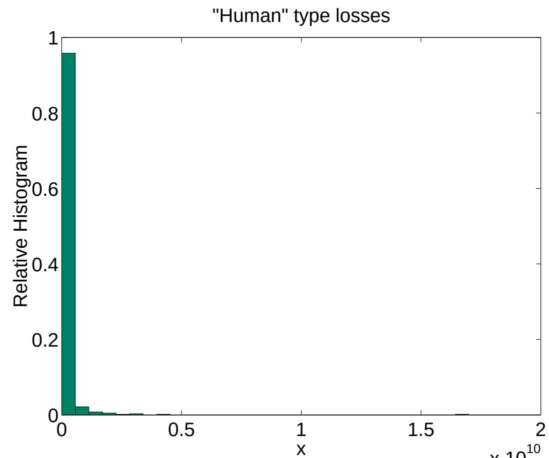
$$\mathrm{k} = \frac{\sum_{j=1}^{n}(x_j - \bar{x})^4}{(n-1)(s^4)}, \tag{5.2}$$

where $\bar{x}$ is the sample mean, s is the sample standard deviation, and n is the sample size. Right-skewness of the operational loss data is captured by the large value of the skewness coefficient; it is also notable that the median value is much lower than the sample mean. A very high degree of kurtosis (for $N(0, 1)$ distribution, the kurtosis coefficient equals 3) indicates that a large mass of the data is concentrated in the lower quantiles of the distribution. The conclusions are supported by the relative histograms of the datasets, depicted in Figure 5.2.

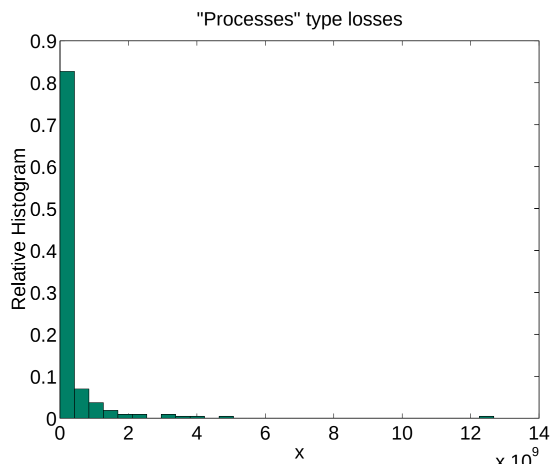
Figure 5.2: Exploratory data analysis of 1980-2002 public operational loss data: relative histograms. (a) “Relationship,” (b) “Human,” (c) “Processes,” (d) “Technology,” and (e) “External.”



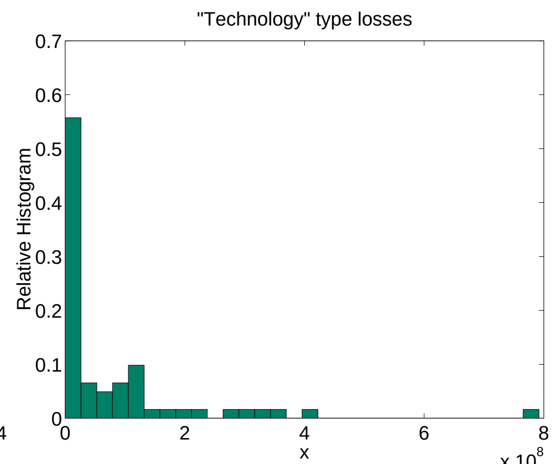
(a) “Relationship”



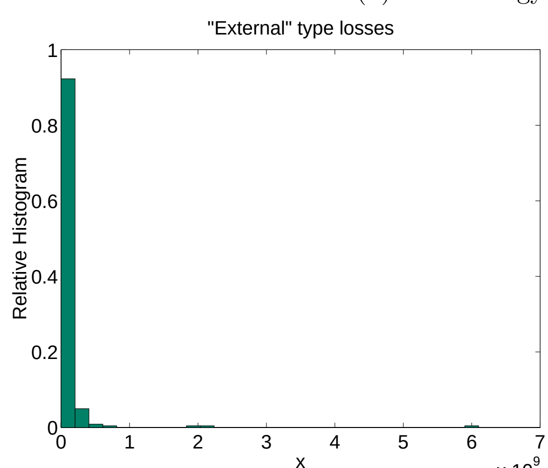
(b) “Human”



(c) “Processes”



(d) “Technology”



(e) “External”

Quantile-Quantile Plots

Quantile-Quantile (QQ) *plots* provide a convenient technique to visually investigate a dataset. QQ-plots plot empirical quantiles against the quantiles of a hypothesized distribution fitted to the data. Figure 5.3 demonstrates QQ plots (on logarithmic scale) for the five datasets against the Exponential quantiles. If the data are to follow the Exponential distribution (that has exponentially fast decaying right tail), then the plot would coincide with the straight 45° line. It is evident that the upper quantiles of the empirical distribution follow a significantly heavier-tailed law than the Exponential law, captured by the qq-plot curving downwards below the 45° line. This suggests that the data are very heavy-tailed.

Mean Excess Plots

Heavy-tailedness in the data is further supported by the behavior of *mean excess plots*.

Definition 2 (*Mean excess function) [62] Let X be a random variable with right endpoint x_F. Then the mean excess function of X is:*

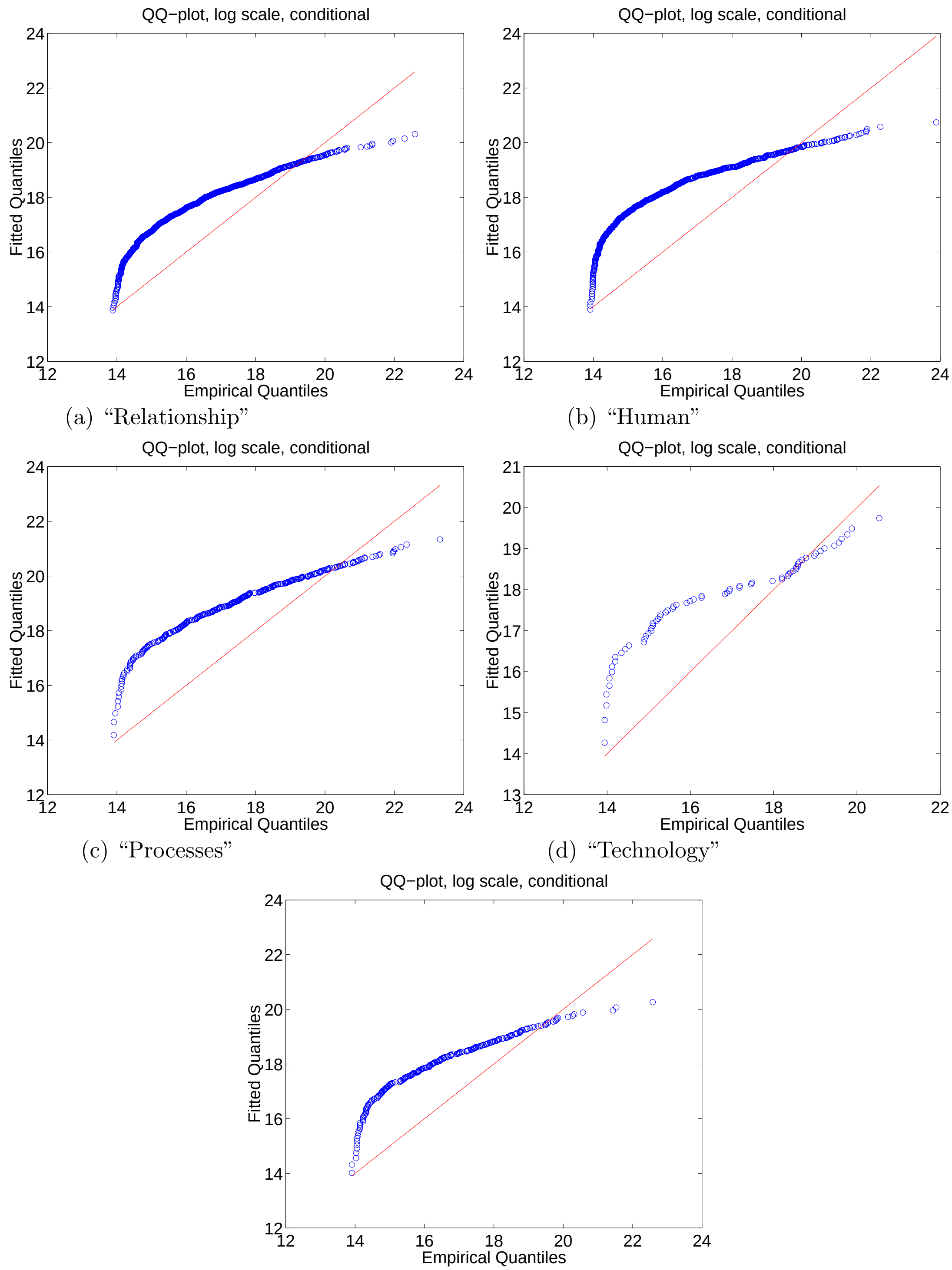
$$e(u) = \mathbb{E}[X - u | X > u], \quad 0 \leq u < x_F. \tag{5.3}$$

The sample mean excess function is then expressed as

$$e_n(u) = \frac{\sum_{j=1}^{n} (x_j - u)_+}{\sum_{j=1}^{n} \mathbb{I}_{\{x_j > u\}}}. \tag{5.4}$$

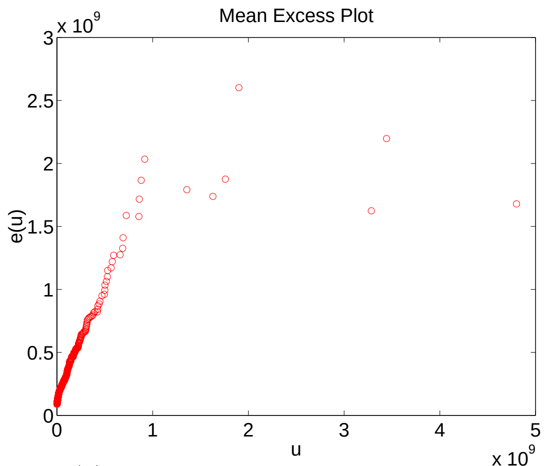
For heavy-tailed data, $e(u)$ typically tends to infinity with an upward-sloping mean excess plot. For example, for the Lognormal and heavy-tailed Weibull ($\alpha < 1$) distribution the 2nd derivative is negative, and for the Pareto distribution it is zero. For the thin-tailed Exponential distribution the mean excess plot is

Figure 5.3: Exploratory data analysis of 1980-2002 public operational loss data: log-transformed QQ-plot against Exponential quantiles. (a) “Relationship,” (b) “Human,” (c) “Processes,” (d) “Technology,” and (e) “External.”

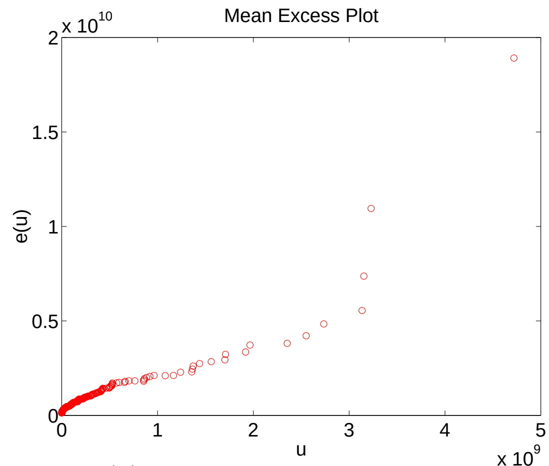


(a) “Relationship” (b) “Human” (c) “Processes” (d) “Technology” (e) “External”

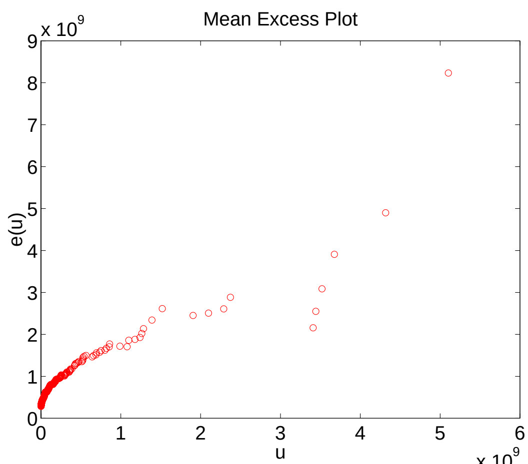
Figure 5.4: Exploratory data analysis of 1980-2002 public operational loss data: sample mean excess ($e_n(u)$) plot. (a) "Relationship," (b) "Human," (c) "Processes," (d) "Technology," and (e) "External."



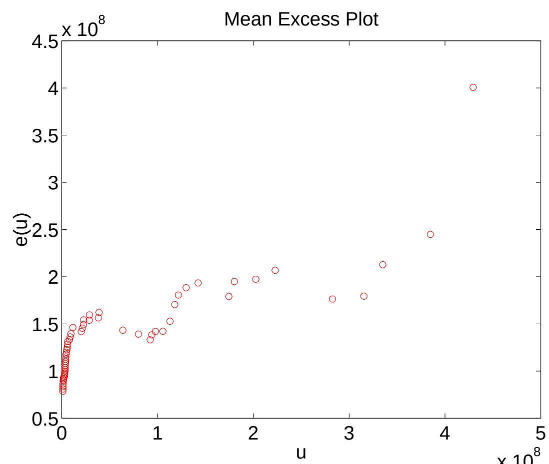
(a) "Relationship"



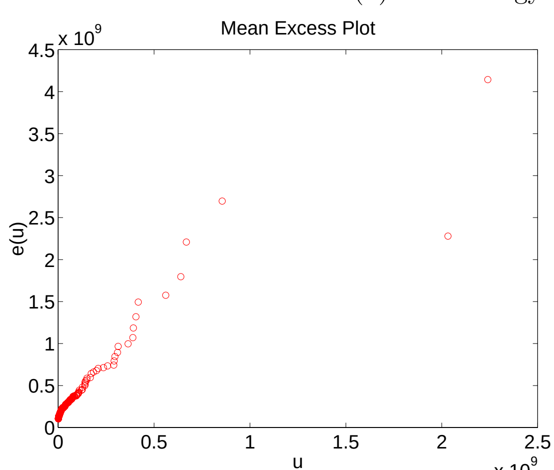
(b) "Human"



(c) "Processes"



(d) "Technology"



(e) "External"

horizontal. Figure 5.4 exhibits sample mean excess plots for the five datasets. The upward-sloping plots indicate that heavy-tailedness is persistent in the operational loss data.

For a detailed empirical analysis with various loss distributions, refer to Chapter 8.

5.3.4 Extreme Value Theory for Extreme Losses

When data are heavy-tailed, two distinct approaches can be undertaken to model the data:

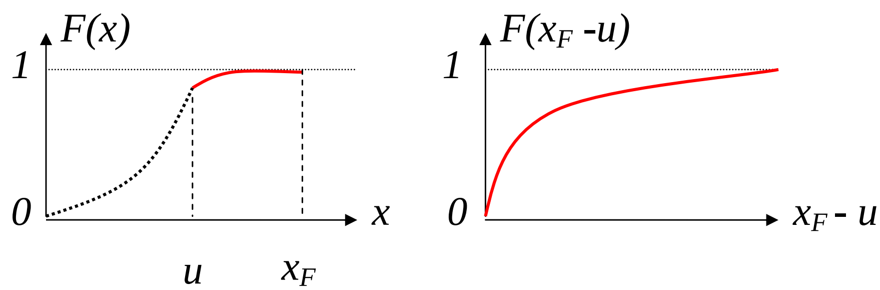
1. An approach that puts equal importance to low-, medium-, and high-scale events. Distributions such as Lognormal, Weibull, Gamma, Pareto, or α-Stable are fitted to the entire dataset.

2. An approach that puts substantial weight on the high quantiles (extreme observations) and treats low and medium quantiles of the loss distribution as less crucial.

Heavy-tailedness of the operational loss data dictates that the extreme observations in the upper quantiles play a significant role in determining the distributional properties of the data. In particular, the Extreme Value Theory (EVT) deals with modeling extremes. An application of EVT to modeling operational risk would be using it to analyze the behavior of losses that exceed a certain high threshold (*Peak Over Threshold* (POT) model) or to analyze the properties of the highest order statistics of the loss distribution in specified time horizons (*Block Maxima* method).

POT focuses on extreme events whose magnitude lies above some pre-specified high threshold. Their distribution is approximated by the *Generalized Pareto Distribution* (GPD) [133].

Figure 5.5: Distribution of X (left) and distribution of the excesses over threshold u (right).



Definition 3 *(Conditional excess distribution function) Let X be a random variable so that $X \sim F$. Let u be a certain threshold and F_u be the excess distribution of X above u, and let x_F be the right endpoint of F so that $x_F \leq \infty$ (see Figure 5.5). Then F_u is called the* conditional excess distribution function*:*

$$P(X - u \leq x | X > u) = \frac{F(u+x) - F(u)}{\overline{F}(u)}, \qquad 0 \leq x < x_F - u. \tag{5.5}$$

For u sufficiently large, $F_u(x)$ is approximated by GPD by the Pickands-Balkema-de Haan Theorem [133].

Definition 4 *(Generalized Pareto Distribution) The GPD distribution for extremes is defined by $G_{\xi;\mu\beta}$:*

$$G_{\xi;\mu\beta} = \begin{cases} 1 - \left(1 + \xi \left(\frac{x-\mu}{\beta}\right)\right)^{-\frac{1}{\xi}} & \textit{if} \quad \xi \neq 0, \\ 1 - e^{-\frac{x-\mu}{\beta}} & \textit{if} \quad \xi = 0, \end{cases} \tag{5.6}$$

where

$$\begin{array}{ll} x \geq 0 & \textit{if} \quad \xi \geq 0 \\ 0 \leq x \leq -\frac{1}{\xi} & \textit{if} \quad \xi < 0. \end{array}$$

and $\mu \in \Re$, $\beta > 0$. Replacing $\frac{x-\mu}{\beta}$ with z, one arrives at a standard GPD distribution G_ξ.

Applying GPD to operational loss data requires choosing a high threshold u. In fact, the choice of such threshold is conventionally performed by visual examination of the mean excess plot, described in Equation (5.3) in §5.3.3, and selecting u as the point x above which the mean excess plot is roughly linear. Using Equations (5.3) and (5.6),

$$e(u) = \frac{\beta}{1-\xi} + \frac{\xi}{1-\xi}u, \tag{5.7}$$

which is a straight line with intercept $\beta/(1-\xi)$ and slope $\xi/(1-\xi)$. The heavier the right tail of the distribution (provided that $0 < \xi < 1$), the steeper the line.

We believe that EVT suffers from several major pitfalls. Some of them are discussed in [52] and [57]. The two drawbacks of EVT that we find important are the following:

- High threshold selection procedure lacks analytical rigor, and the unknown parameters of GPD would be sensitive to the choice of threshold (see, e.g., discussion of this issue in [62]). Quoting [65], *"The optimal value of threshold u to be used is difficult (if not impossible) to obtain."*

- EVT puts most on the distribution of the extremes, and treats low- and medium-scale losses as having a lesser importance. Using the empirical distribution function for the low- and medium-scale losses has become almost a convention in EVT. We believe that all data must equally participate in the statistical analysis of operational loss data (see, e.g., extensive discussion in Chapter 8).

We strongly believe that significant amount of research is required for EVT. Hence, we do not focus on EVT in this dissertation and instead consider loss distributions applicable to entire datasets.

POT model has been examined for operational risk modeling by [59] [61] [124] [123] [127] [34] [35] [36], among others. For general financial loss models, POT has been examined in [62] [118] [119] [121] [122] [120] [52] [65] [57] [58] [80] [79] [98] [115], among others.

5.3.5 The Normality Assumption

The Gaussian distribution is often used to model market risk and credit risk. Despite being easy to work with and having attractive features (such as stability under linear transformations), the Gaussian distribution implies a number of serious assumptions on the loss data, casting doubts regarding its applicability for the operational risk modeling. They include the following:

- The Gaussian distribution is characterized by two parameters, μ and σ, i.e., the mean and the standard deviation. Hence, the Gaussian assumption is useful for modeling the distribution of events that are symmetric around their mean. It has been empirically demonstrated that the operational losses are right-skewed and therefore moments higher than the 2nd become important.

- In most cases (except for the cases when the mean is very high), the use of the Gaussian distribution allows for the occurrence of negative values. This is not a desirable property in the context of operational risk, because negative losses are usually not possible.[3]

[3] Certainly, it is possible to use a truncated (at zero) version of the Gaussian distribution to fit operational losses. Truncated distributions will be presented in Chapter 8.

- More importantly, the Gaussian distribution has an exponential tail decay, $\overline{F}(x) \propto e^{-x^2}$. Such light-tailedness property ensures that the tail events are assigned a near-zero probability. Such assumption is violated in practice.

In this light, it is unlikely that the Gaussian distribution would find much application for the assessment of operational risk. Heavier tailed distributions such as Lognormal, Weibull, and even Pareto and α-Stable, ought to be considered. The large class of α-Stable distributions will be discussed in Chapter 6.

5.4 Compound Poisson Process Model

The LDA assumes an actuarial type model for the aggregated operational losses for a particular "business line/ event type" combination. The losses are assumed to follow a stochastic process $\{S_t\}_{t\geq 0}$ described by:

$$S_t = \sum_{k=0}^{N_t} X_k, \quad X_k \overset{\text{iid}}{\sim} F_\gamma, \tag{5.8}$$

in which the random sequence of loss magnitudes $\{X_k\}$ follows a cumulative distribution function (cdf) F_γ and the density f_γ with the parameter set γ, and in which the counting process N_t is assumed to take a form of a homogeneous Poisson process (HPP) with intensity $\lambda > 0$ (or a non-homogeneous Poisson process (NHPP) with intensity $\lambda(t) > 0$). F_γ belongs to a sufficiently well-behaved parametric family of continuous probability distributions, and f_γ is defined on $\Re_{>0}$. Independence between frequency and severity distributions is assumed under this model, but can be further relaxed (see discussion in Chapter 11). The cdf of the compound Poisson process is given by:

$$P(S_t \leq s) = \begin{cases} \sum_{n=1}^{\infty} P(N_t = n)\ F_{\gamma}^{n*}(s) & s > 0 \\ \\ P(N_t = 0) & s = 0 \end{cases} \tag{5.9}$$

where F_{γ}^{n*} denotes the n-fold convolution of F with itself.

Definition 5 *(Subexponential distribution) A cdf F is subexponential ($F \in \mathcal{S}$) if for all $n \geq 2$*

$$\lim_{x \to \infty} \frac{\overline{F^{n*}}(x)}{\overline{F}(x)} = n. \tag{5.10}$$

Subexponentiality is a property possessed by heavy-tailed distributions, in which the maximum observation $M_n = max\{X_1, X_2, \ldots, X_n\}$ "determine" the behavor of the entire sum $S_n = X_1 + X_2 + \ldots + X_n$: for all $n \geq 2$,

$$P(S_n > x) \sim P(M_n > x), \qquad x \to \infty. \tag{5.11}$$

For $F \in \mathcal{S}$, the following approximation holds:

$$P(S_t > x) \sim \mathbb{E}N_t \cdot \overline{F}(x), \qquad x \to \infty. \tag{5.12}$$

This approximation becomes important in the Value-at-Risk modeling that we introduce in the next section. See also [38].

5.5 Value-at-Risk

Model of Equation (5.8) can be used to determine the required operational capital charge imposed by regulators. Value-at-Risk (VaR) is the predicted worst-case loss at a specific confidence level over a certain period of time [139]. VaR is measured as the $(1-\alpha)\times 100^{\text{th}}$ quantile of the cumulative loss distribution (Equation (5.9)) over a one year period $\triangle t$.

Definition 6 *(Value-at-Risk) For a given confidence level* $1-\alpha$ *and a prespecified time horizon* $\triangle t$*, VaR is defined as:*

$$VaR_{\triangle t,1-\alpha} := G^{-1}_{\triangle t}(1-\alpha) = \sup\{s_{\triangle t}: \; G(s_{\triangle t}) \leq 1-\alpha\}, \tag{5.13}$$

where $G(\cdot)$ *is the cdf of the cumulative loss process.*

Hence, VaR is the solution to:

$$P(S_{\triangle t} > \mathrm{VaR}_{\triangle t,1-\alpha}) = \alpha. \tag{5.14}$$

Combining Equations (5.14) and (5.12), we obtain:

$$\mathrm{VaR}_{\triangle t,1-\alpha} \sim F^{-1}\left(1 - \frac{\alpha}{\mathbb{E}N_{\triangle t}}\right). \tag{5.15}$$

Substituting $F^{-1}(\cdot)$ of a hypothesized individual loss distribution in Equation (5.15), one can easily obtain the desired VaR value for heavy-tailed distributions. See Chapter 8 in which such approximation is used.

In the context of operational risk measurement, VaR was discussed by the Basel Committee (2001-2003) and in works such as [46] [3] [45] [102] [56] [75] [124] [61] [38] [39].

5.6 Conditional Value-at-Risk and Coherent Risk Measures

VaR has been criticized by [163] [162] [160] [161], among others. An alternative risk measure is *Conditional Value-at-Risk* (CVaR).

Definition 7 *(Conditional Value-at-Risk) For a given confidence level* $1-\alpha$ *and a prespecified time horizon* $\triangle t$*, CVaR is defined as:*

$$CVaR_{\triangle t,1-\alpha} := \mathbb{E}\left[S_{\triangle t} | S_{\triangle t} > VaR_{\triangle t,1-\alpha}\right]. \tag{5.16}$$

CVaR is also called *Expected Tail Loss* (ETL) and *Expected Shortfall* (ES). CVaR captures tail events better than VaR [18] [157] [141] and also satisfies all properties of *coherent risk measures.* Denote the risk set by $L = \{X, X_1, X_2, \ldots\}$, and let $r > 0$ be the total return in any state of nature at date T and ρ be the risk measure associated to an acceptance set. Then *coherent risk measures* must satisfy the four axioms [8]:

($A1$) (Translation Invariance)

$\forall X \in L,\ \forall a \in \Re : \quad \rho(X + ar) = \rho(X) - a.$

($A2$) (Subadditivity)

$\forall X_1, X_2 \in L : \quad \rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2).$

($A3$) (Positive Homogeneity)

$\forall X \in L,\ \lambda \geq 0 : \quad \rho(\lambda X) = \lambda\rho(X).$

($A4$) (Monotonicity)

$\forall X_1, X_2 \in L : \quad \rho(X_1) \leq \rho(X_2).$

A1 states that adding (subtracting) the sure initial amount a to the initial position and investing it in the reference instrument decreases (increases) the risk measure by a. A2 states that diversification of business activities or an absence of "firewalls" among different units results in the risk measure at most as high as in the case when these are independent. A3 does not account for netting or diversification effect and says that, for example, two separate firms with an identical position account for risk twice as high.

VaR may fail A2 and may result in overestimation of the capital charge [140] [65] [64]; see also examples in [58] [129]. See applications of VaR and CVaR in Chapter 8 and Chapter 10.

Chapter 6

α-Stable (Paretian) Distributions

> "Once is happenstance. Twice is coincidence. Three times is enemy action."
>
> – Ian L. Fleming (1908-1964)

6.1 Introduction

Operational losses due to errors and omissions, physical loss of securities, natural disasters, and internal fraud are infrequent in nature but can have serious financial consequences for an institution. Such "low frequency/ high severity" operational losses can be extreme in magnitude when compared to the rest of the data. According to BIS [26, Annex 1], *"The internal risk measurement system must capture the impact of infrequent, but potentially severe, operational risk events. That is, the internally generated risk measure must accurately capture the "tail" of the operational risk loss distribution."*

Loss distributions, such as Lognormal, Gamma, and Weibull, are classified as moderately heavy-tailed and thus may not be sufficient to capture the infrequent but potentially severe operational loss events. In this chapter we discuss a wide class of α-Stable distributions.[1] We will review their definition and basic properties and illustrate some empirical studies with operational risk data. The discussion in this chapter closely follows [136].

6.2 Definition of an α-Stable Random Variable

We begin with a definition of an α-Stable random variable. Extensive analysis of α-Stable distributions and their properties can be found in [147] and [137]; see also [153] [154] [138]. Let $X_1, X_2, \ldots, X_n$ be *iid* random variables, independent copies of X.

Definition 8 *(α-Stable random variable) X is said to follow an α-Stable distribution if there exist a positive constant C_n and a real number D_n such that the following relation holds:*

$$X_1 + X_2 + \ldots + X_n \stackrel{d}{=} C_n X + D_n.$$

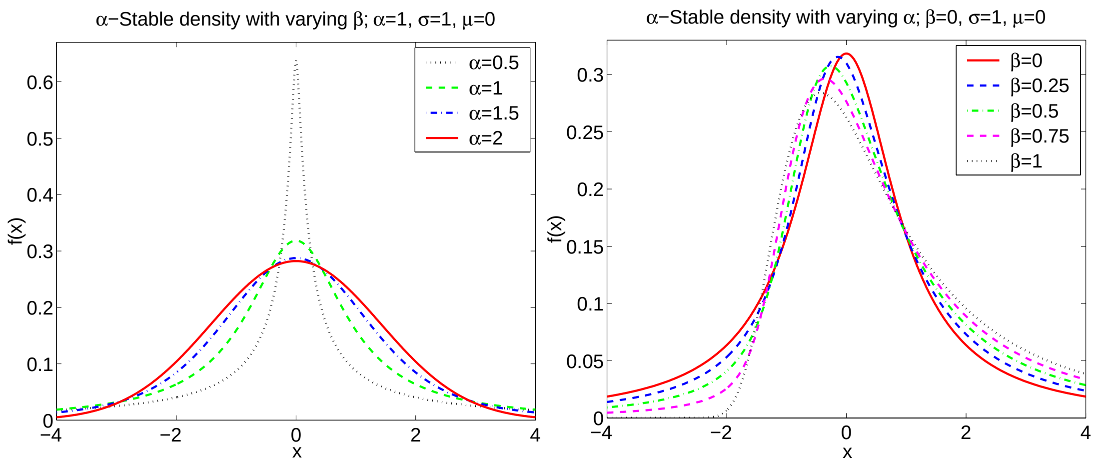
The constant $C_n = n^{1/\alpha}$ dictates the stability property. $\alpha = 2$ refers to the Gaussian case. In subsequent discussions of the α-Stable distributions in this chapter, we restrict ourselves to the *non-Gaussian case*, i.e., $0 < \alpha < 2$. We denote its density by $\mathcal{S}_\alpha(\beta, \sigma, \mu)$.

For the general case, the density does not have a closed form. The distribution is expressed by its characteristic function:

$$\mathbb{E}[e^{itX}] = \begin{cases} \exp\left(-|\sigma t|^\alpha (1 - i\beta(\text{sign } t)\tan\frac{\pi\alpha}{2}) + i\mu t\right), & \alpha \neq 1 \\ \exp\left(-\sigma|t|(1 + i\beta\frac{2}{\pi}(\text{sign } t)\ln|t|) + i\mu t\right), & \alpha = 1, \end{cases} \tag{6.1}$$

[1] α-Stable distributions are often referred to as Stable Paretian distributions.

Figure 6.1: α-Stable densities: effects of changing α and β on the form of the α-Stable density.



(a) Varying α (b) Varying β

where

$$\text{sign } t = \begin{cases} 1 & \text{when} \quad t \geq 0, \\ 0 & \text{when} \quad t = 0, \\ -1 & \text{when} \quad t \leq 0. \end{cases}$$

The distribution is characterized by four parameters $\gamma = \{\alpha, \beta, \sigma, \mu\}$ [147]:[2]

α	$(\alpha \in (0,\ 2))$:	index of stability or the shape parameter,
β	$(\beta \in [-1, +1])$:	skewness parameter,
σ	$(\sigma \in \Re_+)$:	scale parameter,
μ	$(\mu \in \Re)$:	location parameter.

Because of the four parameters, the distribution is highly flexible and suitable for modeling non-symmetric, highly kurtotic, and heavy-tailed data. Figure 6.1

[2]The parameterization of α-Stable distribution is not unique. An overview of the different approaches can be found in [164].

illustrates the effects of the shape and skewness parameters on the shape of the distribution, other parameters kept constant. As is evident from part a) of Figure 6.1, a lower value for α is attributed to heavier tails and higher kurtosis.

The exceptions of closed-form densities are three special cases: the Gaussian case ($\alpha = 2$), Cauchy case ($\alpha = 1, \beta = 0$), and Lévy case ($\alpha = 1/2, \beta = \pm 1$) with the following densities:

Gaussian	$(\alpha = 2)$	$f(x) = \frac{1}{2\sigma\sqrt{\pi}} e^{-\frac{(x-\mu)^2}{4\sigma^2}}$,	$-\infty < x < \infty$,
Cauchy	$(\alpha = 1, \beta = 0)$	$f(x) = \frac{\sigma}{\pi\left((x-\mu)^2+\sigma^2\right)}$,	$-\infty < x < \infty$,
Lévy	$(\alpha = 1/2, \beta = \pm 1)$	$f(x) = \frac{\sqrt{\sigma}}{\sqrt{2\pi}(x-\mu)^{3/2}} e^{-\frac{\sigma}{2(x-\mu)}}$,	$\mu < x < \infty$.

Early applications of the α-Stable distribution to financial data include [68] [70] [69] [71] (stock market prices and portfolio analysis), [137] [135] (financial time series), [43] (non-life insurance), and [31] (real estate market).

6.3 Useful Properties of an α-Stable Random Variable

We briefly present some important properties of the α-Stable distribution [147].[3]

Property 1 *(Power tail) The* power tail decay *property means that the tail of the density function decays like a power function (slower than the exponential decay), which is what allows the distribution to capture extreme events in the tails:*

$$P(|X| > x) \propto C \cdot x^{-\alpha}, \quad x \to \infty,$$

[3]The properties of α-Stable distribution are treated in depth in [147] and [137].

where the constant $C = C_\alpha \frac{1+\beta}{2}\sigma^\alpha$ *with*

$$C_\alpha = \left(\int_0^\infty x^{-\alpha}\sin x dx\right)^{-1} = \begin{cases} \frac{1-\alpha}{\Gamma(2-\alpha)\cos(\pi\alpha/2)} & \text{if} \quad \alpha \neq 1, \\ \frac{2}{\pi} & \text{if} \quad \alpha = 1. \end{cases}$$

Property 2 *(Moments) Raw moments satisfy the property:*

$$\mathbb{E}|X|^p < \infty \quad \text{for any} \quad p \in (0, \alpha),$$

$$\mathbb{E}|X|^p = \infty \quad \text{for any} \quad p \geq \alpha.$$

Property 3 *(Mean) Because of Property 2, the mean is finite only for* $\alpha > 1$*:*

$$\mathbb{E}(X) = \mu \quad \text{for} \quad \alpha \in (1, 2),$$

$$\mathbb{E}(X) = \infty \quad \text{for} \quad \alpha \in (0, 1].$$

Furthermore, the second and higher moments are infinite, implying infinite variance, skewness, and kurtosis.

The next stability property is a useful and convenient property and dictates that the distributional form of the variable is preserved under linear transformations. The stability property is governed by the stability parameter α in the constant C_n (which appeared earlier in the definition of an α-Stable random variable): $C_n = n^{1/\alpha}$. As was stated earlier, smaller values of α refer to a heavier-tailed distribution. The standard Central Limit Theorem (CLT) does not apply to the non-Gaussian case: appropriately standardized large sum of *iid* random variables converges to an α-Stable random variable instead of normal random variable.

Property 4 *(Stability) Suppose that* $X_1, X_2, \ldots, X_n$ *are distributed* $\mathcal{S}_\alpha(\beta_i, \sigma_i, \mu_i)$*,*

$i = 1, 2, \ldots, n$. Then, for some real constant a ($a \neq 0$):

$$
\begin{array}{lcl}
Y := \sum_i^n X_i & \sim & \mathcal{S}_\alpha\left(\frac{\sum_i^n \beta_i \sigma_i^\alpha}{\sum_i^n \sigma_i^\alpha}, \left(\sum_i^n \sigma_i^\alpha\right)^{1/\alpha}, \sum_i^n \mu_i\right); \\
Y := X_1 + a & \sim & \mathcal{S}_\alpha\left(\beta_1, \sigma_1, \mu_1 + a\right); \\
Y := aX_1 & \sim & \begin{cases} \mathcal{S}_\alpha\left((sign\ a)\ \beta_1, |a|\sigma_1, a\mu_1\right) & for \quad \alpha \neq 1, \\ \mathcal{S}_\alpha\left((sign\ a)\ \beta_1, |a|\sigma_1, a\mu_1 - \frac{2}{\pi}a(\ln a)\sigma_1\beta_1\right) & for \quad \alpha = 1. \end{cases} \\
Y := -X_1 & \sim & \mathcal{S}_\alpha\left(-\beta_1, \sigma_1, \mu_1\right).
\end{array}
$$

6.4 Estimating Parameters of the α-Stable Distribution

Since the density of the α-Stable distribution does not exist in closed form, the traditional MLE procedure cannot be applied. Two methodologies are commonly used to estimate the four parameters:

- *Sample Characteristic Function Approach:* Use the observed data to evaluate the sample characteristic function and estimate the unknown parameters such that the distance between the sample and theoretical characteristic functions is minimized.
- *Numerical Approximation of the Density Function Approach:* Approximate the density function using the one-to-one correspondence relation between the characteristic function and the density.

6.4.1 Sample Characteristic Function Approach

Developed by [134], this approach is based on the comparison of the theoretical characteristic function with the sample characteristic function. In the first step of the procedure, for a given sample $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$, the sample characteristic function is calculated as:

$$\hat{\phi}(t) = \frac{1}{n}\sum_{k=1}^{n} e^{itx_k}. \tag{6.2}$$

In the second step, mathematical optimization software is used to fit the theoretical characteristic function to the sample one. The estimates $\hat{\gamma} = \{\hat{\alpha}, \hat{\beta}, \hat{\sigma}, \hat{\mu}\}$ are found so that the distance between the sample and theoretical characteristic functions is minimized. A detailed description of this approach can be found in [110]. See also [100], in particular for the symmetric α-Stable case (see §6.5.1).

6.4.2 Numerical Approximation of the Density Function Approach

Suggested by [54] [55], this approach is based on the numerical approximation of the density function and then using MLE to evaluate the unknown parameters. In the first step of the procedure, the density function is obtained from the characteristic function using the Fourier inversion method:

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-itx}\phi(t)dt. \tag{6.3}$$

This task can be performed using the Fast Fourier Transform (FFT) algorithm [125] [130]. In the second step, MLE of the unknown parameters can be performed using a numerical optimization software, resulting in the choice of the parameter set $\hat{\gamma} = \{\hat{\alpha}, \hat{\beta}, \hat{\sigma}, \hat{\mu}\}$ that maximizes the likelihood function.

6.5 Useful Transformations of α-Stable Random Variables

For $\alpha > 1$ or $|\beta| < 1$, the support of the α-Stable distribution is on $\Re$. It would be unwise to directly apply this distribution to operational loss data that take only positive values. In this light, we suggest using one of three transformations of the α-Stable distribution: symmetric α-Stable distribution, logα-Stable distribution, and truncated α-Stable distribution.

6.5.1 Symmetric α-Stable Random Variable

The symmetric α-Stable distribution (we denote it by $\mathcal{S}_\alpha\mathcal{S}(\sigma)$) is symmetric and centered around zero. Its characteristic function has a simple form:

$$\mathbb{E}[e^{itX}] = e^{-\sigma^\alpha |t|^\alpha}. \tag{6.4}$$

To apply $\mathcal{S}_\alpha\mathcal{S}$ to the operational loss severity data, one can do a simple transformation to the original dataset: $Y = [-X;\ X]$. Then α and σ are the only two parameters of the density $f_X(x) \in \mathcal{S}_\alpha(0, \sigma, 0)$. See Chapter 8 where such distribution is applied to operational loss data.

6.5.2 logα-Stable Random Variable

It is often convenient to work with the natural logarithm transformation of the original data. A random variable X is said to follow a logα-Stable distribution (we denote it by $\log \mathcal{S}_\alpha(\beta, \sigma, \mu)$) if the natural logarithm of the original data follow an α-Stable distribution. Its density is

$$f_X(x) = \frac{g(\log x)}{x}, \quad g \in \mathcal{S}_\alpha(\beta, \sigma, \mu). \tag{6.5}$$

Fitting logα-Stable distribution to data is appropriate when there is reason to believe that the data are very heavy-tailed, and the regular α-Stable distribution may not be sufficient to capture the heavy tails. See Chapter 8 where such distribution is applied to operational loss data.

6.5.3 Truncated α-Stable Random Variable

Another scenario would involve a restriction on the density, rather than a transformation of the original dataset. The support of the α-Stable distribution can be restricted to $\Re_+$ to avoid the possibility of having a positive probability of values below zero in the case when $\beta < 1$. Then, the estimation part would involve fitting a left-truncated α-Stable distribution of the form:

$$f(x) = \frac{g(x)}{1 - G(0)} \times \mathbb{I}_{x>0},$$

where

$$\mathbb{I}_{x>0} = \begin{cases} 1 & \text{if} \quad x > 0 \\ 0 & \text{if} \quad x \leq 0, \end{cases}$$

$g(x)$ is the $\mathcal{S}_\alpha(\beta, \sigma, \mu)$ density, and $G(0)$ is its cdf evaluated at 0. Fitting the left-truncated distribution to the data means fitting the right tail of the distribution.

