The analysis of extreme events with applications to financial risk management

Author(s): Degen, Matthias
Publication Date: 2009
Permanent Link: https://doi.org/10.3929/ethz-a-005978285
Rights / License: In Copyright - Non-Commercial Use Permitted
The Analysis of Extreme Events
With Applications to Financial Risk Management

A dissertation submitted to
ETH ZURICH

for the degree of
Doctor of Sciences

presented by
MATTHIAS DEGEN
Dipl. Math. ETH
citizen of Läufelfingen, BL, Switzerland

accepted on the recommendation of
Prof. Dr. Paul Embrechts, examiner
Prof. Dr. Holger Rootzén, co-examiner
Dr. Mario V. Wüthrich, co-examiner

2009
Acknowledgements

Many people directly and indirectly contributed to the accomplishment of this thesis. First and foremost I wish to express my sincere gratitude to my supervisor Paul Embrechts. I am grateful for him introducing me to the fascinating field of applied extreme value theory research and very much appreciate the confidence he put into me during the years. His support and advice on mathematical as well as personal matters has been of great value to me. Most importantly however I enjoyed the very pleasant and familiar atmosphere he managed to maintain at RiskLab, not only as a nice place to work but also as a very sociable place—our numerous RiskLab activities will always be kept in good memory!

Moreover I would like to thank Holger Rootzén and Mario V. Wüthrich for their willingness to act as co-examiners of my thesis.

Thanks also go to my co-authors Dominik Lambrigger and Johan Segers for the excellent collaboration, to Guus Balkema, Sid Resnick and Holger Rootzén for interesting and stimulating discussions during their visits to ETH as well as to Jan Beirlant and Goedele Dierckx for their invitation to K. U. Leuven. Further thanks go to all my present and former colleagues from RiskLab for the nice time we spent together as well as to all Group 3 members for the good atmosphere in our research group. I am also thankful to all my friends outside the Mathematics Department for providing me with additional energy and vitality through the good times we had together.

Financial support by ETH Zurich and RiskLab Switzerland is gratefully acknowledged.

Finally, I wish to express my deep gratitude to my family and to Petra. Without their confidence, encouragement and patience I would not have come so far. In particular, I am deeply grateful to my parents who taught me what really matters in life. This thesis is dedicated to them.

Zurich, December 3, 2009
Abstract

Enhanced by the 2007-2009 global financial crisis, the discussion on the adequacy of today’s risk management practices within the financial sector has become highly relevant. One important part of managing financial risks involves the calculation of so-called regulatory capital an institution needs to hold in order to safeguard against unexpected large losses. The more advanced capital calculation methods are based on the view that risk can be quantified through the assessment of specific characteristics of an appropriate probability distribution of potential losses. More precisely, for two of the three major risk classes the required capital level is such that the potential total annual loss within that risk class is expected to exceed this level on average only once in 1000 years.

The present thesis is a contribution of mathematical research to financial risk management. It consists of a review together with four self-contained papers. The review provides an introduction to the field and motivates the results of the accompanying papers. The main mathematical methods used in these papers are methods from extreme value theory (EVT). Although the results are written with an emphasis on possible applications to financial risk management, they may be viewed in a broader framework and are not restricted to a particular field of application.

In Paper A we discuss pitfalls an imprudent application of standard EVT methods to specific loss models may bring with it. In particular we emphasize that the accuracy of EVT-based high quantile estimation very much depends on the so-called second-order regular variation behavior of the underlying probability distributions.

In Paper B we consider convergence rates for normalized quantiles. It turns out that the convergence is very slow for certain models relevant for practice. As a consequence, the amount of data needed in order for standard EVT-based estimators to deliver reasonable results may be unrealistically high and largely incompatible with today’s situation for loss databases. The concept of so-called penultimate approximations seems promising in this respect.
Abstract

In Paper C we continue the discussion on penultimate approximations by analyzing different choices of normalizations for quantiles and their influence on the rate of convergence in certain limit laws underlying standard EVT. In particular we show that in certain situations a judicious choice of a power normalization will improve the convergence rate. This gives hope to improve the estimation accuracy for extreme quantiles such as the one-in-thousand years events required for the calculation of regulatory capital.

In Paper D we analyze the concept of diversification of risks. Given the high quantile levels typically of interest for risk management practice, analyzing diversification benefits by means of its empirical counterpart will in general not yield much insight. One is therefore advised to consider (suitable) analytic approximations. The main result in this paper is the derivation of a second-order approximation for diversification benefits.
Kurzfassung


In Paper A diskutieren wir Probleme, welche durch unbedachtes Anwenden von Methoden der Extremwerttheorie auftreten können. Insbesondere zeigen wir auf, dass die Qualität von Schätzern für hohe Quantile basierend auf Methoden der Extremwerttheorie stark vom asymptotischen Verhalten der zugrunde liegenden sogenannten regulär variiernden Wahrscheinlichkeitsverteilungen abhängt.

In Paper B betrachten wir das Konvergenzverhalten von standardisierten Quantilen. Es stellt sich heraus, dass die Konvergenzgeschwindigkeit für gewisse in der Praxis häufig benutzte Modelle sehr langsam ist. Eine Konsequenz davon ist, dass die benötigte Datenmenge um mit Hilfe von herkömmlichen Methoden der Extremwerttheorie vernünftige Schätzresultate zu erhalten, sehr viel höher ist als die derzeit tatsächlich verfügbare Datenmenge. In
diesem Zusammenhang erscheint das Konzept von sogennanten penultimaten Approximationen erfolgversprechend.

In Paper C führen wir die Diskussion zum Thema der penultimaten Approximationen fort und analysieren verschiedene Standardisierungsmethoden für Quantile und deren Einfluss auf die Konvergenzgeschwindigkeit in gewissen asymptotischen Resultaten innerhalb der Extremwerttheorie. Wir zeigen auf, dass in bestimmten Situationen durch eine bedachte Wahl der Standardisierung die Konvergenzgeschwindigkeit verbessert werden kann. Dies gibt Anlass zur Hoffnung, dass dadurch die Schätzungenaugkeit für hohe Quantile verbessert werden kann.

# Contents

Acknowledgements iii

Abstract v

Kurzfassung vii

1 Introduction 1
  1.1 Extreme Events within Financial Systems ................. 2
  1.2 Financial Systems—Too Complex to Regulate? ............. 3
  1.3 Regulatory Capital Charge under Basel II ............... 4

2 Probabilistic Modeling of Extreme Financial Risks 9
  2.1 “How to Model Operational Risk if You Must” ............ 9
  2.2 Accuracy of EVT-based Tail Approximations .......... 13
  2.3 “A Slowly Varying Function with a Sting in its Tail” .... 15

3 Estimation of High Quantiles 21
  3.1 Scaling Rules for High Quantiles ..................... 21
  3.2 Second-Order Asymptotics of Normalized Quantiles ....... 23
  3.3 Penultimate Approximations in Practice ................. 26

4 Analysis of Diversification Benefits 33
  4.1 Estimation—A Brief Outlook .......................... 37
  4.2 Concluding Remarks ................................... 38

Bibliography 39

Accompanying Papers 43
  Paper A—The Quantitative Modeling of Operational Risk .... 43
  Paper B—EVT-Based Estimation of Risk Capital ............ 75
  Paper C—Scaling of High-Quantile Estimators ............. 105
  Paper D—Risk Concentration and Diversification .......... 133

Curriculum Vitae 153
Accompanying Papers


Scaling of High-Quantile Estimators. Submitted.

D Matthias Degen and Dominik D. Lambrigger, Johan Segers (2009).
1. Introduction

In various fields of application, we are witnessing events that can reasonably be referred to as extreme events. Analyzing extreme events requires understanding and quantification of the underlying extreme risks. Inherent to the notion of risk is the concept of randomness and hence the language we shall use is that of probability theory.

Upon occurrence of an extreme event such as for instance a hurricane, an earthquake, a devastating flood or a stock market crash, we often wonder whether anything could have been done either to prevent such an event from happening or at least to have been better prepared for it. We tend to compare an extreme event with events experienced in the past, classify its magnitude and, if possible, implement corresponding measures to protect against future reoccurrence. However, a quantitative analysis of extreme events is not at all straightforward as we cannot draw on a rich wealth of experience due to the rare occurrence of such events. Also, the implementation of protective measures often has to be based on a compromise between safety and cost, that is, between guaranteeing that people/systems survive unscathed when subject to extreme conditions and reasonable economic and social costs. In order to agree on a compromise, a careful (statistical) analysis of extreme events is needed. Natural questions arising in this context include

i) how to estimate the probability of an extreme event based on the data available prior to the actual event,

ii) for a given high magnitude, what is the estimated return period for an event of at least that magnitude, or

iii) what is the magnitude (return level) of a one-in-thousand years event?

An area within mathematics dealing with the quantitative analysis of extreme events is the so-called extreme value theory (EVT). Especially over the recent years we have witnessed a growing interest in the applications of EVT to various fields such as, for instance, environmental (geological, climatological,
INTRODUCTION

hydrological) statistics, telecommunication/data transfer and finance and insurance. Nevertheless, in applying EVT one still runs into theoretical issues which need careful study. In this thesis we discuss some of the mathematical problems that may arise in the modeling of extreme events within finance and insurance.

1.1 Extreme Events within Financial Systems

The current global financial crisis is a strong reminder, if needed, that extreme events with a considerable worldwide impact on the economy do occur inside financial systems. One may ask whether today’s financial systems have become (more) prone to extreme events. To anyone only vaguely familiar with the mechanisms of modern financial markets it is clear that answering this question is not an easy task. We shall nevertheless give reasons why this indeed may be the case.

First, financial systems are complex systems. Slogans like too big to fail or too connected to fail, justifying the repeated government bailouts of insolvent institutions, point this out very clearly. Also, over the years we witnessed an increasing interdependence between multiple counterparties with different interests, trading increasingly complex investment vehicles (e.g. credit derivatives). Industry let the growth in new instruments outstrip the operational capacity to manage them. For example, by the end of 2008 the credit derivatives market was worth about $591 trillion nominal (see Basel Committee [5], p. 103)—roughly ten times the value of the world’s total output in 2008 (world’s GDP).

Second, financial systems are human systems. The dynamics of financial systems is to a great extent governed by humans. The way we act within the system, how we value information, and how we react to changes in the system determines the dynamics of the system. In behavioral finance, terms like irrational behavior or herding have been known for a long time as important features of financial markets. Especially in times of financial distress these features may contribute to amplify moderate events to become extreme events.

Third, financial systems are prone to moral hazard. The belief that governments will bail out insolvent institutions has contributed to irresponsible risk taking. While very obvious during the current crisis, this attitude has been around for a long time—one prominent example is the LTCM hedge fund debacle and its bailout in 1998. No doubt, governmental interventions, or the Greenspan put as some used to call it, are a delicate issue and the margin between trying to stabilize a financial system and giving incentives for excessively risky business practices is small.
1.2. Financial Systems—Too Complex to Regulate?

Altogether, the current crisis has highlighted how the (deliberate?) lack of a careful assessment or understanding of new products together with irresponsible risk taking in complex interconnected markets may lead to a sudden increase of systemic risk (i.e. the risk of failure of one institution implying, like a domino effect, the downfall of others) and to an overall instability of the global financial system.

1.2 Financial Systems—Too Complex to Regulate?

Experts around the world have partly accused the current regulatory framework of Basel II (see Basel Committee [3]) to be one of the major drivers of the global financial crisis. To be fair one should bear in mind that in the United States, the epicenter of the financial crisis, the Basel II rules were not actually applied when the turmoil began. The US regulatory agencies approved the final rules (see Federal Register [20]) to implement the new risk-based capital requirements only in December 2007, involving a very limited number of banks. Moreover, though the legal basis was laid, US institutions as well as regulatory agencies have been very reluctant with the implementation so far. In the EU, the legal basis for Basel II, the so-called Capital Requirements Directive applying Basel II to all banks, credit institutions and investment firms in the EU, was adopted in June 2006; see Table 1.1. Also in the EU, the actual implementation of the new legislation took place later, as many banks exploited the provisions of the Capital Requirements Directive which allowed them to defer the implementation to 2008.

<table>
<thead>
<tr>
<th></th>
<th>2007</th>
<th>2008</th>
<th>2009-2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa</td>
<td>-</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>Americas</td>
<td>-</td>
<td>3</td>
<td>14</td>
</tr>
<tr>
<td>Asia</td>
<td>2</td>
<td>10</td>
<td>18</td>
</tr>
<tr>
<td>Carribe</td>
<td>-</td>
<td>-</td>
<td>8</td>
</tr>
<tr>
<td>Europe</td>
<td>27</td>
<td>34</td>
<td>44</td>
</tr>
<tr>
<td>Middle-East</td>
<td>2</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>Total</td>
<td>31</td>
<td>57</td>
<td>105</td>
</tr>
</tbody>
</table>

Table 1.1: Overview of the Basel II implementation as of August 2008, including Basel Committee member and non-member countries; see Financial Stability Institute [22].
The financial crisis thus developed largely under the old Basel I legislation and in that sense highlighted more the weaknesses of Basel I. There is no doubt though that the crisis clearly also has highlighted some drawbacks of the new Basel II framework—in particular its insufficient risk sensitivity with respect to systemic risks.

Warnings for this were voiced already early on; see for instance “An academic response to Basel II” by Daníelsson et al. [14]. However, their warnings seem to have been completely ignored. Interestingly, many of the weaknesses of Basel II, already pointed out in that report, are now being rediscovered by numerous authors in their analyses of the current crisis; see for instance the recent Turner Review “A regulatory response to the global banking crisis”, published in March 2009 by the Financial Service Authority [21]. The latter report presents an extensive catalogue of challenges needed to be addressed by policymakers and regulators. It seems that the general way forward will lead to higher levels of bank capital than required in the past.

In this discussion, we as academics can contribute by singling out research themes with practical usefulness and by critically looking at the models and assumptions under which certain mathematical results can be/are applied. The present thesis shall be viewed in this light. The mathematical research is motivated by questions concerning quantitative aspects of Basel II or, more precisely, the quantification of regulatory capital.

1.3 Regulatory Capital Charge under Basel II

It should be clear by now that the analysis of extreme events represents an indispensable part in any prudent regulatory framework for financial risk management as it allows to implement protection measures against extreme events. An integral part of such measures within the actual regulatory guidelines of Basel II is the estimation/calculation of regulatory capital, which an institution has to hold as a buffer in order to safeguard against unexpected large losses; see Basel Committee [3], Part 2 on minimum capital requirements (Pillar 1).

**Remark:** Clearly, the calculation of capital charges constitutes only one small part of prudent financial risk management. Rootzén and Klüppelberg [35] aptly sum it up: *a single number can’t hedge against economic catastrophes*. Of increasing importance in today’s interconnected financial systems are the qualitative aspects of risk management (partly covered in Pillars 2 and 3 of Basel II). However, identifying and monitoring or, more importantly, understanding the risks an institution is exposed to is not at all an easy task in today’s risk landscape.
Basel II provides a range of methods of varying degree of sophistication and risk sensitivity for the calculation of regulatory capital charges. Conceptually, the more advanced methods are based on the view that risk can be quantified through the assessment of specific characteristics of an appropriate probability distribution of potential (profits and) losses over a given time horizon. In this thesis we focus on the modeling of downside risk, i.e. the losses. More formally, we will consider risk as a positive random variable \( (rv) \) \( X \) with (unknown) distribution function (df) \( F \). A number of different risk measures have been considered in the literature, one of the most popular being *Value-at-Risk* (VaR).

**Definition 1.1 (Value-at-Risk)** For a rv \( X \) with df \( F \), the quantile or generalized inverse 
\[
F^{-1}(\alpha) = \inf \{ x \in \mathbb{R} : F(x) \geq \alpha \}, \quad \alpha \in (0, 1),
\]
is referred to as the *Value-at-Risk* of \( X \) at level \( \alpha \) and is denoted by \( \text{VaR}_\alpha(X) \) or \( \text{VaR}_\alpha(F) \).

Whether or not VaR is appropriate as a measure to capture extreme financial risks is definitely debatable. It is not our goal however to discuss weaknesses of VaR as a risk measure—others did on numerous occasions; see for instance Artzner et al. \[1\], Danielsson et al. \[14\], Rootzén and Klüppelberg \[35\] or more recently Nešlehová et al. \[29\]. Our interest in VaR as a measure of financial risk is due to the fact that, according to Pillar 1 of Basel II, VaR is the prescribed risk measure for the calculation/estimation of regulatory capital for the three major risk classes credit, market and operational risk. While for the former two risk classes a variety of by now well-established models has been tried and tested, this is not the case for operational risk. Subsequently our focus is therefore on the latter.

### 1.3.1 Operational Risk

According to the regulatory framework of Basel II, operational risk is defined as

“The risk of loss resulting from inadequate or failed internal processes, people and systems or from external events. This definition includes legal risk, but excludes strategic and reputational risk”.

Basel Committee \[3\], § 644.

Operational risk (OR) may thus be viewed as complementary to the widely studied risk classes credit risk (CR) and market risk (MR). Readers not familiar with these concepts are advised to acquire background knowledge as presented for instance in Chapter 1 of McNeil et al. \[28\] or on the Basel website www.bis.org.
Introduction

Whereas CR and MR have been addressed already in the context of the regulatory frameworks of Basel I (1988) and of the Basel I Amendment to Market Risk (1996), it is only under Basel II that OR is incorporated. However, the danger of extreme events due to a failure of properly managing operational risks has been around long before Basel II was put in place; see Table 1.2.

<table>
<thead>
<tr>
<th>Year</th>
<th>Institution</th>
<th>Loss (in billion $)</th>
<th>Type of failure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1994</td>
<td>Orange County</td>
<td>1.7</td>
<td>speculation</td>
</tr>
<tr>
<td>1995</td>
<td>Barings Bank</td>
<td>1.4</td>
<td>external event, fraud, internal control</td>
</tr>
<tr>
<td>1996</td>
<td>Sumimoto Bank</td>
<td>1.8</td>
<td>fraud, internal control</td>
</tr>
<tr>
<td>1998</td>
<td>LTCM</td>
<td>4.0</td>
<td>speculation, leverage</td>
</tr>
<tr>
<td>2001</td>
<td>Enron</td>
<td>2.2</td>
<td>fraud, internal control</td>
</tr>
<tr>
<td>2002</td>
<td>Allied Irish Banks</td>
<td>0.7</td>
<td>fraud, internal control</td>
</tr>
<tr>
<td>2005</td>
<td>Mizuho Securities</td>
<td>0.3</td>
<td>“fat finger” error</td>
</tr>
<tr>
<td>2008</td>
<td>Société Générale</td>
<td>7.0</td>
<td>fraud, internal control</td>
</tr>
<tr>
<td>2008</td>
<td>Madoff Investment Securities</td>
<td>50</td>
<td>fraud (Ponzi scheme)</td>
</tr>
<tr>
<td>2007-?</td>
<td>Financial crisis</td>
<td>10500+</td>
<td>various</td>
</tr>
</tbody>
</table>

Table 1.2: Overview of some historical OR management failures. Loss figures are taken partly from Chernobai [11] and from other media sources.

Remark: In general it is difficult to categorize losses into a specific risk category (CR, MR, OR). In the case of LTCM for instance, the incurred loss of $4 bn was due to a combination of CR and OR management failure. The downturn of LTCM was triggered by the 1998 Russian financial crisis and amplified by speculative trading of highly leveraged positions (dept to equity ratio of up to 25 to 1).

Due to the extremely large losses incurred, some of the institutions in Table 1.2 had to file for bankruptcy (e.g. Barings, Enron, Orange County). Others, especially during the current financial crisis, seem to have been too connected to fail and were bailed out by the government in order to (try to) reduce systemic risk and bring back stability to the financial system. In this manner, the incorporation of OR into regulatory legislations has been overdue. In
1.3. Regulatory Capital Charge under Basel II

the same vein, Lloyd C. Blankfein, CEO of Goldman Sachs, puts it as follows:

“Given the size and interconnected nature of markets, the growth in volumes, the global nature of traders and their cross-asset characteristics, managing operational risk will only become more important.”

Financial Times, February 8, 2009.

Due to the relatively recent incorporation of OR into regulatory guidelines, the estimation of regulatory capital charges for OR is not standard and a more in-dept study of VaR-based estimation methods for OR capital charges is desired.