6.6 Applications to Operational Loss Data

Currently, applications of α-Stable distribution to the operational risk data are limited. Early works suggesting applying the distribution to operational loss data are due to [124] [123].

6.6.1 Empirical Study with 1980-2002 Public Operational Loss Data

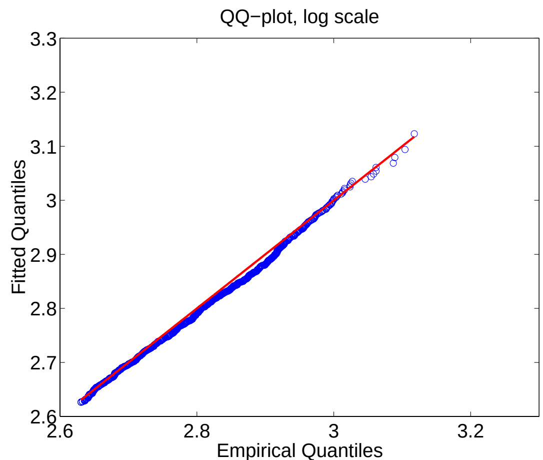
We present results of an empirical application of the α-Stable distribution to modeling operational loss data [39]. The datasets are described in Chapter 5 §5.2 p.31. Histograms of the samples presented in Chapter 5 §5.3.3 reveal the leptokurtic nature of the data: a very high peak is observed close to zero, and an extended right tail indicates the right-skewness and high dispersion of the data values. Heavytailedness is also suggested by the QQ-plots deviating from the straight line in the upper tail and the upward-sloping near-straight-line shape of the mean-excess plots (that suggest Pareto-like tail). This suggests that fitting an α-Stable distribution may be a reasonable approach.

The data are subject to minimum recording thresholds of $1 million in nominal value. Therefore, conditional left-truncated loss distribution was fitted to the data using the method of restricted MLE.[4]

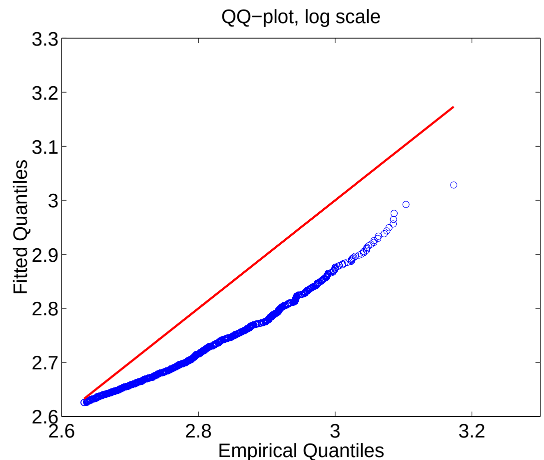
Figures 6.2 and 6.3 illustrate QQ-plots for the fitted logα-Stable and symmetric α-Stable distributions. The former appears to be a remarkably good fit for the "Relationship," "Processes," and "Technology" losses, suggesting that the distribution of these three datasets is severely heavy-tailed. The remaining two loss types – "Human" and "Technology" – are well captured by the symmetric α-Stable distribution. Unfavorable fit of the logα-Stable distribution to the "Hu-

[4] See Chapter 8 for a more comprehensive empirical analysis.

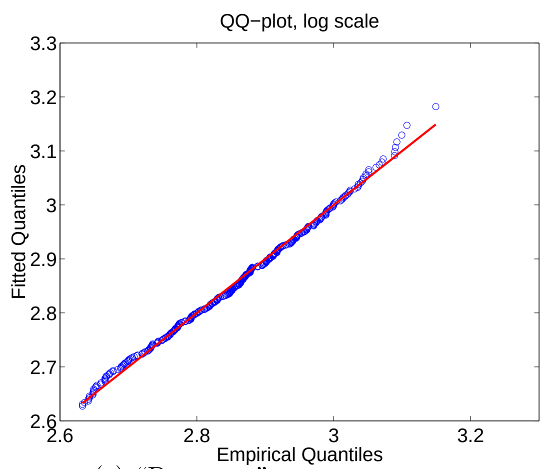
Figure 6.2: QQ-plots (logarithmic scale) for logα-Stable distribution fitted to the 1980-2002 public operational loss data. (a) "Relationship," (b) "Human," (c) "Processes," (d) "Technology," and (e) "External."



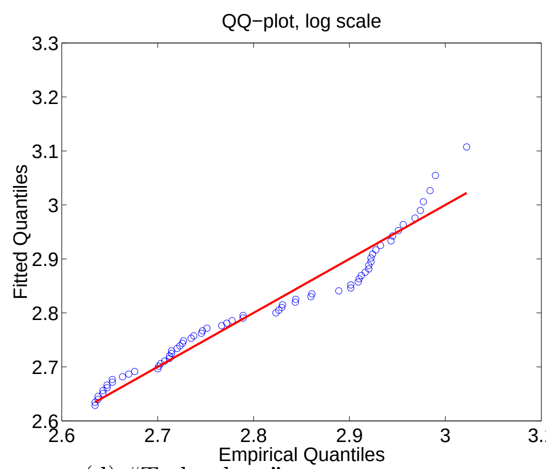
(a) "Relationship"



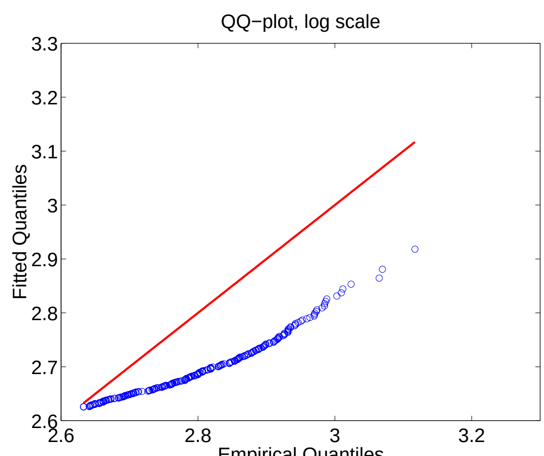
(b) "Human"



(c) "Processes"

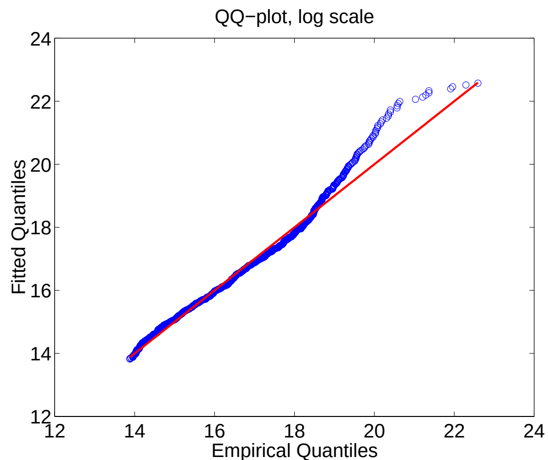


(d) "Technology"

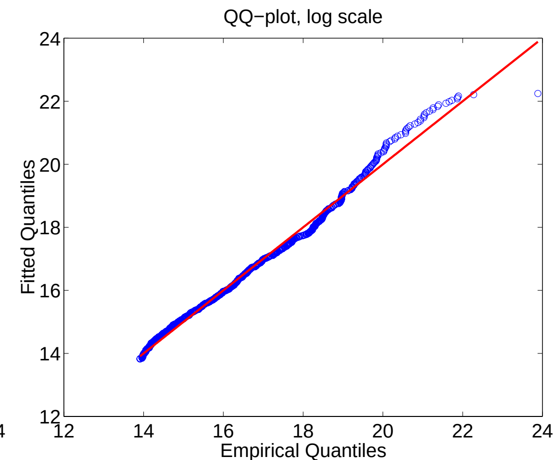


(e) "External"

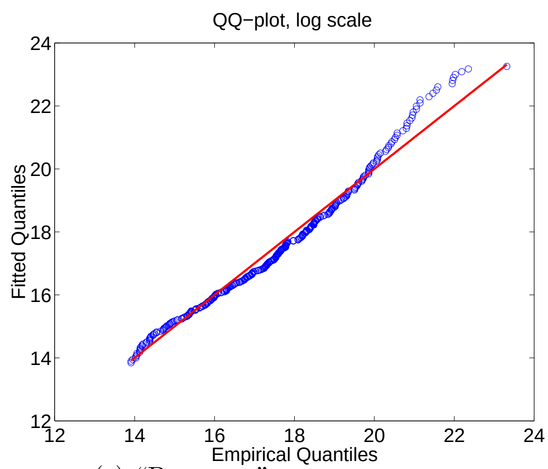
Figure 6.3: QQ-plots (logarithmic scale) for symmetric α-Stable distribution fitted to the 1980-2002 public operational loss data. (a) “Relationship,” (b) “Human,” (c) “Processes,” (d) “Technology,” and (e) “External.”



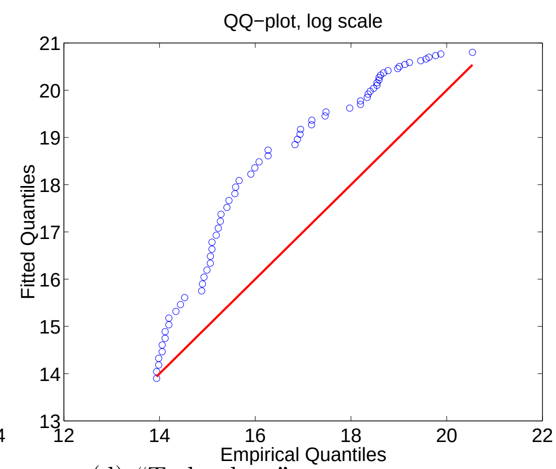
(a) “Relationship”



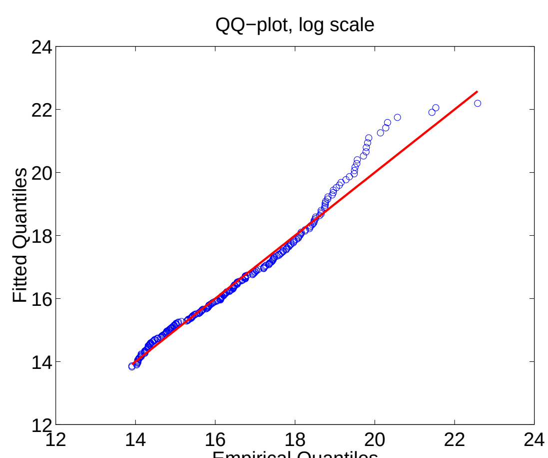
(b) “Human”



(c) “Processes”



(d) “Technology”



(e) “External”

Table 6.1: MLE parameter estimates for logα-Stable and symmetric α-Stable distributions fitted to operational loss data.

	"Relationship"	"Human"	"Processes"	"Technology"	"External"
			$\log \mathcal{S}_\alpha$		
α	1.9340	1.4042	2.0000	2.0000	1.3313
β	-1	-1	0.8195	0.8040	-1
σ	1.5198	2.8957	1.6476	1.9894	2.7031
μ	15.9616	10.5108	17.1535	15.1351	10.1928
			$\mathcal{S}_\alpha\mathcal{S}$		
α	0.6592	0.6061	0.5748	0.1827	0.5905
σ	$1.0 \cdot 10^7$	$0.71 \cdot 10^7$	$1.99 \cdot 10^7$	$0.17 \cdot 10^7$	$0.71 \cdot 10^7$

man" and "Technology" type losses and of the symmetric α-Stable distribution to the "Technology" type losses is likely to be due to Matlab numerical difficulties.

Parameter estimates for logα-Stable and symmetric α-Stable distributions are presented in Table 6.1. Goodness-of-fit (GOF) test statistic and the corresponding p-values for composite GOF tests[5] for these and other loss distributions are summarized in Table 6.2. Their densities are defined in Chapter 8 §8.5.2.

On the basis of the GOF p-values, in majority of cases either logα-Stable or symmetric α-Stable, or even both, resulted in the highest p-values, suggesting the best fit. See also a detailed discussion of the GOF results for a variety of distributions in Chapter 9 §9.4. This supports the conjecture that the overall distribution of operational losses are very heavy-tailed and are well captured by the variations of the α-Stable distribution.

It is notable that due to the infinite second moment (and, for $\alpha \in (0, 1]$, also first moment), Monte Carlo methods may result in generation of samples

[5]See Chapter 9 for the description of the statistics.

Table 6.2: Goodness-of-fit tests for operational loss data. p-values are given in square brackets.

	KS	V	AD	AD_{up}	AD^2	AD^2_{up}	W^2
			"Relationship"				
$\mathcal{E}xp$	11.0868	11.9973	$1.3{\cdot}10^7$	$1.2{\cdot}10^{23}$	344.37	$1.2{\cdot}10^{14}$	50.5365
	[<0.005]	[<0.005]	[<0.005]	[<0.005]	[<0.005]	[<0.005]	[<0.005]
$\mathcal{LN}$	0.8056	1.3341	2.6094	875.40	0.7554	4.6122	0.1012
	[0.082]	[0.138]	[0.347]	[0.593]	[0.043]	[0.401]	[0.086]
$\mathcal{W}eib$	0.5553	1.0821	3.8703	$2.7{\cdot}10^4$	0.7073	24.5068	0.0716
	[0.625]	[0.514]	[0.138]	[0.080]	[0.072]	[0.032]	[0.249]
$\log \mathcal{W}eib$	0.5284	1.0061	3.0718	7332.07	0.4682	10.1322	0.0479
	[0.699]	[0.628]	[0.255]	[0.186]	[0.289]	[0.102]	[0.514]
$\mathcal{GPD}$	1.4797	2.6084	3.5954	374.68	3.7165	22.1277	0.5209
	[<0.005]	[<0.005]	[0.172]	[>0.995]	[<0.005]	[0.048]	[<0.005]
$\mathcal{B}urr$	1.3673	2.4165	3.3069	371.65	3.1371	22.0374	0.4310
	[0.032]	[<0.005]	[0.309]	[0.960]	[<0.005]	[0.019]	[0.011]
$\log \mathcal{S}_\alpha$	1.5929	1.6930	3.8184	1075.30	3.8067	10.1990	0.7076
	[0.295]	[0.295]	[0.275]	[0.041]	[0.290]	[0.288]	[0.292]
$\mathcal{S}_\alpha\mathcal{S}$	1.1634	2.0695	$1.4{\cdot}10^5$	$5.0{\cdot}10^{16}$	4.4723	$2.6{\cdot}10^{14}$	0.3630
	[0.034]	[<0.005]	[>0.995]	[0.971]	[0.992]	[<0.005]	[<0.005]
			"Human"				
$\mathcal{E}xp$	14.0246	14.9145	$2.4{\cdot}10^6$	$1.1{\cdot}10^{22}$	609.15	$3.0{\cdot}10^{12}$	80.3703
	[<0.005]	[<0.005]	[<0.005]	[<0.005]	[<0.005]	[<0.005]	[<0.005]
$\mathcal{LN}$	0.8758	1.5265	3.9829	1086.16	0.7505	4.5160	0.0804
	[0.032]	[0.039]	[0.126]	[0.462]	[0.044]	[0.408]	[0.166]
$\mathcal{W}eib$	0.8065	1.5439	4.3544	$3.2{\cdot}10^4$	0.7908	8.6610	0.0823
	[0.093]	[0.051]	[0.095]	[0.068]	[0.053]	[0.112]	[0.176]
$\log \mathcal{W}eib$	0.9030	1.5771	4.1343	$1.1{\cdot}10^4$	0.7560	4.5125	0.0915
	[0.074]	[0.050]	[0.115]	[0.160]	[0.115]	[0.392]	[0.217]
$\mathcal{GPD}$	1.4022	2.3920	3.6431	374.68	2.7839	23.7015	0.3669
	[<0.005]	[<0.005]	[0.167]	[>0.995]	[<0.005]	[0.051]	[<0.005]
$\mathcal{B}urr$	2.2333	3.1970	4.7780	255.91	7.0968	46.3417	1.2830
	[0.115]	[0.115]	[0.174]	[>0.995]	[0.115]	[0.119]	[0.115]
$\log \mathcal{S}_\alpha$	9.5186	9.5619	36.2617	9846.30	304.61	4198.90	44.5156
	[0.319]	[0.324]	[0.250]	[0.354]	[0.312]	[0.215]	[0.315]
$\mathcal{S}_\alpha\mathcal{S}$	1.1628	2.1537	$5.8{\cdot}10^5$	$4.3{\cdot}10^{17}$	11.9320	$3.3{\cdot}10^{11}$	0.2535
	[0.352]	[0.026]	[0.651]	[0.351]	[0.971]	[0.436]	[0.027]

(Continued on next page)

Table 6.2 *(Continued from previous page)*

	KS	V	AD	AD_{up}	AD^2	AD^2_{up}	W^2
				"Processes"			
$\mathcal{E}xp$	7.6043	8.4160	$3.7{\cdot}10^6$	$1.7{\cdot}10^{22}$	167.61	$6.6{\cdot}10^5$	22.5762
	[<0.005]	[<0.005]	[<0.005]	[<0.005]	[<0.005]	[<0.005]	[<0.005]
$\mathcal{LN}$	0.6584	1.1262	2.0668	272.61	0.4624	4.0556	0.0603
	[0.297]	[0.345]	[0.508]	[0.768]	[0.223]	[0.367]	[0.294]
$\mathcal{W}eib$	0.6110	1.0620	1.7210	2200.75	0.2069	2.2340	0.0338
	[0.455]	[0.532]	[0.766]	[0.192]	[0.875]	[0.758]	[0.755]
$\log \mathcal{W}eib$	0.5398	0.9966	1.6238	658.42	0.1721	1.4221	0.0241
	[0.656]	[0.637]	[0.832]	[0.343]	[0.945]	[0.977]	[0.918]
$\mathcal{GPD}$	1.0042	1.9189	4.0380	148.24	2.6022	13.1082	0.3329
	[<0.005]	[<0.005]	[0.104]	[>0.995]	[<0.005]	[0.087]	[<0.005]
$\mathcal{B}urr$	0.5634	0.9314	1.6075	364.08	0.2639	325.76	0.0323
	[0.598]	[0.800]	[0.841]	[0.429]	[0.794]	[0.844]	[0.840]
$\log \mathcal{S}_\alpha$	0.6931	1.1490	2.0109	272.57	0.4759	328.39	0.0660
	[0.244]	[0.342]	[0.534]	[0.786]	[0.202]	[0.361]	[0.258]
$\mathcal{S}_\alpha\mathcal{S}$	1.3949	1.9537	$3.3{\cdot}10^5$	$2.5{\cdot}10^{17}$	6.5235	$6.8{\cdot}10^{14}$	0.3748
	[0.085]	[0.067]	[0.931]	[0.530]	[0.964]	[0.193]	[0.102]
				"Technology"			
$\mathcal{E}xp$	3.2160	3.7431	27.6434	$1.4{\cdot}10^6$	27.8369	780.50	2.9487
	[<0.005]	[<0.005]	[<0.005]	[<0.005]	[<0.005]	[<0.005]	[<0.005]
$\mathcal{LN}$	1.1453	1.7896	2.8456	41.8359	1.3778	6.4213	0.2087
	[<0.005]	[0.005]	[0.209]	[0.994]	[<0.005]	[0.067]	[<0.005]
$\mathcal{W}eib$	1.0922	1.9004	2.6821	52.5269	1.4536	4.8723	0.2281
	[<0.005]	[<0.005]	[0.216]	[0.944]	[<0.005]	[0.087]	[<0.005]
$\log \mathcal{W}eib$	1.1099	1.9244	2.7553	49.2373	1.5355	5.2992	0.2379
	[<0.005]	[<0.005]	[0.250]	[0.976]	[<0.005]	[0.085]	[<0.005]
$\mathcal{GPD}$	1.2202	1.8390	3.0843	33.4298	1.6182	8.8484	0.2408
	[<0.005]	[<0.005]	[0.177]	[>0.995]	[<0.005]	[0.067]	[<0.005]
$\mathcal{B}urr$	1.1188	1.9374	2.6949	28.4827	2.0320	10.5469	0.3424
	[0.389]	[0.380]	[0.521]	[>0.995]	[0.380]	[0.401]	[0.380]
$\log \mathcal{S}_\alpha$	1.1540	1.7793	2.8728	41.7454	1.3646	6.4919	0.2071
	[<0.005]	[<0.005]	[0.250]	[0.976]	[<0.005]	[0.060]	[<0.005]
$\mathcal{S}_\alpha\mathcal{S}$	2.0672	2.8003	$2.7{\cdot}10^5$	$3.6{\cdot}10^{16}$	19.6225	$7.2{\cdot}10^{10}$	1.4411
	[>0.995]	[>0.995]	[>0.995]	[>0.995]	[>0.995]	[>0.995]	[0.964]

(Continued on next page)

Table 6.2 *(Continued from previous page)*							
	KS	V	AD	AD_{up}	AD^2	AD^2_{up}	W^2
			"External"				
$\mathcal{E}xp$	6.5941	6.9881	$4.4{\cdot}10^6$	$2.0{\cdot}10^{22}$	128.35	$5.0{\cdot}10^7$	17.4226
	[<0.005]	[<0.005]	[<0.005]	[<0.005]	[<0.005]	[<0.005]	[<0.005]
$\mathcal{LN}$	0.6504	1.2144	2.1702	316.20	0.5816	2.5993	0.0745
	[0.326]	[0.266]	[0.469]	[0.459]	[0.120]	[0.589]	[0.210]
$\mathcal{W}eib$	0.4752	0.9498	2.4314	4382.68	0.3470	5.3662	0.0337
	[0.852]	[0.726]	[0.384]	[0.108]	[0.519]	[0.164]	[0.431]
$\log\mathcal{W}eib$	0.6893	1.1020	2.2267	3130.56	0.4711	4.1429	0.0563
	[0.296]	[0.476]	[0.481]	[0.128]	[0.338]	[0.283]	[0.458]
$\mathcal{GPD}$	0.9708	1.8814	2.7742	151.94	1.7091	8.6771	0.2431
	[0.009]	[0.005]	[0.284]	[0.949]	[<0.005]	[0.106]	[<0.005]
$\mathcal{B}urr$	1.3266	2.0385	2.8775	113.13	2.8954	15.4410	0.5137
	[0.050]	[0.048]	[0.329]	[0.989]	[0.048]	[0.064]	[0.048]
$\log\mathcal{S}_\alpha$	7.3275	7.4089	37.4863	4708.71	194.74	3132.60	24.3662
	[0.396]	[0.458]	[0.218]	[0.354]	[0.284]	[0.128]	[0.366]
$\mathcal{S}_\alpha\mathcal{S}$	0.7222	1.4305	$1.1{\cdot}10^5$	$3.4{\cdot}10^{16}$	1.7804	$1.2{\cdot}10^{10}$	0.1348
	[0.586]	[0.339]	[0.990]	[0.797]	[0.980]	[0.841]	[0.265]

with unfeasibly high loss values. This complicates the interpretation of the mean, variance, VaR, and CVaR values. One approach would be to use right-truncated distributions by fixing an upper bound for the admissible support. In the banking industry, such threshold may be determined by, for example, the total value of assets. This remains a topic for future research.

6.6.2 Empirical Study of Inter-Arrival Times with 1950-2002 Public Operational Loss Data

In the second study, α-Stable distribution was fitted to the inter-arrival times. A heavy-tailed inter-arrival times distribution may indicate that there are clusters

Table 6.3: Goodness-of-fit test statistics for the inter-arrival times distribution in empirical study with 1950-2002 public operational loss data.

F	Test	"Relation."	"Human"	"Process."	"Technol."	"External"
$\mathcal{E}xp$	KS	8.1133	6.6606	4.1107	1.5707	4.3121
	AD	39.4554	332.1013	33.4513	19.0028	26.3911
$\mathcal{LN}$	KS	7.2730	6.2306	3.6325	1.5707	4.3121
	AD	26.7241	24.0011	17.4684	16.0836	19.3557
$\mathcal{GPD}$	KS	7.2730	5.4291	3.7369	1.5707	4.3121
	AD	26.1898	11.2737	19.5844	13.0826	24.2408
$\mathcal{S}\alpha$	KS	2.4008	2.3065	1.5376	0.6420	1.2814
	AD	5.3100	5.2508	3.8652	1.8474	2.9118
$\mathcal{C}auchy$	KS	6.2209	6.4260	4.5875	2.1848	3.9933
	AD	42.6918	36.0395	36.3164	26.6083	32.0503
$\mathcal{S}_\alpha\mathcal{S}$	KS	2.9045	2.6095	1.2682	0.5184	1.3712
	AD	5.8113	5.2386	2.5364	1.0479	2.7436

of loss events in time that would violate the simple Poisson assumption on the arrival process.

Table 6.3 reports the goodness-of-fit statistics for several distributions fitted to the inter-arrival times. Since the exact date of occurrence was unavailable for some data points, only those for which such information was available were included into the analysis. This resulted in sample sizes for the "Relationship," "Human," "Processes," "Technology," and "External" data types of 554, 597, 199, 31, and 160 points, respectively. The Cauchy and symmetric α-Stable distributions were fitted to the symmetrized data. It is clear from the Table 6.3 that, based on the low KS statistic values, the α-Stable and symmetric α-Stable distribution fit the data very well compared with alternative candidate distributions. In particular, the Exponential distribution shows a poor fit. This may be an indication that frequency distribution does not follow a homogeneous Poisson process and exhibits a pronounced trend of clustering of events. See also discussion in Chapter 7.

Chapter 7

Modeling Operational Frequency with Cox Processes

7.1 Introduction

We discussed in Chapter 5 that the loss arrival process for operational risk possesses an irregular nature. It is reasonable to assume that, in most situations, operational risk-related events arrive independently from each other. A common model to characterize such a process is a Poisson process. In the simplest scenario, the mean number of events per unit of time is constant in time. In practice, however, it is plausible to expect that the mean number of events in a given time interval does not remain constant, but behaves in a more random fashion or even evolves and changes with time.

In this chapter we review some properties of simple homogeneous and non-homogeneous Poisson processes.

7.2 Cox Processes

We begin with a definition of a Poisson point process.

Definition 9 *(Poisson process) A stochastic process $N_t, t \geq 0$ is called Poisson if it satisfies the following properties:*

1. *N_t has independent increments, i.e., for any $n \in \mathbb{N}$, any $t_0, t_1, \ldots, t_n$ such that $0 \leq t_0 \leq t_1 \leq \ldots \leq t_n < \infty$, the random variables $N_{t_1} - N_{t_0}, N_{t_2} - N_{t_1}, \ldots, N_{t_n} - N_{t_{n-1}}$ are independent;*

2. *N_t is homogeneous, i.e., for any $s \geq 0, t \geq 0$ and $h > 0$, the random variables $N_{t+h} - N_t$ and $N_{s+h} - N_s$ are identically distributed;*

3. *$N_0 = 0$;*

4. *the number of jumps in an interval t is Poisson distributed with mean λt, $\lambda > 0$, i.e., for all $t, s > 0$*

$$P(N_{t+s} - N_s = n) = \frac{(\lambda t)^n e^{-\lambda t}}{n!}, \quad n = 0, 1, \ldots. \tag{7.1}$$

If λ (i.e., the intensity rate) is a constant, we have a *homogeneous Poisson process* (HPP) that has a cumulative intensity λt. The mean of a homogeneous Poisson distribution equals the variance. When λ is not constant, we have a *non-homogeneous Poisson process* (NHPP) or *Cox process* that have a cumulative intensity $\Lambda(t)$. Because the process itself and the measure λ are stochastic, Cox processes are often called *doubly stochastic Poisson processes*.

Definition 10 *(Random measure) A random process $\Lambda(t)$, $t \geq 0$ is called a random measure if it has nondecreasing sample paths and satisfies:*

1. *$\Lambda(0) = 0$;*

2. $\Lambda(t) < \infty$ *a.s. for* $0 < t < \infty$.

Definition 11 *(Cox process) A stochastic process* N_t, $t \geq 0$ *is called a Cox process (or an non-homogeneous Poisson process) with intensity measure* $\Lambda(t)$ *if*

- N_t *has independent increments;*
- *The increment* $N_t - N_s$ *such that* $0 \leq s < t < \infty$ *has the Poisson distribution with parameter* $\Lambda(t) - \Lambda(s)$.

The cumulative intensity $\Lambda(t)$ can be conveniently expressed as:

$$\Lambda(t) = \int_0^t \lambda(\tau) d\tau, \quad \tau > 0, \tag{7.2}$$

for a non-negative instantaneous intensity function $\lambda(\tau)$.

Properties of Cox processes are analyzed in [81] [82] [83] [16], among others. We stress that a necessary requirement for applying such advanced models to operational loss data is the availability of extensive datasets.

7.2.1 Mixed Poisson Processes

If $\Lambda(\cdot)$ has a distribution, then N_t is a *mixed Poisson process*. Due to a non-constant intensity rate, mixed Poisson models allow for an extra variability of the underlying Poisson random variable. For example, it can be easily verified that a mixed Poisson distribution with the intensity rate distributed as a Gamma random variable is a Hypergeometric random variable; the variance of a Hypergeometric random variable is greater than the mean.

7.2.2 Non-Homogeneous Poisson Process with Stochastic Intensity

Unlike mixture distributions in which λ follows a particular distribution of its own, in non-homogeneous Poisson processes with stochastic intensity λ is believed to evolve with time in a fashion that can be expressed by a mathematical function, $\lambda(t)$. For example, a possible cyclical component in the time series of the number of loss events may be captured by a sinusoidal rate function, an upward-sloping tendency may be captured by a quadratic function, and so on. Moreover, deviations from an assumed (or fitted) deterministic model may be further captured by a random stochastic process, such as Brownian motion.

We propose the following algorithms that allow one to determine an optimal stochastic model for operational loss frequency distribution. Later in this chapter we will present two examples with loss data that demonstrate the success of such algorithms.

Algorithm 1:

1. Split the total time frame $[0, t]$, into m small time intervals of equal length, such as days, months, or quarters;[1]

2. Calculate the total number of loss events that has occurred within each interval;

3. Construct the following plot:

   a) On the horizontal axis that represents time, locate the numbers $1 : m$;

   b) On the vertical axis that represents number of events, locate the *cumulative* number of loss events;

[1] Note that the frequency distribution will change depending on the chosen intervals.

4. The resulting plot represents the cumulative intensity function $\Lambda(t) = \int_0^t \lambda(s)ds$;

5. Choose a function that fits best the plot, using the mean square error (MSE) or the mean absolute error (MAE) minimization technique.

Algorithm 2:

1. Sort the data so that the dates of the loss events are in increasing order;

2. Calculate the inter-arrival times between the dates in days and then divide by 365[2] in order to express the inter-arrival times in terms of years;

3. Construct the following plot:

   a) Split the horizontal axis that represents total time frame $[0, t]$, into $n - 1$ small time intervals, the intervals being the *cumulative* inter-arrival times between n total number of loss events;

   b) On the vertical axis that represents number of events, locate the numbers $1 : n$;

4. The resulting plot represents the cumulative intensity function $\Lambda(t) = \int_0^t \lambda(s)ds$;

5. Choose a function that fits best the plot, using the Mean Squared Error (MSE) or the Mean Absolute Error (MAE) minimization technique.

7.3 Renewal Processes

Sometimes, instead of working with the distribution of the number events in a fixed time interval, it may be more convenient to work with the inter-arrival times' distribution.

[2] We use 365 because we are interested in the actual number of days between the occurrences of the events, rather than the number of working days in which case we would use 250 days instead.

Definition 12 *(Renewal process) Let* $\{Y_j\}_{j\geq 1}$ *be independent distances between successive points of a point process* N_t*. If the random variables* $\{Y_j\}_{j\geq 1}$ *are identically distributed, then* N_t *is a renewal process.*

Cox process can be represented with renewal process. Let $V_k = Y_1 + Y_2 + \ldots + Y_k,\ k \geq 1$. Let N_t^1 be a standard Poisson process so that $N_t^1 = N(\Lambda^{-1}(t))$ where $\Lambda^{-1}(t) = \sup\{s :\ \Lambda(s) \leq t\}$. Then $N_t^1 = \sup\{k \geq 1 :\ \Lambda(V_k) \leq t\}$ and $V_k^* := \Lambda(V_k)$ are the jump points of N_t^1. Hence, N_t that is controlled by a random measure $\Lambda(t)$ is a Cox process if and only if the random variables $V_1^*, V_2^* - V_1^*, V_3^* - V_2^*, \ldots$ are *iid* Exponentially distributed.

Modeling operational frequency distribution with renewal processes may plausible when we have reasons to believe that the operational loss events are independent, there is no clustering of events, and there occur no regime switching in time.

7.4 Empirical Studies with Operational Frequency Data

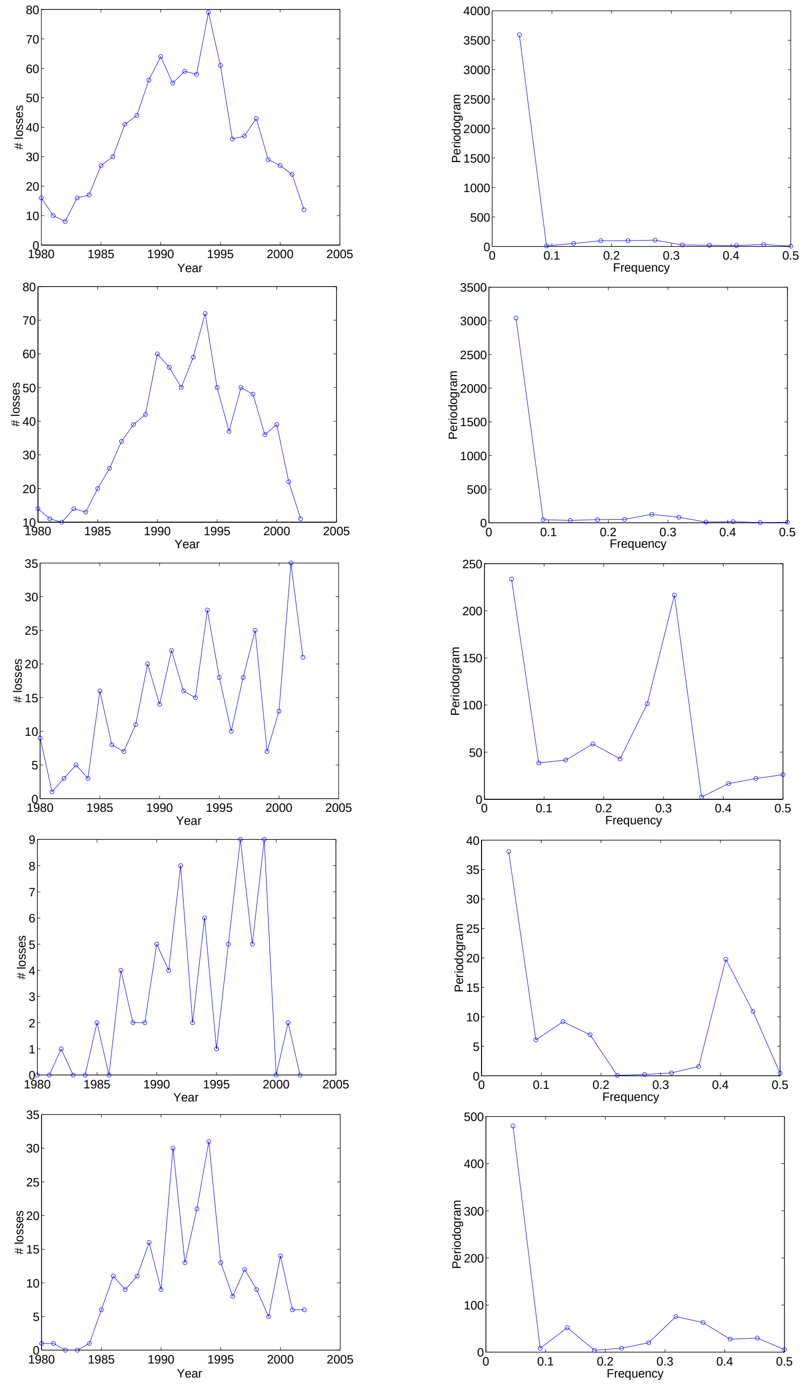
A simple Poisson assumption is prevalent in empirical studies with operational loss data. See, for example, [34] [56] [46] [13] [38] [114] [49] [50]. Hypergeometric distribution has been examined by [46] [127] [50].

7.4.1 Empirical Study with 1980-2002 Public Operational Loss Frequency Data

We examine operational loss data for the period 1980 to 2002 for five loss types obtained from a major European data provider.[3] For the description of the dataset,

[3]Details of this study are presented in [39].

Figure 7.1: Annual number of losses (left) and periodogram (right) of the frequency distribution of 1980-2002 public operational loss data. (a) "Relationship," (b) "Human," (c) "Processes," (d) "Technology," and (e) "External." Periodograms for "Processes" and "Technology" losses reveal distinct peaks at frequency 0.32 and 0.41, suggesting a period of 1/0.33=3.03 years and 1/0.41=2.41 years, respectively.



(a) "Relationship"

(b) "Human"

(c) "Processes"

(d) "Technology"

(d) "External"

we refer the reader to Chapter 5 §5.2 p.31. The time series of the annually aggregated frequency data for each year spanned by the data is illustrated in Figure 7.1, left column.

Following Algorithm 1 described in §7.2.2, we aggregate loss events on the annual basis. There is some evidence of a persistent cyclical component present for the "Processes" and "Technology" type data (see Figure 7.1 right column), with a period of approximately 3.03 years for the "Processes" type and a period of 2.41 years for the "Technology" type. Hence, for the two processes it may be relevant to fit a rate function of a sinusoidal type.[4]

Visual inspection of the plot suggests that, as a general trend, the accumulation resembles a continuous cdf-like process.[5] Including further a cyclical (e.g., sinusoidal) component for the "Processes" and "Technology" type data would involve a greater number of parameters that may result in overfitting. We hence agree to omit the analysis of the cyclicality in the data and focus on the general trend.[6] As a result, we fit two NHPP models with a deterministic intensity function to each of the five datasets:

$$\begin{array}{lll} \text{NHPP Type I} & \text{(Lognormal cdf-like)} & \Lambda(t) = a + b\exp\left\{-\frac{(\log t - d)^2}{2c^2}\right\}(2\pi)^{-1/2}c^{-1}; \\ \text{NHPP Type II} & \text{(Logweibull cdf-like)} & \Lambda(t) = a - b\exp\left\{-c\log^d t\right\}. \end{array}$$

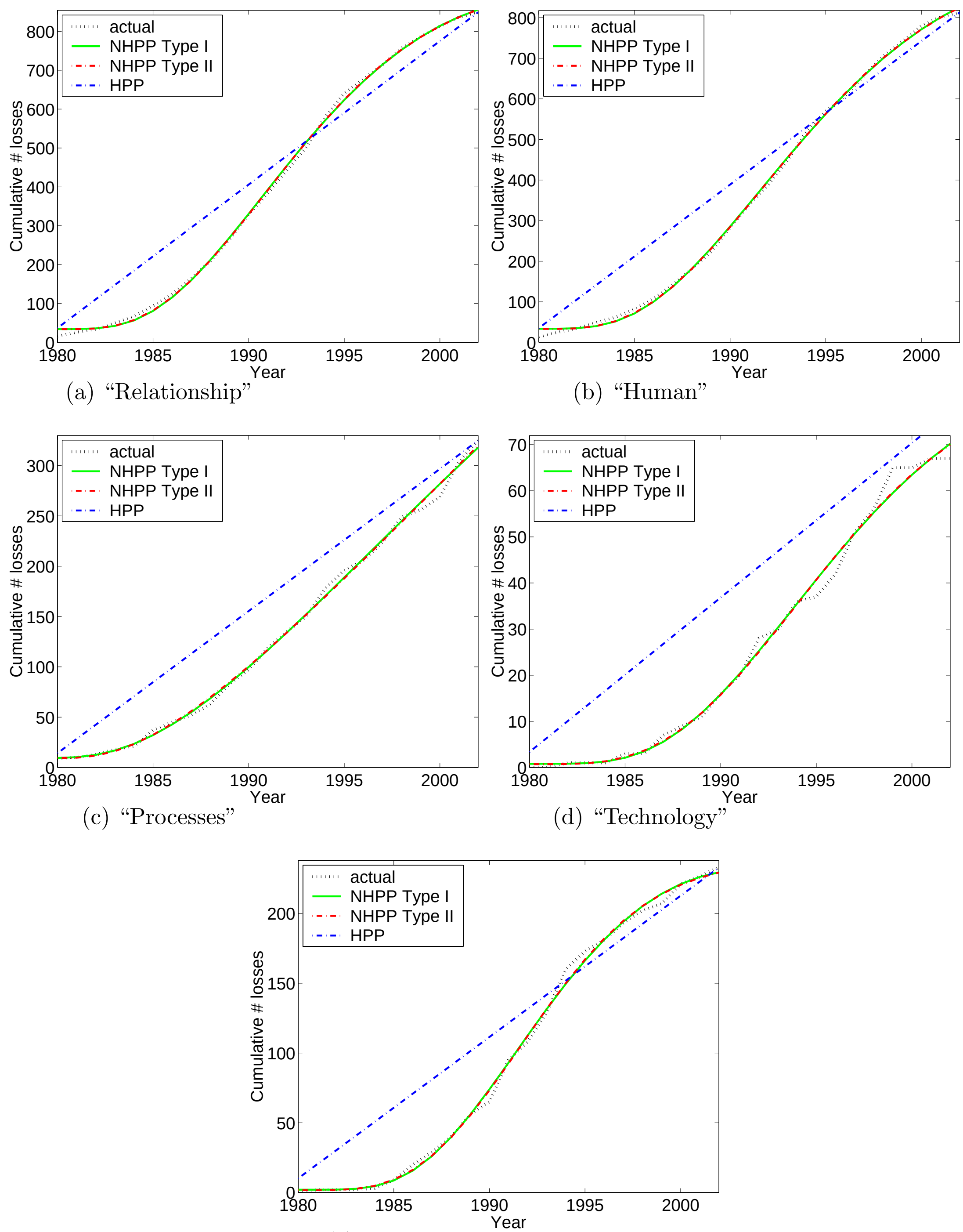
Figure 7.2 plots actual annual frequency of loss events with the fitted HPP and NHPP. Furthermore, Table 7.1 shows the parameter and error estimates for NHPP of both types fitted to the operational losses and compares the fit to HPP.