1.3.2 The Advanced Measurement Approach to Operational Risk

Under the Basel II Advanced Measurement Approach (AMA) to OR, the capital charge to protect against OR losses is such that the total annual OR loss is expected to exceed this capital level only once in 1000 years on average (1000-year return level or 99.9% VaR of the total OR-loss aggregated over a 1-year period). In addition, the internal OR measurement system must be sufficiently granular to capture the main OR-drivers affecting a bank’s business activities; see Basel Committee [3], § 669(c). For simplicity assume the chosen granular model is such that the range of a bank’s business is divided into $d$ sub-units of business or units of measure; see Federal Register [20], p. 69317. With $S_i$ denoting the total OR-loss aggregated over one year within unit of measure $i$, the overall OR-loss at the consolidated level is given by $S_1 + S_2 + \cdots + S_d$. Basel II requires banks to add up the $d$ different risk measures in order to calculate the regulatory capital charge for OR ($RC^{OR}$). In doing so banks may (in principle) allow for a diversification benefit $\delta$, i.e.

$$RC^ {OR} = \text{VaR}_\alpha \left( \sum_{i=1}^{d} S_i \right) = (1 - \delta) \sum_{i=1}^{d} \text{VaR}_\alpha (S_i),$$

at level $\alpha = 99.9\%$; see Basel Committee [3], § 657 and 669(d).

Whether or not Basel II’s AMA to operational risk is feasible to calculate regulatory capital is (more than) debatable. Given the scarcity of available loss data and the resulting modeling uncertainty, prudent estimation of a 99.9\% quantile ($\text{VaR}_{99.9\%}$) seems highly questionable. And yet, the AMA is used by an increasing number of institutions; see Basel Committee [7],
Summary Table I. This leads to the need to perform research in this area and to point out some of the difficulties and pitfalls that may arise. From a more general methodological point of view, we aim at analyzing some of the mathematical problems related to

i) finding appropriate probabilistic models that describe the extreme risks under consideration (e.g. large losses in OR),

ii) the estimation of high quantiles, based on a sample of observations (e.g. realized OR losses), and

iii) the analysis of diversification benefits.

The four papers building the core of the thesis discuss these themes in more detail. Issues around i) are discussed in the accompanying Papers A and B. Parts of these papers are summarized below in Chapter 2 of this review. Problems arising in ii) are the focus of the Papers B and C and an overview is given in Chapter 3. Finally, Paper D is concerned with iii) and is summarized in Chapter 4.

We start by fixing notation and recalling some of the basic concepts from standard EVT. For the tail of a df $F$ we standardly write $F = 1 - F$ and by $x_F \leq \infty$ we denote the upper end-point of the support of $F$. The corresponding tail quantile function is denoted by $U(t) = F^{-1}(1 - 1/t)$, where $F^{-1}$ denotes the (generalized) inverse of $F$; see Definition 1.1. Moreover, we use the notation $U \in RV_\xi$, $\xi \in \mathbb{R}$, for $U(x) = x^{\xi}L(x)$, where $L$ is some slowly varying function in the sense of Karamata, i.e. for all $x > 0$,

$$\lim_{t \to \infty} \frac{L(tx)}{L(t)} = 1.$$ 

Where necessary, we write $L_F$ and $L_U$ for the slowly varying functions associated with $F$ and $U$, respectively. In the context of regular variation (RV), it is often more convenient to work on a log-log scale. In regular cases we will write $L_F$ as

$$L_F(u) = e^{\psi_{L_F}(\log u)}, \quad \psi_{L_F}(\log u) = \int_{1}^{u} \frac{\varepsilon(s)}{s} ds + c,$$

(2.1)

where $\varepsilon(s) = \psi_{L_F}'(\log s) \to 0$ and $c = \log L(1)$.

By $F \in MDA(H_\xi)$ we mean that $F$ belongs to the maximum domain of attraction of the generalized extreme value df $H_\xi$ and refer to $\xi$ as the tail index; see for instance Embrechts et al. [17]. Moreover we refer to a model as being heavy-tailed, if $F \in MDA(H_\xi)$ with $\xi > 0$. In what follows we assume the reader to be familiar with classical univariate EVT as presented for instance in Beirlant et al. [8], Embrechts et al. [17], de Haan and Ferreira [25], Leadbetter et al. [27] or Resnick [33, 34].

2.1 “How to Model Operational Risk if You Must”

Due to the limited data availability and the extreme quantile levels required, finding an appropriate probabilistic model that describes well the underlying
Probabilistic Modeling of Extreme Financial Risks

loss data sample at the (and beyond the) edge, is an inherently difficult task. One of the prime examples where one faces such problems is in the modeling of operational risk.

One of the main goals in the quantitative modeling of OR is the calculation of regulatory capital charges. As before, consider the AMA to operational risk, according to which the capital charge has to be calculated by

$$RC^{OR} = (1 - \delta) \sum_{i=1}^{d} \text{VaR}_{99.9\%}(S_i),$$

with $S_i$ denoting the total OR-loss aggregated over one year within unit of measure $i$. The calculation/estimation of the $d$ stand-alone VaR’s may be carried using different approaches such as

(A1) approximations,

(A2) inversion methods (FFT),

(A3) recursive methods (Panjer),

(A4) Monte Carlo simulation, or

(A5) closed-form methods;

see for instance Embrechts et al. [16] for a description of (A1)-(A4) and Böcker and Klüppelberg [10] for (A5).

One of the most popular methods within the AMA is the so-called loss distribution approach (LDA) based on (A4) and the classical actuarial model for non-life insurance. In such a model, the total OR-loss process $(S(t))_{t \geq 0}$ for a unit of measure $S$ is given by

$$S(t) = \sum_{k=1}^{N(t)} X_k,$$

where the $(X_i)_{i \geq 1}$ are the single OR-losses, assumed to be iid with df $F$ and independent of the claim arrival process $(N(t))_{t \geq 0}$; see Embrechts et al. [17], Chapter 1 for details. Note that in an OR context $t$ is fixed and taken to be one year. For critical remarks on the use of this framework within OR, see for instance Kupiec [26].

The analysis of the 2008 loss data collection exercise (see Basel Committee [6]) showed that—among the institutions using the LDA for operational risk—there seems to be a broad consent in favor of the Poisson distribution as loss frequency model. For severity modeling however there is still an ongoing discussion concerning the use of specific fully parametric models (e.g. GB2,
2.1. “How to Model Operational Risk if You Must”

Lognormal, g-and-h) and the use of semi-parametric approaches based on EVT.

2.1.1 Modeling OR Loss Severity

Stylized facts for OR losses—according to current knowledge—suggest the use of heavy-tailed, skewed, left-truncated severity models. The latter is due to internally applied de minimis data selection thresholds; see Basel Committee [3], § 673. For a loss model \( X > 0 \) with df \( F \) (we write \( X \sim F \)), we thus consider loss data above a threshold \( u > 0 \) and define \( X^u \) as the rv \( X \) conditioned to exceed \( u \). Informally we write

\[
X^u \overset{d}{=} X|X > u.
\]

Commonly considered thresholds for LDA banks range from \( u = \$1'000 \) to \( \$20'000 \); see Basel Committee [7], Table ILD1. What is usually ignored in practice is the fact that the introduction of such collection thresholds creates statistical problems since for \( X \sim F \), the distribution of the exceedances \( X^u \) over thresholds \( u \) will in general not be \( F \) anymore. As a consequence, fitting an unconditional parametric distribution to the truncated loss severities may introduce a large bias at high quantile levels; see Table 2.1. This miss-specification of the severity model in turn leads to an inaccurate estimation of the stand-alone VaR \( \alpha \) of a unit of measure \( S \) at high levels \( \alpha \).

<table>
<thead>
<tr>
<th>Bias</th>
<th>SRMSE</th>
<th>Bias</th>
<th>SRMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>original data</td>
<td></td>
<td>left-truncated at 5% lower order statistics</td>
<td></td>
</tr>
<tr>
<td>Loggamma ((a = 2, l = 1.75))</td>
<td>0.87</td>
<td>17.70</td>
<td>-32.03</td>
</tr>
<tr>
<td>Lognormal ((\mu = 3.5, \sigma = 1.25))</td>
<td>0.50</td>
<td>9.38</td>
<td>-22.52</td>
</tr>
<tr>
<td>g-and-h ((a = 0, b = 1, g = 2, h = 0.2))</td>
<td>-10.76</td>
<td>29.83</td>
<td>-27.94</td>
</tr>
<tr>
<td>Pareto ((x_0 = 1.2, \xi = 0.75))</td>
<td>2.17</td>
<td>16.16</td>
<td>6.13</td>
</tr>
<tr>
<td>Burr ((\alpha = 1, \kappa = 2, \tau = 1.5))</td>
<td>1.69</td>
<td>26.73</td>
<td>90.49</td>
</tr>
<tr>
<td>GB2 ((a = b = 2, p = 1, q = 0.75))</td>
<td>0.22</td>
<td>23.85</td>
<td>18.65</td>
</tr>
</tbody>
</table>

Table 2.1: Bias and standardized root mean squared error (SRMSE) (in %) of VaR\(_{99.9}\) estimates, based on truncated and non-truncated data (200 datasets of 1000 observations) from six different loss models; see Paper C for the respective model parameterizations. The estimation is performed using MLE, except for the g-and-h where letter values are used; see Tukey [36].
Therefore, when using a specific parametric distribution to model OR loss severities, one should fit a conditional instead of an unconditional distribution; see for instance Chernobai et al. [11], Chapter 9. No doubt in all of these parametric approaches, the model risk involved is still considerable. We will therefore restrict our attention to models which exhibit certain stability properties when moving further out in the tail. The importance of such models has always been stressed by non-life and reinsurance actuaries; see for instance Panjer [32].