[4]See, for example, [37] for an example of a sinusoidal NHPP fitted to quarterly aggregated U.S. natural catastrophe frequency data.

[5]Certainly, on a longer time horizon, this would result in a near-zero instantaneous frequency rate function. However, in this particular study, such model appears plausible for this dataset and time frame.

[6]Nevertheless, the presence of cyclicality in operational loss data is an important finding and requires further investigation. This remains a topic for our future research.

Figure 7.2: Frequency distributions of 1980-2002 public operational loss data: cumulative number of loss events over time. The plots reveal a non-homogeneous nature of loss occurrence. (a) “Relationship,” (b) “Human,” (c) “Processes,” (d) “Technology,” and (e) “External.”



(a) “Relationship” (b) “Human”

(c) “Processes” (d) “Technology”

(e) “External”

Table 7.1: Parameter estimates, MSE, and MAE corresponding to fitted NHPP Type I and Type II and HPP processes.

Loss/Process	Parameter Estimates				MSE	MAE
"Relationship"						
NHPP Type I	a	b	c	d		
	34.13	1364.82	0.63	3.32	76.57	7.05
NHPP Type II	a	b	c	d		
	930.29	896.17	0.0010	6.82	69.08	6.57
HPP				λ		
				36.91	5907.45	65.68
"Human"						
NHPP Type I	a	b	c	d		
	33.49	1436.56	0.65	3.43	68.05	6.89
NHPP Type II	a	b	c	d		
	950.20	917.11	0.0008	6.80	61.59	6.60
HPP				λ		
				35.35	6600.38	65.33
"Processes"						
NHPP Type I	a	b	c	d		
	9.44	2098.96	1.04	4.58	22.50	3.64
NHPP Type II	a	b	c	d		
	2034.25	2024.77	0.0007	4.79	23.06	3.65
HPP				λ		
				14.13	1664.82	36.57
"Technology"						
NHPP Type I	a	b	c	d		
	0.79	120.20	0.58	3.47	3.71	1.28
NHPP Type II	a	b	c	d		
	137.68	138.39	0.0006	6.32	4.89	1.67
HPP				λ		
				3.35	217.04	13.42
"External"						
NHPP Type I	a	b	c	d		
	2.02	305.91	0.53	3.21	16.02	2.71
NHPP Type II	a	b	c	d		
	237.66	235.94	0.00025	8.30	14.55	2.74
HPP				λ		
				10.13	947.32	24.67

Clearly, NHPP results in a superior fit as is evident from lower error estimates.

In Chapter 6 §6.6.2 we also presented an empirical study with the inter-arrival times data for the operational losses. It was concluded that the α-Stable and symmetric α-Stable distributions fitted the data best. We refer the reader to p. 66 for further discussions of this study.

Chapter 8

Truncated Loss Models

8.1 Introduction

In an ideal scenario, the data collection process results in all operational loss events being detected and duly recorded. However, the data recording is subject to lower recording thresholds, so that only data above a certain amount enter databases. In this sense, the data available for estimation appears to be left-truncated. Left-truncation of the data must be appropriately addressed in the estimation process, in order to determine a correct capital charge.

In this chapter, we discuss a methodology for the estimation of the parameters of the severity and frequency distributions when some operational loss data are missing from the dataset. We further explore the implications of using a wrong approach (under which the truncation is ignored) and correct approaches (under which the truncation is adequately addressed) on the resulting capital charge and present results of related empirical studies.

8.2 Reporting Bias Problem

Operational loss data are subject to minimum collection threshold which is a fixed pre-determined amount: approximately \$10,000 for banks' internal databases and approximately \$1 million for public (or external) databases; see, e.g., [27]. This creates a so-called *reporting bias*.

There are several reasons for setting such minimum threshold.

1. Data recording may be costly. When the threshold is decreased linearly, the costs of recording data increase exponentially. Furthermore, a large number of small losses may be recorded with mistakes that can result in additional operational losses.

2. Smaller losses are easier to hide, while it is harder to hide larger losses. For example, a trader who has committed a trading error may falsify document and succeed in making small-magnitude losses go unnoticed by bank management. However, large amount are much more difficult to hide.

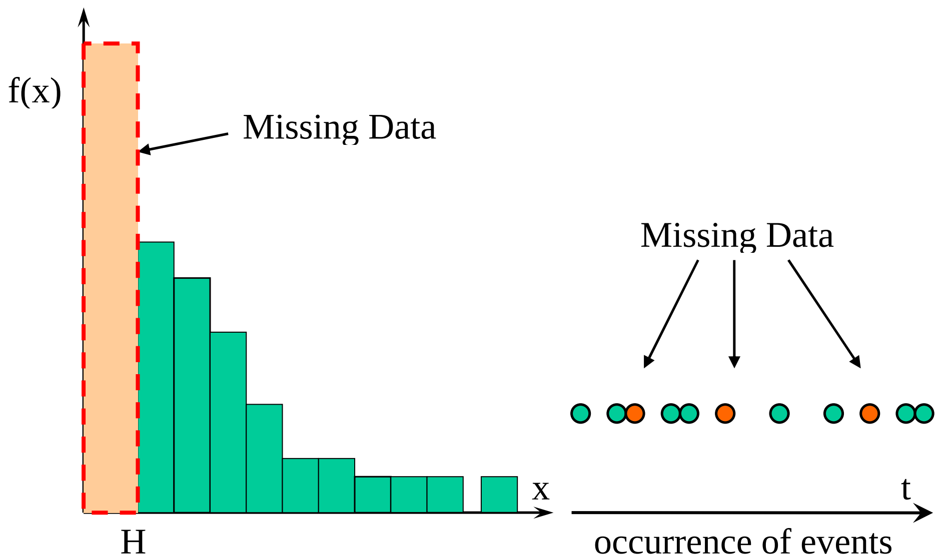
3. Poor operational loss data recording practices in past years may have been such that smaller losses could be left unrecorded, while larger losses were properly reported and recorded.

8.3 Truncated Model for Operational Risk

In the presence of missing data, the recorded operational losses follow a *truncated* compound Poisson process.[1] Correctly specifying the loss and frequency distribution is the key to a correct estimation of the capital charge, given that the assumptions on the severity and frequency are met.

[1] Compound models for operational risk are reviewed in Chapter (Cox).

Figure 8.1: Illustration of incomplete loss and frequency data when the loss data are left-truncated.



8.3.1 Data Specification

Let us denote the minimum threshold above which losses are being recorded by H. The histogram of such observed loss data would represent only the right tail of a fuller loss distribution rather than the entire distribution. Figure 8.1 illustrates the idea.

We identify two distinct approaches – misspecified (the one most frequently used by practitioners) and correctly specified – that one may use:

1. *"Naive" Approach.* Under the "naive" misspecified approach one would treat the observed data as complete and fit an unconditional loss distribution directly to the observed data. The observed frequency is treated as the true frequency distribution.

2. *Conditional Approach.* Under the conditional approach one correctly specifies the distribution by noting the fact that the data are recorded only above

H. The frequency of the observed data is below the true frequency of loss events, and needs to be rescaled.

The first approach, "naive" approach, is incorrect because it ignores the missing data in the lower quantiles of the loss distribution, i.e., $(0, H)$, and both the severity and frequency distributions are misspecified.

The second approach, conditional approach, relies on three assumptions:

Assumption 1 *There is no prior information (neither regarding the frequency nor the severity) on the missing data.*

Assumption 2 *Missing data and recorded data belong to the same family of distributions with identical parameters.*

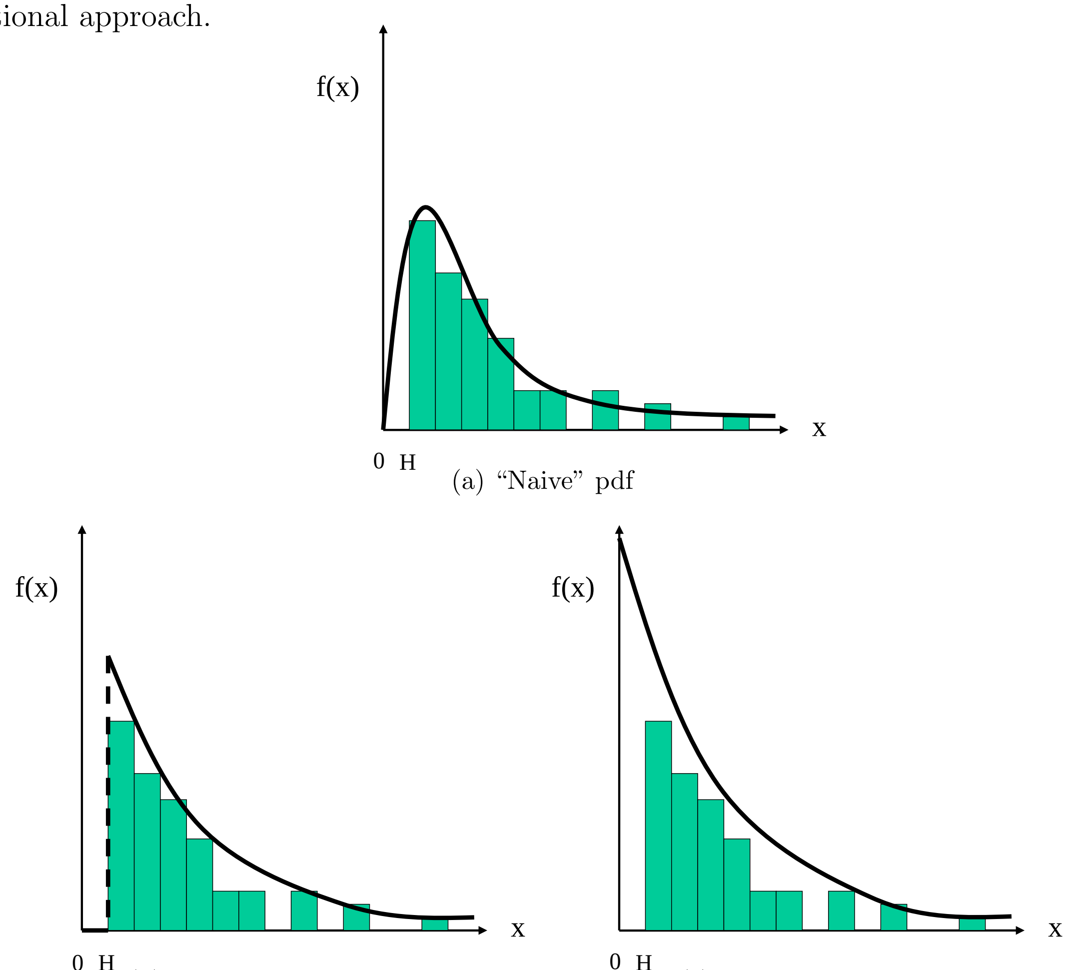
Assumption 3 *Loss magnitudes are independent from the frequency of loss occurrence and can be treated as two independent random processes.*

Under the conditional approach, one needs to fit directly the right tail of the distribution and correctly "scale up" the frequency. This will be treated in §8.3.2. Figure 8.2 illustrates an exemplary histogram of operational loss data with fitted "naive" density (part a), conditional density (part b), and the complete correctly specified density (part c). We will discuss the conditional approach in detail in the following section.

8.3.2 Parameter Estimation

Throughout this section, we follow the notations as those in [38]. Suppose that the available data set collected in the time frame $[T_1, T_2]$ are incomplete due to the non-negative pre-specified threshold H that defines a partition on $\Re_{>0}$ through the events $A_1 = (0, H)$ and $A_2 = [H, \infty)$. Realizations of the losses in A_1 do

Figure 8.2: Illustration of the "naive" approach and the conditional approach. Panel (a) portrays the density estimated under the "naive" approach; panel (b) portrays the conditional density estimated under the conditional approach; panel (c) portrays the unconditional complete-data density estimated under the conditional approach.



(a) "Naive" pdf

(b) Conditional pdf

(c) Complete pdf

not enter the data sample, with neither the frequency nor the severity being recorded. Realizations in A_2 are fully reported, with both the frequency and the loss amounts being specified. Hence, observations in A_1 constitute the missing data, and those in A_2 the observed left-truncated data. The observed sample is of the form $\mathbf{z} = (n, \mathbf{x})$, where n is the number of observations in A_2 and $\mathbf{x}$ is the corresponding observed sample $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$ in A_2. Given that the total number of observations in the complete sample is unknown, *joint* density of $\mathbf{z}$ can be expressed as:

$$g_{\lambda,\gamma}(\mathbf{z}) = \frac{(\triangle t\ \tilde{\lambda})^n}{n!} e^{-\triangle t\ \tilde{\lambda}} \cdot \prod_{j=1}^{n} \frac{f_\gamma(x_j)}{q_{\gamma,2}}, \tag{8.1}$$

where $q_{\gamma,j}$ denotes the probability for a random realization to fall into set A_j, $j = 1, 2$, $\tilde{\lambda}$ is the observed intensity[2] related to the complete-data intensity λ by $\tilde{\lambda} := q_{\gamma,2} \cdot \lambda$, and $\triangle t := T_2 - T_1$ is the length of the sample window. In the representation (8.1), the Poisson process $\tilde{N}_t$ with intensity $\tilde{\lambda}$ that counts only the observed losses exceeding H can be thus interpreted as a thinning of the original process N_t with intensity λ that counts all events in the complete data sample. Due to Assumption 3 of §8.3.1, the maximization of the corresponding log-likelihood function with respect to λ and γ can be divided into two separate maximization problems, each depending on only one parameter:

$$\hat{\gamma}_{\text{MLE}} = \arg\max_{\gamma} \log g_{\lambda,\gamma}(\mathbf{z}) = \arg\max_{\gamma} \log \left(\prod_{j=1}^{n} \frac{f_\gamma(x_j)}{q_{\gamma,2}} \right). \tag{8.2}$$

Given that the $\tilde{N}_t \sim Poisson(\tilde{\lambda})$, using the law of conditional probabilities the complete-data intensity rate can be then obtained by:

$$\hat{\lambda}_{\text{MLE}} = \arg\max_{\lambda} \log g_{\lambda,\hat{\gamma}_{\text{MLE}}}(\mathbf{z}) = \frac{n}{\triangle t \cdot q_{\hat{\gamma}_{\text{MLE}},2}}. \tag{8.3}$$

Assumptions 1, 2, and 3 presented in §8.3.1 guarantee that the models in Equations (8.2) and (8.3) produce correct estimates for the complete loss and frequency distributions.

The MLE estimation of γ can be done in two ways: performing direct numerical

[2] In case of a NHPP, $\lambda \triangle t$ of a HPP is replaced with $\Lambda(t) = \int_0^t \lambda(s) ds$.

maximization of the Constrained Likelihood Function and using the Expectation-Maximization algorithm.

Constrained Maximum Likelihood Function Approach

In Constrained Maximum Likelihood Function approach, one can obtain the parameters of the loss distribution by directly maximizing the likelihood function, as described in Equation (8.2). The frequency parameter can be then obtained using Equation (8.3).

Expectation-Maximization Algorithm Approach

This Expectation-Maximization (EM) Algorithm approach, is aimed at estimating the unknown parameters by maximizing the expected likelihood function using available information on the observed and missing data; see, e.g., [51].

The EM algorithm is a two-step iterative procedure. In the initial step, given an initial guess value $\gamma^{(0)}$ for the unknown parameter set γ, the missing data values in the log-likelihood function are replaced by their expected values. This leads to the guess value for the expected complete log-likelihood function (Expectation step) which is further maximized with respect to the parameter values (Maximization step). The solution is then used as the initial guess in the next iteration of the algorithm, and the Expectation step and the Maximization step are repeated, and so on. For the purpose of clarity, we simplify our earlier notations. The EM algorithm can be thus summarized as follows:

- *Initial step*: choose initial (prior) values $\gamma^{(0)}$. These can be used to estimate the initial guess value $m^{(0)}$ representing the number of missing data.
- *Expectation step* (E-step): Given $\gamma^{(0)}$, calculate the expected log-likelihood

function of complete data. Mathematically, for the left-truncation point H,

$$\mathbb{E}_{\gamma^{(0)}}\left[\log L_\gamma(\mathbf{x}^{\text{complete}}) \mid \mathbf{x}^{\text{observed}} \geq H\right] =$$
$$m^{(0)}\mathbb{E}_{\gamma^{(0)}}\left[\log f_\gamma(x^{\text{missing}})\right] + \sum_{j=1}^{n} \log f_\gamma\big(x^{\text{observed}}\big). \tag{8.4}$$

- *Maximization step* (M-step): find the (posterior) parameter set γ which maximizes the expected log-likelihood function from the previous step and set it equal to the guess value in the next step $\gamma^{(1)}$. Mathematically,

$$\gamma^{(1)} := \arg\max_{\gamma} \mathbb{E}_{\gamma^{(0)}}\left[\log L_\gamma(\mathbf{x}^{\text{complete}}) \mid \mathbf{x}^{\text{observed}} \geq H\right]. \tag{8.5}$$

- *Iteration*: repeat E-step and M-step – the sequence $\{\gamma_k\}_{k>0}$ will converge to the desired Maximum Likelihood estimates $\hat{\gamma}_{\text{MLE}}$ of the distribution describing the complete data sample.

Using the expressions for MLE estimates and rearranging terms, it can be easily shown that for $X \sim LN(\mu, \sigma)$ the$(k+1)$th step yields the new estimates:

$$\begin{aligned} \widehat{\mu}_{\text{MLE}}^{(k+1)} &= (1 - F_{\hat{\gamma}_{\text{MLE}}}(H)^{(k)})\tfrac{1}{n}\sum_{j=1}^{n}\log(x_j) + \int_0^H \log(x) f_{\hat{\gamma}^{(k)}}(x)dx \\ \widehat{\sigma^2}_{\text{MLE}}^{(k+1)} &= (1 - F_{\hat{\gamma}_{\text{MLE}}}(H)^{(k)})\tfrac{1}{n}\sum_{j=1}^{n}\log^2(x_j) + \int_0^H \log^2(x) f_{\hat{\gamma}^{(k)}}(x)dx - \widehat{\mu^2}_{\text{MLE}}^{(k+1)}. \end{aligned} \tag{8.6}$$

Table 8.1 illustrates the convergence results with the EM algorithm for a simple problem with five observed values $\tilde{y} = \{20, 23, 25, 30, 50\}$ and a threshold level $H = 15$. As the start values for the algorithm we chose the MLE estimates of the unconditional Lognormal density, $\hat{\mu} = 3.3327$ and $\hat{\sigma^2} = 0.1011$. This yields

Table 8.1: Iteration results of EM-algorithm for an example with Lognormally distributed sample $\{20, 23, 25, 30, 50\}$. Convergence was achieved after 55 iterations.

iteration step k	$\mu^{(k)}$	$\sigma^{2(k)}$	$\log L_{obs}^{(k)}$
0	3.3327	0.1011	-18.0299
1	3.332665	0.1011401	-17.58342
2	3.314242	0.1129907	-17.40839
3	3.305652	0.1182587	-17.33105
4	3.301268	0.1209048	-17.29199
5	3.29894	0.1223007	-17.27129
6	3.29768	0.1230544	-17.26009
7	3.29699	0.1234662	-17.25396
8	3.29661	0.1236928	-17.25059
9	3.2964	0.1238178	-17.24873
10	3.296284	0.1238869	-17.24770
11	3.29622	0.1239251	-17.24713
12	3.296185	0.1239463	-17.24681
13	3.296165	0.123958	-17.24664
14	3.296154	0.1239645	-17.24654
15	3.296148	0.1239681	-17.24649
16	3.296145	0.1239701	-17.24646
17	3.296143	0.1239712	-17.24644
18	3.296142	0.1239719	-17.24643
19	3.296141	0.1239722	-17.24643
20	3.296141	0.1239724	-17.24642
21	3.296141	0.1239725	-17.24642
22	3.296141	0.1239725	-17.24642
23	3.29614	0.1239726	-17.24642
⋮	⋮	⋮	⋮
55	3.29614	0.1239726	-17.24642

log-likelihood value of -18.0299. The algorithm iterates until

$$\delta = \max_k \{\mu^{(k)} - \mu^{(k-1)},\, \sigma^{2(k)} - \sigma^{2(k-1)}\} < 0.1 \cdot 10^{-15}, \tag{8.7}$$

i.e., until the MLE estimates converge (the value was chosen arbitrarily). The posterior MLE estimates appear in the bottom row of the table. The last column shows the log-likelihood values for the observed portion of the data.

Remark 1 *The convergence of the EM-algorithm to the posterior MLE estimates is achieved regardless of the choice of the initial (prior) guess values.*

Remark 2 *Because in every round of the EM-algorithm the unknown parameters are replaced with the values that are closer to the true values, at every round the value of the likelihood function increases relative to the previous round. (See Figure for the example with Lognormal distribution.)*

The estimated parameters of the severity distribution can be then used in $q_{\hat{\gamma}_{\mathrm{MLE}},2}$ to rescale the intensity rate function in the frequency distribution.

Consequences of Missing Data: Lognormal Example

Many empirical studies found in operational risk literature ignore the missing data and use the "naive" approach. What are the implications of using the "naive" approach on the operational capital charge? We illustrate the implications using a simple example with Lognormally distributed losses.

Our first question is the impact of using the wrong approach on the parameter estimates. Suppose we estimate the parameters of Lognormal distribution to be $\hat{\mu}$ and $\hat{\sigma}$ and the observed intensity rate of the Poisson process is $\hat{\lambda}$[3] under the "naive" approach, while the correct parameters (for the complete data) are μ, σ, and λ, respectively. Then, simple calculations will produce the following bias estimates (i.e., the difference between the estimated parameter value and the true

[3]For a non-homogeneous intensity rate $\lambda(t)$ the procedure remains the same.

value):

$$
\begin{aligned}
&\text{bias}(\hat{\mu}) = \sigma \cdot \frac{\varphi\left(\frac{\log H - \mu}{\sigma}\right)}{1-\Phi\left(\frac{\log H - \mu}{\sigma}\right)} && > 0 \\
&\text{bias}(\hat{\sigma}^2) = \sigma^2 \left(\frac{\log H - \mu}{\sigma} \cdot \frac{\varphi\left(\frac{\log H - \mu}{\sigma}\right)}{1-\Phi\left(\frac{\log H - \mu}{\sigma}\right)} - \left(\frac{\varphi\left(\frac{\log H - \mu}{\sigma}\right)}{1-\Phi\left(\frac{\log H - \mu}{\sigma}\right)} \right)^2 \right) && < 0 \ \text{ since } \log H \text{ is small} \\
&\text{bias}(\hat{\lambda}) = -\lambda \cdot \Phi\left(\frac{\log H - \mu}{\sigma}\right) && < 0,
\end{aligned}
\tag{8.8}
$$

where $\varphi(\cdot)$ and $\Phi(\cdot)$ are the density and distribution function of $N(0,1)$ distribution. Clearly, the location parameter $\hat{\mu}$ is overestimated, and the scale parameter $\hat{\sigma}$ and the intensity rate $\hat{\lambda}$ are underestimated under the "naive" approach.

Our second question is the impact of using the wrong approach on the estimates of the expected aggregate operational loss (EL), Value-at-Risk (VaR), and the Conditional Value-at-Risk (CVaR).[4]

First, we can note that distributions that appear relevant to modeling operational losses belong to the class of sub-exponential distributions. This allows us to use the following approximation:

$$
P(S_{\triangle t} > M) \sim \mathbb{E}N_{\triangle t} \times P(X > M), \qquad \text{as } M \to \infty. \tag{8.9}
$$

Second, substituting M in Equation (8.9) with VaR (see Chapter 5 §5.5) for a given confidence level $(1-\alpha)$ and rearranging, we obtain

$$
\text{VaR}_{\triangle t, 1-\alpha} \sim F^{-1}\left(1 - \frac{\alpha}{\mathbb{E}N_{\triangle t}}\right). \tag{8.10}
$$

[4] The bias estimates for EL, VaR, and CVaR cannot be expressed in a simple closed form. For details see [38] and [39].

Table 8.2: Fraction of missing data for the Lognormal example.

Fraction of missing data $F_{\gamma_0}(H)$ for the Lognormal example with true parameters μ_0, σ_0 and nominal threshold $H = 50$.			
	$\mu_0 = 4$	$\mu_0 = 5$	$\mu_0 = 6.5$
$\sigma_0 = 1.5$	0.48	0.23	0.04
$\sigma_0 = 2$	0.48	0.29	0.10
$\sigma_0 = 2.7$	0.49	0.34	0.17

If losses are Lognormally distributed, Equation (8.10) reduces to:

$$\widehat{\mathrm{VaR}}_{\triangle t,1-\alpha} = \exp(\hat{\mu} + \mathrm{bias}(\hat{\mu}) + (\hat{\sigma} + \mathrm{bias}(\hat{\sigma}))\Phi^{-1}(1 - \tfrac{\alpha}{(\hat{\lambda}+\mathrm{bias}(\hat{\lambda}))\triangle t})). \quad (8.11)$$

Unfortunately, closed-form concise expression for the bias of the VaR estimate cannot be obtained. No simplified expression for CVaR can be determined.

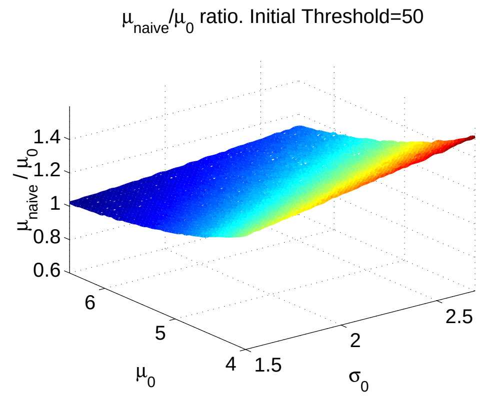
8.4 Simulation Study: Lognormal Example

A simple example can portray how the magnitude of the bias of the estimated parameters increases with an increased fraction of missing data. For an exemplary minimum recording threshold of $H = 50$, and the true parameters μ_0 and σ_0, Table 8.2 reveals the corresponding fractions of missing data $F_{\gamma_0}(H)$.

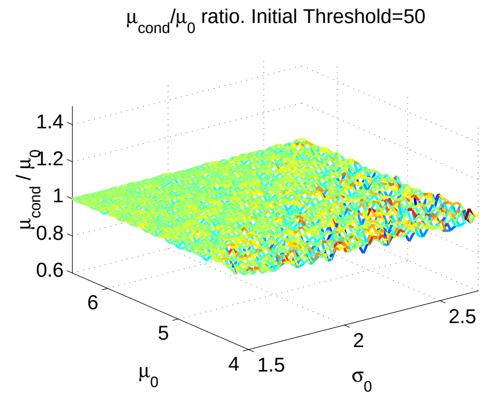
8.4.1 Impact on Parameter Estimates

Figures 8.3 and 8.4 show results of a simulation study of the effects of the biases by illustrating the ratios of the parameters estimated under the "naive" approach to the true parameter values, μ_0, σ_0, and λ_0. The distance between the ratio and unity represents the relative bias for each case. The ratio being equal to unity

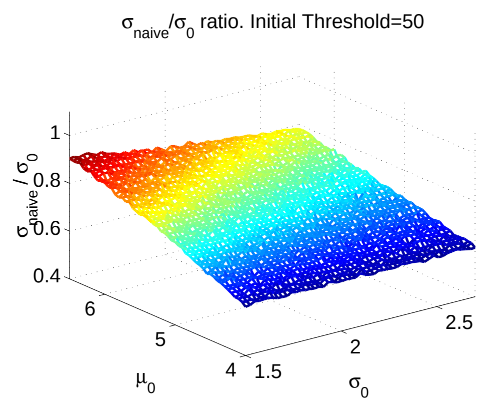
Figure 8.3: Illustration of the bias in the parameters $\hat{\mu}$ and $\hat{\sigma}$ and the fractions of missing data $F_{\hat{\gamma}}(H)$ (denotes by Q), estimated under the "naive" (left) and conditional (right) approaches, under the assumption of Lognormally distributed losses and Poisson frequency, for a range of true values μ_0 and σ_0 and $H = 50$. The figures show the ratio of the estimated parameters to the true parameters of the loss distribution. Bias under the "naive" approach is evident from the discrepancy of the ratios from unity.



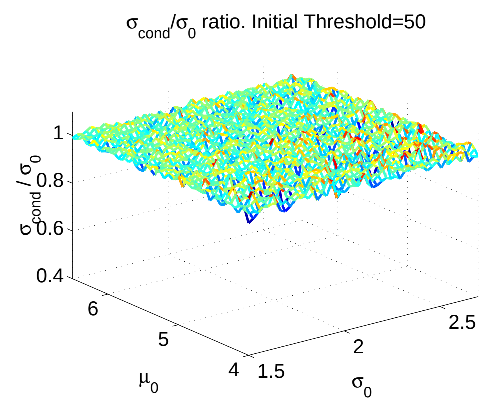
(a) Bias of $\hat{\mu}^{naive}$



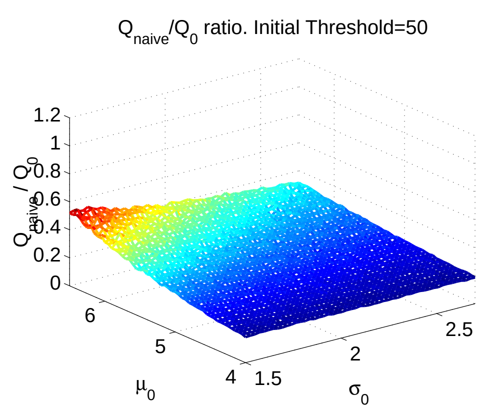
(b) Bias of $\hat{\mu}^{cond}$



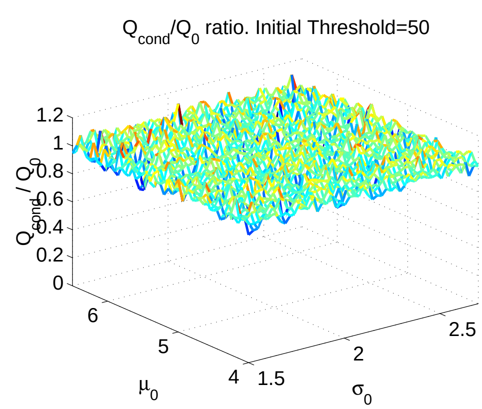
(c) Bias of $\hat{\sigma}^{naive}$



(d) Bias of $\hat{\sigma}^{cond}$

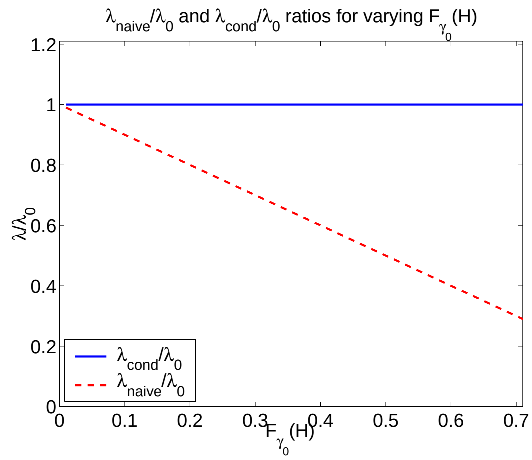


(e) Bias of $F_{\hat{\gamma}^{naive}}(H)$



(f) Bias of $F_{\hat{\gamma}^{cond}}(H)$

Figure 8.4: Illustration of bias of $\hat{\lambda}$ estimated under the "naive" and conditional approaches for varying fractions of missing data ($F_{\gamma_0}(H)$).

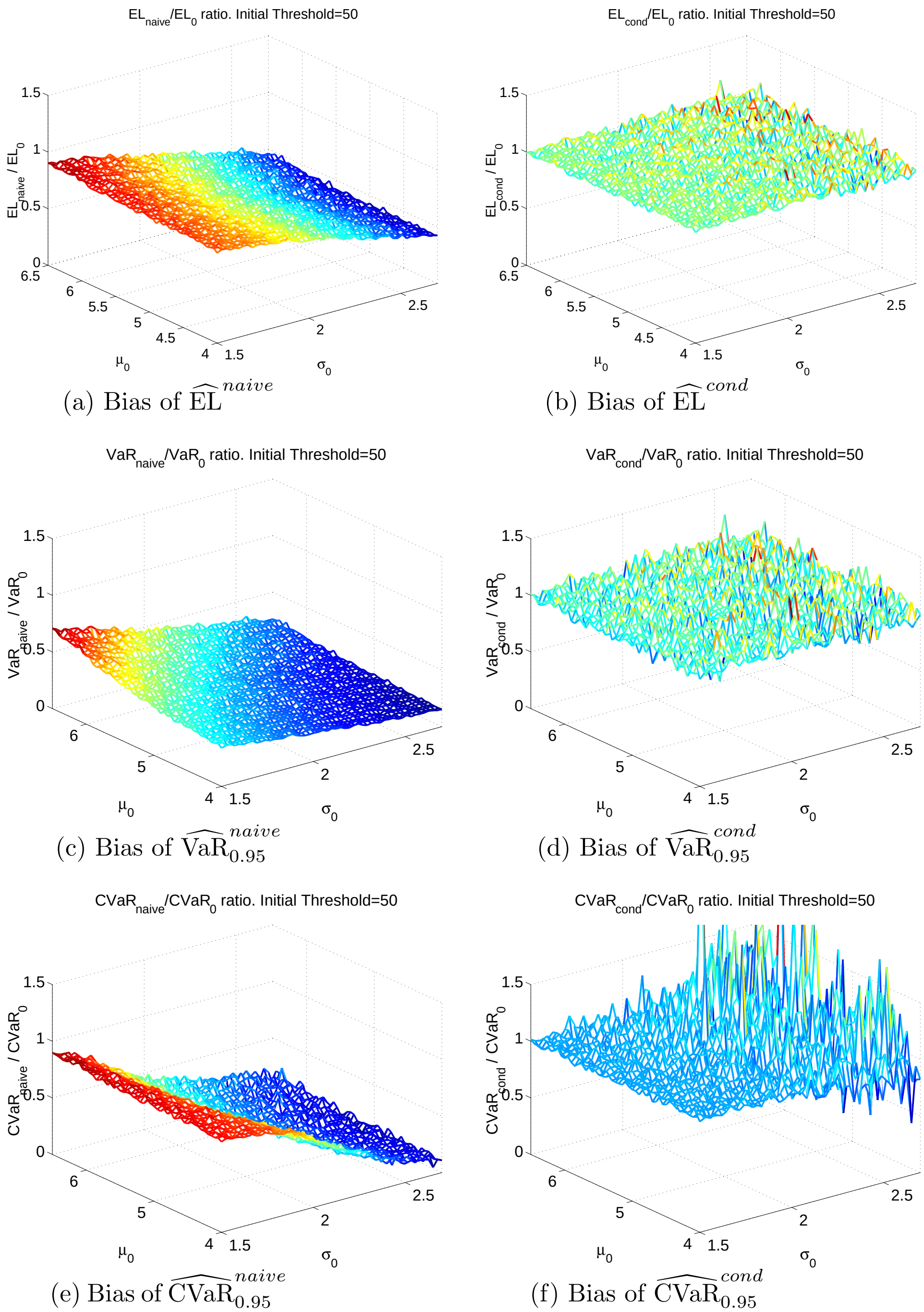


indicates the absence of any bias. Clearly, the bias estimate increases with an increased fraction of missing data. The conditional approach (figures are omitted here) resulted in correct parameter estimates yielding the ratios (at least approximately) equal to one.

8.4.2 Impact on Expected Aggregate Loss, VaR, and CVaR

Using the same example as in §8.4.1, and assuming $\lambda = 100$ and $\alpha = 0.05$, we perform simulations to determine the impact on missing data on various risk estimates. Figure 8.5 demonstrates the ratios of EL, VaR, and CVaR estimated to the true corresponding measures. EL was obtained using $EL = \mathbb{E}N_{\triangle t}\mathbb{E}X$, VaR was estimated using the asymptotic approximation presented in Equation (8.11), and CVaR was estimated using the Monte Carlo technique. Conditional approach resulted in correct estimates for the three measures yielding the ratios equal to one (with some oscillations around unity due to simulation errors). It is clear that if "naive" approach is used to calculate the risk measures, they would be severely

Figure 8.5: Illustration of the bias in the estimates of one-year EL, 95% VaR, and 95% CVaR under the "naive" (left) and conditional (right) approaches, under the assumption of Lognormal distribution for losses and Poisson frequency, for a range of values for μ and σ, $H = 50$, and $\lambda = 100$. The figures show the ratio of the estimated risk measures to the true risk measures. Bias is evident from the discrepancy of the ratios from unity.



(a) Bias of $\widehat{\mathrm{EL}}^{naive}$

(b) Bias of $\widehat{\mathrm{EL}}^{cond}$

(c) Bias of $\widehat{\mathrm{VaR}}_{0.95}^{naive}$

(d) Bias of $\widehat{\mathrm{VaR}}_{0.95}^{cond}$

(e) Bias of $\widehat{\mathrm{CVaR}}_{0.95}^{naive}$

(f) Bias of $\widehat{\mathrm{CVaR}}_{0.95}^{cond}$

underestimated.[5] Serious consequences of using the wrong approach would be a mismatch between the true exposure to operational risk and the estimated capital charge: amount of funds set aside by a financial institution may be insufficient to cover operational losses.

Remark 3 *If the conditional approach is used to estimate the parameters of the loss and frequency distribution, then the capital charge must be invariant to the choice of the initial threshold level.*

This remark is supported by our simulation studies referring to the conditional case. It is worth mentioning that, of course, the invariance property relies on the assumption that the true distribution of loss severity and frequency is correctly identified.

8.5 Empirical Study with Operational Loss Data

8.5.1 Overview of Earlier Studies

The majority of empirical studies with operational loss data use the "naive" approach to estimate the unknown parameters of the loss and frequency distributions, under which the missing data problem is overlooked and not adequately addressed. A common misconception is that using the "naive" approach can severely overestimate the capital charge. See, for example, a brief discussion of this conjecture in [49]. As another example, [127] points out the minimum collection threshold of approximately Euro 10,000, but nevertheless states that it is not necessary to focus on the correct modeling of low and medium magnitude losses since it is the upper quantiles that determine the capital charge; hence, according

[5] For more details on this theoretical study see [38, 39] and [128].

to him, the truncation in the data can be ignored, and he fits unconditional loss distributions to the data. [114] examine publicly reported unauthorized trading operational loss data obtained from the OpVar database (marketed by OpVantage) for the 1980-2001 period. They defined "material" unauthorized trading losses as those events resulting in direct financial losses greater than $ 100,000. They explain the necessity for the threshold by poor data reporting practices prior to 1990. [114] use truncated loss distribution to model the severity and the observed frequency distribution to model the frequency of the losses. By doing so, they leave small and medium-magnitude losses out of the estimation process, basically assuming that only larger-magnitude losses contribute to the loss process. [13] explore truncated operational loss data. In particular, they discuss the importance of this issue when data coming from various institutions is pooled together in public databases. They use truncated loss distributions but perform no adjustments to the frequency. They final results hence indicate that the "naive" estimation procedure overestimates the capital charge. One can draw an important conclusion from this last empirical study: the contribution of the frequency distribution to the operational capital charge is very significant: when the frequency distribution is correctly specified (see our earlier theoretical discussions), then (at least for the Lognormal example) "naive" approach tends to underestimate the capital charge.

8.5.2 Empirical Study with 1980-2002 Public Operational Loss Data

In [38], we discussed theoretical implications of using the "naive" approach for the operational risk modeling. In our later [39] empirical paper, we applied the methodology to 1980-2002 operational loss data from a public database. We remind that the data consist of five loss types: “Relationship”, “Human”, “Processes”, “Technology”, and “External”. The focus of this empirical study was to compare the results obtained under the “naive” and conditional methodologies.

Parameter Estimation

We only consider the Type I intensity rate of the frequency distribution; see Chapter 7 in which the description of such NHPP was described. In the conditional approach, we rescale the intensity rate using the procedure described earlier.

We restrict our attention to the loss distributions that lie on the positive real half-line.[6] The following loss distributions are considered in the study:

Exponential	$\mathcal{E}xp(\lambda)$	$f_X(x) = \lambda e^{-\lambda x}$ $x \geq 0,\ \lambda > 0$
Lognormal	$\mathcal{LN}(\mu, \sigma)$	$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left\{-\frac{(\log x - \mu)^2}{2\sigma^2}\right\}$ $x \geq 0,\ \mu, \sigma > 0$
Gamma	$\mathcal{G}am(\alpha, \beta)$	$f_X(x) = \frac{\beta^\alpha x^{\alpha-1}}{\Gamma(\alpha)} \exp\{-\beta x\}$ $x \geq 0,\ \alpha, \beta > 0$
Weibull	$\mathcal{W}eib(\beta, \alpha)$	$f_X(x) = \alpha\beta x^{\alpha-1} \exp\{-\beta x^\alpha\}$ $x \geq 0,\ \beta, \alpha > 0$
Logweibull	$\log \mathcal{W}eib(\beta, \alpha)$	$f_X(x) = \frac{1}{x}\alpha\beta(\log x)^{\alpha-1} \exp\{-\beta(\log x)^\alpha\}$ $x \geq 0,\ \beta, \alpha > 0$
Generalized Pareto	$\mathcal{GPD}(\xi, \beta)$	$f_X(x) = \beta^{-1}(1 + \xi x \beta^{-1})^{-(1+\frac{1}{\xi})}$ $x \geq 0,\ \beta > 0$

[6]The exception is the symmetric αStable distribution, for which we symmetrized the data: $\mathbf{x}^* = [\mathbf{x}; -\mathbf{x}]$.

Burr	$\mathcal{B}urr(\alpha, \beta, \tau)$	$f_X(x) = \tau\alpha\beta^\alpha x^{\tau-1}(\beta + x^\tau)^{-(\alpha+1)}$
		$x \geq 0,\ \alpha, \beta, \tau > 0$
log-αStable	$\log \mathcal{S}_\alpha(\beta, \sigma, \mu)$	$f_X(x) = \frac{g(\log x)}{x}, g \in \mathcal{S}_\alpha(\beta, \sigma, \mu)$,
		no closed form
		$x > 0,\ \alpha \in (0, 2),\ \beta \in [-1, 1],\ \sigma, \mu > 0$
Symmetric αStable	$\mathcal{S}_\alpha\mathcal{S}(\sigma)$	$f_X(x) \in \mathcal{S}_\alpha(0, \sigma, 0)$, no closed form
		$x = \|y\|,\ \alpha \in (0, 2),\ \sigma > 0$

Tables 8.3, 8.8, 8.9, 8.10, and 8.11 demonstrate the results of the empirical study (the last four tables are presented in the appendix to this chapter). For every distribution considered, quantities of interest were estimated using the misspecified "naive" approach (first row) and correctly specified conditional approach (second row). The quantities of interest are: (1) MLE parameter estimates, (2) fraction of missing data under the estimated parameters, and (3) estimate of log-likelihood function. The unknown parameters were estimated using the Constrained MLE approach. For the conditional approach, the frequency rate was adjusted using the methodology described in §8.3.2. It is evident that a considerable fraction of data appears to be missing (as represented by $F(H)$ under the conditional approach); also, comparison with the corresponding fraction under the "naive" approach indicates that low- and medium-magnitude losses are given a higher weight under the conditional approach. The log-likelihood function is higher under the conditional approach, indicating a better fit of the distributions to the data. Because the Gamma distribution results in near-one fraction of missing data, we exclude it from further analysis.

Table 8.3: Parameter estimates, $F(H)$, and log-likelihood values for the "naive" and conditional approaches for the "Relationship" type losses.

	$\hat{\gamma}$,$F_{\hat{\gamma}}(H)$,$\log L$	"Naive"	Conditional
$\mathcal{E}xp$	$\hat{\lambda}$	$9.6756{\cdot}10^{-9}$	$9.7701{\cdot}10^{-9}$
	$F_{\hat{\gamma}}(H)$	0.0096	0.0097
	$\log L$	-4532.7	-4530.5
$\mathcal{LN}$	$\hat{\mu}$	16.5789	15.7125
	$\hat{\sigma}$	1.7872	2.3639
	$F_{\hat{\gamma}}(H)$	0.0610	0.2111
	$\log L$	-4328.8	-4304.4
$\mathcal{G}am$	$\hat{\alpha}$	0.3574	$1.5392{\cdot}10^{-6}$
	$\hat{\beta}$	$3.4585{\cdot}10^{-9}$	$1.6571{\cdot}10^{-9}$
	$F_{\hat{\gamma}}(H)$	0.1480	≈ 1
	$\log L$	-4166.7	-4129.4
$\mathcal{W}eib$	$\hat{\beta}$	$1.1613{\cdot}10^{-4}$	0.0108
	$\hat{\alpha}$	0.5175	0.2933
	$F_{\hat{\gamma}}(H)$	0.1375	0.4629
	$\log L$	-4361.4	-4303.6
$\log \mathcal{W}eib$	$\hat{\beta}$	$3.1933{\cdot}10^{-12}$	$2.8169{\cdot}10^{-8}$
	$\hat{\alpha}$	9.2660	6.2307
	$F_{\hat{\gamma}}(H)$	0.1111	0.3016
	$\log L$	-15875.8	-15750.2
$\mathcal{GPD}$	$\hat{\xi}$	1.2481	1.5352
	$\hat{\beta}$	$1.2588{\cdot}10^{7}$	$0.7060{\cdot}10^{7}$
	$F_{\hat{\gamma}}(H)$	0.0730	0.1203
	$\log L$	-4333.0	-4310.9
$\mathcal{B}urr$	$\hat{\alpha}$	0.0987	0.1284
	$\hat{\beta}$	$2.5098{\cdot}10^{26}$	$3.2497{\cdot}10^{20}$
	$\hat{\tau}$	4.2672	3.3263
	$F_{\hat{\gamma}}(H)$	0.0145	0.0311
	$\log L$	-4327.1	-4329.6
$\log \mathcal{S}_{\alpha}$	$\hat{\alpha}$	1.8545	1.3313
	$\hat{\beta}$	1	-1
	$\hat{\sigma}$	1.1975	2.7031
	$\hat{\mu}$	16.6536	10.1928
	$F_{\hat{\gamma}}(H)$	0.0331	0.9226
	$\log L$	-15812.8	-15752.5
$\mathcal{S}_{\alpha}\mathcal{S}$	$\hat{\alpha}$	0.6820	0.5905
	$\hat{\sigma}$	$1.1395{\cdot}10^{7}$	$0.7073{\cdot}10^{7}$
	$F_{\hat{\gamma}}(H)$	0.0715	0.1283
	$\log L$	-20807.0	-7800.2

In-Sample Goodness-of-Fit Tests

In this section, we would like to determine which of the considered loss distributions fits the data sample best, based on the in-sample goodness-of-fit tests. We point out that the "naive" approach produced high test statistic values and near-zero p-values for most distributions indicating the inadequacy of the methodology (the figures are omitted for the purpose of saving space). Therefore, we restrict ourselves only to the conditional approach. We test a composite hypothesis that the truncated sample belongs to a hypothesized truncated distribution. For an iid sample drawn from continuous cdf F, for a family of continuous distributions $\mathcal{F}$, the null and alternative hypotheses are summarized as:

$$H_0: \; F(x) \in \mathcal{F}(x) \qquad H_A: \; F(x) \notin \mathcal{F}(x). \tag{8.12}$$

For example, to test whether the sample is drawn from a Lognormal distribution, we formulate the null hypothesis as $H_0: \; F(x) \in \mathcal{F}(x) = \{\mathcal{LN}_{\mu,\sigma}(x): \; \mu \in \Re, \sigma > 0\}$.

We consider five statistics for the measure of the distance between the empirical and hypothesized cdf: Kolmogorov-Smirnov (KS), Kuiper (V), supremum and quadratic Anderson-Darling (AD and AD^2), and Cramér-von Mises (W^2).

$$KS = \max\left\{KS^+, KS^-\right\}, \tag{8.13}$$

$$V = KS^+ + KS^-, \tag{8.14}$$

$$AD = \sqrt{n}\sup_x \left| \frac{F_n(x) - \widehat{F}(x)}{\sqrt{\widehat{F}(x)\big(1-\widehat{F}(x)\big)}} \right|, \tag{8.15}$$

$$AD^2 = n\int_{-\infty}^{\infty} \frac{(F_n(x) - \widehat{F}(x))^2}{\widehat{F}(x)(1-\widehat{F}(x))} d\widehat{F}(x), \tag{8.16}$$

Table 8.4: Goodness-of-fit test statistics and corresponding p-values (in square brackets) for the conditional loss distributions fitted to "Relationship" loss data.

	KS	V	AD	AD^2	W^2
$\mathcal{E}xp$	11.0868	11.9973	$1.3{\cdot}10^7$	344.37	50.5365
	[<0.005]	[<0.005]	[<0.005]	[<0.005]	[<0.005]
$\mathcal{LN}$	0.8056	1.3341	2.6094	0.7554	0.1012
	[0.082]	[0.138]	[0.347]	[0.043]	[0.086]
$\mathcal{W}eib$	0.5553	1.0821	3.8703	0.7073	0.0716
	[0.625]	[0.514]	[0.138]	[0.072]	[0.249]
$\log\mathcal{W}eib$	0.5284	1.0061	3.0718	0.4682	0.0479
	[0.699]	[0.628]	[0.255]	[0.289]	[0.514]
$\mathcal{GPD}$	1.4797	2.6084	3.5954	3.7165	0.5209
	[<0.005]	[<0.005]	[0.154]	[<0.005]	[<0.005]
$\mathcal{B}urr$	1.3673	2.4165	3.3069	3.1371	0.4310
	[0.032]	[<0.005]	[0.309]	[<0.005]	[0.011]
$\log\mathcal{S}_\alpha$	1.5929	1.6930	3.8184	3.8067	0.7076
	[0.295]	[0.295]	[0.275]	[0.290]	[0.292]
$S\alpha S$	1.1634	2.0695	$1.4{\cdot}10^5$	4.4723	0.3630
	[0.034]	[<0.005]	[>0.995]	[0.992]	[<0.005]

$$W^2 = n\int_{-\infty}^{\infty}(F_n(x) - \widehat{F}(x))^2 d\widehat{F}(x), \tag{8.17}$$

where $KS^+ = \sqrt{n}\sup_x\{F_n(x) - \widehat{F}(x)\}$ and $KS^- = \sqrt{n}\sup_x\{\widehat{F}(x) - F_n(x)\}$. Scaling factors $\sqrt{n}$ and n were used for the supremum class and the quadratic class statistics, respectively, to make them comparable across samples of different size. The limiting distributions of the test statistics are not parameter-free, so the p-values and the critical values were obtained with Monte Carlo simulations, as described in [144]. p-values suggest how likely it is that the data comes from a considered class of distributions; they were obtained following the four steps: (a) generate 1,000 samples from fitted distribution, of the same size as the original sample, (b) fit the distribution to each sample, (c) estimate the statistic value for each sample, and (d) find the proportion of time the statistic values from the simulated samples exceed the original statistic value.

The goodness-of-fit test statistics and the corresponding p-values for the conditional approach are presented in Tables 8.4, 8.12, 8.13, 8.14, and 8.15 (the last four tables are given in the appendix to this chapter). Weibull and Logweibull show the best overall fit around the center of the data, based on the KS and V tests, and more heavy-tailed distributions such as Burr, Pareto, and symmetric αStable, suggest the best fit around the tails. In general, the figures suggest that none of the data provides with the best overall fit: the data is best described by distributions with a moderate tail around the center and by those with a very heavy-tail around the upper tail. The Exponential distribution produces a very poor fit; we therefore exclude this distribution from future analysis.

One-Year Ahead EL, VaR, and CVaR Forecasts

To estimate EL, VaR, and CVaR, we undertake a forward-looking approach and use the functional form of the frequency and the parameters of the remaining severity distribution, obtained from the historical data over the available 23 year period, to forecast EL, VaR, and CVaR one year ahead. Tables 8.5, 8.16, 8.17, 8.18, and 8.19 present the estimates of EL, 95% and 99% VaR, and 95% and 99% CVaR. EL was calculated as $EL = \mathbb{E}N_{\triangle t}\mathbb{E}X$, and VaR and CVaR were estimated via 50,000 Monte Carlo samples. Clearly, ignoring the missing data results in (often highly significant) underestimation of the expected aggregate loss, VaR, and CVaR, whenever the wrong approach is used, in some instances up to five times.[7]

[7]The estimates of EL and CVaR were infinite (denoted by "-" in the table) whenever the first moment of the loss distribution was infinite. One way to fix this problem would be to fit doubly-truncated loss distribution to the data, by defining the upper bound U for the loss amount. In this case, the conditional distribution must be specified as:
$f_\gamma(x \mid H < x < U) = \frac{f_\gamma(x \ ; \ H<x<U)}{F_\gamma(U)-F_\gamma(H)}\mathbb{I}_{x<U}$,
that results in the maximum likelihood parameters $\hat{\gamma}_{\mathrm{MLE}}$. Then, the unconditional distribution

Table 8.5: Estimates of one-year ahead EL, VaR, and CVaR ($\times 10^{10}$) for the "naive" and conditional approaches for "Relationship" type losses.

	EL	$\mathrm{VaR}_{0.95}$	$\mathrm{VaR}_{0.99}$	$\mathrm{CVaR}_{0.95}$	$\mathrm{CVaR}_{0.99}$
			$\mathcal{LN}$		
"Naive"	0.1105	0.2832	0.5386	0.4662	0.8685
Conditional	0.1634	0.4662	1.0644	0.9016	1.9091
			$\mathcal{W}eib$		
"Naive"	0.1065	0.2203	0.2996	0.2700	0.3505
Conditional	0.1284	0.3187	0.5121	0.4430	0.6689
			$\log \mathcal{W}eib$		
"Naive"	-	0.2235	0.3193	-	-
Conditional	-	0.3332	0.5902	-	-
			$\mathcal{GPD}$		
"Naive"	-	0.8240	4.1537	-	-
Conditional	-	1.5756	11.3028	-	-
			$\mathcal{B}urr$		
"Naive"	-	2.8595	31.5637	-	-
Conditional	-	1.5713	11.5519	-	-
			$\log \mathcal{S}_\alpha$		
"Naive"	-	1.9124	7488.08	-	-
Conditional	-	0.4359	0.9557	-	-
			$\mathcal{S}_\alpha\mathcal{S}$		
"Naive"	-	2.1873	17.3578	-	-
Conditional	-	4.5476	56.2927	-	-

Backtesting

In this section, we conduct an out-of-sample backtesting of the models. We split our data sample into two parts: (1) the first sample consists of all data points in 1980-1995 and will be used for calibration, and (2) the second sample consists of the remaining data in 1996-2002. We use the first sample and the obtained

---

becomes:
$f_{\hat{\gamma}_{\mathrm{MLE}}}(x \mid x < U) = \frac{f_{\hat{\gamma}_{\mathrm{MLE}}}(x\ ;x<U)}{F_{\hat{\gamma}_{\mathrm{MLE}}}(U)} \mathbb{I}_{x<U}$.
The scaling factor for the frequency distribution becomes $(F_{\hat{\gamma}_{\mathrm{MLE}}}(U) - F_{\hat{\gamma}_{\mathrm{MLE}}}(H))^{-1}$. One possibility for U would be determining the worst potential loss or, alternatively, the total value of assets. See also [142], footnote 15, for a brief discussion of a similar issue: they suggest Winsorizing loss data at the point equal to 1,000 standard deviations.

truncated loss distributions' parameter estimates to analyze our models' predicting power regarding the data belonging to the second sample. We assume that our model has a one-step ahead predicting power, with one step equal to one year (due to a scarcity of data, it would be unreasonable to use smaller intervals). In the primary step we use the data from 1980 until 1995 to conduct the forecasting about 1996 losses.

First, we estimate the unknown parameters of truncated distributions. Next, to obtain the distribution of the annually aggregated losses we repeat the following a large number (10,000) of times: use the estimated parameters to simulate N losses exceeding the \$1 million threshold, where N is the random number of losses in the year that we perform forecasting on as dictated by the fitted frequency function, and aggregate them. At each forecasting step (seven steps total) we shift the window by one year forward and repeat the above procedure. In this way we test the model for both the severity and frequency distributions. We have observed that both Type I and II models fit the data very well; in this section we only focus on the Type I model. Since the observed data is incomplete, we are only able to compare the forecasting power regarding the truncated (rather than complete) data.

The analysis was carried out in two parts. In part one, we compared several quantiles (25, 50, 75, 95, 99, and 99.9) of the forecasted aggregated loss distribution with the corresponding bootstrapped (non-parametric) quantiles of the realized loss distribution.[8] Table 8.6 presents the MSE and MAE error estimates for the forecasted quantiles relative to the corresponding bootstrapped quantiles (left) and relative to the realized total loss (middle), and the errors of the simulated relative to the actual aggregate loss (right), for the "Relationship" type losses. (Corresponding tables 8.20, 8.21, 8.22, and 8.23 are given in the ap-

[8]The use of bootstrapping and Monte Carlo was suggested by the Basel Committee [26, 28].

pendix.) Figures were obtained based on 50,000 Monte Carlo samples. Errors around the 25th and 75th quantiles show the errors around the central bulk of the data, 50th quantile corresponds to the median, and the errors around the highest quantiles (95, 99, and 99.9) are the errors at the far right end of the distribution. Clearly the Weibull model provides the lowest estimates for the errors, followed by the Logweibull and log-αStable models.

In the second part of the analysis, we tested the severity distribution models (without checking for the frequency) via the Likelihood Ratio (LR) test suggested in [17]. While non-parametric tests like the Kolmogorov-Smirnov, Kuiper, or Cramér-von Mises are rather data-intensive [44], the LR test is especially useful for small data samples. It is based on the following methodology. Assume that we are interested in a stochastic process x_t, $t > 0$, which is being forecasted at time $t-1$. Let further the probability density of x_t be $f(x_t)$ and the associated distribution function be $F(x_t) = \int_{-\infty}^{x_t} f(u)du$. To conduct the test, we estimate the parameters of the loss distribution $\hat{F}$ from the historical data in the calibration period. If $\hat{F}$ is the correct loss distribution, then based on the so-called [143] transformation:

$$y_t = \int_{-\infty}^{x_t} \hat{f}(u)du = \hat{F}(x_t). \tag{8.18}$$

y_t are *iid* and distributed $\mathcal{U}[0,1]$. Further, an *iid* series z_t, such that $z_t \sim N(0,1)$, can be generated from the original data x_t with:

$$z_t = \Phi^{-1}(y_t) = \Phi^{-1}\Big(\int_{-\infty}^{x_t} \hat{f}(u)du\Big). \tag{8.19}$$

If $\hat{F}$ is correctly specified, z_t will be *iid* $N(0,1)$. To test whether the obtained

Table 8.6: Average forecast errors for "Relationship" type aggregated losses. *Left:* errors between corresponding quantiles; *middle:* errors of forecasted quantiles relative to realized loss; *right:* overall error between forecasted and realized loss.

	Forecasted quantiles vs. bootstrapped quantiles		Forecasted quantiles vs. actual loss		Overall error: forecasted vs. actual loss	
%	MSE ($\times 10^{20}$)	MAE ($\times 10^{10}$)	MSE ($\times 10^{20}$)	MAE ($\times 10^{10}$)	MAE ($\times 10^{20}$)	MSE ($\times 10^{10}$)
			$\mathcal{LN}$			
25	0.0038	0.0485	0.0302	0.1310		
50	0.0127	0.0823	0.0155	0.0891		
75	0.0260	0.1219	0.0155	0.1115	0.1357	0.1812
95	0.1342	0.3467	0.2140	0.4442		
99	0.8866	0.8618	1.4125	1.1760		
99.9	13.6999	3.6286	16.8095	4.0731		
			$\mathcal{W}eib$			
25	0.0018	0.0390	0.0017	0.0367		
50	0.0026	0.0446	0.0025	0.0439		
75	0.0039	0.0509	0.0055	0.0556	0.0052	0.0552
95	0.0083	0.0729	0.0170	0.1181		
99	0.0151	0.1069	0.0340	0.1733		
99.9	0.0288	0.1583	0.0667	0.2498		
			$\log \mathcal{W}eib$			
25	0.0036	0.0462	0.0295	0.1297		
50	0.0131	0.0822	0.0161	0.0890		
75	0.0278	0.1140	0.0120	0.0977	0.0397	0.1402
95	0.0800	0.2312	0.0766	0.2559		
99	0.2002	0.4232	0.3186	0.5532		
99.9	1.0261	0.9527	1.6881	1.2759		
			$\mathcal{GPD}$			
25	0.0034	0.0466	0.0272	0.1212		
50	0.0105	0.0807	0.0120	0.0828		
75	0.0892	0.2429	0.1195	0.3047	$3.1 \cdot 10^5$	6.9486
95	8.5160	2.7851	9.7921	2.9941		
99	475.35	21.3313	490.17	21.6451		
99.9	$1.8 \cdot 10^5$	404.99	$1.8 \cdot 10^5$	405.43		
			$\mathcal{B}urr$			
25	0.0031	0.0389	0.0240	0.1116		
50	0.0142	0.1049	0.0146	0.1069		
75	0.2266	0.4093	0.2875	0.4809	$7.9 \cdot 10^5$	13.1930
95	32.1950	5.2165	34.6466	5.4256		
99	2885.1	48.2596	2917.4	48.5730		
99.9	$2.5 \cdot 10^6$	1275.7	$2.5 \cdot 10^6$	1276.1		
			$\log \mathcal{S}_\alpha$			
25	0.0038	0.0488	0.0300	0.1308		
50	0.0129	0.0834	0.0158	0.0902		
75	0.0270	0.1208	0.0155	0.1104	0.1357	0.1746
95	0.1266	0.3286	0.1893	0.4083		
99	0.7405	0.7785	1.2092	1.0679		
99.9	8.3578	2.7464	10.7800	3.1893		
			$\mathcal{S}_\alpha\mathcal{S}$			
25	0.0051	0.0471	0.0167	0.0969		
50	0.0483	0.1636	0.0463	0.1542		
75	0.7872	0.8014	0.9111	0.8727	$1.6 \cdot 10^9$	277.43
95	115.14	10.1901	120.01	10.3991		
99	$1.3 \cdot 10^4$	106.08	$1.3 \cdot 10^4$	106.40		
99.9	$2.6 \cdot 10^7$	4229.7	$2.6 \cdot 10^7$	4230.2		

Table 8.7: Averaged p-values for "Relationship" type aggregated losses in the 7-year forecast period.

	$\mathcal{LN}$	$\mathcal{W}eib$	$\log \mathcal{W}eib$	$\mathcal{GPD}$	$\mathcal{B}urr$	$\log \mathcal{S}_\alpha$	$\mathcal{S}_\alpha\mathcal{S}$
Ave. p-value	0.4965	0.5922	0.5585	0.4217	0.4239	0.4130	0.4830

series z_t is independent across observations and standard Normal, we follow [17]. The used test statistic is $LR = -2(l_0 - l_1)$ where l_0 and l_1 are, respectively, the log-likelihood estimates under the null parameters ($\mu = 0$ and $\sigma = 1$) and under the parameters μ_{z_t} and σ_{z_t} estimated via MLE. The p-values are obtained by referring to the χ^2 distribution table and using 2 degrees of freedom. This LR test has a number of desirable statistical properties and can be considered a powerful test even for small sample sizes [17]. Thus, we consider this methodology as an adequate method to investigate whether the realized losses have come from a particular estimated distribution.

Table 8.7 presents the results for the "Relationship" losses. (Tables 8.24, 8.25, 8.26, and 8.27 for the other four loss types are presented in the Appendix.) The log-αStable, Burr, and Pareto distributions show the lowest 7-year average p-values, and the Logweibull and Weibull distributions gave the highest. The results are roughly consistent with those in Table 8.6. Overall we conclude that Logweibull and Weibull seem to be most appropriate for forecasting considered "Relationship" losses. Comparable results were obtained for the remaining four loss types.

Certainly, the true fraction of missing data and, hence, the correct amount of the capital charge, are dependent to the loss distribution that is selected for modeling the losses. Therefore, in-sample and out-of-sample goodness-of-fit tests become essential for this purpose.

Not surprisingly, goodness-of-fit tests for this study resulted in near-zero p-values for most of the goodness-of-fit statistics whenever the "naive" approach was used (results are omitted here), and in high p-values whenever the conditional approach was used. Again, this supports the idea that the "naive" approach is misspecified for the data in hand.

8.6 Appendix: Results of Empirical Study

Table 8.8: Parameter estimates, $F(H)$, and log-likelihood values for the "naive" and conditional approaches for the "Human" type losses.

	$\hat{\gamma}$,$F_{\hat{\gamma}}(H)$,$\log L$	"Naive"	Conditional
$\mathcal{E}xp$	$\hat{\lambda}$	$7.2216\cdot 10^{-9}$	$7.2741\cdot 10^{-9}$
	$F_{\hat{\gamma}}(H)$	0.0072	0.0073
	$\log L$	-16053.7	-16047.8
$\mathcal{LN}$	$\hat{\mu}$	16.5878	15.4627
	$\hat{\sigma}$	1.8590	2.5642
	$F_{\hat{\gamma}}(H)$	0.0679	0.2603
	$\log L$	-15143.6	-15045.3
$\mathcal{G}am$	$\hat{\alpha}$	0.3167	$6.9763\cdot 10^{-8}$
	$\hat{\beta}$	$2.2869\cdot 10^{-9}$	$1.1679\cdot 10^{-9}$
	$F_{\hat{\gamma}}(H)$	0.1628	≈ 1
	$\log L$	-14626.2	-14481.8
$\mathcal{W}eib$	$\hat{\beta}$	0.0002	0.0240
	$\hat{\alpha}$	0.4841	0.2526
	$F_{\hat{\gamma}}(H)$	0.1501	0.5441
	$\log L$	-15274.3	-15044.9
$\log \mathcal{W}eib$	$\hat{\beta}$	$14.3254\cdot 10^{-12}$	$30.7344\cdot 10^{-8}$
	$\hat{\alpha}$	9.8946	7.0197
	$F_{\hat{\gamma}}(H)$	0.1221	0.3718
	$\log L$	-15217.8	-15044.6
$\mathcal{GPD}$	$\hat{\xi}$	1.3761	1.6562
	$\hat{\beta}$	$1.1441\cdot 10^{7}$	$0.6135\cdot 10^{7}$
	$F_{\hat{\gamma}}(H)$	0.0792	0.1344
	$\log L$	-15145.2	-15060.6
$\mathcal{B}urr$	$\hat{\alpha}$	0.0938	0.0922
	$\hat{\beta}$	$5.1819\cdot 10^{27}$	$2.8463\cdot 10^{27}$
	$\hat{\tau}$	4.4823	4.4717
	$F_{\hat{\gamma}}(H)$	0.0131	0.0195
	$\log L$	-15108.7	-15112.0
$\log \mathcal{S}_\alpha$	$\hat{\alpha}$	1.6294	1.4042
	$\hat{\beta}$	1	-1
	$\hat{\sigma}$	1.1395	2.8957
	$\hat{\mu}$	16.8464	10.5108
	$F_{\hat{\gamma}}(H)$	0.0083	0.8793
	$\log L$	-15219.0	-15417.0
$\mathcal{S}_\alpha\mathcal{S}$	$\hat{\alpha}$	0.6724	0.6061
	$\hat{\sigma}$	$1.1126\cdot 10^{7}$	$0.7143\cdot 10^{7}$
	$F_{\hat{\gamma}}(H)$	0.0742	0.1241
	$\log L$	-72453.5	-27583.3

Table 8.9: Parameter estimates, $F(H)$, and log-likelihood values for the "naive" and conditional approaches for the "Processes" type losses.

	$\hat{\gamma}$,$F_{\hat{\gamma}}(H)$,$\log L$	"Naive"	Conditional
$\mathcal{E}xp$	$\hat{\lambda}$	$3.5020{\cdot}10^{-9}$	$3.5143{\cdot}10^{-9}$
	$F_{\hat{\gamma}}(H)$	0.0035	0.0035
	$\log L$	-6652.7	-6651.6
$\mathcal{LN}$	$\hat{\mu}$	17.5163	17.1600
	$\hat{\sigma}$	2.0215	2.3249
	$F_{\hat{\gamma}}(H)$	0.0336	0.0751
	$\log L$	-6382.7	-6366.9
$\mathcal{G}am$	$\hat{\alpha}$	0.3450	0.0247
	$\hat{\beta}$	$1.2082{\cdot}10^{-9}$	$0.5480{\cdot}10^{-9}$
	$F_{\hat{\gamma}}(H)$	0.1104	≈ 1
	$\log L$	-6126.5	-6088.5
$\mathcal{W}eib$	$\hat{\beta}$	0.0001	0.0021
	$\hat{\alpha}$	0.4938	0.3515
	$F_{\hat{\gamma}}(H)$	0.0923	0.2338
	$\log L$	-6412.1	-6364.9
$\log \mathcal{W}eib$	$\hat{\beta}$	$2.4894{\cdot}10^{-12}$	$0.1091{\cdot}10^{-8}$
	$\hat{\alpha}$	9.1693	7.1614
	$F_{\hat{\gamma}}(H)$	0.0687	0.1479
	$\log L$	-6397.3	-6364.8
$\mathcal{GPD}$	$\hat{\xi}$	1.4754	1.6147
	$\hat{\beta}$	$2.9230{\cdot}10^{7}$	$2.2886{\cdot}10^{7}$
	$F_{\hat{\gamma}}(H)$	0.0328	0.0413
	$\log L$	-6391.5	-6379.4
$\mathcal{B}urr$	$\hat{\alpha}$	0.8661	14.3369
	$\hat{\beta}$	$4.3835{\cdot}10^{6}$	$1.1987{\cdot}10^{4}$
	$\hat{\tau}$	0.8884	0.3829
	$F_{\hat{\gamma}}(H)$	0.0405	0.2097
	$\log L$	-15108.7	-15112.0
$\log \mathcal{S}_{\alpha}$	$\hat{\alpha}$	2.0000	2.0000
	$\hat{\beta}$	0.9697	0.8195
	$\hat{\sigma}$	1.4294	1.6476
	$\hat{\mu}$	17.5163	17.1535
	$F_{\hat{\gamma}}(H)$	0.0336	0.0760
	$\log L$	-6382.7	-6366.8
$\mathcal{S}_{\alpha}\mathcal{S}$	$\hat{\alpha}$	0.5902	0.5478
	$\hat{\sigma}$	$2.7196{\cdot}10^{7}$	$1.9925{\cdot}10^{7}$
	$F_{\hat{\gamma}}(H)$	0.0358	0.0536
	$\log L$	-29830.8	-12042.0

Table 8.10: Parameter estimates, $F(H)$, and log-likelihood values for the "naive" and conditional approaches for the "Technology" type losses.

	$\hat{\gamma}$,$F_{\hat{\gamma}}(H)$,$\log L$	"Naive"	Conditional
$\mathcal{E}xp$	$\hat{\lambda}$	$1.2914{\cdot}10^{-8}$	$1.3083{\cdot}10^{-8}$
	$F_{\hat{\gamma}}(H)$	0.0128	0.0130
	$\log L$	-1284.1	-1283.2
$\mathcal{LN}$	$\hat{\mu}$	16.6176	15.1880
	$\hat{\sigma}$	1.9390	2.7867
	$F_{\hat{\gamma}}(H)$	0.0742	0.3112
	$\log L$	-1252.8	-1243.7
$\mathcal{G}am$	$\hat{\alpha}$	0.4217	$7.5176{\cdot}10^{-6}$
	$\hat{\beta}$	$5.4458{\cdot}10^{-9}$	$2.3538{\cdot}10^{-9}$
	$F_{\hat{\gamma}}(H)$	0.1250	≈ 1
	$\log L$	-1189.8	-1180.9
$\mathcal{W}eib$	$\hat{\beta}$	$6.3668{\cdot}10^{-5}$	0.0103
	$\hat{\alpha}$	0.5490	0.2938
	$F_{\hat{\gamma}}(H)$	0.1177	0.4485
	$\log L$	-1256.7	-1242.1
$\log\mathcal{W}eib$	$\hat{\beta}$	$1.9309{\cdot}10^{-12}$	$11.0647{\cdot}10^{-8}$
	$\hat{\alpha}$	9.4244	5.7555
	$F_{\hat{\gamma}}(H)$	0.1023	0.3329
	$\log L$	-1254.8	-1242.7
$\mathcal{GPD}$	$\hat{\xi}$	1.5823	2.0925
	$\hat{\beta}$	$1.0470{\cdot}10^{7}$	$0.3446{\cdot}10^{7}$
	$F_{\hat{\gamma}}(H)$	0.0851	0.2029
	$\log L$	-1256.0	-1247.5
$\mathcal{B}urr$	$\hat{\alpha}$	0.0645	0.0684
	$\hat{\beta}$	$1.7210{\cdot}10^{35}$	$8.7406{\cdot}10^{20}$
	$\hat{\tau}$	5.8111	5.2150
	$F_{\hat{\gamma}}(H)$	0.0227	0.8042
	$\log L$	-1251.3	-1358.7
$\log\mathcal{S}_\alpha$	$\hat{\alpha}$	2.0000	2.0000
	$\hat{\beta}$	0.7422	0.8040
	$\hat{\sigma}$	1.3715	1.9894
	$\hat{\mu}$	16.6181	15.1351
	$F_{\hat{\gamma}}(H)$	0.0747	0.3195
	$\log L$	-1252.8	-1243.6
$\mathcal{S}_\alpha\mathcal{S}$	$\hat{\alpha}$	0.1827	0.1827
	$\hat{\sigma}$	$0.1676{\cdot}10^{7}$	$0.1676{\cdot}10^{7}$
	$F_{\hat{\gamma}}(H)$	0.3723	0.3723
	$\log L$	-5038.4	-1449.2

Table 8.11: Parameter estimates, $F(H)$, and log-likelihood values for the "naive" and conditional approaches for the "External" type losses.

	$\hat{\gamma}$,$F_{\hat{\gamma}}(H)$,$\log L$	"Naive"	Conditional
$\mathcal{E}xp$	$\hat{\lambda}$	$9.6756{\cdot}10^{-9}$	$9.7701{\cdot}10^{-9}$
	$F_{\hat{\gamma}}(H)$	0.0096	0.0097
	$\log L$	-4532.7	-4530.5
$\mathcal{LN}$	$\hat{\mu}$	16.5789	15.7125
	$\hat{\sigma}$	1.7872	2.3639
	$F_{\hat{\gamma}}(H)$	0.0610	0.2111
	$\log L$	-4328.8	-4304.4
$\mathcal{G}am$	$\hat{\alpha}$	0.3574	$1.5392{\cdot}10^{-6}$
	$\hat{\beta}$	$3.4585{\cdot}10^{-9}$	$1.6571{\cdot}10^{-9}$
	$F_{\hat{\gamma}}(H)$	0.1480	≈ 1
	$\log L$	-4166.7	-4129.4
$\mathcal{W}eib$	$\hat{\beta}$	$1.1613{\cdot}10^{-4}$	0.0108
	$\hat{\alpha}$	0.5175	0.2933
	$F_{\hat{\gamma}}(H)$	0.1375	0.4629
	$\log L$	-4361.4	-4303.6
$\log\mathcal{W}eib$	$\hat{\beta}$	$3.1933{\cdot}10^{-12}$	$2.8169{\cdot}10^{-8}$
	$\hat{\alpha}$	9.2660	6.2307
	$F_{\hat{\gamma}}(H)$	0.1111	0.3016
	$\log L$	-4347.2	-4303.7
$\mathcal{GPD}$	$\hat{\xi}$	1.2481	1.5352
	$\hat{\beta}$	$1.2588{\cdot}10^{7}$	$0.7060{\cdot}10^{7}$
	$F_{\hat{\gamma}}(H)$	0.0730	0.1203
	$\log L$	-4333.0	-4310.9
$\mathcal{B}urr$	$\hat{\alpha}$	0.0987	0.1284
	$\hat{\beta}$	$2.5098{\cdot}10^{26}$	$3.2497{\cdot}10^{20}$
	$\hat{\tau}$	4.2672	3.3263
	$F_{\hat{\gamma}}(H)$	0.0145	0.0311
	$\log L$	-4327.1	-4329.6
$\log\mathcal{S}_{\alpha}$	$\hat{\alpha}$	1.8545	1.3313
	$\hat{\beta}$	1	-1
	$\hat{\sigma}$	1.1975	2.7031
	$\hat{\mu}$	16.6536	10.1928
	$F_{\hat{\gamma}}(H)$	0.0331	0.9226
	$\log L$	-4330.2	-4569.0
$\mathcal{S}_{\alpha}\mathcal{S}$	$\hat{\alpha}$	0.6820	0.5905
	$\hat{\sigma}$	$1.1395{\cdot}10^{7}$	$0.7073{\cdot}10^{7}$
	$F_{\hat{\gamma}}(H)$	0.0715	0.1283
	$\log L$	-20807.0	-7800.2

Table 8.12: Goodness-of-fit test statistics and corresponding p-values (in square brackets) for the conditional loss distributions fitted to the "Human" loss data.

	KS	V	AD	AD^2	W^2
$\mathcal{E}xp$	14.0246	14.9145	$2.4{\cdot}10^6$	609.15	80.3703
	[<0.005]	[<0.005]	[<0.005]	[<0.005]	[<0.005]
$\mathcal{LN}$	0.8758	1.5265	3.9829	0.7505	0.0804
	[0.032]	[0.039]	[0.126]	[0.044]	[0.166]
$\mathcal{W}eib$	0.8065	1.5439	4.3544	0.7908	0.0823
	[0.103]	[0.051]	[0.095]	[0.068]	[0.188]
$\log\mathcal{W}eib$	0.9030	1.5771	4.1343	0.7560	0.0915
	[0.074]	[0.050]	[0.115]	[0.115]	[0.217]
$\mathcal{GPD}$	1.4022	2.3920	3.6431	2.7839	0.3669
	[<0.005]	[<0.005]	[0.167]	[<0.005]	[<0.005]
$\mathcal{B}urr$	2.2333	3.1970	4.7780	7.0968	1.2830
	[0.115]	[0.115]	[0.174]	[0.115]	[0.115]
$\log\mathcal{S}_\alpha$	9.5186	9.5619	36.2617	304.61	44.5156
	[0.319]	[0.324]	[0.250]	[0.312]	[0.315]
$S\alpha S$	1.1628	2.1537	$5.8{\cdot}10^5$	11.9320	0.2535
	[0.352]	[0.026]	[0.651]	[0.971]	[0.027]

Table 8.13: Goodness-of-fit test statistics and corresponding p-values (in square brackets) for the conditional loss distributions fitted to the "Processes" loss data.

	KS	V	AD	AD^2	W^2
$\mathcal{E}xp$	7.6043	8.4160	$3.7{\cdot}10^6$	167.60	22.5762
	[<0.005]	[<0.005]	[<0.005]	[<0.005]	[<0.005]
$\mathcal{LN}$	0.6584	1.1262	2.0668	0.4624	0.0603
	[0.297]	[0.345]	[0.508]	[0.223]	[0.294]
$\mathcal{W}eib$	0.6110	1.0620	1.7210	0.2069	0.0338
	[0.455]	[0.532]	[0.766]	[0.875]	[0.755]
$\log\mathcal{W}eib$	0.5398	0.9966	1.6238	0.1721	0.0241
	[0.656]	[0.637]	[0.832]	[0.945]	[0.918]
$\mathcal{GPD}$	1.0042	1.9189	4.0380	2.6022	0.3329
	[0.005]	[<0.005]	[0.128]	[<0.005]	[<0.005]
$\mathcal{B}urr$	0.5634	0.9314	1.6075	0.2639	0.0323
	[0.598]	[0.800]	[0.841]	[0.794]	[0.840]
$\log\mathcal{S}_\alpha$	0.6931	1.1490	2.0109	0.4759	0.0660
	[0.244]	[0.342]	[0.534]	[0.202]	[0.258]
$S\alpha S$	1.3949	1.9537	$3.3{\cdot}10^5$	6.5235	0.3748
	[0.085]	[0.067]	[0.931]	[0.964]	[0.102]

Table 8.14: Goodness-of-fit test statistics and corresponding p-values (in square brackets) for the conditional loss distributions fitted to "Technology" loss data.