2.1.2 Tail Stability

The problem of fitting models to truncated data is bypassed if one considers models $F$ which are stable under successive conditioning over increasing thresholds. However, there is only one heavy-tailed loss model, the exact Pareto distribution, defined by $F(x) = 1 - (x/x_0)^{-1/\xi}$ with $x > x_0 > 0$ and $\xi > 0$, which is stable under conditioning in the sense that for every $u > x_0$ the df of $X^u$ is again of the type $F$ and one has

$$X \sim \text{Pareto}(x_0, \xi) \Rightarrow X^u \sim \text{Pareto}(u, \xi).$$

An analogous tail stability property holds for the excess distribution function $F_u$, the df of the excesses $X^u - u$ over $u$. The only df $F$ for which $F_u$ is again of the type $F$ is the generalized Pareto distribution (GPD), defined by $G_\xi(x) = 1 - (1+\xi x)^{-1/\xi}$ for $1+\xi x/\beta > 0$. Indeed, for $u$ such that $1+\xi u/\beta > 0$, one has

$$X \sim G_{\xi, \beta} \Rightarrow (X^u - u) \sim G_{\xi, \beta+\xi u}, \quad (2.2)$$

where $G_{\xi, \beta}(x) = G_{\xi}(x/\beta)$; see also Embrechts et al. [17], Theorem 3.4.13 (c). This tail stability of the GPD is a key property in EVT. The following result stands at the heart of univariate EVT and highlights the important role of the GPD within EVT. It is occasionally referred to as the “fundamental theorem of (excess-of-loss) reinsurance”. It shows that, although the GPD is the only df for which (2.2) holds in exact form, the GPD serves as approximation of $F_u$ (for large values of $u$) for a very broad class of models $F$.

**Theorem 2.1 (Pickands-Balkema-de Haan)**

For $\xi \in \mathbb{R}$ the following are equivalent.

i) $F \in \text{MDA}(H_\xi)$

ii) There exists a measurable function $\beta(.) > 0$ such that

$$\lim_{u \uparrow x_F} \sup_{x \in (0, x_F-u)} |F_u(x) - \overline{G}_{\gamma, \beta(u)}(x)| = 0. \quad (2.3)$$


2.2 Accuracy of EVT-based Tail Approximations

Based on Theorem 2.1, EVT provides powerful techniques to extrapolate into regions beyond the edge of a data sample. Clearly, it cannot do magic, but what EVT is doing is making the best use of whatever data you do have about extreme phenomena—in the words of R. L. Smith; see Embrechts et al. [17], p. VII.

Especially in presence of very heavy-tailed data (as is the case for OR), one might not be willing to just fit a certain parametric model to the underlying data due to the possibly considerable model risk; see for instance Paper C, Table 2. In this case, modeling the body of the underlying model separately from the tail may be helpful. While for the body one has several possibilities (e.g. parametric, empirical), (2.3) suggests the following approximation for the tail region:

\[ F(x) \approx F(u) \bar{G}_{\xi,\beta(u)}(x-u), \]  

(2.4)

for large values of \( u \) and \( x > u \). Because (2.4) holds for a broad class of relevant dfs \( F \in \text{MDA}(H_\xi) \), it avoids a stark concentration on a very specific parametric model. Moreover, the issue of data truncation is of less concern as (2.4) is based on a specific consideration of extreme events. In order to appreciate (2.4), it is important to analyze the accuracy of such an EVT-based tail approximation.

2.2 Accuracy of EVT-based Tail Approximations

The theory of second-order regular variation (2RV) provides a convenient methodological framework to analyze the accuracy of the tail approximation (2.4). To do so, we aim for a function \( A(\cdot) \) with \( \lim_{u \to \infty} A(u) = 0 \) such that

\[ \lim_{u \to \infty} \frac{F_u(x\beta(u)) - \bar{G}_\xi(x)}{A(u)} = K(x), \]  

(2.5)

for some non-degenerate limit \( K \).

Remark: Obviously \( A(\cdot) \) will depend on the choice of the scaling function \( \beta(\cdot) \). One particular choice of \( \beta(\cdot) \) and the resulting second-order properties for different underlying dfs \( F \) is discussed more in detail in Section 3 of the accompanying Paper A. Motivated by this, we compare in Paper C the influence of different choices of normalizations \( \beta(\cdot) \) in the more general setting of second-order extended regular variation for tail quantile functions.
The following example gives a flavor of how to derive second-order regular variation results of the type (2.5). In particular it highlights the important role of the associated slowly varying functions.

**Example 2.1 (Extended Regular Variation)**

Assume $F \in \text{MDA}(H_\xi)$ with $\xi > 0$, or equivalently, $F \in \text{RV}_{-1/\xi}$. In this case, it is easy to see that necessarily $\beta(u) \sim \xi u$, $u \to \infty^1$, and hence one may simply choose $\beta(u) = \xi u$. With this choice of $\beta(\cdot)$, the convergence rate in (2.5) solely depends on the rate of $L(tx)/L(t) \to 1$, where $L$ is the slowly varying function associated with $F \in \text{RV}_{-1/\xi}$. With the notation of (2.1) we obtain

\[
\frac{F_u(x\beta(u)) - G_\xi(x)}{\varepsilon(u)} = y^{-1/\xi} \frac{L(uy)}{L(u)} - 1 = y^{-1/\xi} \frac{\exp \left( \int_u^y \frac{\varepsilon(s)}{s} ds \right) - 1}{\varepsilon(u)}
\]

\[
\sim y^{-1/\xi} \int_1^y \frac{\varepsilon(uz)}{\varepsilon(u)} \frac{1}{z} dz, \quad u \to \infty,
\]

where $y = 1 + \xi x$. Without going into details on extended regular variation (ERV) at this point, we note that in regular cases one obtains for $y > 0$

\[
\lim_{u \to \infty} \frac{L(uy)}{L(\varepsilon(u))} = \frac{y^\tau - 1}{\tau}, \quad (2.6)
\]

for some $\tau \leq 0$, with the obvious interpretation for $\tau = 0$. Note that the mere existence of a limit already implies that it is necessarily of the above form. If (2.6) holds we write $L \in \text{ERV}_\tau$ or, equivalently, $F \in \text{2RV}_{-1/\xi,\tau}$. For further reading, see for instance de Haan and Ferreira [25], Appendix B.2.

\[\square\]

In the case of Example 2.1, the convergence rate in (2.5) may be expressed through $A(u) = \varepsilon(u) = \Psi'_L(\log u)$ and hence can be viewed as a measure of how fast the slope in the log-log plot of $L$ tends to 0. The accuracy of the tail approximation (2.4) is thus crucially dependent on the second-order behavior of the associated slowly varying function. Note that by (2.1) one may always construct theoretical examples of $L$’s that are arbitrarily “unpleasant”. In the accompanying Papers A and B we highlight that, somewhat surprisingly, this may manifest itself also in fairly straightforward, and increasingly used, parametric models such as for instance the so-called g-and-h distribution. In the next section we give an overview of the main issues that arise from an EVT perspective for g-and-h-like models.

---

\(^1\)Throughout we mean by $f_1(u) \sim f_2(u)$ as $u \to u_0$, for two functions $f_1$ and $f_2$ that $f_1(u)/f_2(u) \to 1$ as $u \to u_0$. 

14
2.3 “A Slowly Varying Function with a Sting in its Tail”

On a more general methodological note we first want to mention that EVT-based tail and quantile approximation methods are in some or other form linked to the assumption of ultimate linearity of log-log plots. However, depending on the underlying model $F$, ultimate linearity of $\Psi_{L_F}$ may arrive very far in the tail only—possibly way beyond regions reasonable for practical applications. As a consequence, EVT-based estimators may show a severe bias whenever data do not come from an exact Pareto model. Often the presence of an estimation bias is easily detected using graphical tools. In certain situations however the issue is more subtle.

In the accompanying Paper A we introduce an example of a particular parametric loss model with “bad” second-order behavior. We focus on the so-called g-and-h distribution, which was originally proposed in Tukey [36] and recently advocated by Dutta and Perry [15] to model operational risk losses. Recall that a rv $X$ is said to have a g-and-h distribution, if $X$ satisfies

$$X = a + b e^{gZ} - \frac{1}{g} e^{hZ^2/2}, \quad a, g, h \in \mathbb{R}, b > 0,$$

where $Z \sim N(0, 1)$. The linear transformation parameters $a$ and $b$ are of minor importance for our analysis. Unless stated otherwise we therefore restrict our attention to the standard case $a = 0$ and $b = 1$.

We show that the g-and-h with $g, h > 0$ satisfies $F \in RV_{-1/h}$ and discuss a convergence rate result of the type (2.5). It turns out that for the g-and-h (due to the associated slowly varying function) this convergence is extremely slow: see Paper A, Lemma 3.1 and Table 1. A key message from that paper is that g-and-h-like models provide a class of heavy-tailed models for which the tail behavior can be rather intricate. Based on empirical evidence we conclude in Paper A that the precise behavior of $L$ may profoundly affect the statistical properties of EVT-based estimators. The empirical findings in Paper A are then made analytically precise in Section 4 of the accompanying Paper B. At this point, we prefer not to go into technical details but rather to present some of the main issues by means of graphical tools.

We show in Figure 2.1 the MLE-POT of the tail index $\xi$ as a function of the number of exceedances used, based on $n = 10^4$ observations from a g-and-h model with for OR reasonable parameter values $g = 2$ and $h = 0.2$ (see Dutta and Perry [15], p. 43).

At first glance, Figure 2.1 suggests that the data follows nearly perfectly an exact Pareto law. Indeed, the deviation from exact power law behavior seems to vanish quickly as the MLE behaves stably and is flat over a large
region of thresholds. As a consequence one would accept an estimate of the tail index of around $\hat{\xi} \approx 0.85$.

Now, compare this to Figure 2.2, where we again plot the MLE-POT of the tail index $\xi$ for the exact same g-and-h model ($g = 2$, $h = 0.2$) but this time with $n = 10^7$ simulations.

![MLE-POT estimates of tail index](image)

Figure 2.2: MLE-POT estimates of $\xi$ together with the 95% confidence bounds for g-and-h data with $g = 2$ and $h = 0.2$, based on $n = 10^7$ observations.

Based on Figure 2.2 one would accept a tail index estimate of around $\hat{\xi} \approx 0.6$
as opposed to the $\hat{\xi} \approx 0.85$ suggested by Figure 2.1. In this sense it seems as if g-and-h-like models fool EVT as the estimation bias in this case is indeed rather intricate. And interestingly, we are still far off the true theoretical value of $\xi = 0.2$ (recall that for the g-and-h with $g, h > 0$, $F \in RV_{-1/h}$). In both cases, infinite variance models are suggested, whereas the data are simulated from a model with finite moments up to order five.

The behavior in the Figures 2.1 and 2.2 is to be compared with the situation of Figure 2.3. In the latter figure we show the same situation as in the former two figures except for changing the parameter value of $h$ from 0.2 to 0.5.