	KS	V	AD	AD^2	W^2
$\mathcal{E}xp$	3.2160	3.7431	27.6434	27.8369	2.9487
	[<0.005]	[<0.005]	[<0.005]	[<0.005]	[<0.005]
$\mathcal{LN}$	1.1453	1.7896	2.8456	1.3778	0.2087
	[<0.005]	[0.005]	[0.209]	[<0.005]	[<0.005]
$\mathcal{W}eib$	1.0922	1.9004	2.6821	1.4536	0.2281
	[<0.005]	[<0.005]	[0.216]	[<0.005]	[<0.005]
$\log \mathcal{W}eib$	1.1099	1.9244	2.7553	1.5355	0.2379
	[<0.005]	[<0.005]	[0.250]	[<0.005]	[<0.005]
$\mathcal{GPD}$	1.2202	1.8390	3.0843	1.6182	0.2408
	[<0.005]	[<0.005]	[0.177]	[<0.005]	[<0.005]
$\mathcal{B}urr$	1.1188	0.9374	2.6949	2.0320	0.3424
	[0.389]	[0.380]	[0.521]	[0.380]	[0.380]
$\log \mathcal{S}_\alpha$	1.1540	1.7793	2.8728	1.3646	0.2071
	[<0.005]	[0.007]	[0.208]	[<0.005]	[<0.005]
$S\alpha S$	2.0672	2.8003	$2.7{\cdot}10^5$	19.6225	1.4411
	[>0.995]	[>0.995]	[>0.995]	[>0.995]	[0.964]

Table 8.15: Goodness-of-fit test statistics and corresponding p-values (in square brackets) for the conditional loss distributions fitted to "External" loss data.

	KS	V	AD	AD^2	W^2
$\mathcal{E}xp$	6.5941	6.9881	$4.4{\cdot}10^6$	128.35	17.4226
	[<0.005]	[<0.005]	[<0.005]	[<0.005]	[<0.005]
$\mathcal{LN}$	0.6504	1.2144	2.1702	0.5816	0.0745
	[0.326]	[0.266]	[0.469]	[0.120]	[0.210]
$\mathcal{W}eib$	0.4752	0.9498	2.4314	0.3470	0.0337
	[0.852]	[0.726]	[0.384]	[0.519]	[0.781]
$\log \mathcal{W}eib$	0.6893	1.1020	2.2267	0.4711	0.0563
	[0.296]	[0.476]	[0.481]	[0.338]	[0.458]
$\mathcal{GPD}$	0.9708	1.8814	2.7742	1.7091	0.2431
	[0.009]	[<0.005]	[0.284]	[<0.005]	[<0.005]
$\mathcal{B}urr$	1.3266	2.0385	2.8775	2.8954	0.5137
	[0.050]	[0.048]	[0.328]	[0.048]	[0.048]
$\log \mathcal{S}_\alpha$	7.3275	7.4089	37.4863	194.74	24.3662
	[0.396]	[0.458]	[0.218]	[0.284]	[0.366]
$S\alpha S$	0.7222	1.4305	$1.1{\cdot}10^5$	1.7804	0.1348
	[0.586]	[0.339]	[0.990]	[0.980]	[0.265]

Table 8.16: Estimates of one-year ahead EL, VaR, and CVaR ($\times 10^{10}$) for the "naive" and conditional approaches for "Human" type losses.

	EL	$\mathrm{VaR}_{0.95}$	$\mathrm{VaR}_{0.99}$	$\mathrm{CVaR}_{0.95}$	$\mathrm{CVaR}_{0.99}$
			$\mathcal{LN}$		
"Naive"	0.1981	0.4970	0.9843	0.8534	1.6652
Conditional	0.4171	1.2161	3.4190	3.3869	9.4520
			$\mathcal{W}eib$		
"Naive"	0.1993	0.4017	0.5507	0.4945	0.6456
Conditional	0.2881	0.7997	1.5772	1.3232	2.3746
			$\log \mathcal{W}eib$		
"Naive"	-	0.4174	0.6184	-	-
Conditional	-	0.8672	1.8603	-	-
			$\mathcal{GPD}$		
"Naive"	-	3.9831	33.5741	-	-
Conditional	-	12.1150	168.64	-	-
			$\mathcal{B}urr$		
"Naive"	-	85.5620	2690.44	-	-
Conditional	-	94.8281	3042.32	-	-
			$\log \mathcal{S}_\alpha$		
"Naive"	-	$1.9 \cdot 10^7$	$7.2 \cdot 10^{24}$	-	-
Conditional	-	2.2737	4.2319	-	-
			$\mathcal{S}_\alpha\mathcal{S}$		
"Naive"	-	6.2811	77.4762	-	-
Conditional	-	14.5771	203.24	-	-

Table 8.17: Estimates of one-year ahead EL, VaR, and CVaR ($\times 10^{10}$) for the "naive" and conditional approaches for "Processes" type losses.

	EL	$\mathrm{VaR}_{0.95}$	$\mathrm{VaR}_{0.99}$	$\mathrm{CVaR}_{0.95}$	$\mathrm{CVaR}_{0.99}$
			$\mathcal{LN}$		
"Naive"	0.5622	1.5508	3.5665	3.1201	6.9823
Conditional	0.8457	2.5610	6.5625	5.7823	13.9079
			$\mathcal{W}eib$		
"Naive"	0.4170	0.8800	1.2102	1.0891	1.4311
Conditional	0.5131	1.2761	2.1308	1.8257	2.8578
			$\log \mathcal{W}eib$		
"Naive"	-	0.9611	1.4498	-	-
Conditional	-	1.4780	2.6511	-	-
			$\mathcal{GPD}$		
"Naive"	-	12.5930	131.25	-	-
Conditional	-	20.8700	262.52	-	-
			$\mathcal{B}urr$		
"Naive"	-	6.8569	52.0391	-	-
Conditional	-	1.7987	4.1859	-	-
			$\log \mathcal{S}_\alpha$		
"Naive"	-	1.5613	3.5159	-	-
Conditional	-	2.5394	6.7070	-	-
			$\mathcal{S}_\alpha\mathcal{S}$		
"Naive"	-	38.7627	529.99	-	-
Conditional	-	74.9073	1280.02	-	-

Table 8.18: Estimates of one-year ahead EL, VaR, and CVaR ($\times 10^{10}$) for the "naive" and conditional approaches for "Technology" type losses.

	EL	$\mathrm{VaR}_{0.95}$	$\mathrm{VaR}_{0.99}$	$\mathrm{CVaR}_{0.95}$	$\mathrm{CVaR}_{0.99}$
			$\mathcal{LN}$		
"Naive"	0.0324	0.1202	0.3593	0.2970	0.7303
Conditional	0.0958	0.2898	1.2741	1.5439	5.4865
			$\mathcal{W}eib$		
"Naive"	0.0226	0.0798	0.1368	0.1159	0.1795
Conditional	0.0358	0.1454	0.3625	0.2958	0.6180
			$\log \mathcal{W}eib$		
"Naive"	-	0.0861	0.1683	-	-
Conditional	-	0.1670	0.4747	-	-
			$\mathcal{GPD}$		
"Naive"	-	0.4415	5.6954	-	-
Conditional	-	1.6249	54.4650	-	-
			$\mathcal{B}urr$		
"Naive"	-	2.8840	158.94	-	-
Conditional	-	9.0358	855.78	-	-
			$\log \mathcal{S}_\alpha$		
"Naive"	-	0.1222	0.3560	-	-
Conditional	-	0.2990	1.2312	-	-
			$\mathcal{S}_\alpha\mathcal{S}$		
"Naive"	-	$4.9 \cdot 10^5$	$3.2 \cdot 10^9$	-	-
Conditional	-	$7.1 \cdot 10^6$	$6.9 \cdot 10^{10}$	-	-

Table 8.19: Estimates of one-year ahead EL, VaR, and CVaR ($\times 10^{10}$) for the "naive" and conditional approaches for "External" type losses.

	EL	$\text{VaR}_{0.95}$	$\text{VaR}_{0.99}$	$\text{CVaR}_{0.95}$	$\text{CVaR}_{0.99}$
			$\mathcal{LN}$		
"Naive"	0.0157	0.0613	0.1697	0.1450	0.3451
Conditional	0.0327	0.1126	0.4257	0.3962	1.1617
			$\mathcal{W}eib$		
"Naive"	0.0151	0.0613	0.1190	0.0975	0.1628
Conditional	0.0208	0.0885	0.2494	0.2025	0.4509
			$\log \mathcal{W}eib$		
"Naive"	-	0.0611	0.1309	-	-
Conditional	-	0.0839	0.2489	-	-
			$\mathcal{GPD}$		
"Naive"	-	0.1190	0.8381	-	-
Conditional	-	0.2562	2.6514	-	-
			$\mathcal{B}urr$		
"Naive"	-	0.4072	8.7417	-	-
Conditional	-	0.7165	15.8905	-	-
			$\log \mathcal{S}_\alpha$		
"Naive"	-	0.1054	3.7687	-	-
Conditional	-	0.3879	0.8064	-	-
			$\mathcal{S}_\alpha\mathcal{S}$		
"Naive"	-	0.1730	1.8319	-	-
Conditional	-	0.4714	7.6647	-	-

Table 8.20: Average forecast errors for "Human" type aggregated losses. *Left:* errors between corresponding quantiles; *middle:* errors of forecasted quantiles relative to realized loss; *right:* overall error between forecasted and realized loss.

	Forecasted quantiles vs. bootstrapped quantiles		Forecasted quantiles vs. actual loss		Overall error: forecasted vs. actual loss	
%	MSE ($\times 10^{20}$)	MAE ($\times 10^{10}$)	MSE ($\times 10^{20}$)	MAE ($\times 10^{10}$)	MAE ($\times 10^{20}$)	MSE ($\times 10^{10}$)
			$\mathcal{LN}$			
25	0.0234	0.1082	0.1340	0.2867		
50	0.0552	0.1847	0.0696	0.2070		
75	0.0887	0.2432	0.0719	0.2314	35.7234	0.5449
95	0.9726	0.8818	1.9093	1.3553		
99	16.1004	3.9229	22.1948	4.6771		
99.9	518.30	21.9305	562.41	22.9933		
			$\mathcal{W}eib$			
25	0.0226	0.1077	0.1315	0.2842		
50	0.0594	0.1937	0.0748	0.2159		
75	0.1016	0.2524	0.0527	0.1937	0.2545	0.3286
95	0.2373	0.4028	0.5274	0.6930		
99	1.1885	1.0042	2.8841	1.6796		
99.9	9.9539	2.9937	16.8131	4.0560		
			$\log \mathcal{W}eib$			
25	0.0231	0.1074	0.1330	0.2862		
50	0.0585	0.1907	0.0738	0.2129		
75	0.0971	0.2429	0.0534	0.1970	0.4104	0.3584
95	0.3158	0.4950	0.7258	0.8181		
99	2.4188	1.4129	4.8112	2.1663		
99.9	32.7795	5.6141	45.4473	6.6756		
			$\mathcal{GPD}$			
25	0.0017	0.1070	0.1026	0.2493		
50	0.0062	0.1991	0.0604	0.2010		
75	1.8709	1.3121	2.3376	1.4844	$1.1 \cdot 10^{11}$	1996.5
95	755.17	26.8771	783.21	27.3695		
99	$2.0 \cdot 10^{5}$	430.82	$2.0 \cdot 10^{5}$	431.57		
99.9	$5.9 \cdot 10^{8}$	$2.1 \cdot 10^{4}$	$5.9 \cdot 10^{8}$	$2.1 \cdot 10^{4}$		
			$\mathcal{B}urr$			
25	0.0073	0.2000	0.0697	0.2146		
50	1.4999	1.0885	1.4432	1.0600		
75	65.7098	7.6095	68.3058	7.7819	$8.1 \cdot 10^{17}$	$4.2 \cdot 10^{6}$
95	$8.6 \cdot 10^{4}$	272.76	$8.6 \cdot 10^{4}$	273.25		
99	$6.4 \cdot 10^{7}$	7733.8	$6.4 \cdot 10^{7}$	7734.5		
99.9	$3.5 \cdot 10^{12}$	$1.5 \cdot 10^{6}$	$3.5 \cdot 10^{12}$	$1.5 \cdot 10^{6}$		
			$\log \mathcal{S}_\alpha$			
25	0.0230	0.1081	0.1322	0.2839		
50	0.0545	0.1865	0.0689	0.2090		
75	0.0876	0.2528	0.0683	0.2144	4.1107	0.4811
95	0.7972	0.7504	1.6497	1.2440		
99	11.7047	3.1938	16.8163	3.9486		
99.9	353.01	17.2799	389.68	18.3389		
			$\mathcal{S}_\alpha\mathcal{S}$			
25	0.0190	0.1137	0.0964	0.2457		
50	0.0874	0.2288	0.0833	0.2202		
75	2.1481	1.3854	2.6382	1.5585	$2.5 \cdot 10^{8}$	248.42
95	620.66	24.3374	645.52	24.8316		
99	$9.3 \cdot 10^{4}$	298.99	$9.3 \cdot 10^{4}$	299.74		
99.9	$1.9 \cdot 10^{8}$	$1.4 \cdot 10^{4}$	$1.9 \cdot 10^{8}$	$1.4 \cdot 10^{4}$		

Table 8.21: Average forecast errors for "Processes" type aggregated losses. *Left:* errors between corresponding quantiles; *middle:* errors of forecasted quantiles relative to realized loss; *right:* overall error between forecasted and realized loss.

	Forecasted quantiles vs. bootstrapped quantiles		Forecasted quantiles vs. actual loss		Overall error: forecasted vs. actual loss	
%	MSE ($\times 10^{20}$)	MAE ($\times 10^{10}$)	MSE ($\times 10^{20}$)	MAE ($\times 10^{10}$)	MAE ($\times 10^{20}$)	MSE ($\times 10^{10}$)
			$\mathcal{LN}$			
25	0.1845	0.3767	0.7813	0.6932		
50	0.5499	0.5703	0.6100	0.6027		
75	1.0214	0.6848	0.4277	0.4759	4.8553	0.8211
95	2.5537	1.4027	2.1662	1.3522		
99	17.5320	3.9941	22.9917	4.5137		
99.9	372.16	17.7825	423.24	19.5002		
			$\mathcal{W}eib$			
25	0.1792	0.3718	0.7686	0.6856		
50	0.5613	0.5787	0.6226	0.6116		
75	1.1229	0.7448	0.4656	0.5072	0.6909	0.6418
95	2.2488	0.9302	0.5310	0.6656		
99	3.4129	1.3507	1.7793	1.2427		
99.9	7.3804	2.3555	8.5875	2.7014		
			$\log \mathcal{W}eib$			
25	0.1815	0.3737	0.7748	0.6895		
50	0.5588	0.5765	0.6212	0.6096		
75	1.1028	0.7324	0.4556	0.4948	0.8050	0.6653
95	2.2039	0.9722	0.6675	0.7635		
99	4.0523	1.7245	3.5640	1.7009		
99.9	19.5854	4.1441	26.7134	4.7749		
			$\mathcal{GPD}$			
25	0.1808	0.3723	0.7738	0.6878		
50	0.4797	0.5170	0.5314	0.5491		
75	1.2251	1.0102	1.0405	0.9499	$1.3 \cdot 10^{10}$	684.34
95	348.86	16.6486	371.07	17.4765		
99	$7.7 \cdot 10^4$	241.57	$7.8 \cdot 10^4$	242.82		
99.9	$4.7 \cdot 10^8$	$1.6 \cdot 10^4$	$4.7 \cdot 10^8$	$1.6 \cdot 10^4$		
			$\mathcal{B}urr$			
25	0.1912	0.3823	0.7968	0.7014		
50	0.5625	0.5791	0.6237	0.6119		
75	0.9859	0.6680	0.4165	0.4989	$3.3 \cdot 10^4$	2.6236
95	5.5129	2.1175	6.8680	2.2928		
99	204.42	12.6032	232.35	13.8325		
99.9	$3.3 \cdot 10^4$	162.51	$3.4 \cdot 10^4$	164.23		
			$\log \mathcal{S}_\alpha$			
25	0.1835	0.3758	0.7795	0.6919		
50	0.5532	0.5729	0.6136	0.6059		
75	1.0425	0.6940	0.4320	0.4572	3.1026	0.7709
95	2.3285	1.2779	1.6237	1.1215		
99	14.6743	3.2239	18.9137	3.7639		
99.9	327.02	13.9361	367.72	15.6366		
			$\mathcal{S}_\alpha\mathcal{S}$			
25	0.1554	0.3542	0.7137	0.6587		
50	0.4121	0.4724	0.4522	0.4945		
75	3.4738	1.7482	3.8926	1.7777	$8.1 \cdot 10^{12}$	$1.4 \cdot 10^4$
95	1949.9	42.5619	2010.8	43.3890		
99	$7.0 \cdot 10^5$	785.20	$7.0 \cdot 10^5$	786.43		
99.9	$2.4 \cdot 10^9$	$4.7 \cdot 10^4$	$2.4 \cdot 10^9$	$4.7 \cdot 10^4$		

Table 8.22: Average forecast errors for "Technology" type aggregated losses. *Left:* errors between corresponding quantiles; *middle:* errors of forecasted quantiles relative to realized loss; *right:* overall error between forecasted and realized loss.

	Forecasted quantiles vs. bootstrapped quantiles		Forecasted quantiles vs. actual loss		Overall error: forecasted vs. actual loss	
%	MSE ($\times 10^{20}$)	MAE ($\times 10^{10}$)	MSE ($\times 10^{20}$)	MAE ($\times 10^{10}$)	MAE ($\times 10^{20}$)	MSE ($\times 10^{10}$)
			$\mathcal{LN}$			
25	0.0005	0.0146	0.0011	0.0205		
50	0.0008	0.0251	0.0008	0.0256		
75	0.0037	0.0514	0.0040	0.0534	29.1823	0.1734
95	0.1968	0.4199	0.2118	0.4381		
99	3.8728	1.8794	3.9690	1.9052		
99.9	154.56	12.1107	155.26	12.1468		
			$\mathcal{W}eib$			
25	0.0005	0.0149	0.0010	0.0202		
50	0.0007	0.0246	0.0008	0.0251		
75	0.0023	0.0440	0.0023	0.0422	0.0129	0.0537
95	0.0295	0.1628	0.0346	0.1812		
99	0.1884	0.4253	0.2090	0.4512		
99.9	1.1553	1.0500	1.2259	1.0858		
			$\log \mathcal{W}eib$			
25	0.0005	0.0148	0.0010	0.0204		
50	0.0007	0.0243	0.0008	0.0248		
75	0.0024	0.0454	0.0025	0.0432	0.0271	0.0612
95	0.0404	0.1924	0.0466	0.2106		
99	0.3151	0.5512	0.3425	0.5770		
99.9	3.2761	1.7834	3.3963	1.8192		
			$\mathcal{GPD}$			
25	0.0005	0.0140	0.0011	0.0206		
50	0.0008	0.0254	0.0008	0.0259		
75	0.0178	0.1108	0.0191	0.1176	$4.0 \cdot 10^{10}$	1599.5
95	44.4984	5.3895	44.6978	5.4077		
99	$6.9 \cdot 10^{4}$	185.52	$6.9 \cdot 10^{4}$	185.55		
99.9	$9.1 \cdot 10^{8}$	$2.2 \cdot 10^{4}$	$9.1 \cdot 10^{8}$	$2.2 \cdot 10^{4}$		
			$\mathcal{B}urr$			
25	0.0005	0.0138	0.0011	0.0205		
50	0.0010	0.0288	0.0010	0.2927		
75	0.1000	0.2734	0.1037	0.2803	$4.6 \cdot 10^{20}$	$9.7 \cdot 10^{7}$
95	1838.9	34.7495	1840.1	34.7676		
99	$2.2 \cdot 10^{7}$	3360.1	$2.2 \cdot 10^{7}$	3360.2		
99.9	$3.6 \cdot 10^{13}$	$3.9 \cdot 10^{16}$	$3.6 \cdot 10^{13}$	$3.9 \cdot 10^{6}$		
			$\log \mathcal{S}_\alpha$			
25	0.0005	0.0145	0.0011	0.0204		
50	0.0008	0.0251	0.0008	0.0256		
75	0.0037	0.0515	0.0041	0.0541	4.6680	0.1680
95	0.1965	0.4220	0.2117	0.4403		
99	4.4264	1.9974	4.5241	2.0233		
99.9	305.95	15.3729	306.88	15.4089		
			$\mathcal{S}_\alpha\mathcal{S}$			
25	79.7415	4.1267	79.6492	4.1191		
50	$1.1 \cdot 10^{6}$	443.75	$1.1 \cdot 10^{6}$	443.75		
75	$1.0 \cdot 10^{11}$	$1.3 \cdot 10^{5}$	$1.0 \cdot 10^{11}$	$1.3 \cdot 10^{5}$	$4.1 \cdot 10^{63}$	$2.4 \cdot 10^{29}$
95	$4.9 \cdot 10^{20}$	$8.4 \cdot 10^{9}$	$4.9 \cdot 10^{20}$	$8.4 \cdot 10^{9}$		
99	$1.1 \cdot 10^{30}$	$4.0 \cdot 10^{14}$	$1.1 \cdot 10^{30}$	$4.0 \cdot 10^{14}$		
99.9	$1.4 \cdot 10^{45}$	$1.4 \cdot 10^{22}$	$1.4 \cdot 10^{45}$	$1.4 \cdot 10^{22}$		

Table 8.23: Average forecast errors for "External" type aggregated losses. *Left:* errors between corresponding quantiles; *middle:* errors of forecasted quantiles relative to realized loss; *right:* overall error between forecasted and realized loss.

	Forecasted quantiles vs. bootstrapped quantiles		Forecasted quantiles vs. actual loss		Overall error: forecasted vs. actual loss	
%	MSE ($\times 10^{20}$)	MAE ($\times 10^{10}$)	MSE ($\times 10^{20}$)	MAE ($\times 10^{10}$)	MAE ($\times 10^{20}$)	MSE ($\times 10^{10}$)
			$\mathcal{LN}$			
25	0.0006	0.0222	0.0016	0.0353		
50	0.0016	0.0287	0.0016	0.0295		
75	0.0088	0.0731	0.0106	0.0836	0.6723	0.1351
95	0.2182	0.4012	0.2563	0.4449		
99	2.5172	1.4166	2.7096	1.4796		
99.9	60.6060	6.9971	62.0201	7.0857		
			$\mathcal{W}eib$			
25	0.0006	0.0219	0.0015	0.0348		
50	0.0015	0.0281	0.0016	0.0289		
75	0.0065	0.0619	0.0078	0.0724	0.0360	0.0841
95	0.0780	0.2426	0.1004	0.2862		
99	0.4396	0.6138	0.5223	0.6768		
99.9	3.1658	1.6437	3.4673	1.7296		
			$\log \mathcal{W}eib$			
25	0.0006	0.0218	0.0015	0.0350		
50	0.0015	0.0282	0.0015	0.0289		
75	0.0064	0.0608	0.0078	0.0713	0.0716	0.0899
95	0.0938	0.2569	0.1177	0.3007		
99	0.6237	0.7115	0.7186	0.7750		
99.9	7.4055	2.4166	7.8048	2.5032		
			$\mathcal{GPD}$			
25	0.0007	0.0230	0.0017	0.0359		
50	0.0028	0.0349	0.0028	0.0349		
75	0.0540	0.1674	0.0059	0.1779	$0.56{\cdot}10^{10}$	309.80
95	19.5035	3.2362	19.8077	3.2804		
99	6988.76	54.7280	6995.41	54.7908		
99.9	$3.1{\cdot}10^7$	3054.17	$3.1{\cdot}10^7$	3054.26		
			$\mathcal{B}urr$			
25	0.0008	0.0244	0.0017	0.0373		
50	0.0194	0.0714	0.0109	0.0713		
75	0.5632	0.5450	0.5812	0.5555	$40.0{\cdot}10^{10}$	6604.93
95	1286.89	25.4530	1289.38	25.4968		
99	$0.2{\cdot}10^7$	991.46	$0.2{\cdot}10^7$	991.52		
99.9	$4.5{\cdot}10^{10}$	$1.4{\cdot}10^5$	$4.5{\cdot}10^{10}$	$1.4{\cdot}10^5$		
			$\log \mathcal{S}_\alpha$			
25	0.0006	0.0223	0.0015	0.0353		
50	0.0016	0.0284	0.0016	0.0291		
75	0.0064	0.0626	0.0078	0.0731	0.0844	0.0937
95	0.0899	0.2662	0.1150	0.3099		
99	0.6948	0.7844	0.7980	0.8476		
99.9	8.5024	2.7599	8.9913	2.8457		
			$\mathcal{S}_\alpha\mathcal{S}$			
25	0.0007	0.0223	0.0015	0.0354		
50	0.0038	0.0429	0.0038	0.0427		
75	0.0782	0.2175	0.0851	0.2280	$0.03{\cdot}10^{10}$	130.60
95	26.1532	4.1342	26.5507	4.1781		
99	6397.98	63.2193	6406.84	63.2823		
99.9	$2.8{\cdot}10^7$	3339.51	$2.8{\cdot}10^7$	3339.60		

Table 8.24: Averaged p-values for "Human" type aggregated losses in the 7-year forecast period.

	$\mathcal{LN}$	$\mathcal{W}eib$	$\log \mathcal{W}eib$	$\mathcal{GPD}$	$\mathcal{B}urr$	$\log \mathcal{S}_\alpha$	$\mathcal{S}_\alpha\mathcal{S}$
Ave. p-value	0.4993	0.5310	0.5204	0.4530	0.4659	0.4115	0.0923

Table 8.25: Averaged p-values for "Processes" type aggregated losses in the 7-year forecast period.

	$\mathcal{LN}$	$\mathcal{W}eib$	$\log \mathcal{W}eib$	$\mathcal{GPD}$	$\mathcal{B}urr$	$\log \mathcal{S}_\alpha$	$\mathcal{S}_\alpha\mathcal{S}$
Ave. p-value	0.2462	0.2392	0.2431	0.2526	0.2433	0.1339	0.2103

Table 8.26: Averaged p-values for "Technology" type aggregated losses in the 7-year forecast period.

	$\mathcal{LN}$	$\mathcal{W}eib$	$\log \mathcal{W}eib$	$\mathcal{GPD}$	$\mathcal{B}urr$	$\log \mathcal{S}_\alpha$	$\mathcal{S}_\alpha\mathcal{S}$
Ave. p-value	0.5238	0.5165	0.5107	0.5185	0.5210	0.3354	0.3247

Table 8.27: Averaged p-values for "External" type aggregated losses in the 7-year forecast period.

	$\mathcal{LN}$	$\mathcal{W}eib$	$\log \mathcal{W}eib$	$\mathcal{GPD}$	$\mathcal{B}urr$	$\log \mathcal{S}_\alpha$	$\mathcal{S}_\alpha\mathcal{S}$
Ave. p-value	0.4739	0.4617	0.4751	0.4682	0.4210	0.0156	0.4768

Chapter 9

Goodness-of-Fit Tests for Heavy-Tailed Loss Data

9.1 Introduction

In most loss models, the central attention is devoted to studying the distributional properties of the loss data. The shape of the dispersion of the data determines the vital statistics such as the expected loss, variance, and ruin probability, Value-at-Risk or Conditional Value-at-Risk where the shape in the right tail is crucial. Parametric procedures for testing the goodness of fit (GOF) include the Likelihood Ratio test and Chi-squared test; non-parametric tests include visual examination of QQ-plots and mean excess plots. A standard semi-parametric procedure to test how well a hypothesized distribution fits the data involves applying the in-sample GOF tests that provide a comparison of the fitted distribution to the empirical distribution (EDF). These tests, referred to as EDF tests, include the Kolmogorov-Smirnov test, Kuiper test, Anderson-Darling test, and the Cramér-von Mises tests. Related works include [6] [7] [47] [149].

In this chapter we consider tests of a *composite* hypothesis that the empirical distribution function of an observed loss data sample belongs to a family of hypothesized distributions (with parameters not specified). For an *iid* sample drawn from continuous cdf F, for a family of continuous distributions $\mathcal{F}$, the null and alternative hypotheses are summarized as:

$$H_0: \; F(x) \in \mathcal{F}(x) \qquad H_A: \; F(x) \notin \mathcal{F}(x). \tag{9.1}$$

For example, to test whether the sample is drawn from a Lognormal distribution, we formulate the null hypothesis as $H_0: \; F(x) \in \mathcal{F}(x) = \{\mathcal{LN}_{\mu,\sigma}(x): \; \mu \in \Re, \sigma > 0\}$.

Under the null Equation (9.1), $\mathcal{F}(X) \sim \mathcal{U}[0,1]$, and the null is rejected if the p-value is lower than the level α, such as α from 5% to 10%. Letting D be the observed value of a GOF statistic and d the critical value for a given level α, the p-value is computed as p-value $= P(D \geq d)$. Since the distribution of the statistic is not parameter-free, one way to compute the p-values and the critical values is by means of Monte Carlo simulation, for each hypothesized fitted distribution [144]. Under the procedure, the observed value D is computed. Then, for a given level α the following algorithm is applied:

1. Generate large number of samples (e.g. I =1,000) from the fitted truncated distribution of size n equal to the number of observed data (such that all these points are above or equal to H);

2. Fit *truncated* distribution and estimate conditional parameters $\hat{\theta}$ for each sample $i = 1, 2, ...I$;

3. Estimate the GOF statistic value D_i for each sample $i = 1, 2, ...I$;

4. Calculate p-value as the proportion of times the sample statistic values exceed the observed value D of the original sample;

5. Reject $\mathbf{H_0}$ if the p-value is smaller than α.

In this chapter we propose two new EDF statistics to be used for the situations when the fit in the upper tail is of the central concern. Application of the modified EDF tests to operational loss data obtained from European public operational loss datatabase, and the USA natural catastrophe insurance claims data obtained from Insurance Services Office Inc. Property Claim Services, is presented in §9.4. Necessary derivations are provided in the Appendix.

9.2 Overview of Common EDF-Based Goodness-of-Fit Statistics

The EDF statistics are based on the vertical differences between the empirical and fitted distribution function.

Definition 13 *(Empirical distribution function) Let* $\{X_{(j)}\}_{1\le j\le n}$ *be a vector of the order statistics, such that* $X_{(1)} \le X_{(2)} \le ... \le X_{(n)}$. *Empirical distribution function (EDF) of the sample is defined as*

$$\begin{aligned} F_n(x) &:= n^{-1}\sum_{k=1}^{n} \mathbb{I}_{\{X_k \le x\}} \\ &= \begin{cases} 0 & x < x_{(1)} \\ \frac{j}{n} & x_{(j)} \le x < x_{(j+1)}, \quad j = 1, 2, ..., n-1 \\ 1 & x \ge x_n. \end{cases} \end{aligned} \tag{9.2}$$

EDF statistics are divided into two classes:

1. *Supremum class* includes statistics that represent a measure based on the maximum distance between the fitted and empirical cdf.

2. *Quadratic class* includes statistics that represent a measure based on the integral of the distance between the fitted and empirical cdf on the entire range of the support of X.

We review some commonly used statistics. The first three statistics described below belong to the supremum class.

Definition 14 *(Kolmogorov-Smirnov statistic) The Kolmogorov-Smirnov (KS) statistic is defined by:*

$$KS = \sqrt{n} \sup_x \left| F_n(x) - F(x) \right| . \tag{9.3}$$

Definition 15 *(Kuiper statistic) The Kuiper (V) statistic is defined by:*

$$V = KS^+ + KS^-, \tag{9.4}$$

where

$$\begin{aligned} KS^+ &= \sqrt{n} \sup_x \left\{ F_n(x) - F(x) \right\} , \\ KS^- &= \sqrt{n} \sup_x \left\{ F(x) - F_n(x) \right\} . \end{aligned} \tag{9.5}$$

Definition 16 *(supremum class Anderson-Darling statistic) The supremum class Anderson-Darling (AD) statistic is defined by:*

$$AD = \sqrt{n} \sup_x \left| \frac{F_n(x) - F(x)}{\sqrt{F(x)\big(1 - F(x)\big)}} \right| . \tag{9.6}$$

The quadratic statistics for complete data samples are grouped under the Cramér-von Mises family.

Definition 17 *(Cramér statistic) The Cramér family of quadratic statistics is defined by:*

$$Q = n \int_{-\infty}^{\infty} \big(F_n(x) - F(x)\big)^2 \psi(F(x)) dF(x), \tag{9.7}$$

in which the weight function $\psi(F(x))$ is assigned to give a certain weight to different observations.

Common quadratic class statistics defined by different weight function are as follows:

$$\begin{aligned} &\text{Cramér-von Mises statistic } (W^2)\text{:} && \psi(F(x)) = 1, \\ &\text{Anderson-Darling statistic } (AD^2)\text{:} && \psi(F(x)) = \big\{F(x)\big(1 - F(x)\big)\big\}^{-1}. \end{aligned}$$

Definition 18 *(Cramér-von Mises statistic) The Cramér-von Mises (W^2) statistic is defined by:*

$$W^2 = n \int_{-\infty}^{\infty} \big(F_n(x) - F(x)\big)^2 dF(x). \tag{9.8}$$

Definition 19 *(quadratic class Anderson-Darling statistic) The quadratic class Anderson-Darling (AD^2) statistic is defined by:*

$$AD^2 = n \int_{-\infty}^{\infty} \frac{\big(F_n(x) - F(x)\big)^2}{F(x)\big(1 - F(x)\big)} dF(x). \tag{9.9}$$

It is notable that in financial loss models, the dataset analyzed is often incomplete, in the sense that the observations that are present in the loss database only if they exceed a pre-determined threshold level.[1] While this problem is usually absent in risk models involving market risk and credit risk, in operational risk, banks' internal databases are subject to a minimum recording thresholds of roughly \$6,000-\$10,000, and external databases usually collect operational losses starting from \$1 million [27]. Similarly, in non-life insurance models, the thresholds are set at \$5 million, \$25 million, or other levels. With left-truncated data it is inappropriate to employ standard GOF tests, and the form of the cdf needs to be appropriately adjusted:

$$F(x|x > H) = \frac{F(x) - F(H)}{1 - F(H)}, \tag{9.10}$$

where H is the left-truncation point. GOF tests for truncated and censored data have been studied by [53] [77] [85] among others.

9.3 "Upper Tail" Anderson-Darling Statistic

In practice, often situations arise when it is necessary to test whether a distribution fits the data well mainly in the upper tail, and the fit in the lower tail or around the median is of little or less importance. Examples include operational risk and insurance claims modelling, in which goodness of the fit in the tails determines the Value-at-Risk or Conditional Value-at-Risk measures and the ruin probabilities. Given the Basel II recommendations, under the LDA, the operational risk capital charge is derived from the Value-at-Risk measure, which requires an accurate estimate of the upper tail of the loss distribution. Similarly, in insurance, the upper tail of the claim size distribution is vital to obtain the right estimates of ruin probability. For this purpose, we introduce a statistic, which we refer to as

[1] See Chapter 8 for the discussion of left-truncated distributions.

the *"upper tail" Anderson-Darling* statistic. We propose two different versions of it: the supremum class *"upper tail" Anderson-Darling* statistic (AD_{up}) and the quadratic class *"upper tail" Anderson-Darling* statistic (AD^2_{up}).

9.3.1 Supremum Class "Upper Tail" Anderson-Darling Statistic

The first version of the *"upper tail" Anderson-Darling* statistic belongs to the supremum class EDF statistics. Each observation of the KS statistic is assigned a weight of $\psi(F(x)) = \Big\{\big(1 - F(x)\big)\Big\}^{-1}$. Under this specification, the observations in the upper tail are assigned a higher weight than those in the lower tail. Let $\{X_{(j)}\}_{1\le j\le n}$ be the vector of the order statistics, such that $X_{(1)} \le X_{(2)} \le ... \le X_{(n)}$.

Definition 20 *(supremum class "upper tail" Anderson-Darling statistic) The supremum class "upper tail" Anderson-Darling (AD_{up}) statistic is defined as*

$$AD_{up} = \sqrt{n} \sup_x \left| \frac{F_n(x) - F(x)}{1 - F(x)} \right|. \tag{9.11}$$

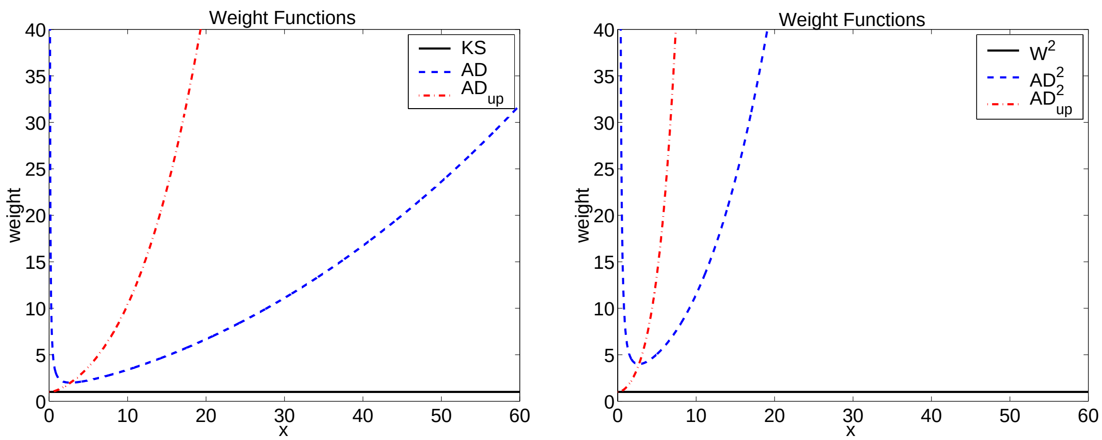
Derivation of the computing formula makes use of Equation (9.2) and involves the Probability Integral Transformation (PIT) technique [47]. Denoting $z_j := \widehat{F}_\gamma(x_{(j)})$, the computing formula is derived from:

$$\begin{aligned} AD^+_{up} &= \sqrt{n} \sup_j \left\{ \frac{\frac{j}{n} - z_j}{1 - z_j} \right\}, \\ AD^-_{up} &= \sqrt{n} \sup_j \left\{ \frac{z_j - \frac{j-1}{n}}{1 - z_j} \right\}, \end{aligned}$$

and becomes

$$AD_{up} = \max\{AD^+_{up},\ AD^-_{up}\}. \tag{9.12}$$

Figure 9.1: Weights associated with goodness-of-fit statistics assigned to observations in x, for a Lognormal($\mu = 1, \sigma = 1$) example. Left panel: supremum class statistics, right panel: quadratic class statistics.



(a) Supremum class

(b) Quadratic class

9.3.2 Quadratic Class "Upper Tail" Anderson-Darling Statistic

Another way to define the *"upper tail" Anderson-Darling* statistic is by an integral of the Cramér-von Mises family (see Equation (9.7)) with the weighting function of the form $\psi(F(x)) = \{1 - F(x)\}^{-2}$, that gives a higher weight to the upper tail and a lower weight to the lower tail. Let $\{X_{(j)}\}_{1 \le j \le n}$ be the vector of the order statistics, such that $X_{(1)} \le X_{(2)} \le ... \le X_{(n)}$.

Definition 21 *(quadratic class "upper tail" Anderson-Darling statistic) The quadratic class "upper tail" Anderson-Darling (AD^2_{up}) statistic is defined as*

$$AD^2_{up} = n \int_{-\infty}^{\infty} \frac{\big(F_n(x) - F(x)\big)^2}{\big(1 - F(x)\big)^2} dF(x). \tag{9.13}$$

Figure 9.1 compares the weight functions (as a function of x) associated with each of the GOF statistics.

Proposition 1 *Let* $z_j := \widehat{F}_\gamma(x_{(j)})$. *Then the computing formula for the* AD^2_{up} *statistic can be expressed as:*

$$AD^2_{up} = 2\sum_{j=1}^{n} \log\big(1 - z_j\big) + \frac{1}{n}\sum_{j=1}^{n}\Big(1 + 2(n-j)\Big)\frac{1}{1-z_j}. \tag{9.14}$$

Proof:

$$AD^2_{\text{up}} = n\int_{-\infty}^{+\infty} \frac{\big(F_n(x) - \widehat{F}(x)\big)^2}{\big(1-\widehat{F}(x)\big)^2} d\widehat{F}(x) \overset{PIT}{=} n\int_0^1 \frac{\big(F_n(z) - z\big)^2}{(1-z)^2} dz$$

by the PIT method, where $F_n(Z) = F(F_n(X)) = F_n(X)$ is the EDF of $Z = \widehat{F}_\gamma(X) = F(\widehat{F}_\gamma(X))$ so that $F(\cdot) \sim \mathcal{U}[0,1]$.

Using Equation (9.2), the computing formula is expressed in terms of $z_j := \widehat{F}_\gamma(x_{(j)}) = F(\widehat{F}_\gamma(x_{(j)})), j = 1, 2, ..., n$ as

$$n^{-1}AD^2_{\text{up}} = \underbrace{\int_0^{z_1} \frac{z^2}{(1-z)^2} dz}_{\text{A}} + \underbrace{\sum_{j=1}^{n-1}\int_{z_j}^{z_{j+1}} \frac{\big(\frac{j}{n} - z\big)^2}{(1-z)^2} dz}_{\text{B}} + \underbrace{\int_{z_n}^{1} \frac{\big(1-z\big)^2}{(1-z)^2} dz}_{\text{C}}$$

Separately solving for A, B and C yields

$$\text{A} = z_1 - 1 + \frac{1}{1-z_1} + 2\log(1-z_1);$$

$$
\begin{aligned}
\mathrm{B} &= z_n - z_1 - \frac{1}{n^2}\sum_{j=1}^{n-1}(n-j)^2\Big(\frac{1}{1-z_j} - \frac{1}{1-z_{j+1}}\Big) - ... \\
&- 2\frac{1}{n}\sum_{j=1}^{n-1}(n-j)\Big(\log(1-z_j) - \log(1-z_{j+1})\Big) \\
&= z_n - z_1 - \frac{1}{1-z_1} + \frac{1}{n^2}\sum_{j=1}^{n}\Big(1+2(n-j)\Big)\frac{1}{1-z_j} - ... \\
&- 2\log(1-z_1) + 2\frac{1}{n}\sum_{j=1}^{n}\log(1-z_j);
\end{aligned}
$$

$$
\mathrm{C} = 1 - z_n.
$$

Summing the terms A, B and C, multiplying by n, and simplifying yields the resulting computing formula:

$$
AD^2_{up} = 2\sum_{j=1}^{n}\log\big(1-z_j\big) + \frac{1}{n}\sum_{j=1}^{n}\Big(1+2(n-j)\Big)\frac{1}{1-z_j}.
$$

■

Table 9.1 summarizes the EDF statistics and their computing formulae.

9.4 Application to Financial Loss Data

In this section we apply the GOF testing procedure to (1) operational loss data extracted from a public database and (2) catastrophe insurance claims data. The operational loss data set was described in Chapter 5 §5.2 p.31. We remind the reader that the dataset covered 1980-2002 losses of five types: "Relationship,"

Table 9.1: Description of EDF statistics and computing formulae. We use notations: $z_j := \widehat{F}_\gamma(x_{(j)})$, $j = 1, 2, ..., n$.

Statistic	Description and computing formula
KS	$KS := \sqrt{n} \sup_x \lvert F_n(x) - F(x) \rvert$ Computing formula: $KS = \sqrt{n} \max \left\{ \sup_j \left\{ \frac{j}{n} - z_j \right\},\ \sup_j \left\{ z_j - \frac{j-1}{n} \right\} \right\}$
V	$V := \sqrt{n} \big(\sup_x \{F_n(x) - F(x)\} + \sup_x \{F(x) - F_n(x)\}$ Computing formula: $V = \sqrt{n} \left(\sup_j \left\{ \frac{j}{n} - z_j \right\} + \sup_j \left\{ z_j - \frac{j-1}{n} \right\} \right)$
AD	$AD := \sqrt{n} \sup_x \left\lvert \frac{F_n(x) - F(x)}{\sqrt{F(x)\big(1 - F(x)\big)}} \right\rvert$ Computing formula: $AD = \sqrt{n} \max \left\{ \sup_j \left\{ \frac{\frac{j}{n} - z_j}{\sqrt{z_j\big(1 - z_j\big)}} \right\},\ \sup_j \left\{ \frac{z_j - \frac{j-1}{n}}{\sqrt{z_j\big(1 - z_j\big)}} \right\} \right\}$
AD_{up}	$AD_{up} := \sqrt{n} \sup_x \left\lvert \frac{F_n(x) - F(x)}{1 - F(x)} \right\rvert$ Computing formula: $AD_{up} = \sqrt{n} \max \left\{ \sup_j \left\{ \frac{\frac{j}{n} - z_j}{1 - z_j} \right\},\ \sup_j \left\{ \frac{z_j - \frac{j-1}{n}}{1 - z_j} \right\} \right\}$
AD^2	$AD^2 := n \int_{-\infty}^{\infty} \frac{\big(F_n(x) - F(x)\big)^2}{F(x)\big(1 - F(x)\big)} dF(x)$ Computing formula: $AD^2 = -n + \frac{1}{n} \sum_{j=1}^{n} \big(1 - 2j\big) \log z_j - \frac{1}{n} \sum_{j=1}^{n} \big(1 + 2(n - j)\big) \log(1 - z_j)$
W^2	$W^2 := n \int_{-\infty}^{\infty} \big(F_n(x) - F(x)\big)^2 dF(x)$ Computing formula: $W^2 = \frac{n}{3} + \frac{1}{n} \sum_{j=1}^{n} \big(1 - 2j\big) z_j + \sum_{j=1}^{n} {z_j}^2$
AD^2_{up}	$AD^2_{up} := n \int_{-\infty}^{\infty} \frac{\big(F_n(x) - F(x)\big)^2}{\big(1 - F(x)\big)^2} dF(x)$ Computing formula: $AD^2_{up} = \frac{1}{n} \sum_{j=1}^{n} \big(1 + 2(n - j)\big) \frac{1}{(1 - z_j)} + 2 \sum_{j=1}^{n} \log(1 - z_j)$

"Human," "Processes," "Technology," and "External". The insurance claims data set covers claims resulting from natural catastrophe events occurred in the United States over the time period from 1990 to 1996. It was obtained from Insurance Services Office Inc. Property Claim Services (PCS). The data set includes 222 losses. The data recording is subject to \$25 million minimum threshold and so the observations are greater than \$25 million in nominal value. The data were adjusted for inflation using the Consumer Price Index of the U.S. Department of Labor.

Left-truncated distributions of eight types were fitted to each of the data set: Exponential, Lognormal, Weibull, Logweibull,[2] GPD, Burr, log-αStable, and symmetric αStable. The densities are defined in Chapter 8 §8.5.2. Table 9.2 presents the observed statistic values and the p-values for the six data sets (five operational loss datasets and insurance claims dataset).

The results reported in Table 9.2 lead us to conclude the following.

- *"Relationship" type operational losses.* KS and V tests indicate (both based on the statistic values and p-values) that the Weibull and Logweibull models are the most optimal for the data. AD tests suggest that, based on the p-values, that the symmetric αStable model is the best, followed by log-αStable and Logweibull. The Logweibull model is also supported by the W^2 test. However, the AD_{up} and AD^2_{up} tests are in favor of the Lognormal model; AD_{up} also resulted in high p-values for the GPD, Burr, and symmetric αStable models. It is notable that while we may tend to reject the heavy-tailed GPD and Burr models based on conventional KS test, they are strongly supported by the "upper tail" tests.

[2]Estimation of the Logweibull model parameters with the restricted MLE resulted in unacceptable fit for the insurance claims data. Therefore, the corresponding figures are missing from the table.

Table 9.2: Goodness-of-fit tests for operational and insurance loss data. p-values are given in square brackets.

	KS	V	AD	AD_{up}	AD^2	AD^2_{up}	W^2
			"Relationship"				
$\mathcal{E}xp$	11.0868	11.9973	$1.3{\cdot}10^7$	$1.2{\cdot}10^{23}$	344.37	$1.2{\cdot}10^{14}$	50.5365
	[<0.005]	[<0.005]	[<0.005]	[<0.005]	[<0.005]	[<0.005]	[<0.005]
$\mathcal{LN}$	0.8056	1.3341	2.6094	875.40	0.7554	4.6122	0.1012
	[0.082]	[0.138]	[0.347]	[0.593]	[0.043]	[0.401]	[0.086]
$\mathcal{W}eib$	0.5553	1.0821	3.8703	$2.7{\cdot}10^4$	0.7073	24.5068	0.0716
	[0.625]	[0.514]	[0.138]	[0.080]	[0.072]	[0.032]	[0.249]
$\log\mathcal{W}eib$	0.5284	1.0061	3.0718	7332.07	0.4682	10.1322	0.0479
	[0.699]	[0.628]	[0.255]	[0.186]	[0.289]	[0.102]	[0.514]
$\mathcal{GPD}$	1.4797	2.6084	3.5954	374.68	3.7165	22.1277	0.5209
	[<0.005]	[<0.005]	[0.172]	[>0.995]	[<0.005]	[0.048]	[<0.005]
$\mathcal{B}urr$	1.3673	2.4165	3.3069	371.65	3.1371	22.0374	0.4310
	[0.032]	[<0.005]	[0.309]	[0.960]	[<0.005]	[0.019]	[0.011]
$\log\mathcal{S}_\alpha$	1.5929	1.6930	3.8184	1075.30	3.8067	10.1990	0.7076
	[0.295]	[0.295]	[0.275]	[0.041]	[0.290]	[0.288]	[0.292]
$\mathcal{S}_\alpha\mathcal{S}$	1.1634	2.0695	$1.4{\cdot}10^5$	$5.0{\cdot}10^{16}$	4.4723	$2.6{\cdot}10^{14}$	0.3630
	[0.034]	[<0.005]	[>0.995]	[0.971]	[0.992]	[<0.005]	[<0.005]
			"Human"				
$\mathcal{E}xp$	14.0246	14.9145	$2.4{\cdot}10^6$	$1.1{\cdot}10^{22}$	609.15	$3.0{\cdot}10^{12}$	80.3703
	[<0.005]	[<0.005]	[<0.005]	[<0.005]	[<0.005]	[<0.005]	[<0.005]
$\mathcal{LN}$	0.8758	1.5265	3.9829	1086.16	0.7505	4.5160	0.0804
	[0.032]	[0.039]	[0.126]	[0.462]	[0.044]	[0.408]	[0.166]
$\mathcal{W}eib$	0.8065	1.5439	4.3544	$3.2{\cdot}10^4$	0.7908	8.6610	0.0823
	[0.093]	[0.051]	[0.095]	[0.068]	[0.053]	[0.112]	[0.176]
$\log\mathcal{W}eib$	0.9030	1.5771	4.1343	$1.1{\cdot}10^4$	0.7560	4.5125	0.0915
	[0.074]	[0.050]	[0.115]	[0.160]	[0.115]	[0.392]	[0.217]
$\mathcal{GPD}$	1.4022	2.3920	3.6431	374.68	2.7839	23.7015	0.3669
	[<0.005]	[<0.005]	[0.167]	[>0.995]	[<0.005]	[0.051]	[<0.005]
$\mathcal{B}urr$	2.2333	3.1970	4.7780	255.91	7.0968	46.3417	1.2830
	[0.115]	[0.115]	[0.174]	[>0.995]	[0.115]	[0.119]	[0.115]
$\log\mathcal{S}_\alpha$	9.5186	9.5619	36.2617	9846.30	304.61	4198.90	44.5156
	[0.319]	[0.324]	[0.250]	[0.354]	[0.312]	[0.215]	[0.315]
$\mathcal{S}_\alpha\mathcal{S}$	1.1628	2.1537	$5.8{\cdot}10^5$	$4.3{\cdot}10^{17}$	11.9320	$3.3{\cdot}10^{11}$	0.2535
	[0.352]	[0.026]	[0.651]	[0.351]	[0.971]	[0.436]	[0.027]

(Continued on next page)

Table 9.2 *(Continued from previous page)*

	KS	V	AD	AD_{up}	AD^2	AD^2_{up}	W^2
				"Processes"			
$\mathcal{E}xp$	7.6043	8.4160	$3.7{\cdot}10^6$	$1.7{\cdot}10^{22}$	167.61	$6.6{\cdot}10^5$	22.5762
	[<0.005]	[<0.005]	[<0.005]	[<0.005]	[<0.005]	[<0.005]	[<0.005]
$\mathcal{LN}$	0.6584	1.1262	2.0668	272.61	0.4624	4.0556	0.0603
	[0.297]	[0.345]	[0.508]	[0.768]	[0.223]	[0.367]	[0.294]
$\mathcal{W}eib$	0.6110	1.0620	1.7210	2200.75	0.2069	2.2340	0.0338
	[0.455]	[0.532]	[0.766]	[0.192]	[0.875]	[0.758]	[0.755]
$\log \mathcal{W}eib$	0.5398	0.9966	1.6238	658.42	0.1721	1.4221	0.0241
	[0.656]	[0.637]	[0.832]	[0.343]	[0.945]	[0.977]	[0.918]
$\mathcal{GPD}$	1.0042	1.9189	4.0380	148.24	2.6022	13.1082	0.3329
	[<0.005]	[<0.005]	[0.104]	[>0.995]	[<0.005]	[0.087]	[<0.005]
$\mathcal{B}urr$	0.5634	0.9314	1.6075	364.08	0.2639	325.76	0.0323
	[0.598]	[0.800]	[0.841]	[0.429]	[0.794]	[0.844]	[0.840]
$\log \mathcal{S}_\alpha$	0.6931	1.1490	2.0109	272.57	0.4759	328.39	0.0660
	[0.244]	[0.342]	[0.534]	[0.786]	[0.202]	[0.361]	[0.258]
$\mathcal{S}_\alpha\mathcal{S}$	1.3949	1.9537	$3.3{\cdot}10^5$	$2.5{\cdot}10^{17}$	6.5235	$6.8{\cdot}10^{14}$	0.3748
	[0.085]	[0.067]	[0.931]	[0.530]	[0.964]	[0.193]	[0.102]
				"Technology"			
$\mathcal{E}xp$	3.2160	3.7431	27.6434	$1.4{\cdot}10^6$	27.8369	780.50	2.9487
	[<0.005]	[<0.005]	[<0.005]	[<0.005]	[<0.005]	[<0.005]	[<0.005]
$\mathcal{LN}$	1.1453	1.7896	2.8456	41.8359	1.3778	6.4213	0.2087
	[<0.005]	[0.005]	[0.209]	[0.994]	[<0.005]	[0.067]	[<0.005]
$\mathcal{W}eib$	1.0922	1.9004	2.6821	52.5269	1.4536	4.8723	0.2281
	[<0.005]	[<0.005]	[0.216]	[0.944]	[<0.005]	[0.087]	[<0.005]
$\log \mathcal{W}eib$	1.1099	1.9244	2.7553	49.2373	1.5355	5.2992	0.2379
	[<0.005]	[<0.005]	[0.250]	[0.976]	[<0.005]	[0.085]	[<0.005]
$\mathcal{GPD}$	1.2202	1.8390	3.0843	33.4298	1.6182	8.8484	0.2408
	[<0.005]	[<0.005]	[0.177]	[>0.995]	[<0.005]	[0.067]	[<0.005]
$\mathcal{B}urr$	1.1188	1.9374	2.6949	28.4827	2.0320	10.5469	0.3424
	[0.389]	[0.380]	[0.521]	[>0.995]	[0.380]	[0.401]	[0.380]
$\log \mathcal{S}_\alpha$	1.1540	1.7793	2.8728	41.7454	1.3646	6.4919	0.2071
	[<0.005]	[<0.005]	[0.250]	[0.976]	[<0.005]	[0.060]	[<0.005]
$\mathcal{S}_\alpha\mathcal{S}$	2.0672	2.8003	$2.7{\cdot}10^5$	$3.6{\cdot}10^{16}$	19.6225	$7.2{\cdot}10^{10}$	1.4411
	[>0.995]	[>0.995]	[>0.995]	[>0.995]	[>0.995]	[>0.995]	[0.964]

(Continued on next page)

Table 9.2 *(Continued from previous page)*							
	KS	V	AD	AD_{up}	AD^2	AD^2_{up}	W^2
				"External"			
$\mathcal{E}xp$	6.5941	6.9881	$4.4{\cdot}10^6$	$2.0{\cdot}10^{22}$	128.35	$5.0{\cdot}10^7$	17.4226
	[<0.005]	[<0.005]	[<0.005]	[<0.005]	[<0.005]	[<0.005]	[<0.005]
$\mathcal{LN}$	0.6504	1.2144	2.1702	316.20	0.5816	2.5993	0.0745
	[0.326]	[0.266]	[0.469]	[0.459]	[0.120]	[0.589]	[0.210]
$\mathcal{W}eib$	0.4752	0.9498	2.4314	4382.68	0.3470	5.3662	0.0337
	[0.852]	[0.726]	[0.384]	[0.108]	[0.519]	[0.164]	[0.431]
$\log\mathcal{W}eib$	0.6893	1.1020	2.2267	3130.56	0.4711	4.1429	0.0563
	[0.296]	[0.476]	[0.481]	[0.128]	[0.338]	[0.283]	[0.458]
$\mathcal{GPD}$	0.9708	1.8814	2.7742	151.94	1.7091	8.6771	0.2431
	[0.009]	[0.005]	[0.284]	[0.949]	[<0.005]	[0.106]	[<0.005]
$\mathcal{B}urr$	1.3266	2.0385	2.8775	113.13	2.8954	15.4410	0.5137
	[0.050]	[0.048]	[0.329]	[0.989]	[0.048]	[0.064]	[0.048]
$\log\mathcal{S}_\alpha$	7.3275	7.4089	37.4863	4708.71	194.74	3132.60	24.3662
	[0.396]	[0.458]	[0.218]	[0.354]	[0.284]	[0.128]	[0.366]
$\mathcal{S}_\alpha\mathcal{S}$	0.7222	1.4305	$1.1{\cdot}10^5$	$3.4{\cdot}10^{16}$	1.7804	$1.2{\cdot}10^{10}$	0.1348
	[0.586]	[0.339]	[0.990]	[0.797]	[0.980]	[0.841]	[0.265]
				Natural Catastrophe			
$\mathcal{E}xp$	5.5543	5.9282	$9.0{\cdot}10^6$	$4.1{\cdot}10^{22}$	72.2643	$6.1{\cdot}10^{13}$	13.1717
	[<0.005]	[<0.005]	[<0.005]	[<0.005]	[<0.005]	[<0.005]	[<0.005]
$\mathcal{LN}$	0.6854	1.1833	5.3860	$1.1{\cdot}10^4$	0.7044	27.4651	0.0912
	[0.243]	[0.307]	[0.064]	[0.053]	[0.068]	[0.023]	[0.111]
$\mathcal{W}eib$	0.8180	1.5438	5.6345	$1.5{\cdot}10^4$	1.3975	15.8416	0.1965
	[0.096]	[0.041]	[0.043]	[0.028]	[0.007]	[0.025]	[0.006]
$\log\mathcal{W}eib$	-	-	-	-	-	-	-
	-	-	-	-	-	-	-
$\mathcal{GPD}$	0.4841	0.8671	2.4299	1277.28	0.3528	4.3053	0.0390
	[0.799]	[0.837]	[0.369]	[0.239]	[0.490]	[0.235]	[0.645]
$\mathcal{B}urr$	0.4604	0.8668	1.9408	477.59	0.2772	2.3411	0.0342
	[0.822]	[0.793]	[0.565]	[0.350]	[0.560]	[0.716]	[0.659]
$\log\mathcal{S}_\alpha$	0.8961	1.2111	2.0143	215.77	0.8062	2.9276	0.1535
	[0.456]	[0.470]	[0.700]	[0.726]	[0.484]	[0.642]	[0.444]
$\mathcal{S}_\alpha\mathcal{S}$	0.5282	1.0383	$1.0{\cdot}10^5$	$8.7{\cdot}10^{16}$	0.5587	$1.08{\cdot}10^{10}$	0.0328
	[0.560]	[0.415]	[0.093]	[0.103]	[0.265]	[0.158]	[0.642]

- *"Human" type operational losses.* Conventional KS test would reject most of the models, with the p-values barely exceeding 35% for the symmetric αStable and log-αStable models. The symmetric αStable distribution appears to be suitable in the tails of the distribution, as indicated by the AD and AD^2 tests. W^2 does not strongly support any of the models. However, based on the "upper tail" tests, the Lognormal, GPD, Burr, and symmetric αStable assumptions are well supported.

- *"Processes" type operational losses.* Weibull, Logweibull, and Burr models fit the data well around the center, as suggested by high p-values of the KS and V tests. All except the Exponential and GPD assumptions result in high p-values of the AD and AD^2 tests, making the judgements regarding the choice of an optimal model complicated; these, combined with the W^2 test results, suggest that the Weibull, Logweibull, and Burr models are optimal. A look at the "upper tail" tests gives a slightly different picture, with the GPD model showing a very good fit, both on the basis of high p-value and low statistic value (AD_{up}^2 test). In general, the "upper tail" tests are in favor of very heavy-tailed model assumptions.

- *"Technology" type operational losses.* The KS and V tests reject all except the symmetric αStable models. The AD test weakly supports the Burr assumption and strongly supports the symmetric αStable assumption; similar conclusion is drawn from the W^2 test. Judging from the abovementioned conventional GOF tests, the symmetric αStable model seems the only reasonable model. However, the AD_{up} and AD_{up}^2 tests also support several other models, making them valid candidates for modeling the losses.

- *"External" type operational losses.* Weibull models explains the data best around the center, as is evident from the high p-values of the KS and V tests. In the tails, the symmetric αStable model appears the best, while GPD and Burr also appear to fit the data well in the upper tail.

- *Natural Catastrophe claims.* GPD and Burr are the most appropriate to model the center of the data, while log-αStable is also a good candidate for the tails. In the upper quantiles, highest p-values correspond to the log-αStable and Burr models. Overall, it seems that the GPD, Burr, and log-αStable models are the best for the data, with the latter two the best for the upper tail.

Chapter 10

Robust Modeling

> "Robust methods, in one form or another (and be it a glance at the data), are necessary; those who still don't use them are either careless or ignorant."
>
> – F. R. Hampel (1973)

10.1 Introduction

In 2001, the Basel Committee made the following recommendation [25, Annex 6]:

> *"...data will need to be collected and "robust" estimation techniques (for event impact, frequency, and aggregate operational loss) will need to be developed."*

In recent years, the methods of robust statistics have been applied to tackle important issues in finance. [107] use robust statistics to empirically demonstrate that some of the factors in the well-known three-factor Fama-French model (1992) are not significant once outliers are eliminated. [105] apply robust estimators to examine the properties of the S&P500 index returns. [12] investigate the performance of portfolio return distribution using robust statistic and and conclude that the resulting forecasts outperform those under a conventional classical analysis. [132] use robust estimates for mean and the covariance matrix in the mean-variance portfolio selection problem.

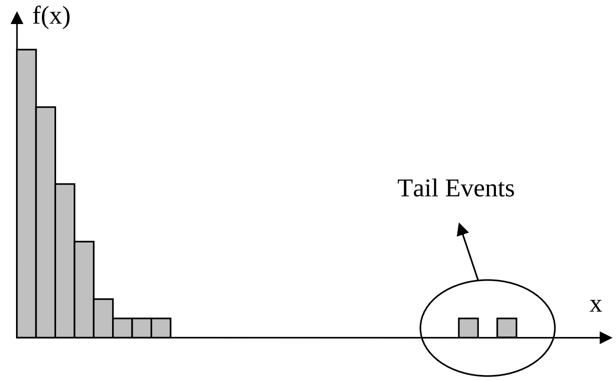
In this chapter we review the basic concepts of robust statistics and examine potential applications to operational loss data. Commonly used classical estimators of model parameters may be sub-optimal under minor departures of data from the model assumptions. Operational loss data are characterized by a very heavy right tail of the loss distribution attributed to several "low frequency/ high severity" events. Classical estimators may produce biased estimates of parameters leading to unreasonably high estimates of mean, variance, and the operational risk VaR and CVaR measures. The main objective of robust methods is to focus the analysis on the fundamental properties of the bulk of the data, without being distorted by outliers. We argue that further comparison of results obtained under the classical and robust procedures can serve as a basis for the VaR sensitivity analysis and can lead to an understanding of the economic role played by these extreme events. An empirical study with 1980-2002 public operational loss data reveals that the highest 5% of losses account for beyond 50% of the operational risk capital charge. The discussion in this chapter closely follows [40].

10.2 Outliers in Operational Loss Data

Existing empirical evidence suggests that the general pattern of operational loss severity data is characterized by high kurtosis, severe right-skewness, and a very heavy right tail created by several outlying events. Figure 10.1 portrays an illustrative example of operational loss severity data.

One approach to calibrate operational losses is to fit a parametric family of

Figure 10.1: Exemplary histogram of operational loss data. Extreme events appear as a distinctive tail in the far right of the distribution.



common loss distributions. One drawback of using these distributions is that they may not be optimal in fitting well both the center and the tails.[1] An alternative approach, discussed in Chapter 5 §5.3.4, is to use EVT to fit a GPD distribution to extreme losses exceeding a high pre-specified threshold. Both approaches may be severely influenced by one or more outlying observations. Classical estimators that assign equal importance to all available data are highly sensitive to extreme losses and in the presence of just a few such losses can produce arbitrarily large estimates of the mean, variance, and the capital charge. For example, high mean and standard deviation values for operational loss data do not provide an indication as to whether this is due to generally large values of observations or just one high-scale event, and it may be difficult to give the right interpretation to such result.

The presence of "low frequency/ high severity" events in the operational loss data creates the following paradox. On the one hand, the tail events correspond to the losses that are infrequent but often are the most destructive for an institution. In this sense, tail events cannot be ignored as they convey important information

[1]See, for example, discussion in [39].

regarding the loss generation process and may signal important flaws in the system. On the other hand, as stated earlier, recent empirical findings suggest that classical methods will frequently fit neither the bulk of the operational loss data nor the tail events well, and the center and the tails of the data appear to conform to different laws. The tail events possess the properties of outliers. In this light, applying classical methods that use all available data may not be the best approach, and using robust methods that focus on the dominant portion of the data may be a better approach. Robust methods take into account the underlying structure of the data and "separate" the bulk of the data from outlying events, thereby avoiding the upward bias in the vital statistics and forecasts.

Due to this paradox, the classical model and the robust model are not competitors and both models can be used as important complements to each other. The results from both approaches are not expected to be the same, as they explain different phenomena dictated by the original data: the general tendency (the robust method) and the conservative view (the classical method).

10.3 Some Dangers of Using the Classical Approach

There are dangers of using only the classical approach in modeling operational risk. Here we provide two examples.

Suppose, a risk expert constructs a one quarter ahead forecast of the total operational loss based on the historic data of his institution. Further, assume the data include the events of the order of magnitude of "9/11" or Hurricane Andrew (1992) and the Hurricane Katrina (2005). Would his forecast be robust? Most likely, his forecasts would indicate that his bank will have little reserves left if it decides to cover the potential loss.

As another example, suppose a risk analyst fits a heavy-tailed loss distribution to full data which include several “low frequency/ high severity” data points. The estimate of the aggregate expected loss (EL) is likely to be very high. In particular, if the fitted distribution is very heavy-tailed, such as some cases of Pareto or α-Stable, he may get an infinite mean and infinite second and higher moments' estimates. Occasionally, EL may even exceed VaR. Ongoing discussions by the Risk Management Group of the BIS suggest excluding the EL amount from the total estimated capital charge (e.g., VaR or CVaR) and setting the charge on the basis of the marginal unexpected loss[2] (UL), provided that the bank can demonstrate its ability to effectively monitor expected operational losses. The danger of treating outliers equally with the rest of the data is that the resulting UL-based capital charge may appear insufficient to cover the true exposure to the risk.

10.4 Overview of Robust Statistics Methodology

Robust statistics is the generalization of the classical theory: it takes into account the possibility of model misspecification, and the inferences remain valid not only at the parametric model but also in the neighborhood. The pioneering work on robust statistics is due to [95] and [86]. The objectives of robust statistics are as follows [90, Ch. 1]:

- To describe the structure best fitting the bulk of the data;
- To identify deviating data points (outliers) or deviating substructures for

[2]BIS defines UL as VaR – EL.

further treatment, if desired;

- To identify and give a warning about highly influential data points ("leverage points");

- To deal with unsuspected serial correlations, or more generally, with deviations from the assumed correlation structures.

5-10% of wrong values in the data appear to be the rule rather than the exception [87]. Outliers may appear in data due to (a) gross errors, (b) wrong classification of the data (outlying observations may not belong to the model followed by the bulk of the data), (c) grouping, and (d) correlation in the data [90].

10.4.1 Formal Model for Robust Statistics

Let $(1-\varepsilon)$ be the probability of well-behaved data, and ε be the probability of data being contaminated by "bad" observations. If $H(x)$ is an arbitrary distribution defining a neighborhood of the parametric model F_γ, then G is the two-point mixture of the parametric model and the contamination distribution:

$$G(x) = (1-\varepsilon)F_\gamma(x) + \varepsilon H(x). \tag{10.1}$$

10.4.2 Traditional Methods of Outlier Detection

Under traditional robust models, outliers are exogenously detected and excluded from the dataset, and the classical analysis is performed on the "cleaned" data. Data editing, screening, truncation, censoring, Winsorizing, and trimming are various methods for data cleaning. Such procedures for outlier detection are referred to as "forwards-stepping rejection," or "outside-in rejection" of outliers [150] [151].

Outlier detection methods in the forwards-stepping rejection procedure can be of two types: informal and formal. The former approach is rather subjective: a visual inspection of the database may be performed by a risk expert, and data points that clearly do not follow “the rule of the majority” are excluded. A risk expert may further conduct a background analysis of extreme losses, analyze whether they follow a pattern, and decide whether they are likely to repeat in the future. Which losses and how many to exclude is left up to his subjective judgment. For example, [127] examines the operational loss data[3] and excludes 1 outlier from the Retail Brokerage loss data (that consists of a total of 3,267 observations) and 5 outliers from the Commercial Banking loss database (that consists of a total of 3,414 observations).

Formal approaches to discriminate outliers include trimming and Winsorizing data. For example, $(\delta, 1-\gamma)$-trimmed data have the lowest δ and the highest γ fractions of the original data removed. For symmetrically contaminated data, $\delta = \gamma$. In the context of operational risk, contamination is asymmetric (on the right) and $\delta = 0$.

Unlike trimming, Winsorizing data does not suffer from loss of efficiency. For an original sample $x_j, j = 1, \ldots, n$ of size n, define $L_n = \lfloor n\delta \rfloor$ and $U_n = \lfloor n\gamma \rfloor$, and let $x_{(k)}$ denote the kth order statistic such that $x_{(1)} \leq \ldots \leq x_{(n)}$. The Winsorized sample $y_j, j = 1, \ldots, n$ is obtained by transforming $x_j, j = 1, \ldots, n$ in the following way:

$$y_j = \begin{cases} x_{(L_n+1)} & j \leq L_n \\ x_{(j)} & L_n + 1 \leq j \leq U_n, \quad j = 1, \ldots, n \\ x_{(U_n)} & j \geq U_n + 1. \end{cases} \tag{10.2}$$

[3]The data are taken from the second Loss Data Collection Exercise (Quantitative Impact Study 3), see also [27].

Other outlier rejection principles are based on kurtosis, largest Studentized residual, Studentized range, Shapiro-Wilk statistic, and Dixon's rule. A variety of outlier rejection methods have been discussed by [87] [89] [90] [150] [151] [152] [48], to name a few. The main criticism of the outlier rejection approach is that information is lost due to discarding several data points. One possibility is to choose to allow a fixed efficiency loss of, say, 5% or 10% [90]. [89] also showed that outside-in outlier rejection procedures possess low *breakdown points*[4] and estimators can be severily affected by a relatively small number of extreme observations, which means that estimators are not robust to heavy contamination. Nevertheless, "any way of treating outliers which is not totally inappropriate, prevents the worst" [87].

10.4.3 Examples of Non-Robust Estimators

Examples of non-robust estimators include the arithmetic mean, standard deviation, mean deviation and range, covariance and correlation, ordinary least squares (OLS). Robust measures of center include median, trimmed mean, and Winsorized mean. Robust measures of spread include inter-quartile range (IQR), median absolute deviation (MAD), mean absolute deviation, and Winsorized standard deviation; more estimators of scale were proposed by [145]. Robust estimators of skewness were studied by [104] [30] [92] and [84]. Robust estimators of kurtosis for heavy-tailed distributions were proposed by [93] [94]; others are due to [126] [93] [94] [84].

[4]The breakdown of an estimator is the maximum fraction of outliers that an estimator can tolerate.

10.4.4 Outlier Detection Approach Based On Influence Functions

Under more modern robust models, instead of being simply discarded, outliers are given a further treatment using a "backwards-stepping" or "inside-out" rejection procedure. One approach is based on the *influence functions* (IF), proposed by [86] [88]; see also [90]. IF measures the differential effect of an infinitesimal amount of contamination in an uncontaminated sample on the value of the estimator T at a point x, standardized by the amount of contamination:

$$\mathrm{IF}(x; T, F_\gamma) := \lim_{\varepsilon \searrow 0} \frac{T(G) - T(F_\gamma)}{\varepsilon}. \tag{10.3}$$

IF can be used to measure the *gross-error sensitivity* (GES) – the worst (approximate) influence which a small amount of contamination of fixed size can have on the value of the estimator T [90]:

$$\mathrm{GES}(T, F_\gamma) = \sup_x |\mathrm{IF}(x; T, F_\gamma)|. \tag{10.4}$$

GES can be used as a tool to detect the observations having a large influence on the value of the estimator. "Inside-out" outlier rejection rules have high breakdown points and the estimators can tolerate up to 50% of contamination [150]. Further discussion on IF and "inside-out" outlier treatment procedures can be found in [96] [150] [151]. Other references on robust statistics include [146] [116] [105] [9] [97] [131].

10.4.5 Outlier Rejection Approach and Stress Tests

Performing robust or classical analysis of the data is a trade-off between safety and efficiency: although some information may be lost while discarding or diminishing the contribution of the outlying events, one can significantly improve forecasts and produce more reliable estimates by applying a robust methodology.

An important application of robust statistics is using them as a *diagnostic technique* for evaluating the sensitivity of the inference conducted under the classical model to the rare events and to reveal their possible economic role [107]. Such analysis can be performed by comparing the results obtained under the classical and robust procedures.

We note a parallel between data trimming and stress tests that are widely applied in the operational risk modeling. The underlying mechanism of stress tests is to add several extreme observations to the dataset. By doing so, a risk analyst seeks to examine the incremental effect of potentially hazardous events on VaR and other risk measures. In contrast, with the robust methodology, instead of adding potential events, already existing but potentially improbable events are excluded from the database. The purpose is to investigate the fundamental properties of the main subset of the data in the absence of these unlikely events, as well as to study their incremental impact on risk measures.

Decisions about whether to include (stress tests) or exclude high-magnitude events (robust method), or whether to perform both tests, as well as how many points and of what magnitudes to include or exclude, can be left up to the subjective judgment of the risk expert or can be performed using one of the formal (objective) procedures discussed earlier.

10.5 Literature Review: Applications of Robust Models in Finance

Applications of robust analysis can be found in a variety of recent finance literature and are dominant in regression analysis. A classical example is the study of stock return anomalies by [67]: they argue that there appears to be risk premia associated with the size of firm and book-to-market. [107], however, demonstrated that the results were driven by a small portion of firms, and use instead least trimmed squares (LTS) as a robust regression technique to trim the few outlying observations, and then perform OLS on the remainder of the data.

Another recent study is due to [105] who apply robust estimators to examine the properties of the S&P500 index returns. They find evidence that the S&P500 index returns are composed of a mixture of two components, with a predominant component being nearly symmetric with mild kurtosis, and a relatively rare component generating extreme anomalies.

[12] investigates the performance of portfolio return distribution using robust and quantile-based methods, and conclude that the resulting forecasts outperform those under a conventional classical analysis.

[132] uses robust estimates for mean and the covariance matrix in the mean-variance portfolio selection problem. They show that the robust portfolio outperforms the classical one, and the outlying observations that account for 12.5% of the dataset can have serious influence on portfolio selection under the classical approach.

10.6 Application of Robust Methods to Operational Loss Data

In this section, we apply the simple data trimming technique to historic operational loss data.

10.6.1 Empirical Study with 1980-2002 Operational Loss Data

The datasets used for the analysis are described in Chapter 5 §5.2 p.31.

Contamination of the data is of a non-symmetric nature and is located in the far right tail of the loss distribution. We consistently trim the original datasets by cutting off the *highest* 5% of losses. Table 10.1 summarizes the descriptive statistics of the full and cleaned data.

A dramatic change in the statistics is evident when the robust methodology is applied: the mean, standard deviation, skewness and kurtosis coefficients have decreased significantly. Note that the robust measures of center and spread – median and median absolute deviation (MAD), respectively – remain practically unaffected.

In the next step, we fit loss distributions to both complete and trimmed datasets. The densities are defined in Chapter 8 §8.5.2. Table 10.3 (in the Appendix) exhibits parameter estimates. In majority of cases, outlier rejection has resulted in significantly less heavy-tailed distributions, as is evident from reduced location parameters and increased shape parameters. As expected, in majority of cases, such changes resulted in lower figures for the mean and variance.

GOF statistics and corresponding p-values are presented in Table 10.4 (in the Appendix). For the "External" loss example we can conclude that medium-tailed

Table 10.1: Descriptive sample statistics of full and top-5%-trimmed operational loss data.

	"Relation."	"Human"	"Process."	"Techn."	"Extern."
1. Classical Approach					
n	849	813	325	67	233
min (\$ $\times 10^6$)	1.07	1.10	1.10	1.13	1.1
max (\$ $\times 10^6$)	6,480	23,630	13,334	830	6,384
mean (\$ $\times 10^6$)	89.86	138.47	285.55	77.43	103.35
median (\$ $\times 10^6$)	14.63	12.32	39.98	11.60	12.89
st.dev. (\$ $\times 10^6$)	360.45	901.51	955.52	136.65	470.24
MAD (\$ $\times 10^6$)	12.47	10.40	37.08	10.42	11.17
skewness	11.6429	22.2416	9.1070	3.1761	11.0320
kurtosis	169.9732	570.1188	112.5151	15.7230	140.8799
2. Robust Approach					
n	806	772	304	63	221
min (\$ $\times 10^6$)	1.07	1.10	1.10	1.13	1.1
max (\$ $\times 10^6$)	427.09	855.32	1,178	830.00	364.80
mean (\$ $\times 10^6$)	39.63	45.47	113.31	74.40	39.7515
median (\$ $\times 10^6$)	13.50	11.12	33.61	11.60	11.40
st.dev. (\$ $\times 10^6$)	59.78	85.41	188.66	134.77	63.84
MAD (\$ $\times 10^6$)	11.21	9.14	30.26	10.42	9.64
skewness	2.4998	3.4703	2.6273	3.4060	2.5635
kurtosis	10.1200	19.6686	10.5012	17.5666	10.0539

distributions such as Lognormal and Weibull fit the trimmed data well, in contrast to heavy-tailed laws in the case when all data are included into the analysis.

Next, we examine the aggregated 1-year EL, $\mathrm{VaR}_{0.95}$, $\mathrm{VaR}_{0.99}$, $\mathrm{CVaR}_{0.95}$, and $\mathrm{CVaR}_{0.99}$. The estimates are based on out-of-sample 1-year ahead forecast. For the frequency distribution, a Cox process with a non-homogeneous intensity rate function was used (see Chapter 7). We note, however, that robust methods have a negligible effect on the parameters of the frequency distribution. Tables 10.5 (in the Appendix) reports the estimates for the risk measures. The estimates of the risk measures are considerably lower under the robust method in all cases.

Hence, robust methods can prevent over-estimation of the capital charge. Table 10.2 also reports the incremental effect inflicted on these measures by the top 5% observations. The marginal impact was computed by

$$\triangle = \frac{T_{class.} - T_{robust}}{T_{class.}} \times 100\%, \tag{10.5}$$

with T being the appropriate measure – one of EL, VaR, and CVaR. For example, from Table 10.2 it is evident that the twelve extreme data points of the "External" type losses account for up to 58% of the total EL, and up to 76% of the total operational risk capital charge (VaR or CVaR). For further details of this empirical study, see [39].

Regarding backtesting, results for the classical approach were presented in Chapter 8 §8.5.2 p. 8.5.2. We here reproduce the results for the robust approach. Tables 10.6, 10.7, 10.8, 10.9, and 10.10 (in the Appendix) present the MSE and MAE of the forecasts. Clearly, the accuracy of the forecasts has remarkably improved. For all distributions forecasted quantiles of the loss distribution are much closer to the bootstrapped quantiles and actual losses. This is especially true for the high quantiles, as expected, since extreme losses are not extracted from the analysis. We conclude that with the robust approach the general tendency of the losses is captured adequately. Further, the approach reveals the sensitivity of the risk measures VaR and CVaR to the biggest losses in the data. Table 10.11 (in the Appendix) presents average p-values of the forecasts. Considering the choice of the right distribution, both tests (forecast error estimates, and the LR test) converge in their indication of the best model with the in-sample goodness of fit tests: for example, for the "External" type losses, the robust approach confirms that the Logweibull distribution has the best forecasting power with Weibull being the second best choice.

Table 10.2: Sensitivity of classical VaR to outliers. Figures indicate incremental VaR as the percentage of classical VaR attributed to top 5% of data. *Note:* for the *"Technology"* type losses, excluding the outliers in the data resulted in heavier-tailed distributions than under the classical analysis. This may be explained by the estimation inaccuracy due to the small dataset (63 points).