Compared with the situation before, the estimates of $\xi$ for this choice of the $h$ parameter value seem to be more stable under an increase of the number of observations to $n = 10^7$; see Figure 2.3.

Altogether, the Figures 2.1 to 2.3 raise questions on how to value the information provided these (and similar) figures about the shape of the underlying distribution tail/data structure. While the notion of the tail index is ubiquitous and of crucial importance within the asymptotic theory of extremes, its role in statistical applications of EVT needs a proper assessment. Indeed, for (finite) samples, an estimate of the tail index as a measure of heavy-tailedness of the underlying model has to be interpreted with care, as its value may heavily depend on the size of the sample. While this is obvious for the EVT expert, it is maybe not for the more applied risk management end user. In order to avoid comparing apples to oranges (i.e. comparing tail index estimates based on different sample sizes), we prefer the more suggestive notion/concept of local tail index and local heavy-tailedness. The meaning of the term local will be made precise below.
2.3.1 Local Heavy-Tailedness

One of the key elements in the accompanying Paper B is the introduction of the notion of local heavy-tailedness which, in our opinion, allows an easy and intuitive approach to the rather technical framework of second-order (extended) regular variation (2ERV). In particular it explains the prima facie confusing tail behavior showed in the Figures 2.1 to 2.3.

**Definition 2.1 (Local heavy-tailedness)** For a df $F$ with $F \in \text{RV}_{-\alpha}$, $\alpha \geq 0$, we write $\log F(x) = -\Psi_F(\log x) = -\alpha \log x + \Psi_{L_F}(\log x)$. We may thus interpret $-\Psi'_F(\log x) = -\alpha + \Psi'_{L_F}(\log x)$ as a measure of local heavy-tailedness at points $x$. It consists of the ultimate heavy-tailedness of the model, measured through the RV-index $-\alpha = -\Psi'_F(\infty)$ and an additional source of heavy-tailedness $\Psi'_{L_F}$ due to the slowly varying part $L_F$ of $F$. Note that $\Psi'_{L_F}$ tends to zero and measures the speed at which the influence of the slowly varying nuisance part vanishes. Similarly, in the tail quantile function setting of Paper C we introduce the notion of a local tail index through $\log U(t) = \varphi(\log t) = \xi \log t + \varphi_{L_U}(\log t)$ for $U \in \text{RV}_\xi$, $\xi \geq 0$.

Using the notion of local heavy-tailedness, the empirical findings of Paper A (partly summarized in the Figures 2.1 to 2.3) about the influence of slowly varying functions on the performance of EVT-based estimators are made analytically more precise in Paper B. In Section 4 of that paper the behavior of the g-and-h slowly varying function is discussed.

**Proposition 2.1** Let $F \sim g$-and-$h$ with $a = 0, b = 1$ and $g, h > 0$. Then $F(x) = x^{-1/h} L_F(x)$, with $L_F \in \text{SV}$ and given by

$$L_F(x) = c e^{\frac{2g}{\sqrt{2h}} x^{3/2} / \sqrt{\log x}} \left( 1 + O\left( \frac{1}{\sqrt{\log x}} \right) \right), \quad x \to \infty,$$

where $c = \frac{1}{2} \sqrt{\frac{h}{\pi}} g^{-1/h} e^{-\frac{g^2}{h^2}}$.

**Proof:** See Paper B, Appendix A.

By Proposition 2.1 the g-and-h slowly varying function $L_F$ tends to $\infty$ at a growth rate of the order $e^{\sqrt{\log x} / \sqrt{\log x}}$. We study the influence of this slowly varying function on the statistical properties of EVT-based estimators by means analyzing the associated additional heavy-tailedness $\Psi'_{L_F}$. We show that

$$\Psi'_{L_F}(s) = \frac{g}{\sqrt{2h^{3/2}}} \frac{1}{\sqrt{s}} - \frac{1}{2s} + O\left( \frac{1}{s^{3/2}} \right), \quad s \to \infty,$$
yielding a rate of convergence in (2.5) of the order $A(u) = O\left(1/\sqrt{\log u}\right)$ as $u \to \infty$. Due to this slow convergence rate of $\Psi'_{L_F}$ to 0, the slowly varying function $L_F$ behaves “regularly varying-like” over large ranges, putting a significant amount of additional heavy-tailedness to the model. Moreover, the constant $g h^{-3/2}$ in the leading term shows how the parameters $g$ and $h$ influence the shape of the slowly varying function $L_F$. This is illustrated in Figure 2.4, where we plot $\Psi'_{L_F}$ for the g-and-h distribution for the case of $g = 2$ with $h = 0.2$ (left panel) and $h = 0.5$ (right panel).

Figure 2.4: Slope of the log-log plot of the g-and-h slowly varying function for $g = 2, h = 0.2$ (left panel) and $g = 2, h = 0.5$ (right panel) in a range relevant for OR-practice.

Approximating $\Psi'_{L_F}$ by constant pieces helps to explain the behavior of the g-and-h model in the Figures 2.1 and 2.2 as well as in Figure 2.3. For illustrative purposes we focus on situation in the Figures 2.1 and 2.2. In the former case, the $k = 1000$ upper order statistics taken into account for estimation range from 90% up to the 99.99% quantile ($n = 10^4$). Over that range, the slope of $\Psi_{L_F}$ is close to constant and hence $L_F$ grows approximately like a power function $x^{\eta_1}$ with (averaged) $\eta_1 \approx 3.8$; see Figure 2.4. Similarly, in the case of Figure 2.2, in the range of the $k = 1000$ upper order statistics ($n = 10^7$), $L_F$ grows approximately like a power $x^{\eta_2}$ with (averaged) $\eta_2 \approx 3.4$. Altogether, the asymptotic tail decay of $F \in \text{RV}_{-5}$ together with the local power-like growth of $L_F$ leads to a (local) regular variation index of around $-1.2$ and $-1.6$ in the respective cases. In terms of tail indices, this means we arrive locally at a tail index for $F \in \text{RV}_{-1/\xi}$ of around 0.83 or 0.625. Compare these values with the estimated $\hat{\xi} \approx 0.85$ or $\approx 0.6$ of the Figures 2.1 and 2.2.

To sum up, on the one hand the tail approximation (2.4) suggests that $F$, above a high enough threshold $u$, is well approximated by an exact Pareto distribution $K(x) = cx^{1/\xi}$ with tail index $\xi$ and for some constant $c = c(u)$. On the other hand, according to the above, it might be preferable in certain cases to approximate $F$ by a series of exact Pareto distributions with
threshold-dependent tail indices. Such an approximation is referred to as a penultimate approximation to $F$—as opposed to the ultimate tail approximation (2.4). The idea of penultimate approximations (which goes back to Fisher and Tippett [24]) is reviewed below in a more general tail quantile setting and discussed in detail in the accompanying Papers B and C.
3. Estimation of High Quantiles

There is a considerable amount of literature discussing new methodological developments in high-quantile estimation together with specific applications to such fields as environmental statistics, telecommunication, insurance and finance. Examples of textbooks include Balkema and Embrechts [2], Beirlant et al. [8], de Haan and Ferreira [25], Embrechts et al. [17], Finkenstädt and Rootzén [23], McNeil et al. [28] or Resnick [34].

As for financial risk management, recall that Basel II’s more advanced methods for the calculation of regulatory capital require the estimation of quantiles at very high levels of confidence—for MR usually 99% at a 10-day horizon, for CR and OR 99.9% and for economic capital even 99.97%, all three of them at a 1-year horizon. The credit crisis prompted the introduction of an extra 99.9%, 1-year capital charge for MR, the so-called Incremental Risk Charge; see Basel Committee [4].

Especially in the case of OR, the scarcity of loss databases together with the heavy-tailedness of losses make estimation at such high levels an inherently difficult task. One method increasingly championed in practice estimates quantiles at a lower level (e.g. 99%) and then scales up to the desired higher level (e.g. 99.9%) according to some scaling procedure to be specified. Below we summarize parts of the accompanying Paper C in which scaling properties of high quantiles are discussed.

3.1 Scaling Rules for High Quantiles

From a methodological perspective, it is the framework of second-order extended regular variation (2ERV) that is most useful for our purposes; see for instance de Haan and Ferreira [25], Appendix B, for an introduction to 2ERV. It allows for a unified treatment of the for quantitative financial risk management important cases (tail index \( \xi > 0 \) and \( \xi = 0 \)).

Recall from Theorem 2.1 that for a large class of dfs, the df \( F_u \) of normalized excesses \((X^u - u)/\beta(u)\) is, above high thresholds \( u \), well approximated by the df \( G_\xi \) of the GPD. Translating this to the quantile setting, we expect excess quantiles (properly normalized) to be well approximated by the quantiles...
Estimation of High Quantiles

of the GPD. Mathematically this is assured by the convergence properties of (generalized) inverse functions; see for instance Resnick [33], Proposition 0.1. Taking inverses in (2.3) one readily obtains, for $x > 0$,

$$\frac{U(tx) - U(t)}{a(t)} \to \frac{x^{\xi} - 1}{\xi}, \quad t \to \infty, \quad (3.1)$$

where we set $u = U(t)$ and $a(t) = \beta(u)$ in (2.3). Recall that $U$ satisfying (3.1) is said to be of extended regularly variation with index $\xi$ and auxiliary function $a(\cdot)$ and we write $U \in \text{ERV}_\xi(a)$. In fact $U \in \text{ERV}_\xi$ is equivalent to $F \in \text{MDA}(H_\xi)$; see for instance de Haan and Ferreira [25], Theorem 1.1.6. In the cases of interest from a financial risk management perspective (i.e. heavy- and semi heavy-tailed case $\xi > 0$ and $\xi = 0$), we have the following result.

**Proposition 3.1 (Corollary 3.1, Paper C)** Assume that $U \in \text{ERV}_\xi(a)$ for some $\xi \geq 0$ and some auxiliary function $a(\cdot)$. Then, $\log U \in \text{ERV}_0(b)$ and hence

$$\left(\frac{U(tx)}{U(t)}\right)^{1/b(t)} \to x, \quad t \to \infty, \quad (3.2)$$

where $b(t) = a(t)/U(t) \to \xi$. □

For $U \in \text{ERV}_\xi(a)$, $\xi \geq 0$, the limit relations (3.1) and (3.2) give rise to different scaling procedures for high quantiles. For $x > 1$, high quantiles $U(tx)$ may be approximated by scaling up lower quantiles $U(t)$ according to the following rules

$$i) \quad U(tx) \approx U(t) + a(t)\frac{x^{\xi} - 1}{\xi}, \quad (3.3)$$

$$ii) \quad U(tx) \approx x^{b(t)}U(t), \quad (3.4)$$

$$iii) \quad U(tx) \approx x^{\xi}U(t), \quad (3.5)$$

where $b(t) = a(t)/U(t)$.