	$\mathcal{LN}$	$\mathcal{W}eib$	$\log \mathcal{W}eib$	$\mathcal{GPD}$	$\mathcal{B}urr$	$\log \mathcal{S}_\alpha$	$\mathcal{S}_\alpha\mathcal{S}$
"Relationship"							
$\frac{\triangle \mathrm{VaR}_{0.95}}{\mathrm{VaR}_{0.95}^{class.}} \times 100$	56%	59%	59%	77%	89%	52%	66%
$\frac{\triangle \mathrm{VaR}_{0.99}}{\mathrm{VaR}_{0.99}^{class.}} \times 100$	63%	66%	67%	88%	97%	58%	78%
"Human"							
$\frac{\triangle \mathrm{VaR}_{0.95}}{\mathrm{VaR}_{0.95}^{class.}} \times 100$	68%	70%	71%	85%	89%	82%	67%
$\frac{\triangle \mathrm{VaR}_{0.99}}{\mathrm{VaR}_{0.99}^{class.}} \times 100$	75%	78%	79%	92%	95%	79%	77%
"Processes"							
$\frac{\triangle \mathrm{VaR}_{0.95}}{\mathrm{VaR}_{0.95}^{class.}} \times 100$	65%	65%	67%	88%	78%	65%	-
$\frac{\triangle \mathrm{VaR}_{0.99}}{\mathrm{VaR}_{0.99}^{class.}} \times 100$	71%	71%	73%	94%	87%	72%	-
"Technology"							
$\frac{\triangle \mathrm{VaR}_{0.95}}{\mathrm{VaR}_{0.95}^{class.}} \times 100$	10%	8%	9%	-	-	6%	-
$\frac{\triangle \mathrm{VaR}_{0.99}}{\mathrm{VaR}_{0.99}^{class.}} \times 100$	7%	8%	8%	-	-	-	-
"External"							
$\frac{\triangle \mathrm{VaR}_{0.95}}{\mathrm{VaR}_{0.95}^{class.}} \times 100$	48%	60%	53%	63%	91%	85%	53%
$\frac{\triangle \mathrm{VaR}_{0.99}}{\mathrm{VaR}_{0.99}^{class.}} \times 100$	61%	71%	65%	79%	98%	79%	68%

The magnitude of the impact of extreme events on the operational risk capital charge can serve as an important guideline for a bank to decide whether and at what price it should use insurance against extreme losses.

10.7 Appendix: Results of Empirical Study

Table 10.3: Estimated parameters for loss distributions fitted to the full and top-5%-trimmed 1980-2002 operational loss data.

	Classical	Robust
	"Relationship"	
$\mathcal{LN}$	$\mu = 16.1911$ $\sigma = 2.0654$	$\mu = 16.1722$ $\sigma = 1.7476$
$\mathcal{W}eib$	$\beta = 0.0032$ $\alpha = 0.3538$	$\beta = 0.0003$ $\alpha = 0.4952$
$\log \mathcal{W}eib$	$\beta = 0.27 \cdot 10^{-8}$ $\alpha = 7.0197$	$\beta = 1.0 \cdot 10^{-11}$ $\alpha = 9.1858$
$\mathcal{GPD}$	$\xi = 1.2852$ $\beta = 1.06 \cdot 10^{7}$	$\xi = 0.9352$ $\beta = 1.1 \cdot 10^{7}$
$\mathcal{B}urr$	$\alpha = 5.1242$ $\beta = 1.02 \cdot 10^{4}$	$\alpha = 2.6845$ $\beta = 4.1 \cdot 10^{5}$
	$\tau = 0.4644$	$\tau = 0.7242$
$\log \mathcal{S}_\alpha$	$\alpha = 1.9340$ $\beta = -1$	$\alpha = 2$ $\beta = 0.9936$
	$\sigma = 1.5198$ $\mu = 15.9616$	$\sigma = 1.2392$ $\mu = 16.1656$
$\mathcal{S}_\alpha\mathcal{S}$	$\alpha = 0.6592$ $\sigma = 9.97 \cdot 10^{6}$	$\alpha = 0.7532$ $\sigma = 9.6 \cdot 10^{6}$
	"Human"	
$\mathcal{LN}$	$\mu = 15.4627$ $\sigma = 2.5642$	$\mu = 15.6905$ $\sigma = 2.0691$
$\mathcal{W}eib$	$\beta = 0.0240$ $\alpha = 0.2526$	$\beta = 0.0030$ $\alpha = 0.3679$
$\log \mathcal{W}eib$	$\beta = 30.73 \cdot 10^{-8}$ $\alpha = 7.0197$	$\beta = 1.8 \cdot 10^{-9}$ $\alpha = 7.2258$
$\mathcal{GPD}$	$\xi = 1.6562$ $\beta = 0.61 \cdot 10^{7}$	$\xi = 1.2808$ $\beta = 0.7 \cdot 10^{7}$
$\mathcal{B}urr$	$\alpha = 0.0922$ $\beta = 2.85 \cdot 10^{27}$	$\alpha = 0.3288$ $\beta = 1.6 \cdot 10^{11}$
	$\tau = 4.4717$	$\tau = 1.7551$
$\log \mathcal{S}_\alpha$	$\alpha = 1.4042$ $\beta = -1$	$\alpha = 2$ $\beta = -0.3944$
	$\sigma = 2.8957$ $\mu = 10.5108$	$\sigma = 1.4700$ $\mu = 15.6746$
$\mathcal{S}_\alpha\mathcal{S}$	$\alpha = 0.6061$ $\sigma = 0.71 \cdot 10^{7}$	$\alpha = 0.6750$ $\sigma = 6.7 \cdot 10^{6}$
	"Processes"	
$\mathcal{LN}$	$\mu = 17.1600$ $\sigma = 2.3249$	$\mu = 17.0090$ $\sigma = 1.9917$
$\mathcal{W}eib$	$\beta = 0.0021$ $\alpha = 0.3515$	$\beta = 0.0003$ $\alpha = 0.4671$
$\log \mathcal{W}eib$	$\beta = 0.11 \cdot 10^{-8}$ $\alpha = 7.1614$	$\beta = 9.0 \cdot 10^{-12}$ $\alpha = 8.8672$
$\mathcal{GPD}$	$\xi = 1.6147$ $\beta = 2.29 \cdot 10^{7}$	$\xi = 1.1848$ $\beta = 2.4 \cdot 10^{7}$
$\mathcal{B}urr$	$\alpha = 14.3369$ $\beta = 1.20 \cdot 10^{4}$	$\alpha = 48.4907$ $\beta = 4.2 \cdot 10^{5}$
	$\tau = 0.3829$	$\tau = 0.5125$
$\log \mathcal{S}_\alpha$	$\alpha = 2.0000$ $\beta = 0.8195$	$\alpha = 2$ $\beta = -0.1606$
	$\sigma = 1.6476$ $\mu = 17.1535$	$\sigma = 1.4096$ $\mu = 17.0067$
$\mathcal{S}_\alpha\mathcal{S}$	$\alpha = 0.5478$ $\sigma = 1.99 \cdot 10^{7}$	$\alpha = 0.6087$ $\sigma = 19.9 \cdot 10^{6}$

(Continued on next page)

Table 10.3 *(Continued from previous page)*		
	Classical	Robust
	"Technology"	
$\mathcal{LN}$	$\mu = 15.1880 \quad \sigma = 2.7867$	$\mu = 15.0313 \quad \sigma = 2.8285$
$\mathcal{W}eib$	$\beta = 0.0103 \quad \alpha = 0.2938$	$\beta = 0.0120 \quad \alpha = 0.2870$
$\log \mathcal{W}eib$	$\beta = 11.06 \cdot 10^{-8} \quad \alpha = 5.7555$	$\beta = 7.7 \cdot 10^{-8} \quad \alpha = 5.8818$
$\mathcal{GPD}$	$\xi = 2.0925 \quad \beta = 0.34 \cdot 10^{7}$	$\xi = 2.1207 \quad \beta = 0.3 \cdot 10^{7}$
$\mathcal{B}urr$	$\alpha = 0.0684 \quad \beta = 8.74 \cdot 10^{20}$	$\alpha = 0.1643 \quad \beta = 0.2 \cdot 10^{5}$
	$\tau = 5.2150$	$\tau = 2.2048$
$\log \mathcal{S}_\alpha$	$\alpha = 2.0000 \quad \beta = 0.8040$	$\alpha = 2 \quad \beta = -0.4694$
	$\sigma = 1.9894 \quad \mu = 15.1351$	$\sigma = 2.0239 \quad \mu = 14.9627$
$\mathcal{S}_\alpha\mathcal{S}$	$\alpha = 0.1827 \quad \sigma = 1.70 \cdot 10^{6}$	$\alpha = 0.1773 \quad \sigma = 5.7 \cdot 10^{6}$
	"External"	
$\mathcal{LN}$	$\mu = 15.7125 \quad \sigma = 2.3639$	$\mu = 15.8095 \quad \sigma = 1.9705$
$\mathcal{W}eib$	$\beta = 0.0108 \quad \alpha = 0.2933$	$\beta = 0.0012 \quad \alpha = 0.4178$
$\log \mathcal{W}eib$	$\beta = 2.82 \cdot 10^{-8} \quad \alpha = 6.2307$	$\beta = 0.21 \cdot 10^{-9} \quad \alpha = 7.9597$
$\mathcal{GPD}$	$\xi = 1.5352 \quad \beta = 0.71 \cdot 10^{7}$	$\xi = 1.1813 \quad \beta = 7.7 \cdot 10^{6}$
$\mathcal{B}urr$	$\alpha = 0.1284 \quad \beta = 3.25 \cdot 10^{20}$	$\alpha = 1.1642 \quad \beta = 8.6 \cdot 10^{5}$
	$\tau = 3.3263$	$\tau = 0.8490$
$\log \mathcal{S}_\alpha$	$\alpha = 1.3313 \quad \beta = -1$	$\alpha = 2 \quad \beta = 0.4377$
	$\sigma = 2.7031 \quad \mu = 10.1928$	$\sigma = 1.3992 \quad \mu = 15.7960$
$\mathcal{S}_\alpha\mathcal{S}$	$\alpha = 0.5905 \quad \sigma = 0.70 \cdot 10^{7}$	$\alpha = 0.6598 \quad \sigma = 0.68 \cdot 10^{7}$

Table 10.4: Goodness-of-fit test statistics and corresponding p-values (in square brackets) for the loss distributions fitted to the 1980-2002 operational loss data, under the robust approach.

	KS	V	AD	AD_{up}	AD^2	AD^2_{up}	W^2
				"Relationship"			
$\mathcal{LN}$	1.3111	2.1614	4.5274	209.12	2.8289	34.5294	0.3485
	[<0.005]	[<0.005]	[0.094]	[>0.995]	[<0.005]	[0.031]	[<0.005]
$\mathcal{W}eib$	1.0407	1.7907	3.1282	327.36	1.5822	17.8496	0.2144
	[0.005]	[0.007]	[0.241]	[>0.995]	[<0.005]	[0.051]	[<0.005]
$\log \mathcal{W}eib$	1.0827	1.9746	3.35808	298.45	2.1510	20.5534	0.2989
	[0.005]	[<0.005]	[0.209]	[>0.995]	[<0.005]	[0.061]	[<0.005]
$\mathcal{GPD}$	1.6949	3.1270	4.8998	186.58	6.4187	43.6995	0.8247
	[<0.005]	[<0.005]	[0.072]	[>0.995]	[<0.005]	[0.024]	[<0.005]
$\mathcal{B}urr$	1.4346	2.6549	4.1987	251.92	4.3188	30.0690	0.5892
	[<0.005]	[<0.005]	[0.091]	[>0.995]	[<0.005]	[<0.005]	[<0.005]
$\log \mathcal{S}_\alpha$	1.3409	2.1544	4.5217	209.27	2.8492	34.7768	0.3579
	[<0.005]	[<0.005]	[0.078]	[>0.995]	[<0.005]	[0.006]	[<0.005]
$S\alpha S$	1.4187	2.7793	6.3995	144.74	5.5682	59.9109	0.6432
	[<0.005]	[<0.005]	[>0.995]	[>0.995]	[0.444]	[>0.995]	[<0.005]
				"Human"			
$\mathcal{LN}$	1.2655	2.0577	3.8877	263.20	1.9615	25.5394	0.2088
	[<0.005]	[<0.005]	[0.129]	[>0.995]	[<0.005]	[0.032]	[0.005]
$\mathcal{W}eib$	1.1172	1.9160	3.9489	400.50	1.4831	16.1407	0.1682
	[<0.005]	[<0.005]	[0.124]	[0.991]	[<0.005]	[0.044]	[0.2702]
$\log \mathcal{W}eib$	1.1910	2.0574	3.7219	375.73	2.1002	17.9457	0.2712
	[<0.005]	[<0.005]	[0.156]	[>0.995]	[<0.005]	[0.065]	[<0.005]
$\mathcal{GPD}$	1.4888	2.7433	4.8677	191.47	4.5564	39.6902	0.5588
	[<0.005]	[<0.005]	[0.091]	[>0.995]	[<0.005]	[0.019]	[<0.005]
$\mathcal{B}urr$	2.0257	3.4016	6.1192	149.34	7.6553	57.4970	1.1822
	[0.006]	[0.006]	[0.062]	[>0.995]	[0.006]	[0.011]	[0.006]
$\log \mathcal{S}_\alpha$	1.2826	2.0270	3.8906	262.57	1.9443	25.8006	0.2071
	[<0.005]	[<0.005]	[0.122]	[>0.995]	[<0.005]	[0.032]	[<0.005]
$S\alpha S$	1.0613	2.0876	5.1613	258.41	3.3279	38.9658	0.3708
	[0.032]	[<0.005]	[>0.995]	[>0.995]	[0.990]	[>0.995]	[<0.005]

(Continued on next page)

Table 10.4 *(Continued from previous page)*

	KS	V	AD	AD_{up}	AD^2	AD^2_{up}	W^2
				"Processes"			
$\mathcal{LN}$	0.9080	1.5288	2.9178	107.11	1.4125	14.0184	0.1825
	[0.034]	[0.040]	[0.249]	[>0.995]	[<0.005]	[0.070]	[<0.005]
$\mathcal{W}eib$	0.5271	1.0392	1.8810	163.49	0.5123	6.6792	0.0572
	[0.707]	[0.582]	[0.624]	[>0.995]	[0.201]	[0.137]	[0.377]
$\log \mathcal{W}eib$	0.5576	1.1087	2.1457	146.86	0.7221	8.0808	0.0811
	[0.593]	[0.436]	[0.499]	[>0.995]	[0.069]	[0.096]	[0.162]
$\mathcal{GPD}$	1.1004	2.1675	4.5813	96.7443	3.7234	18.1558	0.4483
	[<0.005]	[<0.005]	[0.097]	[>0.995]	[<0.005]	[0.046]	[<0.005]
$\mathcal{B}urr$	0.8113	1.4379	2.3716	185.52	1.0810	6.6638	0.1363
	[0.118]	[0.109]	[0.336]	[0.959]	[0.021]	[0.039]	[0.050]
$\log \mathcal{S}_\alpha$	0.9352	1.5433	2.9087	107.42	1.4381	14.0480	0.1915
	[0.028]	[0.026]	[0.218]	[>0.995]	[<0.005]	[0.056]	[0.008]
$S\alpha S$	1.1692	2.2717	4.3264	77.5942	2.7614	26.4799	0.3115
	[0.016]	[<0.005]	[>0.995]	[>0.995]	[0.964]	[>0.995]	[0.016]
				"Technology"			
$\mathcal{LN}$	1.0796	1.7451	2.7127	41.1440	1.3364	5.9777	0.1978
	[<0.005]	[0.005]	[0.217]	[0.989]	[<0.005]	[0.080]	[<0.005]
$\mathcal{W}eib$	1.0368	1.8359	2.7551	51.8632	1.4171	4.5168	0.2150
	[<0.005]	[<0.005]	[0.210]	[0.929]	[<0.005]	[0.087]	[<0.005]
$\log \mathcal{W}eib$	1.0358	1.9068	2.9926	50.0593	1.6058	4.7010	0.2408
	[0.005]	[<0.005]	[0.179]	[0.971]	[<0.005]	[0.097]	[<0.005]
$\mathcal{GPD}$	1.1362	1.7691	2.8950	32.5285	1.5441	8.3021	0.2295
	[<0.005]	[<0.005]	[0.213]	[>0.995]	[<0.005]	[0.070]	[<0.005]
$\mathcal{B}urr$	1.1179	1.8744	2.5384	28.0818	1.8242	9.7598	0.3061
	[0.356]	[0.344]	[0.522]	[>0.995]	[0.345]	[0.361]	[0.346]
$\log \mathcal{S}_\alpha$	1.0877	1.7429	2.7385	40.9965	1.3202	6.0542	0.1961
	[<0.005]	[0.006]	[0.222]	[0.990]	[0.006]	[0.068]	[0.006]
$S\alpha S$	2.8693	2.9544	6.0980	33.9452	19.9170	28.8310	3.7892
	[0.918]	[0.990]	[>0.995]	[>0.995]	[>0.995]	[>0.995]	[0.084]

(Continued on next page)

Table 10.4 *(Continued from previous page)*

	KS	V	AD	AD_{up}	AD^2	AD^2_{up}	W^2
"External"							
$\mathcal{LN}$	0.8005	1.5985	2.5289	89.9172	1.2314	10.3629	0.1581
	[0.074]	[0.017]	[0.331]	[>0.995]	[<0.005]	[0.080]	[0.013]
$\mathcal{W}eib$	0.8193	1.3842	2.1208	118.00	0.8992	6.3242	0.1149
	[0.074]	[0.108]	[0.469]	[>0.995]	[0.038]	[0.102]	[0.069]
$\log \mathcal{W}eib$	0.9288	1.5545	2.3070	114.96	1.1789	6.9924	0.1550
	[0.030]	[0.034]	[0.430]	[>0.995]	[0.005]	[0.124]	[0.017]
$\mathcal{GPD}$	1.0889	2.1497	3.2082	78.7580	2.4537	14.2314	0.3238
	[<0.005]	[<0.005]	[0.193]	[>0.995]	[<0.005]	[0.063]	[<0.005]
$\mathcal{B}urr$	1.0552	2.0537	2.8205	85.6792	2.1547	13.0326	0.2922
	[0.106]	[0.005]	[0.362]	[>0.995]	[<0.005]	[0.006]	[0.019]
$\log \mathcal{S}_\alpha$	0.8213	1.5891	2.5280	89.9499	1.2311	10.4590	0.1605
	[0.046]	[0.012]	[0.302]	[>0.995]	[<0.005]	[0.088]	[0.008]
$S\alpha S$	0.8182	1.6214	3.4638	67.2664	1.7561	17.3323	0.2046
	[0.076]	[0.034]	[>0.995]	[>0.995]	[0.850]	[>0.995]	[0.020]

Table 10.5: Estimated 1-year EL, VaR, and CVaR values ($\times 10^{10}$) for the full and top-5%-trimmed 1980-2002 operational loss data.

		EL	$\mathrm{VaR}_{0.95}$	$\mathrm{VaR}_{0.99}$	$\mathrm{CVaR}_{0.95}$	$\mathrm{CVaR}_{0.99}$
			"Relationship"			
$\mathcal{LN}$	Classical	0.1634	0.4662	1.0644	0.9016	1.9091
	Robust	0.0826	0.2068	0.3947	0.3450	0.6560
$\mathcal{W}eib$	Classical	0.1284	0.3187	0.5121	0.4430	0.6689
	Robust	0.0638	0.1307	0.1766	0.1604	0.2090
$\log \mathcal{W}eib$	Classical	-	0.3332	0.5902	-	-
	Robust	-	0.1355	0.1924	-	-
$\mathcal{GPD}$	Classical	-	1.5756	11.3028	-	-
	Robust	0.3013	0.3627	1.4156	5.2570	23.7687
$\mathcal{B}urr$	Classical	-	1.5713	11.5519	-	-
	Robust	0.0732	0.1715	0.3333	0.3131	0.6709
$\log \mathcal{S}_\alpha$	Classical	-	0.4359	0.9557	-	-
	Robust	-	0.2106	0.4006	-	-
$\mathcal{S}_\alpha\mathcal{S}$	Classical	-	4.5476	56.2927	-	-
	Robust	-	1.5339	12.2508	-	-
			"Human"			
$\mathcal{LN}$	Classical	0.4171	1.2161	3.4190	3.3869	9.4520
	Robust	0.1497	0.3953	0.8625	0.7443	1.5569
$\mathcal{W}eib$	Classical	0.2881	0.7997	1.5772	1.3232	2.3746
	Robust	0.1095	0.2377	0.3504	0.3096	0.4385
$\log \mathcal{W}eib$	Classical	-	0.8672	1.8603	-	-
	Robust	-	0.2527	0.3958	-	-
$\mathcal{GPD}$	Classical	-	12.1150	168.64	-	-
	Robust	-	1.7932	13.4743	-	-
$\mathcal{B}urr$	Classical	-	94.8281	3042.32	-	-
	Robust	-	10.6753	161.59	-	-
$\log \mathcal{S}_\alpha$	Classical	-	2.2737	4.2319	-	-
	Robust	-	0.4074	0.8983	-	-
$\mathcal{S}_\alpha\mathcal{S}$	Classical	-	14.5771	203.24	-	-
	Robust	-	4.7612	47.7160	-	-

(Continued on next page)

Table 10.5 *(Continued from previous page)*

		EL	$\text{VaR}_{0.95}$	$\text{VaR}_{0.99}$	$\text{CVaR}_{0.95}$	$\text{CVaR}_{0.99}$
			"Processes"			
$\mathcal{LN}$	Classical	0.8457	2.5610	6.5625	5.7823	13.9079
	Robust	0.3188	0.8921	1.9222	1.6858	3.5673
$\mathcal{W}eib$	Classical	0.5131	1.2761	2.1308	1.8257	2.8578
	Robust	0.2121	0.4439	0.6272	0.5568	0.7447
$\log\mathcal{W}eib$	Classical	-	1.4780	2.6511	-	-
	Robust	-	0.4911	0.7287	-	-
$\mathcal{GPD}$	Classical	-	20.8700	262.52	-	-
	Robust	-	2.5224	15.4268	-	
$\mathcal{B}urr$	Classical	-	1.7987	4.1859	-	-
	Robust	0.1987	0.4053	0.5646	0.5043	0.6731
$\log\mathcal{S}_\alpha$	Classical	-	2.5394	6.7070	-	-
	Robust	-	0.8842	1.8991	-	-
$\mathcal{S}_\alpha\mathcal{S}$	Classical	-	74.9073	1280.02	-	-
	Robust	-	$2.04\cdot10^{11}$	$28.9\cdot10^{11}$	-	-
			"Technology"			
$\mathcal{LN}$	Classical	0.0958	0.2898	1.2741	1.5439	5.4865
	Robust	0.0921	0.2617	1.1838	1.3261	4.5793
$\mathcal{W}eib$	Classical	0.0358	0.1454	0.3625	0.2958	0.6180
	Robust	0.0334	0.1336	0.3326	0.2787	0.5962
$\log\mathcal{W}eib$	Classical	-	0.1670	0.4747	-	-
	Robust	-	0.1514	0.4355	-	-
$\mathcal{GPD}$	Classical	-	1.6249	54.4650	-	-
	Robust	-	$1.47\cdot10^{10}$	$45.1\cdot10^{10}$	-	-
$\mathcal{B}urr$	Classical	-	9.0358	855.78	-	-
	Robust	-	$7.73\cdot10^{10}$	$801\cdot10^{10}$	-	-
$\log\mathcal{S}_\alpha$	Classical	-	0.2990	1.2312	-	-
	Robust	-	0.2802	1.2870	-	-
$\mathcal{S}_\alpha\mathcal{S}$	Classical	-	$7.1\cdot10^{6}$	$6.9\cdot10^{10}$	-	-
	Robust	-	$5.89\cdot10^{17}$	$5.60\cdot10^{21}$	-	-

(Continued on next page)

Table 10.5 *(Continued from previous page)*						
		EL	$\text{VaR}_{0.95}$	$\text{VaR}_{0.99}$	$\text{CVaR}_{0.95}$	$\text{CVaR}_{0.99}$
			"External"			
$\mathcal{LN}$	Classical	0.0327	0.1126	0.4257	0.3962	1.1617
	Robust	0.0154	0.0580	0.1642	0.1397	0.3334
$\mathcal{W}eib$	Classical	0.0208	0.0885	0.2494	0.2025	0.4509
	Robust	0.0088	0.0354	0.0715	0.0599	0.1066
$\log\mathcal{W}eib$	Classical	-	0.0839	0.2489	-	-
	Robust	-	0.0395	0.0865	-	-
$\mathcal{GPD}$	Classical	-	0.2562	2.6514	-	-
	Robust	-	0.0943	0.5604	-	-
$\mathcal{B}urr$	Classical	-	0.7165	15.8905	-	-
	Robust	-	0.0676	0.3246	-	-
$\log\mathcal{S}_\alpha$	Classical	-	0.3879	0.8064	-	-
	Robust	-	0.0570	0.1695	-	-
$\mathcal{S}_\alpha\mathcal{S}$	Classical	-	0.4714	7.6647	-	-
	Robust	-	0.2234	2.4408	-	-

Table 10.6: Average forecast errors for "Relationship" type aggregated losses: robust approach. *Left:* errors between corresponding quantiles; *middle:* errors of forecasted quantiles relative to realized loss; *right:* overall error between forecasted and realized loss.

	Forecasted quantiles vs. bootstrapped quantiles		Forecasted quantiles vs. actual loss		Overall error: forecasted vs. actual loss	
%	MSE ($\times 10^{20}$)	MAE ($\times 10^{10}$)	MSE ($\times 10^{20}$)	MAE ($\times 10^{10}$)	MAE ($\times 10^{20}$)	MSE ($\times 10^{10}$)
			$\mathcal{LN}$			
25	0.0022	0.0417	0.0018	0.0394		
50	0.0040	0.0509	0.0039	0.0502		
75	0.0092	0.0753	0.0120	0.0921	0.0274	0.0859
95	0.0431	0.1916	0.0629	0.2368		
99	0.1704	0.3952	0.2283	0.4617		
99.9	1.1851	1.0771	1.3920	1.1685		
			$\mathcal{W}eib$			
25	0.0018	0.0390	0.0017	0.0367		
50	0.0026	0.0446	0.0025	0.0439		
75	0.0039	0.0509	0.0055	0.0556	0.0052	0.0552
95	0.0083	0.0729	0.0170	0.1181		
99	0.0151	0.1069	0.0340	0.1733		
99.9	0.0288	0.1583	0.0667	0.2498		
			$\log \mathcal{W}eib$			
25	0.0017	0.0384	0.0016	0.0361		
50	0.0025	0.0442	0.0024	0.0436		
75	0.0040	0.0512	0.0056	0.0572	0.0055	0.0561
95	0.0090	0.0782	0.0183	0.1234		
99	0.0182	0.1227	0.0394	0.1892		
99.9	0.0475	0.2101	0.0951	0.3017		
			$\mathcal{GPD}$			
25	0.0023	0.0425	0.0019	0.0403		
50	0.0056	0.0563	0.0054	0.0557		
75	0.0205	0.1200	0.0249	0.1368	10.6607	0.2400
95	0.3315	0.5390	0.3838	0.5842		
99	5.3368	2.1998	5.6442	2.2663		
99.9	345.46	17.1380	348.73	17.2293		
			$\mathcal{B}urr$			
25	0.0018	0.0384	0.0016	0.0362		
50	0.0033	0.0465	0.0031	0.0459		
75	0.0090	0.0781	0.0120	0.0948	1.7318	0.1225
95	0.0855	0.2745	0.1132	0.3196		
99	0.9772	0.9112	1.1087	0.9777		
99.9	31.5765	4.9530	32.5110	5.0443		
			$\log \mathcal{S}_\alpha$			
25	0.0006	0.0223	0.0015	0.0353		
50	0.0016	0.0284	0.0016	0.0291		
75	0.0064	0.0626	0.0078	0.0731	0.0844	0.0937
95	0.0899	0.2662	0.1150	0.3099		
99	0.6948	0.7844	0.7980	0.8476		
99.9	8.5024	2.7599	8.9913	2.8457		
			$\mathcal{S}_\alpha\mathcal{S}$			
25	0.0072	0.0643	0.0053	0.0586		
50	0.0326	0.1493	0.0321	0.1478		
75	0.2309	0.4306	0.2462	0.4474	$1.0 \cdot 10^5$	6.7593
95	13.7673	3.4414	14.0927	3.4866		
99	869.34	27.2307	873.16	27.2970		
99.9	$5.1 \cdot 10^5$	674.49	$5.1 \cdot 10^5$	674.58		

Table 10.7: Average forecast errors for "Human" type aggregated losses: robust approach. *Left:* errors between corresponding quantiles; *middle:* errors of forecasted quantiles relative to realized loss; *right:* overall error between forecasted and realized loss.

	Forecasted quantiles vs. bootstrapped quantiles		Forecasted quantiles vs. actual loss		Overall error: forecasted vs. actual loss	
%	MSE ($\times 10^{20}$)	MAE ($\times 10^{10}$)	MSE ($\times 10^{20}$)	MAE ($\times 10^{10}$)	MAE ($\times 10^{20}$)	MSE ($\times 10^{10}$)
			$\mathcal{LN}$			
25	0.0023	0.0356	0.0029	0.0468		
50	0.0050	0.0479	0.0046	0.0465		
75	0.0161	0.1099	0.0221	0.1364	0.1233	0.1392
95	0.1414	0.3660	0.1997	0.4399		
99	0.8455	0.9129	1.0574	1.0231		
99.9	10.0861	3.1153	11.0224	3.2675		
			$\mathcal{W}eib$			
25	0.0021	0.0357	0.0030	0.0477		
50	0.0032	0.0413	0.0031	0.0408		
75	0.0061	0.0558	0.0090	0.0785	0.0123	0.0789
95	0.0221	0.1322	0.0455	0.2062		
99	0.0587	0.2302	0.1190	0.3403		
99.9	0.2154	0.4556	0.3734	0.6083		
			$\log \mathcal{W}eib$			
25	0.0022	0.0362	0.0030	0.0481		
50	0.0034	0.0411	0.0032	0.0406		
75	0.0069	0.0589	0.0100	0.0829	0.0144	0.0828
95	0.0269	0.1468	0.0526	0.2207		
99	0.0820	0.2745	0.1529	0.3846		
99.9	0.3523	0.5741	0.5514	0.7270		
			$\mathcal{GPD}$			
25	0.0033	0.0361	0.0029	0.0428		
50	0.0171	0.1118	0.0163	0.1085		
75	0.1576	0.3784	0.1783	0.4050	$4.9 \cdot 10^{6}$	15.7218
95	10.6869	3.1887	11.1722	3.2626		
99	736.10	26.7003	742.06	26.8102		
99.9	$2.6 \cdot 10^{5}$	455.70	$2.6 \cdot 10^{5}$	455.85		
			$\mathcal{B}urr$			
25	0.0174	0.1131	0.0116	0.0828		
50	0.2081	0.4244	0.2052	0.4212		
75	4.0533	1.9078	4.1594	1.9344	$3.1 \cdot 10^{11}$	4224.2
95	1583.8	38.2283	1589.6	38.302		
99	$5.9 \cdot 10^{5}$	721.19	$5.9 \cdot 10^{5}$	721.30		
99.9	$4.5 \cdot 10^{9}$	$6.0 \cdot 10^{4}$	$4.5 \cdot 10^{9}$	$6.0 \cdot 10^{4}$		
			$\log \mathcal{S}_{\alpha}$			
25	0.0024	0.0364	0.0031	0.0484		
50	0.0044	0.0431	0.0042	0.0427		
75	0.0111	0.0726	0.0151	0.0993	0.0430	0.0994
95	0.0651	0.2067	0.1008	0.2807		
99	0.3245	0.4564	0.4411	0.5663		
99.9	3.5175	1.3713	3.9992	1.5242		
			$\mathcal{S}_{\alpha}\mathcal{S}$			
25	0.0090	0.0731	0.0059	0.0489		
50	0.0619	0.2265	0.0604	0.2232		
75	0.6697	0.7782	0.7125	0.8048	$5.3 \cdot 10^{7}$	54.9360
95	71.5082	8.1565	72.7595	8.5327		
99	7551.45	85.2174	7570.7	2661.2		
99.9	$7.6 \cdot 10^{6}$	2661.0	$7.6 \cdot 10^{6}$	54.9360		

Table 10.8: Average forecast errors for “Processes” type aggregated losses: robust approach. *Left:* errors between corresponding quantiles; *middle:* errors of forecasted quantiles relative to realized loss; *right:* overall error between forecasted and realized loss.

	Forecasted quantiles vs. bootstrapped quantiles		Forecasted quantiles vs. actual loss		Overall error: forecasted vs. actual loss	
%	MSE ($\times 10^{20}$)	MAE ($\times 10^{10}$)	MSE ($\times 10^{20}$)	MAE ($\times 10^{10}$)	MAE ($\times 10^{20}$)	MSE ($\times 10^{10}$)
			$\mathcal{LN}$			
25	0.0331	0.1494	0.0516	0.1785		
50	0.0392	0.1687	0.0400	0.1704		
75	0.0538	0.1963	0.0485	0.1893	0.2313	0.2548
95	0.2980	0.4759	0.3717	0.5666		
99	1.9817	1.3538	2.3595	1.5047		
99.9	22.4811	4.5575	24.1880	4.7621		
			$\mathcal{W}eib$			
25	0.0332	0.1492	0.0519	0.1783		
50	0.0406	0.1697	0.0416	0.1714		
75	0.0499	0.1893	0.0385	0.1697	0.0537	0.1869
95	0.0791	0.2440	0.0750	0.2330		
99	0.1437	0.3271	0.1833	0.3760		
99.9	0.3465	0.5078	0.5118	0.6844		
			$\log \mathcal{W}eib$			
25	0.0335	0.1498	0.0524	0.1789		
50	0.0405	0.1702	0.0415	0.1720		
75	0.0499	0.1893	0.0390	0.1706	0.0593	0.1931
95	0.0877	0.2569	0.0924	0.2660		
99	0.2033	0.3856	0.2751	0.4826		
99.9	0.7410	0.7835	1.0267	0.9885		
			$\mathcal{GPD}$			
25	0.0033	0.0361	0.0029	0.0428		
50	0.0171	0.1118	0.0163	0.1085		
75	0.1576	0.3784	0.1783	0.4050	$4.9 \cdot 10^{6}$	15.7218
95	10.6869	3.1887	11.1722	3.2626		
99	736.10	26.7003	742.06	26.8102		
99.9	$2.6 \cdot 10^{5}$	455.70	$2.6 \cdot 10^{5}$	455.85		
			$\mathcal{B}urr$			
25	0.0174	0.1131	0.0116	0.0828		
50	0.2081	0.4244	0.2052	0.4212		
75	4.0533	1.9078	4.1594	1.9344	$3.1 \cdot 10^{11}$	4224.2
95	1583.8	38.2283	1589.6	38.302		
99	$5.9 \cdot 10^{5}$	721.19	$5.9 \cdot 10^{5}$	721.30		
99.9	$4.5 \cdot 10^{9}$	$6.0 \cdot 10^{4}$	$4.5 \cdot 10^{9}$	$6.0 \cdot 10^{4}$		
			$\log \mathcal{S}_\alpha$			
25	0.0024	0.0364	0.0031	0.0484		
50	0.0044	0.0431	0.0042	0.0427		
75	0.0111	0.0726	0.0151	0.0993	0.0430	0.0994
95	0.0651	0.2067	0.1008	0.2807		
99	0.3245	0.4564	0.4411	0.5663		
99.9	3.5175	1.3713	3.9992	1.5242		
			$\mathcal{S}_\alpha\mathcal{S}$			
25	0.0090	0.0731	0.0059	0.0489		
50	0.0619	0.2265	0.0604	0.2232		
75	0.6697	0.7782	0.7125	0.8048	$5.3 \cdot 10^{7}$	54.9360
95	71.5082	8.1565	72.7595	8.5327		
99	7551.45	85.2174	7570.7	2661.2		
99.9	$7.6 \cdot 10^{6}$	2661.0	$7.6 \cdot 10^{6}$	54.9360		

Table 10.9: Average forecast errors for "Technology" type aggregated losses: robust approach. *Left:* errors between corresponding quantiles; *middle:* errors of forecasted quantiles relative to realized loss; *right:* overall error between forecasted and realized loss.

	Forecasted quantiles vs. bootstrapped quantiles		Forecasted quantiles vs. actual loss		Overall error: forecasted vs. actual loss	
%	MSE ($\times 10^{20}$)	MAE ($\times 10^{10}$)	MSE ($\times 10^{20}$)	MAE ($\times 10^{10}$)	MAE ($\times 10^{20}$)	MSE ($\times 10^{10}$)
			$\mathcal{LN}$			
25	0.0005	0.0144	0.0011	0.0206		
50	0.0008	0.0247	0.0008	0.0252		
75	0.0036	0.0512	0.0039	0.0516	1.2010	0.1449
95	0.1906	0.4121	0.2055	0.4303		
99	3.9387	1.8837	4.0368	1.9095		
99.9	186.07	12.7354	186.96	12.7711		
			$\mathcal{W}eib$			
25	0.0005	0.0146	0.0011	0.0206		
50	0.0007	0.0241	0.0008	0.0246		
75	0.0020	0.0427	0.0020	0.0392	0.0131	0.0523
95	0.0277	0.1572	0.0326	0.1754		
99	0.1928	0.4277	0.2130	0.4534		
99.9	1.2291	1.0954	1.3054	1.1315		
			$\log \mathcal{W}eib$			
25	0.0005	0.0147	0.0010	0.0204		
50	0.0007	0.0241	0.0008	0.0246		
75	0.0021	0.0429	0.0021	0.0401	0.0256	0.0590
95	0.0368	0.1830	0.0429	0.2012		
99	0.3080	0.5413	0.3367	0.5670		
99.9	3.1803	1.7301	3.3058	1.7656		
			$\mathcal{GPD}$			
25	0.0005	0.0140	0.0011	0.0206		
50	0.0008	0.0255	0.0009	0.0260		
75	0.0205	0.1151	0.0220	0.1219	$2.1 \cdot 10^{13}$	$2.4 \cdot 10^{4}$
95	54.4298	5.6635	54.6411	5.6817		
99	$1.1 \cdot 10^{5}$	228.25	$1.1 \cdot 10^{5}$	228.27		
99.9	$2.9 \cdot 10^{9}$	$3.5 \cdot 10^{4}$	$2.9 \cdot 10^{9}$	$3.5 \cdot 10^{4}$		
			$\mathcal{B}urr$			
25	0.0005	0.0138	0.0011	0.0204		
50	0.0010	0.0293	0.0011	0.2979		
75	0.1533	0.3034	0.1576	0.3103	$1.0 \cdot 10^{21}$	$1.3 \cdot 10^{8}$
95	2873.8	39.8687	2875.3	39.8869		
99	$3.1 \cdot 10^{7}$	3789.3	$3.1 \cdot 10^{7}$	3789.3		
99.9	$1.8 \cdot 10^{13}$	$2.4 \cdot 10^{16}$	$1.8 \cdot 10^{13}$	$2.4 \cdot 10^{6}$		
			$\log \mathcal{S}_\alpha$			
25	0.0005	0.0144	0.0011	0.0206		
50	0.0008	0.0244	0.0008	0.0249		
75	0.0035	0.0505	0.0037	0.0509	4.4229	0.1518
95	0.1749	0.4035	0.1896	0.4219		
99	3.6596	1.8473	3.7559	1.8731		
99.9	186.17	12.0747	187.10	12.1103		
			$\mathcal{S}_\alpha\mathcal{S}$			
25	570.47	13.0921	570.22	13.0862		
50	$4.9 \cdot 10^{8}$	$1.2 \cdot 10^{4}$	$4.9 \cdot 10^{8}$	$1.2 \cdot 10^{4}$		
75	$1.7 \cdot 10^{16}$	$6.9 \cdot 10^{7}$	$1.7 \cdot 10^{16}$	$6.9 \cdot 10^{7}$	$7.3 \cdot 10^{78}$	$1.0 \cdot 10^{37}$
95	$1.4 \cdot 10^{31}$	$2.0 \cdot 10^{15}$	$1.4 \cdot 10^{31}$	$2.0 \cdot 10^{15}$		
99	$4.4 \cdot 10^{45}$	$3.0 \cdot 10^{22}$	$4.4 \cdot 10^{45}$	$3.0 \cdot 10^{22}$		
99.9	$1.3 \cdot 10^{65}$	$1.6 \cdot 10^{32}$	$1.3 \cdot 10^{65}$	$1.6 \cdot 10^{32}$		

Table 10.10: Average forecast errors for "External" type aggregated losses: robust approach. *Left:* errors between corresponding quantiles; *middle:* errors of forecasted quantiles relative to realized loss; *right:* overall error between forecasted and realized loss.