While the Approximations (3.3) and (3.5) are well studied, see for instance Beirlant et al. [8], Chapter 5, or de Haan and Ferreira [25], Chapter 3, the so-called penultimate Approximation (3.4) has received little attention in the literature so far. It is the object of main interest in the accompanying Paper C, where we compare the goodness of the three Approximations (3.3)–(3.5) from a second-order extended regular variation perspective.
3.2 Second-Order Asymptotics of Normalized Quantiles

In Paper C our main focus is on the derivation and comparison of second-order results for (3.1) and (3.2). We consider functions $A(\cdot)$ and $B(\cdot)$ with

\[ \lim_{t \to \infty} A(t) = \lim_{t \to \infty} B(t) = 0, \]

which for $\xi \geq 0$ and $x > 0$ satisfy

\[
\frac{U(tx) - U(t)}{a(t)} - D_\xi(x) \to S(x), \quad t \to \infty, \tag{3.6}
\]

and

\[
D_\xi \left( \frac{U(tx)}{U(t)} \right)^{1/b(t)} - D_\xi(x) \to T(x), \quad t \to \infty, \tag{3.7}
\]

with $b(t) = a(t)/U(t)$ and for some non-trivial limits $S(\cdot)$ and $T(\cdot)$. A key message from that paper is that in certain cases relevant for practice, a judicious choice of a non-constant power normalization $b(\cdot)$ instead of a linear normalization $a(\cdot)$ may improve the convergence rate in the sense that $B(t)/A(t) \to 0$ as $t \to \infty$. Second-order results of the type (3.6) and (3.7) for three different choices of normalizations $b(\cdot)$ are derived in Paper C. The underlying general assumption for all of these results is that

(A1) the von Mises condition holds, i.e. $\frac{tU''(t)}{U'(t)} \to \xi - 1$, for some $\xi \geq 0$.

For the simplest choice $b(t) = \xi > 0$, no improvement of the convergence rate is possible. The two limit relations (3.6) and (3.7) coincide and we have the following second-order result.

**Proposition 3.2 (Proposition 3.3, Paper C)** Suppose $U(t) = e^{\phi(\log t)}$ is twice differentiable and let $A(t) = ta'(t)/a(t) - \xi$ with $a(t) = \xi U(t)$ for some $\xi > 0$. Assume that (A1) holds, that $\phi'$ is ultimately monotone, and that

\[ \lim_{t \to \infty} \frac{\phi''(t)}{(\phi'(t) - \xi)} = \rho, \text{ for some } \rho \leq 0. \]

Then, for $x > 0$,

\[ \lim_{t \to \infty} \frac{U(tx) - U(t)}{a(t)} - D_\xi(x) = \frac{1}{\xi} S_{\xi,\rho}(x), \]

where $D_\xi(x) = (x^\xi - 1)/\xi$ and

\[ S_{\xi,\rho}(x) = \begin{cases} x^\xi \frac{x^\rho - 1}{\rho}, & \rho < 0, \\ x^\xi \log x, & \rho = 0. \end{cases} \]
Estimation of High Quantiles

In order to illustrate how to find power normalizations \(b(\cdot)\) which allow for an improvement of the convergence rate (i.e. \(B(t)/A(t) \to 0\) as \(t \to \infty\)), consider the following argument given in Paper B, Section 3.2. Recall that in the case of \(b(t) = \xi\) high quantiles \(U(tx)\) are approximated by \(x^\xi U(t)\). The latter corresponds to the tail quantile function of an exact Pareto distribution with tail index \(\xi\) and left endpoint \(U(t)\). In a log-log setting with \(U(t) = e^{\varphi(\log t)}\) this means that the graph of \(\varphi\) is at high quantile levels approximated by a straight line with slope \(\xi\), i.e.

\[
\varphi(r + s) \approx \varphi(r) + \xi s, \quad (3.8)
\]

where \(r = \log t\) and \(s = \log x\). A better approximation may be achieved, if for large values of \(r\), \(\varphi\) is still approximated linearly, but now by its tangent line in the respective points \(r = \log t\), i.e. by a straight line with slope \(\varphi'\), leading to

\[
\varphi(r + s) \approx \varphi(r) + \varphi'(r)s. \quad (3.9)
\]

Note that (3.9) means that we consider a series of exact Pareto distributions (so-called penultimate approximation) with threshold-dependent tail index \(\varphi'(\log t)\) to approximate the true model \(U\).

One of the contributions in the accompanying Paper B is to make the above heuristic idea mathematically precise and we derive second-order results for (3.8) and (3.9). We show that under certain conditions on \(\varphi\) and its derivatives one obtains the following second-order results.

\[
\lim_{r \to \infty} \frac{\varphi(r + s) - (\varphi(r) + \xi s)}{\varphi'(r) - \xi} = \frac{e^{\rho s} - 1}{\rho},
\]

for some \(\rho \leq 0\), and

\[
\lim_{r \to \infty} \frac{\varphi(r + s) - (\varphi(r) + \varphi'(r)s)}{\varphi''(r)} = \frac{1}{2} s^2.
\]

see Paper B, Theorems 3.1 and 3.2 and the respective proofs for details.

The findings in Paper B motivate the consideration of a power normalization \(b(t) = \varphi'(\log t) = tU'(t)/U(t)\) in Paper C. One of the main results in the accompanying Paper C is the derivation of a second-order result for the power normalization case (3.7) for this choice of \(b(\cdot)\). We present it at this point to give a flavor.

**Proposition 3.3** (Proposition 3.5, Paper C) Suppose \(U(t) = e^{\varphi(\log t)}\) is three times differentiable and let \(B(t) = tb'(t)/b(t)\) with \(b(t) = tU'(t)/U(t)\). Assume that (A1) holds for some \(\xi \geq 0\), that \(\varphi''\) is ultimately monotone, and
3.2. Second-Order Asymptotics of Normalized Quantiles

that \( \lim_{t \to \infty} \varphi'''(t)/\varphi''(t) = \rho \), for some \( \rho \leq 0 \). Then, for \( x > 0 \),

\[
\lim_{t \to \infty} \frac{D_\xi \left( \left( \frac{U(tx)}{U(t)} \right)^{1/b(t)} \right) - D_\xi(x)}{B(t)} = T_{\xi,\rho}(x),
\]

where \( D_\xi(x) = (x^\xi - 1)/\xi \) and

\[
T_{\xi,\rho}(x) = \begin{cases} \frac{x^\xi}{\rho} \left( \frac{x^\rho - 1}{\rho} - \log x \right), & \rho < 0, \\ \frac{x^\xi}{\rho^2} (\log x)^2, & \rho = 0. \end{cases}
\]

We compare this result with the corresponding result for the linear normalization case (3.6). The latter is found for instance de Haan and Ferreira [25], Theorem 2.3.12. We show that under certain conditions, the application of the power normalization \( b(t) = tU'(t)/U(t) \) may indeed improve the rate of convergence when compared to the case of the linear normalization, i.e. for the choice \( a(t) = tU'(t) \) the function \( A(\cdot) \) in (3.6) is such that \( B(t)/A(t) \to 0 \) as \( t \to \infty \). Without going into details at this point we remark that this necessitates the the second-order regular variation parameter \( \rho \) to be equal to 0; see Paper C, Section 3.2.

The third choice of a power normalization \( b(\cdot) \) considered in Paper C may be obtained by the following reasoning. If one does not want to assume differentiability a priori, \( \varphi' \) need not exist. In that case, by Karamata’s Theorem, \( \varphi(\log t) \) is of the same order as its average \( \tilde{\varphi}(\log t) := \frac{1}{t} \int_{t_0}^t \varphi(\log s)ds \), for some \( 0 < t_0 < t \), i.e. we have \( \tilde{\varphi}(\log t)/\varphi(\log t) \to 1 \) as \( t \to \infty \). However, unlike \( \varphi \), \( \tilde{\varphi} \) is always differentiable and hence one may choose \( b(t) = \tilde{\varphi}(\log t) := \varphi(\log t) - \frac{1}{t} \int_{t_0}^t \varphi(\log s)ds \) as power normalization in (3.2). We may think of \( \tilde{\varphi} \) as a kind of local “pseudo” slope of \( \varphi \). Second-order results of the type (3.6) and (3.7) under this choice of power normalization are derived in the Propositions 3.6 and 3.7 of Paper C. Similar to before, a comparison of the two second-order results yields that under certain conditions we may obtain \( B(t)/A(t) \to 0 \) as \( t \to \infty \). Again, this is only possible if \( \rho \) equals 0.

In conclusion, from a second-order extended regular variation perspective it is thus exactly in the worst case \( \rho = 0 \) where we may hope for an (asymptotic) improvement of the convergence rate by applying a judicious power normalization in (3.2). Furthermore, the asymptotic results on power norming of high quantiles provide the methodological basis for the application of penultimate approximations as given in (3.4). Finally, the application of penultimate approximations may well lead to a considerable improvement of the approximation accuracy for high quantiles as we show next.
3.3 Penultimate Approximations in Practice

In this section we illustrate the potential usefulness of the penultimate Approximation (3.4) for practical applications for the case $b(t) = \varphi'(\log t)$ along the lines of Paper B, Section 3.3 and Paper C, Section 4. We compare the performance of the penultimate Approximation (3.4) with the two other Approximations (3.3) and (3.5) for five OR loss severity models. To do so, we consider the (theoretical) relative approximation error at the 99.9% quantile as a function of the level $t$ of the scaled up lower quantile $U(t)$, i.e.

$$e(t) = \left| \frac{1 + b(t) \left( \frac{x}{\xi} \right)^{\xi-1} U(t)}{U(x)} - 1 \right|, \quad t \leq x = 1000,$$

for Approximation (3.3) and

$$e(t) = \left| \frac{\left( \frac{x}{\xi} \right)^{b(t)} U(t)}{U(x)} - 1 \right|, \quad t \leq x = 1000,$$

with $b(t) = \varphi'(\log t)$ for the penultimate Approximation (3.4) and with $b(t) \equiv \xi$ for the ultimate Approximation (3.5); see Figure 3.1.

Figure 3.1: Relative approximation errors for the 99.9% quantile using the ultimate Approximation (3.5) (dashed), the penultimate Approximation (3.4) (full) and the Approximation (3.3) (dotted) for the g-and-h ($g = 2$, $h = 0.2$; labels $u_1$, $p_1$, $l_1$), the lognormal ($\mu = 3.5$, $\sigma = 1.25$; labels $u_2$, $p_2$, $l_2$) and the loggamma distribution ($\alpha = 2$, $\beta = 1.25$; labels $u_3$, $p_3$, $l_3$) and the Burr ($\alpha = 1$, $\kappa = 2$, $\tau = 1.5$, labels $u_4$, $p_4$, $l_4$) and the GB2 distribution ($a = 2$, $b = 2$, $p = 1$, $q = 0.75$; labels $u_5$, $p_5$, $l_5$). For the respective parameterizations, see Paper B.
3.3. Penultimate Approximations in Practice

From the left panel of Figure 3.1 we may conclude that the use of a penultimate approximation to approximate high quantiles may indeed be beneficial for certain loss models. The approximation error under a penultimate approximation with power normalization \( b(t) = \varphi'(\log t) \) is lower than for the two Approximations (3.3) and (3.5). The situation in the right panel of Figure 3.1 is similar in that the ultimate Approximation (3.5) performs worst. Also, the errors for the dfs in the right panel vanish rather quickly compared with the left panel.

To complement these results quantitatively, we show in Table 3.1 the approximation errors when scaling from a lower quantile level of 99%, say, up to the desired level of 99.9% using the three Approximations (3.3)–(3.5). Note that we disregard parameter estimation uncertainty for the moment as the tail index \( \xi \) and the local tail index \( \varphi'(\log t) \) are assumed to be known.

<table>
<thead>
<tr>
<th>Approximation</th>
<th>Approx. (3.5)</th>
<th>Approx. (3.4)</th>
<th>Approx. (3.3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>g-and-h</td>
<td>77.43</td>
<td>46.94</td>
<td>21.83</td>
</tr>
<tr>
<td>Lognormal</td>
<td>61.51</td>
<td>19.95</td>
<td>13.32</td>
</tr>
<tr>
<td>Loggamma</td>
<td>15.39</td>
<td>6.06</td>
<td>3.16</td>
</tr>
<tr>
<td>Burr</td>
<td>4.76</td>
<td>0.91</td>
<td>3.71</td>
</tr>
<tr>
<td>GB2</td>
<td>0.10</td>
<td>0.07</td>
<td>0.23</td>
</tr>
</tbody>
</table>

Table 3.1: Relative approximation error (in %) for the five loss models of Figure 3.1 due to quantile scaling from lower level 99% to higher level 99.9% using the Approximations (3.3)–(3.5).

From Figure 3.1 and Table 3.1 we may draw the following conclusions. Firstly, the g-and-h, the lognormal and the loggamma show substantial approximation errors under the ultimate approximation—even at very high quantile levels. Secondly, the theoretical approximation errors—especially for these three models—are substantially lower when a penultimate approximation is used. These conclusions do not come as a surprise and are due to the fact that the three models g-and-h, lognormal and loggamma, belong to the “worst” second-order case, i.e. \( U \in \mathcal{ERV}_{\xi,\rho} \) with \( \rho = 0 \). It is exactly in this case where an improvement of the convergence rate in (3.2) over the rate in (3.1) is possible; see Paper C, Propositions 3.5 and 3.6. and subsequent discussion in Section 3.2.
3.3.1 Estimation under Penultimate Approximations

The introduction of power normalizations and the resulting penultimate approximations motivate the estimation of high quantiles $U(x)$ through

$$\hat{U}(x) = \left(\frac{x^k}{n}\right)^{\hat{b}(\frac{x}{k})} X_{n-k,n},$$

where $\hat{b}$ is an estimate of $b$ and where $X_{n-k,n}$ denotes the k-th largest order statistic of an underlying iid sample $X_1, \ldots, X_n \sim F$.

For the estimation of the power normalization $b(\cdot)$, different approaches may be considered. One possibility is to take $b(\cdot) = \varphi'(\log \cdot)$ as above, i.e. as the slope in the log-log plot of $U$. A convenient way to estimate the slope of a curve is through local polynomial regression (LPR), i.e. $\hat{b}_{LPR}(\frac{n}{k})$ is the derivative of the fitted local polynomial, evaluated at the point $t = n/k$. This approach is discussed more in detail in the accompanying Paper C, Section 4. As we work with $b(t) = a(t)/U(t)$, we may alternatively consider

$$\hat{b}(\frac{n}{k}) = \hat{a}(\frac{n}{k}) X_{n-k,n}.$$

This approach is not discussed in Paper C and so we give some more notational details here. Different well-studied estimators for $a(\cdot)$ are discussed for instance in Beirlant et al. [8], Section 5.3, or de Haan and Ferreira [25], Chapter 3. We adopt the notation of the latter and define

$$P_n := \frac{1}{k} \sum_{i=1}^{k-1} (X_{n-i,n} - X_{n-k,n}), \quad Q_n := \frac{1}{k} \sum_{i=1}^{k-1} \frac{1}{n} (X_{n-i,n} - X_{n-k,n}) \quad \text{and} \quad M_n^{(j)} := \frac{1}{k} \sum_{i=1}^{k-1} (\log X_{n-i,n} - \log X_{n-k,n})^j, \quad j = 1, 2.$$

We obtain a probability-weighted moment estimator, a moment estimator and a maximum likelihood estimator for $b(\frac{n}{k})$, respectively given by

$$\hat{b}_{PWM}(\frac{n}{k}) = \frac{P_n}{X_{n-k,n}} \left(\frac{P_n}{2Q_n} - 1\right)^{-1},$$

$$\hat{b}_M(\frac{n}{k}) = \frac{1}{2} M_n^{(1)} \left(1 - \left(\frac{M_n^{(1)}}{M_n^{(2)}}\right)^2\right)^{-1}, \quad \text{and}$$

$$\hat{b}_{MLE}(\frac{n}{k}) = \frac{\hat{\sigma}_{MLE}(\frac{n}{k})}{X_{n-k,n}};$$

see de Haan and Ferreira [25], Section 3.4, for details on $\hat{\sigma}_{MLE}$. 

28
3.3. Penultimate Approximations in Practice

3.3.2 A Small Simulation Study

We first stress again that the main goal in Paper C was to explore the influence of power norming as opposed to linear norming on the respective second-order regular variation properties rather than to do an in-depth comparative statistical analysis for the different resulting estimators $\hat{b}$. Nevertheless, in order to give a flavor we compare the performance of the different Approximations (3.3)–(3.5) by means of a small simulation study. The simulation study is based on 200 samples of sizes $n = 500$ and $n = 250$ of simulated data from a loggamma, a g-and-h and a Burr model, we estimate the 99.9\% quantile by scaling up the 98\% quantile according to the scaling rules (3.3)–(3.5); see Table 3.2.

<table>
<thead>
<tr>
<th></th>
<th>$n = 500$</th>
<th></th>
<th>$n = 250$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias</td>
<td>SRMSE</td>
<td>Bias</td>
<td>SRMSE</td>
</tr>
<tr>
<td>Loggamma ($\alpha = \beta = 1.25$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3.5) with $\hat{\xi}_{POT}$</td>
<td>1.98</td>
<td>81.02</td>
<td>-7.61</td>
<td>135.29</td>
</tr>
<tr>
<td>(3.3) with $\hat{a}<em>M$ and $\hat{\xi}</em>{POT}$</td>
<td>9.51</td>
<td>84.28</td>
<td>-9.74</td>
<td>117.11</td>
</tr>
<tr>
<td>(3.4) with $\hat{b}_M$</td>
<td>24.43</td>
<td>96.64</td>
<td>-15.20</td>
<td>81.54</td>
</tr>
<tr>
<td>(3.4) with $\hat{b}_{LPR}$</td>
<td>-5.57</td>
<td>38.77</td>
<td>-12.71</td>
<td>53.98</td>
</tr>
<tr>
<td>g-and-h ($a = b = 1.5, g = 0.6, h = 0.8$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3.5) with $\hat{\xi}_{POT}$</td>
<td>23.75</td>
<td>121.25</td>
<td>41.30</td>
<td>270.46</td>
</tr>
<tr>
<td>(3.3) with $\hat{a}<em>M$ and $\hat{\xi}</em>{POT}$</td>
<td>24.41</td>
<td>117.63</td>
<td>16.74</td>
<td>208.65</td>
</tr>
<tr>
<td>(3.4) with $\hat{b}_M$</td>
<td>24.98</td>
<td>110.35</td>
<td>23.50</td>
<td>101.66</td>
</tr>
<tr>
<td>(3.4) with $\hat{b}_{LPR}$</td>
<td>-7.19</td>
<td>44.65</td>
<td>-15.14</td>
<td>58.29</td>
</tr>
<tr>
<td>Burr ($\alpha = \kappa = 0.8, \tau = 2$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3.5) with $\hat{\xi}_{POT}$</td>
<td>-17.51</td>
<td>49.39</td>
<td>-20.67</td>
<td>68.95</td>
</tr>
<tr>
<td>(3.3) with $\hat{a}<em>M$ and $\hat{\xi}</em>{POT}$</td>
<td>-7.96</td>
<td>48.08</td>
<td>-16.91</td>
<td>61.06</td>
</tr>
<tr>
<td>(3.4) with $\hat{b}_M$</td>
<td>4.49</td>
<td>51.72</td>
<td>-13.74</td>
<td>54.05</td>
</tr>
<tr>
<td>(3.4) with $\hat{b}_{LPR}$</td>
<td>-0.75</td>
<td>40.44</td>
<td>3.73</td>
<td>58.84</td>
</tr>
</tbody>
</table>

Table 3.2: Estimation of the 99.9\% quantile by scaling up the 98\% quantile using the Approximations (3.3)–(3.5). Estimation bias and SRMSE (in \%) of the four estimators are based on 200 samples of 500 and 250 observations from a Loggamma, a g-and-h and a Burr model; see Paper C for the parameterization of the respective models.