	Forecasted quantiles vs. bootstrapped quantiles		Forecasted quantiles vs. actual loss		Overall error: forecasted vs. actual loss	
%	MSE ($\times 10^{20}$)	MAE ($\times 10^{10}$)	MSE ($\times 10^{20}$)	MAE ($\times 10^{10}$)	MAE ($\times 10^{20}$)	MSE ($\times 10^{10}$)
			$\mathcal{LN}$			
25	0.0002	0.0143	0.0004	0.0173		
50	0.0004	0.0171	0.0004	0.0171		
75	0.0012	0.0256	0.0017	0.0326	0.2235	0.0486
95	0.0179	0.1146	0.0254	0.1437		
99	0.1615	0.3686	0.1946	0.4108		
99.9	2.6093	1.4848	2.7830	1.5420		
			$\mathcal{W}eib$			
25	0.0002	0.0145	0.0004	0.0169		
50	0.0003	0.0157	0.0003	0.0157		
75	0.0006	0.0199	0.0009	0.0238	0.0019	0.0285
95	0.0037	0.0484	0.0071	0.0752		
99	0.0141	0.1041	0.0244	0.1464		
99.9	0.0591	0.2240	0.0868	0.2810		
			$\log \mathcal{W}eib$			
25	0.0002	0.0142	0.0004	0.0169		
50	0.0003	0.0153	0.0003	0.0153		
75	0.0006	0.0195	0.0009	0.0232	0.0021	0.0285
95	0.0037	0.0480	0.0071	0.0760		
99	0.0149	0.1097	0.0256	0.1518		
99.9	0.0775	0.2561	0.1084	0.3130		
			$\mathcal{GPD}$			
25	0.0002	0.0143	0.0005	0.0182		
50	0.0006	0.0205	0.0006	0.0205		
75	0.0047	0.0486	0.0059	0.0577	$0.23{\cdot}10^5$	1.6344
95	0.6101	0.5945	0.6453	0.6236		
99	45.2582	5.1661	45.7041	5.2080		
99.9	$0.3{\cdot}10^5$	124.80	$0.3{\cdot}10^5$	124.85		
			$\mathcal{B}urr$			
25	0.0003	0.0161	0.0005	0.0183		
50	0.0019	0.0344	0.0019	0.0344		
75	0.0568	0.1692	0.0611	0.1785	$0.31{\cdot}10^{10}$	401.06
95	36.6345	4.1927	36.8775	4.2218		
99	$0.2{\cdot}10^5$	90.7056	$0.2{\cdot}10^5$	90.7477		
99.9	$2.8{\cdot}10^8$	$0.1{\cdot}10^5$	$2.8{\cdot}10^8$	$0.1{\cdot}10^5$		
			$\log \mathcal{S}_\alpha$			
25	0.0002	0.0141	0.0004	0.0176		
50	0.0004	0.0168	0.0004	0.0168		
75	0.0012	0.0253	0.0016	0.0307	0.0154	0.0416
95	0.0153	0.0972	0.0215	0.1262		
99	0.1256	0.2958	0.1510	0.3377		
99.9	1.7147	1.0895	1.8363	1.1464		
			$\mathcal{S}_\alpha\mathcal{S}$			
25	0.0003	0.0176	0.0005	0.0174		
50	0.0014	0.0305	0.0014	0.0305		
75	0.0200	0.1080	0.0229	0.1171	$2.34{\cdot}10^7$	29.9384
95	3.5661	1.5601	3.6621	1.5892		
99	620.34	19.6938	622.08	19.7359		
99.9	$12.5{\cdot}10^5$	812.08	$12.5{\cdot}10^5$	812.14		

Table 10.11: Averaged p-values for aggregated losses in the 7-year 1996-2002 forecast period, under the robust approach.

$\mathcal{LN}$	$\mathcal{W}eib$	$\log \mathcal{W}eib$	$\mathcal{GPD}$	$\mathcal{B}urr$	$\log \mathcal{S}_\alpha$	$\mathcal{S}_\alpha\mathcal{S}$
			"Relationship"			
0.5080	0.5260	0.4975	0.4246	0.4410	0.3679	0.5180
			"Human"			
0.4821	0.4947	0.4774	0.4401	0.4730	0.1443	0.4841
			"Processes"			
0.2190	0.2008	0.2060	0.2526	0.2433	0.1339	0.2103
			"Technology"			
0.3373	0.3266	0.3278	0.3303	0.3428	0.3392	0.2148
			"External"			
0.4876	0.5131	0.5290	0.4730	0.4395	0.3272	0.4704

Chapter 11

Conclusions and Future Research Directions

11.1 Conclusions

In Chapter 2, we defined operational risk and summarized its place among other risks faced by financial institutions. Although the official definition of operational risk has been widely accepted by the industry, the idiosyncratic nature of operational risk requires that banks adjust this definition to account for peculiarities of their business profile, capital structure, and other specifics.

In Chapter 3 we focused on the significance of operational risk and its potential dangers. We discussed the effects of globalization and financial deregulation on the risk exposure and further illustrated some examples of large-scale operational losses in the financial industry from the last two decades.

In Chapter 4 we discussed the regulatory requirements of the Basel II regarding the operational risk capital charge. The Loss Distribution Approach is of our primary interest, as it suggests an estimation of the capital charge that relies on the statistical techniques for modeling operational risk.

Chapter 5 looked at exploratory data analysis of the operational loss data. We concluded that the loss amounts need to be modeled with a heavy-tailed distribution, and that the frequency distribution follows a non-homogeneous Poisson process. We also discussed the aggregation of the frequency and severity distributions and talked about VaR and CVaR as two alternative measures of risk to be used as a proxy for the operational risk capital charge.

In Chapter 6 we focused on a wide class of α-Stable distributions that possess attractive features making them applicable to operational risk modeling: (1) high flexibility due to four parameters, (2) stability under linear transformations, and (3) the power-law tail decay that captures heavy tails. Variations of the α-Stable distributions were considered before being applied to the operational loss data. They include symmetric α-Stable distribution, logα-Stable distribution, and left-truncated α-Stable distribution. Applications of the α-Stable distribution to actual operational loss data provided empirical evidence that the operational loss data are severely heavy-tailed and are well captured by the variations of the α-Stable distributions.

Chapter 7 was devoted to modeling the frequency distribution with a Cox process with stochastic intensity. In related literature, majority of empirical studies rely on the simple homogeneous Poisson assumption for the frequency distribution of the losses. In the chapter we refuted such assumption and proposed analytic formulae for the non-homogeneous frequency rate functions. Necessary tests were performed against real operational frequency data and the assumption on non-homogeneity was strongly supported.

In Chapter 8 we discussed the reporting bias problem. In the first half of the chapter, we described theoretical implications of ignoring unrecorded data and proposed an effective solution to account for the resulting bias. Under the "naive" approach that ignores the missing data (and the one most frequently used by the industry), the loss distribution becomes misspecified and the frequency of losses is underestimated. We suggested a conditional approach that allows one to capture the portion of missing data and to correctly estimate the unknown parameters of both loss and frequency distributions. Theoretical studies, supplemented by extensive Monte Carlo simulations, revealed that using "naive" approach may have serious financial consequences by underestimating the capital charge. In the second half of the chapter, we tested the methodology against real operational loss data. A comprehensive empirical study supported the theoretical findings that EL, VaR, and CVaR are severely understated whenever a wrong approach is used.

In Chapter 9 we proposed two modified versions of the Anderson-Darling goodness-of-fit statistic in which the weighting function is increasing in the direction of the upper tail of the distribution. A supremum and quadratic versions of the statistic were proposed. The statistic was applied to check the goodness of fit of a number of distributions using operational loss data and catastrophe insurance claims data sets. From the empirical analysis we concluded that in the upper quantiles heavier-tailed distributions better fit the data than Lognormal or thinner-tailed distributions in many instances.

Chapter 10 was devoted to robust methods and their applications to operational loss data. Mathematically elegant classical estimation procedures may behave poorly under minor departures in the data from the model assumptions and in the presence of outliers classical procedures may produce biased estimates of the model parameters and vital statistics. We applied robust methods to investigate the marginal contribution of extreme low-frequency events to EL, VaR, and CVaR. Empirical findings demonstrated that the incremental contribution of extreme events that account for the highest 5% of data stands well beyond 50% of the annual aggregate expected loss and the operational risk regulatory capital charge.

11.2 Future Research Directions

Operational risk modelling and management is not limited to financial institutions, and similar models will find application in private corporations. Future research interests reside in the following key areas.

Integrated Risk Management

We plan to undertake a global approach to modelling financial risk in banks – *Integrated Risk Management* (IRM). In this approach, a bank's loss data would be analyzed in a single integrated model. Relevant directions for the research include:

- Examination of a bank's loss data in a multiple regression model with explanatory variables spanned by a number of market risk, credit risk, and operational risk associated factors. In this *factor analysis model*, the contribution of market, credit, and operational risk to the total risk is captured by the beta coefficients. Currently available estimates suggest that roughly 60%, 15%, and 25% of bank's regulatory capital is allocated for credit, market, and operational risks, respectively (proportional to the shares of financial risk attributed to the risk type). The integrated regression model would enable one to formalize the results. The regression error is expected to be attributed to political, reputational, and other risks outside the scope of credit, market, and operational risks.

- *Principal component analysis* and development of *model reduction* tools upon the assessment of the sensitivity of a bank's losses to various risks. In particular, factor analysis performed with operational loss data would serve the purpose of detecting key drivers behind the operational risk exposure.

- Checking for possible inter-dependence between market, credit, and operational risk types. Since multivariate distributions rely on a linear dependence structure, they may not be optimal to sufficiently capture the dependence when the underlying data is heavy-tailed. Hence, alternative models such as copulas (t-copulas, Gumbel copulas, semi- parametric copulas, etc.) would be brought into study.

Smoothly Truncated α-Stable Distributions

A significant portion of the empirical and theoretical analysis in this dissertation involved αStable distributions. αStable distributions have a number of highly attractive features making them applicable to operational risk modeling: (1) stability under affine transformations, (2) flexibility due to four parameters (location, scale, shape, and skewness), and (3) ability to capture heavy tails. We considered variations of αStable distributions: log-αStable, symmetric αStable (fitted to symmetrized data), and the regular αStable – they were supported by our empirical findings.

Apart from the case when $\alpha = 2$ (Gaussian), the second moment is infinite and the first moment is finite only for $\alpha > 1$. We plan to examine the following modified versions of the α-Stable distribution: (1) right-truncated α-Stable, in which a truncated loss distribution is defined on the support bounded above by a prespecified (or random) threshold, (2) smoothly truncated α-Stable in which heavy tails are replaced by Exponential tails (the Lévy flight), and (3) smoothly truncated α-Stable, in which heavy tails are replaced by Gaussian tails [125]. As a further generalization, one may replace Gaussian tails by Weibull or Gamma tails, and test the models with real data problems.

Extreme Value Theory

Many existing operational loss models rely on Extreme Value Theory. Under this approach, the parametric assessment of the loss distribution is concerned only with the extreme events (modeled with the Generalized Pareto Distribution (GPD)), while lower-magnitude losses are given a lesser attention (and are often modeled with the empirical distribution function). In our opinion, this approach, in its current development, has two significant drawbacks. First, the use of the empirical loss distribution for low-magnitude and medium-magnitude losses can be justified only by a comprehensive dataset (interpreted as the population). For example, the Hill estimator can really work for sample of size 500,000, even if the sample is taken from alpha-stable distribution with index alpha=1.8 [137]. However, scarcity of available data would not allow for such an assumption to be valid. Second, existing methodologies on detecting the high threshold (above which the losses are to follow the GPD) do not possess sufficient analytical grounds and are rather *ad hoc*: the common procedure relies on a visual inspection of the mean excess plot and selection of the threshold above which the graph follows a fairly straight line. The parameters estimated with the GPD distribution are highly sensitive to the choice of the threshold. In this light, relevant research direction includes the following:

- Developing formal statistical procedures for choosing an optimal high threshold. One possibility would be to view the problem as a *constrained minimization problem*, in which the threshold point is chosen such that the distance between the empirical and fitted distribution functions is minimized in the upper tail (above the threshold), according to a particular measure: MSE, KS, AD, or another measure.

- Analyzing models in which the dataset is split into two subsets, with the losses below the high threshold modelled with a parametric (rather than empirical) loss distribution, and the losses above the threshold modelled with the GPD distribution.

- Studying *mixture models* applied to the full data to simultaneously fit low-magnitude and high-magnitude losses and extreme losses. Due to the scarcity in the data, an optimal statistical procedure would be to use the *Expectation-Maximization algorithm*. Certainly, all competing models should be analyzed and tested against real data.

Dependent Random Sum Models

In Chapter 5 we discussed the aggregation of the frequency and severity distributions. Actuarial models currently used are limited due to simplifying assumptions on the loss and frequency distributions: *iid* and stationarity. We intend to relax these assumptions and evaluate *generalized random sum models* with non-stationary and possibly dependent claims correlated with stochastic frequency through a particular dependence scheme. For example, it is possible that operational losses of a low-magnitude are serially correlated: say, if a bank hires new employees in April then it may observe cyclicality effects in the losses falling into the category of human errors, with a cycle of length one year (evidence of cyclical behavior was presented in Chapter 7). Likewise, it is less likely that large-magnitude operational losses (such as terrorist attacks or large internal frauds) would be dependent in time.

Aggregation of several loss processes remains an important task. Although this issue was not covered in this dissertation, future work will include modeling complex dependence structure of operational loss data. The AMA proposed by the Basel II suggest estimating operational risk capital separately for each "business line/ event type" combination. This is done in order to capture the differences in the degrees and nature of operational risk exposure across bank's various business units. Such procedure is not common in market risk and credit risk models. A simple sum of the capital charges across all "business line/ event type" combinations may result in over-estimation of the total capital charge. To prevent this from happening, it is essential to account for dependence between these combinations. *Covariance* and *correlation* are the simplest measures of dependency, but they assume a linear type of dependence, and therefore can produce misleading results if the linearity assumption is not true. If the groups are *perfectly correlated* (i.e., have a perfect positive correlation), then the total capital charge is a simple sum of the individual capital charges across groups. An alternative approach would involve using *copulas* that are more flexible about the form of the dependence structure that can exist between different groups. Examples include the Gaussian copula, t-copula, Gumbel copula, Frank copula, etc. An attractive property of copulas is their ability to capture the tail dependence between random variables, that is preserved under linear transformations of the variables. Applying copulas to operational risk modeling remains a topic for future research and is not treated in this dissertation. Dependent models in operational risk have been discussed by [56] [63] [60] [29] [64].

Glossary of Acronyms

AD	Anderson-Darling
AMA	Advanced Measurement Approaches
Basel I	The Basel I Capital Accord
Basel II	The Basel II Capital Accord
BCBS	Basel Committee for Banking Supervision
BIA	Basic Indicator Approach
BIS	Bank of International Settlements
CAPM	Capital Asset Pricing Model
cdf	Cumulative distribution function
CLT	Central Limit Theorem
CVaR	Conditional Value-at-Risk
EDF	Empirical distribution function
EL	Expected aggregate loss
EM	Expectation-Maximization
ES	Expected Shortfall
ETL	Expected Tail Loss
EVT	Extreme Value Theory
FFT	Fast Fourier Transform
GOF	Goodness of fit

GES	Gross Error Sensitivity
GPD	Generalized Pareto Distribution
HPP	Homogeneous Poisson Process
IF	Influence function
iid	Independent and identically distributed
KS	Kolmogorov-Smirnov
LDA	Loss Distribution Approach
LDCA	Loss Data Collection Exercise
LR	Likelihood Ratio
LTS	Least Trimmed Squares
MAE	Mean Absolute Error
M&A	Mergers and Acquisitions
MLE	Maximum Likelihood Estimator
MRC	Minimum regulatory capital
MSE	Mean Squared Error
NHPP	Non-homogeneous Poisson process
OLS	Ordinary Least Squares
pdf	Probability density function
PIT	Probability Integral Transformation
POT	Peak Over Threshold
pmf	Probability mass function
TSA	Standardized Approach
UL	Unexpected aggregate loss
VaR	Value-at-Risk

Bibliography

[1] J. R. Adams and D. Frantz. *A Full Service Bank: How BCCI Stole Billions Around the World*. Simon & Schuster, United Kingdom, 1993.

[2] R. J. Adler, R. E. Feldman, and M. S. Taqqu, editors. *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*, Boston, 1998. Birkhäuser.

[3] C. Alexander, editor. *Operational Risk: Regulation, Analysis and Management*, Great Britain, 2003. Prentice Hall.

[4] L. Allen, J. Boudoukh, and A. Saunders. *Understanding Market, Credit, and Operational Risk: The Value-at-Risk Approach*. Blackwell Publishig, Oxford, 2004.

[5] G. Alvarez. Operational Risk Event Classification. *http://www.garp.com*, 2002.

[6] T. W. Anderson and D. A. Darling. Asymptotic Theory of Certain 'Goodness of Fit' Criteria Based on Stochastic Processes. *The Annals of Mathematical Statistics*, 23(2):193–212, 1952.

[7] T. W. Anderson and D. A. Darling. A Test of Goodness of Fit. *Journal of the American Statistical Association*, 49(268):765–769, 1954.

[8] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9:203–228, 1999.

[9] L. Aucremanne, G. Brys, M. Hubert, P. J. Rousseeuw, and A. Struyf. A Study of Belgian Inflation, Relative Prices and Nominal Rigidities Using New Robust Measures of Skewness and Tail Weight. In Hubert et al. [97], pages 13–25.

[10] Bank of England. Report of the Banking Supervision. Inquiry into the Circumstances of the Collapse of Barings. Technical report, Her Majesty's Stationery Office, London, 1995.

[11] Bank of England. The Bank of England Report into the Collapse of Barings Bank. *http://www.numa.com/ref/barings/bar00.htm*, 1995.

[12] G. Bassett, G. Gerber, and P. Rocco. Robust Strategies for Quantitative Investment Management. In Hubert et al. [97], pages 27–37.

[13] N. Baud, A. Frachot, and T. Roncalli. Internal Data, External Data, and Consortium Data for Operational Risk Measurement: How to Pool Data Properly? Technical report, Groupe de Recherche Opérationelle, Crédit Lyonnais, France, 2002.

[14] BBA/ISDA/RMA. Operational Risk Management - The Next Frontier. *Journal of Lending & Risk Management*, March 2000:38–44, 2000.

[15] J. Beaty and S. C. Gwynne. *The Outlaw Bank: A Wild Ride into the Secret Heart of BCCI.* Random House Inc, Beard Books, United Kingdom, 1993.

[16] V. E. Bening and V. Y. Korolev. *Generalized Poisson Models and their Applications in Insurance and Finance.* VSP International Science Publishers, Utrecht, Boston, 2002.

[17] J. Berkowitz. Testing Density Forecasts with Applications to Risk Management. Technical report, University of California Irvine, 2000.

[18] J. Berkowitz and J. O'Brien. How accurate are Value-at-Risk models at commercial banks? *Journal of Finance*, 57:1093–1112, 2002.

[19] BIS. International Covergence of Capital Measurement and Capital Standard. *http://www.bis.org*, 1988.

[20] BIS. Amendment to the Capital Accord to Incorporate Market Risks. *http://www.bis.org*, 1996.

[21] BIS. Overview to the Amendment to the Capital Accord to Incorporate Market Risks. *http://www.bis.org*, 1998a.

[22] BIS. Operational Risk Management. *http://www.bis.org*, 1998b.

[23] BIS. A New Capital Adequacy Framework. *http://www.bis.org*, 1999.

[24] BIS. Electronic Banking Group Initiatives and White Papers. *http://www.bis.org*, 2000.

[25] BIS. Consultative Document: Operational Risk. *http://www.bis.org*, 2001a.

[26] BIS. Working Paper on the Regulatory Treatment of Operational Risk. *http://www.bis.org*, 2001b.

[27] BIS. The 2002 Loss Data Collection Exercise for Operational Risk: Summary of the Data Collected. *http://www.bis.org*, 2003.

[28] BIS. International Convergence of Capital Measurement and Capital Standards. *http://www.bis.org*, 2004.

[29] P. Blum, A. Dias, and P. Embrechts. The ART of Dependence Modelling: the Latest Advances in Correlation Analysis. In Morton Lane, editor, *Alternative Risk Strategies*, pages 339–356, London, 2002. RISK Books.

[30] A. L. Bowley. *Elements of Statistics*. Charles Scribner's Sons, New York, 1920.

[31] R. J. Brown. *Private Real Estate Investment: Data Analysis and Decision Making*. Elsevier Academic Press, San Diego, 2005.

[32] R. Bryce. *Pipe Dreams: Greed, Ego, and the Death of Enron*. PublicAffairs, USA, 2002.

[33] S. Carillo Menéndez. Operational Risk. *presentation*, International Summer School on Risk Measurement and Control, Rome, June 2005, 2005.

[34] A. Chapelle, Y. Crama, G. Hübner, and J.-P. Peters. Measuring and Managing Operational Risk in the Financial Sector: An Integrated Framework. Technical report, National Bank of Belgium, 2005.

[35] V. Chavez-Demoulin and P. Embrechts. Advanced Extremal Models for Operational Risk. Technical report, ETH Zürich, 2004.

[36] V. Chavez-Demoulin, P. Embrechts, and J. Nešlehová. Quantitative Models for Operational Risk: Extremes, Dependence, and Aggregation. Technical report, ETH Zürich, 2005.

[37] A. Chernobai, K. Burnecki, S. T. Rachev, S. Trück, and R. Weron. Modelling Catastrophe Claims with Left-Truncated Severity Distributions. *Journal of Computational Statistics*, 3, 2006.

[38] A. Chernobai, C. Menn, S. T. Rachev, and S. Trück. A Note on the Estimation of the Frequency and Severity Distribution of Operational Losses. *Mathematical Scientist*, 30(2):87–97, 2005a.

[39] A. Chernobai, C. Menn, S. T. Rachev, and S. Trück. Estimation of Operational Value-at-Risk in the Presence of Minimum Collection Thresholds. Technical report, University of California at Santa Barbara, 2005b.

[40] A. Chernobai and S. T. Rachev. Applying robust methods to operational risk modeling. *Journal of Operational Risk*, 1(1):27–42, 2006.

[41] A. Chernobai, S. T. Rachev, and F. J. Fabozzi. *Operational Risk: A Guide to Basel II Capital Requirements, Models and Analysis*. John Wiley & Sons, preprint, 2006.

[42] L. Chew. Not Just One Man - Barings. Technical report, IFCI Risk Institute report,http://riskinstitute.ch/137550.htm.

[43] P. Cizek, W. Härdle, and R. Weron, editors. *Statistical Tools for Finance and Insurance*, Heidelberg, 2005. Springer.

[44] C. Crnkovic and J. Drachman. *Quality Control in VaR: Understanding and Applying Value-at-Risk*. Risk Books, London, 1997.

[45] M. Crouhy, D. Galai, and R. Mark. *Risk Management*. McGraw-Hill, New York, 2001.

[46] M. G. Cruz. *Modeling, Measuring and Hedging Operational Risk*. John Wiley & Sons, New York, Chichester, 2002.

[47] R. D'Agostino and M. Stephens. *Goodness-of-Fit Techniques*. Marcel Dekker, Inc., New York and Basel, 1986.

[48] H. A. David. *Order Statistics.* John Wiley & Sons, New York, second edition, 1981.

[49] P. De Fontnouvelle, V. DeJesus-Rueff, J. Jordan, and E. Rosengren. Using Loss Data to Quantify Operational Risk. Technical report, Federal Reserve Bank of Boston, 2003.

[50] P. De Fontnouvelle, E. Rosengren, and J. Jordan. Implications of Alternative Operational Risk Modelling Techniques. Technical report, Federal Reserve Bank of Boston and FitchRisk, 2005.

[51] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society, Series B (Methodological)*, 39(1):1–38, 1977.

[52] F. X. Diebold, T. Schuermann, and J. D. Stroughair. Pitfalls and Opportunities in the Use of Extreme Value Theory in Risk Management. Technical report, The Wharton School, University of Pennsylvania, 1998.

[53] R. Dufour and U. Maag. Distribution Results for Modified Kolmogorov-Smirnov Statistics for Truncated or Censored Samples. *Technometrics*, 20:29–32, 1978.

[54] W. DuMouchel. Stable Distributions in Statistical Inference. *doctoral dissertation, Yale University*, 1971.

[55] W. DuMouchel. Stable Distributions in Statistical Inference: 1. Symmetric Stable Distribution Compared to Other Symmetric Long-Tailed Distributions. *Journal of the American Statistical Association*, 68:469–477, 1973.

[56] S. Ebnöther, P. Vanini, A. J. McNeil, and P. Antolinez-Fehr. Modelling Operational Risk. Technical report, ETH Zürich, 2001.

[57] P. Embrechts. Extreme Value Theory: Potential and Limitations as an Integrated Risk Management Tool. *Derivatives Use, Trading and Regulation*, 6:449–456, 2000.

[58] P. Embrechts. Extremes in Economics and Economics of Extremes. *Extreme Values in Finance, Telecommunications and the Environment,* (eds. Eds. B. Finkenstädt and H. Rootzén), Chapman and Hall CRC, London, pages 169–183, 2004.

[59] P. Embrechts, H. Furrer, and R. Kaufmann. Quantifying Regulatory Capital For Operational Risk. *Derivatives Use, Trading, and Regulation*, 9(3):217–233, 2003.

[60] P. Embrechts, A. Hoeing, and A. Juri. Using Copulae to Bound the Value-at-Risk for Functions of Dependent Risks. *Finance and Stochastics*, 7(2):145–167, 2003.

[61] P. Embrechts, P. Kaufmann, and G. Samorodnitsky. Ruin Theory Revisited: Stochastic Models for Operational Risk. In C. Bernadell *et al*, editor, *Risk Management for Central Bank Foreign Reserves*, pages 243–261, Frankfurt, 2004. European Central Bank.

[62] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modeling Extremal Events for Insurance and Finance.* Springer-Verlag, Berlin, 1997.

[63] P. Embrechts, F. Lindskog, and A. McNeil. Modelling Dependence with Copulas and Applications to Risk Management. In Rachev [135], pages 329–384.

[64] P. Embrechts, A. McNeil, and D. Straumann. Correlation and Dependence in Risk Management: Properties and Pitfalls. In M. A. H. Dempster, editor,

Risk Management: Value-at-Risk and Beyond, pages 176–223, Cambridge, 2002. Cambridge University Press.

[65] P. Embrechts, S. I. Resnick, and G. Samorodnitsky. Extreme Value Theory as a Risk Management Tool. *North American Actuarial Journal*, 3:30–41, 1999.

[66] Erisk. Learning Center, Case Studies, Operational Risk.

[67] E. Fama and K. French. The Cross-Section of Expected Stock Returns. *The Journal of Finance*, 47:427–466, 1992.

[68] E. F. Fama. Mandelbrot and the Stable Paretian Hypothesis. *Journal of Business*, 36:420–429, 1963.

[69] E. F. Fama. Portfolio Analysis in a Stable Paretian Market. *Management Science*, January, 1965.

[70] E. F. Fama. The Behavior of Stock-Market Prices. *Journal of Business*, 38(1):34–105, 1965.

[71] E. F. Fama. Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance*, 25:383–417, 1970.

[72] S. Fay. *The Collapse Of Barings*. W. W. Norton & Company, USA, first edition, 1997.

[73] FDIC. Regulators terminate the US operations of Daiwa Bank, Japan. *http://www.fdic.gov*, 1995.

[74] L. Fox. *Enron: The Rise and Fall*. John Wiley & Sons, Hoboken, New Jersey, 2003.

[75] A. Frachot, O. Moudoulaut, and T. Roncalli. Loss Distribution Approach in Practice. Technical report, Groupe de Recherche Opérationelle, Crédit Lyonnais, France, 2003.

[76] S. M. Frost. *The Bank Analyst's Handbook: Money, Risk and Conjuring Tricks*. John Wiley & Sons, Chichester, 2004.

[77] T. Gastaldi. A Kolmogorov-Smirnov Test Procedure Involving a Possibly Censored or Truncated Sample. *Communications in Statistics: Theory and Method*, 22:31–39, 1993.

[78] H. Geiger. Die Risikopolitik der Banken, Teil 1 und Teil 2. *Der Schweizer Treuhander*, 73(6/7 und 8):555 – 564, 713 – 718, 1999.

[79] G. D. Gettinby, C. D. Sinclair, D. M. Power, and R. A. Brown. An Analysis of the Distribution of Extreme Share Returns in the UK from 1975 to 2000. *Journal of Business Finance and Accounting*, 31(5,6):607–645, 2004.

[80] M. Gilli and E. Këllezi. An Application of Extreme Value Theory for Measuring Risk. Technical report, University of Geneva and FAME, 2003.

[81] J. Grandell. *Doubly Stochastic Poisson Processes. In: Lecture Notes in Mathematics*, volume 529. Springer-Verlag, Berlin-New York, 1976.

[82] J. Grandell. *Aspects of Risk Theory*. Springer-Verlag, New York, 1991.

[83] J. Grandell. *Mixed Poisson Processes*. Chapman & Hall, London, 1997.

[84] R. A. Groeneveld and G. Meeden. Measuring Skewness and Kurtosis. *The Statistician*, 33:391–399, 1984.

[85] O. Guilbaud. Exact Kolmogorov-Type Test for Left-Truncated and/or Right-Censored Data. *Journal of American Statistical Association*, 83:213–221, 1998.

[86] F. R. Hampel. Contribution to the Theory of Robust Estimation. Technical report, Ph.D. Thesis, University of California Berkeley, 1968.

[87] F. R. Hampel. Robust Estimation: A Condensed Partial Survey. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 27:87–104, 1973.

[88] F. R. Hampel. The Influence Curve and its Role in Robust Estimation. *Journal of the American Statistical Association*, 69:383–393, 1974.

[89] F. R. Hampel. The Breakdown Points of the Mean Combined with Some Rejection Rules. *Technometrics*, 27:95–107, 1985.

[90] F. R. Hampel, E. M. Ronchetti, R. J. Rousseeuw, and W. A. Stahel. *Robust Statistics: The Approach Based on Influence Functions*. John Wiley & Sons, 1986.

[91] M. Haubenstock. The Operational Risk Management Framework. In Alexander [3], pages 241–261.

[92] D. V. Hinkley. On Power Transformations to Symmetry. *Biometrika*, 62:101–111, 1975.

[93] R. V. Hogg. More Light on the Kurtosis and Related Statistics. *Journal of the American Statistical Association*, 67:422–424, 1972.

[94] R. V. Hogg. Adaptive Robust Procedures: A Partial Review and Some Suggestions for Future Applications and Theory. *Journal of the American Statistical Association*, 69:909–923, 1974.

[95] P. J. Huber. Robust Estimation of a Location Parameter. *Annals of Mathematical Statistics*, 35:73–101, 1964.

[96] P. J. Huber. *Robust Statistics*. John Wiley & Sons, New York, 1981.

[97] M. Hubert, G. Pison, A. Struyf, and S. van Aelst, editors. *Theory and Applications of Recent Robust Methods*, Basel, 2004. Birkhäuser.

[98] R. Huisman, K. G. Koedijk, C. J. M. Kool, and F. Palm. Tail-Index Estimates in Small Samples. *Journal of Business and Economic Statistics*, 19(1):208–216, 2001.

[99] R. Irving. County in Crisis. *Risk*, 1995.

[100] S. R. Jammalamadaka and S. Iyer. A Simple Estimate of the Index of Stability for Symmetric Stable Distributions. *Stochastic Modelling and Applications*, 6:9–19, 2003.

[101] P. Jorion. Orange County Case: Using Value-at-Risk to Control Financial Risk. *http://www.gsm.uci.edu/ jorion/oc/case.html*, 1998.

[102] P. Jorion. *Value-at-Risk: The New Benchmark for Managing Financial Risk*. McGraw-Hill, New York, 2nd edition, 2000.

[103] P. Jorion and R. Roper. *Big Bets Gone Bad: Derivatives and Bankruptcy in Orange County*. Academic Press, 1995.

[104] M. G. Kendall and A. Stuart. *The Advanced Theory of Statistics*, volume 1. Charles Griffin and Co., London, 1977.

[105] T.-H. Kim and H. White. On More Robust Estimation of Skewness and Kurtosis: Simulation and Application to the S&P500 Index. Technical report, University of California at San Diego, 2003.

[106] J. L. King. *Operational Risk: Measurement and Modelling*. John Wiley & Sons, New York, 2001.

[107] P. J. Knez and M. J. Ready. On the Robustness of Size and Book-to-Market in Cross-Sectional Regressions. *Journal of Finance*, 52:1355–1382, 1997.

[108] J. Koernert. The Collapse of Barings 1995. Financial Derivatives, Banking Crises and Contagion Effects. *Freiberg Working Papers, Vol. 96/2*, 1996.

[109] J. Koernert. Consequences of the Crisis of Barings. Technical Report Vol. 97/1, Freiberg Working Papers, 1997.

[110] S. M. Kogon and B. W. William. Characteristic Function Based Estimation of Stable Distribution Parameters. In Adler et al. [2], pages 311–336.

[111] Lectric Law Library. *11/95 Criminal Complaint & Indictment Against Daiwa Bank*. http://www.lectlaw.com/files/cas60.htm, 1995.

[112] N. Leeson. *Rogue Trader*. Time Warner Paperbacks, -, 1997.

[113] N. Leeson and I. Tyrrell. *Back from the Brink: Coping with Stress*. Virgin Books, -, 2005.

[114] C. M. Lewis and Y. Lantsman. What is a Fair Price to Transfer the Risk of Unauthorised Trading? A Case Study on Operational Risk. In E. Davis, editor, *Operational risk: Practical Approaches to Implementation*, pages 315–355, London, 2005. Risk Books.

[115] C. Marinelli, S. d'Addona, and S. T. Rachev. A Comparison of Some Univariate Models for Value-at-Risk and Expected Shortfall. Technical report, Universität Bonn, 2006.

[116] R. D. Martin and T. T. Simin. Outlier Resistant Estimates of Beta. *Financial Analysts Journal*, 59:56–69, 2003.

[117] B. McLean and P. Elkind. *Smartest Guys in the Room: The Amazing Rise and Scandalous Fall of Enron*. Penguin Books, USA, 2003.

[118] A. J. McNeil. Estimating the Tails of Loss Severity Distributions Using Extreme Value Theory. Technical report, ETH Zürich, 1996.

[119] A. J. McNeil. Extreme Value Theory for Risk Managers. Technical report, ETH Zürich, 1999.

[120] A. J. McNeil and R. Frey. Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: an Extreme Value Approach. Technical report, ETH Zürich, 1999.

[121] A. J. McNeil and T. Saladin. The Peaks over Thresholds Method for Estimating High Quantiles of Loss Distributions. Technical report, ETH Zürich, 1997.

[122] A. J. McNeil and T. Saladin. Developing Scenarios for Future Extreme Losses Using the POT Model. Technical report, ETH Zürich, 1998.

[123] E. A. Medova. Operational Risk Capital Allocation and Integration of Risks. Technical report, Center for Financial Research, University of Cambridge, UK, 2002.

[124] E. A. Medova and M. N. Kuriacou. Extremes in Operational Risk Management. Technical report, Center for Financial Research, University of Cambridge, UK, 2001.

[125] C. Menn and S. T. Rachev. Calibrated FFT-Based Density Approximations for Alpha-Stable Distributions. *Computational Statistics and Data Analysis*, forthcoming, 2005.

[126] J. J. A. Moors. A Quantile Alternative for Kurtosis. *The Statistician*, 37:25–32, 1988.

[127] M. Moscadelli. The Modelling of Operational Risk: Experience with the Analysis of the Data Collected by the Basel Committee. Technical report, Bank of Italy, 2004.

[128] M. Moscadelli, A. Chernobai, and S. T. Rachev. Treatment of missing data in the field of operational risk: Effects on parameter estimates, EL, UL and CVaR measures. *Operational Risk*, 2005.

[129] J. Neslehová, P. Embrechts, and V. Chavez-Demoulin. Infinite Mean Models and the LDA for Operational Risk. Technical report, ETH Zürich, 2006.

[130] J. P. Nolan. Numerical Calculation of Stable Densities and Distribution Functions. *Communications in Statistics – Stochastic Models*, 13:759–774, 1997.

[131] D. J. Olive. *Applied Robust Statistics.* preprint, , 2005.

[132] C. Perret-Gentil and M. Victoria-Feser. Robust Mean-Variance Portfolio Selection. Technical Report 140, International Center for Financial Asset Management and Engineering, 2005.

[133] J. III Pickands. Statistical Inference Using Extreme Order Statistics. *Annals of Statistics*, 3:119–131, 1975.

[134] S. J. Press. Estimation of Univariate and Multivariate Stable Distributions. *Journal of the American Statistical Association*, 67:842–846, 1972.

[135] S. T. Rachev, editor. *Handbook of Heavy Tailed Distributions in Finance*, The Netherlands, 2003. Elsevier.

[136] S. T. Rachev, A. Chernobai, and C. Menn. Empirical examination of operational loss distribution. In M. Morlock *et al*, editor, *Perspectives on Operational Research*, pages 379–402, Wiesbaden, Germany, 2006. Deutscher Unversitäts Verlag/GWV Fachverlage GmbH.

[137] S. T. Rachev and S. Mittnik. *Stable Paretian Models in Finance.* John Wiley & Sons, New York, 2000.

[138] S. T. Rachev, S. Mittnik, and M. Paolella. Stable Paretian Modeling in Finance: Some Empirical and Theoretical Aspects. In Adler et al. [2], pages 79–110.

[139] RiskMetrics. Risk Management: A Practical Guide. *http://www.riskmetrics.com*, 1999.

[140] RiskMetrics. Return to RiskMetrics: The Evolution of a Standard. *http://www.riskmetrics.com*, 2001.

[141] R. T. Rockafellar and S. Uryasev. Conditional Value-at-Risk for General Loss Distributions. *Journal of Banking and Finance*, 26:1443–1471, 2002.

[142] J. V. Rosenberg and T. Schuermann. A General Approach to Integrated Risk Management with Skewed, Fat-Tailed Risks. Technical report, Federal Reserve Bank of New York, 2004.

[143] M. Rosenblatt. Remarks on a Multivariate Transformation. *Annals of Mathematical Statistics*, 23:470–472, 1952.

[144] S. Ross. *Simulation*. Academic Press, Boston, 3rd edition, 2001.

[145] P. J. Rousseeuw and C. Croux. Alternatives to the Median Absolute Deviation. *Journal of American Statistics Association*, 88:1273–1283, 1993.

[146] P. J. Rousseeuw and A. M. Leroy. *Robust Regression and Outlier Detection*. John Wiley & Sons, New York, 1987.

[147] G. Samorodnitsky and M. S. Taqqu. *Stable Non-Gaussian Random Processes. Stochastic Models with Infinite Variance*. Chapman & Hall, London, 1994.

[148] D. Shirreff. Lessons from NatWest. *Euromoney*, 1997.

[149] G. R. Shorack and J. A. Wellner. *Empirical Processes with Applications to Statistics*. John Wiley & Sons, New York, 1986.

[150] J. S. Simonoff. The Breakdown and Influence Properties of Outlier-Rejection-Plus-Mean Procedures. *Communications in Statistics: Theory and Methods*, 16:1749–1769, 1987a.

[151] J. S. Simonoff. Outlier Detection and Robust Estimation of Scale. *Journal of Statistical Computation and Simulation*, 27:79–92, 1987b.

[152] S. M. Stigler. Simon Newcomb, Percy Daniell, and the History of Robust Estimation 1885-1920. *Journal of the American Statistical Association*, 68:872–878, 1973.

[153] S. Stoyanov and B. Racheva-Iotova. Univariate Stable Laws in the Field of Finance – Approximation of Density and Distribution Functions. *Journal of Concrete and Applicable Mathematics*, 2(1), 2004a.

[154] S. Stoyanov and B. Racheva-Iotova. Univariate Stable Laws in the Field of Finance – Parameter Estimation. *Journal of Concrete and Applicable Mathematics*, 2(4), 2004b.

[155] M. Swartz and S. Watkins. *Power Failure: The Inside Story of the Collapse of Enron*. Random House, USA, 2003.

[156] C. S. Tapiero. *Risk and Financial Management: Mathematical and Computational Methods*. John Wiley & Sons, Chichester, 2004.

[157] S. Uryasev. Conditional Value-at-Risk: Optimization Algorithms and Applications. *Financial Engineering News*, 14:1–5, 2000.

[158] H. van Greuning and S. B. Bratanovic. *Analyzing and Managing Banking Risk: A Framework for Assessing Corporate Governance and Financial Risk*. The World Bank, Washington, D.C., second edition, 2003.

[159] E. J. Vaughan and T. Vaughan. *Fundamentals of Risk and Insurance*. John Wiley & Sons, 9 edition, 2003.

[160] Y. Yamai and T. Yoshiba. Comparative Analyses of Expected Shortfall and Value-at-Risk (2): Expected Utility Maximization and Tail Risk. *Monetary and Economic Studies*, April:95–115, 2002.

[161] Y. Yamai and T. Yoshiba. Comparative Analyses of Expected Shortfall and Value-at-Risk (2): Their Validity under Market Stress. *Monetary and Economic Studies*, October:181–237, 2002.

[162] Y. Yamai and T. Yoshiba. Comparative Analyses of Expected Shortfall and Value-at-Risk: Their Estimation, Decomposition, and Optimization. *Monetary and Economic Studies*, January:87–121, 2002.

[163] Y. Yamai and T. Yoshiba. On the Validity of Value-at-Risk: Comparative Analyses with Expected Shortfall. *Monetary and Economic Studies*, January:57–85, 2002.

[164] V. M. Zolotarev. *One-dimensional Stable Distributions*. Translations of Mathematical Monographs, vol. 65. American Mathematical Society, Providence, 1986.