Concerning estimation in Table 3.2, the lower quantile (98\%) is in each case estimated as the empirical quantile. The tail index $\xi$ appearing in the Approximations (3.3) and (3.5) is estimated using the MLE-POT method based
Estimation of High Quantiles

on the 10% largest order statistics. The scale functions $a(\cdot)$ and $b(\cdot)$ in the Approximations (3.3) and (3.4) are estimated using the moment estimator. Finally, the estimation of $\hat{b}_{LPR}$ is carried out with the function “locfit” provided in R. In this local polynomial regression procedure we fit a local quadratic polynomial (with a tricube weight function (default kernel) and smoothing parameter of 1/2); see Paper C and references therein for details.

Besides the moment estimators $\hat{a}_M$ and $\hat{b}_M$, other estimators have been considered as well as other loss models such as the Pareto, the lognormal or the GB2. For the sake of brevity we refrain from showing all these results here. Moreover, for $\hat{b}_{LPR}$ we have also tested local linear polynomials as well as different weight functions but the influence seems to be marginal.

The Achilles’ heel of standardly used kernel smoothing procedures (i.e. local polynomials of degree 0) is the sensitivity of the estimation results to the choice of the bandwidth as well as the well-known boundary problem. However, using non-constant polynomials one may circumvent the boundary problem and also the bandwidth (nearest neighborhood fraction) selection problem does not seem to be a too delicate issue; see Figure 3.2. In that figure we show the sensitivity of changes of the bandwidth/smoothing parameter in the estimation errors of the 99.9% quantile under Approximation (3.4) with $\hat{b}_{LPR}$ using local quadratic regression with a tricube weight function and a smoothing parameter of 1/2.

\begin{center}
\begin{figure}
\begin{subfigure}{0.45\textwidth}
\centering
\includegraphics[width=\textwidth]{bias_curve}
\caption{Bias (in %) vs. bandwidth for the 99.9% quantile estimates under (3.4) with $\hat{b}_{LPR}$ using local quadratic regression. The estimation is based on 200 samples of 250 observations of loggamma (full), g-and-h (dotted) and Burr (dashed) data with parameter values as in Table 3.2.}
\end{subfigure}
\begin{subfigure}{0.45\textwidth}
\centering
\includegraphics[width=\textwidth]{srmse_curve}
\caption{SRMSE (in %) vs. bandwidth for the 99.9% quantile estimates under (3.4) with $\hat{b}_{LPR}$ using local quadratic regression.}
\end{subfigure}
\end{figure}
\end{center}

According to Figure 3.2 the performance of high quantile estimation using $\hat{b}_{LPR}$ seems to be relatively stable (or at least stays within an acceptable
range) under bandwidth changes over a large range of different bandwidth values (around 0.4 up to 0.9).

In summary, the empirical results from Table 3.2 are in line with the theoretical findings of Table 3.1. In particular we see that Approximation (3.3) and the penultimate Approximation (3.4) outperform the ultimate Approximation (3.5). Comparing the performance results of Approximation (3.3) with those of the penultimate Approximation (3.4) there seems to be a slight preference for the latter—at least in the case of $\hat{b}_{LPR}$. The potential usefulness of the latter is further supported by Figure 3.2. Finally, we want to re-emphasize that the empirical findings on penultimate approximations presented in this section are by no means intended as a justification to generally prefer one method over the other but shall rather be seen as a motivation to perform further statistical research on penultimate approximations.
4. Analysis of Diversification Benefits

Recall that according to the Basel II guidelines, the calculation of the regulatory capital charge $RC_{OR}$ for operational risk (OR) requires banks to divide the range of their business activities into $d$, say, units of measure, to add up the $d$ different risk measures and to allow for a diversification benefit $\delta$, i.e.

$$RC_{OR} = \text{VaR}_\alpha \left( \sum_{i=1}^{d} S_i \right) = (1 - \delta) \sum_{i=1}^{d} \text{VaR}_\alpha (S_i),$$

at level $\alpha = 99.9\%$; see Basel Committee [3], § 657 and 669(d).

**Remark:** At this point it is worthwhile to remember that VaR is not a coherent risk measure so that one may encounter $\delta > 0$ (subadditivity) as well as $\delta < 0$ (superadditivity).

A bank is not allowed to take into account a diversification benefit $\delta > 0$ unless it can demonstrate adequate support of its dependence assumptions; see for instance Federal Register [20], p. 69317. From a mathematical viewpoint requiring $\delta = 0$, i.e. adding up quantiles, is correct if the risks $(S_i)_{1 \leq i \leq d}$ are comonotonic which in turn is the case, if the dependence measure Kendall’s $\tau$ attains its maximal value 1 for every pair $(S_i, S_j)$, $1 \leq i, j \leq d$; see for instance Embrechts et al. [19]. There is empirical evidence however that OR data shows surprisingly little dependence. In particular Kendall’s $\tau$ at business line level seems to be far off from its maximal value 1; see Cope and Antonini [13], Figure 3. Therefore the assumption of comonotonicity and hence also the requirement $\delta = 0$ may not be appropriate.

**Remark (Practical Relevance):** No doubt, the incorporation of diversification benefits $\delta > 0$ to lower the regulatory capital charge $RC_{OR}$ would be convenient from an economic viewpoint. So far not enough evidence has been provided though to convince regulators to allow $\delta > 0$.  

33
The empirical findings of Cope and Antonini [13] motivate a rigorous mathematical analysis of diversification. In the accompanying Paper D we consider the case $S_i \sim F, 1 \leq i \leq d$, and analyze $\delta$ as a function of

i) the quantile level $\alpha$,

ii) the single loss model $F$, and

ii) the number of risks $d$.

The underlying general assumption in Paper D is that of heavy-tailed loss models, i.e. $F \in RV_{-1/\xi}, \xi > 0$, or equivalently $U \in RV_\xi$ for its tail quantile function. In that case, $F$ is subexponential so that, with $G(x) = F^d(x)$ denoting the $d$-fold convolution of $F$, we have

$$\frac{G(x)}{F(x)} \to d, \quad x \to \infty;$$

see for instance Embrechts et al. [17], Corollary 1.3.2. Setting $G^{-1}(\alpha) = x$, one obtains for $\alpha \to 1$,

$$1 - \delta(\alpha) = \frac{\text{VaR}_\alpha \left( \sum_{i=1}^d S_i \right)}{\sum_{i=1}^d \text{VaR}_\alpha (S_i)} = \frac{1}{d} \frac{U \left( \frac{1}{F(x)} \right)}{U \left( \frac{F(x)}{G(x)} 1/F(x) \right)} \to d^{\xi-1}, \quad (4.1)$$

due to the Uniform Convergence Theorem for regularly varying functions; see for instance Bingham et al. [9], Theorem 1.5.2.

Given the high $\alpha$-levels typically of interest for risk management practice, analyzing diversification benefits by means of their empirical counterparts will in general not yield much insight. One is therefore advised to consider (suitable) analytic approximations of $\delta(\cdot)$. In the simple case of $d$ regularly varying iid losses, (4.1) gives rise to a first-order approximation $1 - \delta_1(\alpha) = d^{\xi-1}$ of the diversification benefit $\delta(\alpha)$ for large values of $\alpha$.

To get a grasp we show in Figure 4.1 the behavior of diversification benefits for the simple case of $d$ iid Pareto rvs. In the left panel we fix $d = 2$ and vary the tail index $\xi$ whereas in the right panel we keep $\xi = 0.5$ fixed and vary the numbers of risks $d$.

From Figure 4.1 we see that in the case of a Pareto loss model, diversification benefits tend to increase with increasing number of risks. On the other hand, for a fixed number of risks, benefits seem to decrease (and even become negative) with increasing heavy-tailedness. In particular Figure 4.1 also shows that diversification benefits are not constant as a function of $\alpha$ and, consequently, the first-order approximation $1 - \delta_1(\alpha) \equiv d^{\xi-1}$ may be
One of the main problems discussed in the accompanying Paper D is the fact that the convergence in (4.1) may be arbitrarily slow. As illustrated in that paper, the behavior of $\delta(\alpha)$ at levels of $\alpha$ close to 1 (typically of interest for applications) may be very sensitive—small changes of $\alpha$ lead to large changes of $\delta(\alpha)$. In economic terms this means that while we may well expect diversification benefits of considerable size at a certain level $\alpha$, this may change rather drastically and even lead to non-diversification (superadditivity of VaR) once we move away only a little from that level.

Altogether this motivates the consideration of a second-order approximation for the diversification benefit $\delta$. Concerning methodology, (4.1) suggests to draw on the theories of second-order regular variation and second-order subexponentiality; see for instance de Haan and Ferreira [25] and Omey and Willekens [30, 31] for an introduction. The main result of Paper D is a second-order result of the form

$$1 - \delta(\alpha) = d^{\xi-1} + A(\alpha)K + o(A(\alpha)), \quad \alpha \to 1,$$

for a suitable function $A(\cdot)$ and some constant $K$; see Paper D, Theorem 3.1 for details. The main assumption needed in order for Theorem 3.1 to hold is that of second-order regular variation, i.e. $F \in 2RV_{-1/\xi,\rho/\xi}$, for some $\xi > 0$ and some $\rho \leq 0$ or, equivalently, $U \in 2RV_{\xi,\rho}$.

A key message from our main result is that two different asymptotic regimes may arise leading to two different expansions of the type (4.2). Either it is the behavior in the regular variation part or in the subexponential part.

Figure 4.1: Behavior of diversification benefits $1 - \delta(\alpha)$ for $d$ iid Pareto$(1, \xi)$ rvs where in the left panel $d = 2$ is fixed with varying $\xi$ (theoretical $\delta$, $G^-$ numerically inverted). In the right panel $\xi = 0.5$ is fixed and $d = 2, 4, 6, 8, 10$ (empirical $\delta$, based on $n = 10^7$ simulations).

too crude. This may be more pronounced for other loss models depending on their convergence properties in (4.1).
Analysis of Diversification Benefits

in (4.1) that dominates in the limit. In terms of first- and second-order regular variation parameters this is the case if \( \rho > -(1 \wedge \xi) \) or \( \rho < -(1 \wedge \xi) \) respectively. In Paper D we refer to these two cases as the slow convergence and fast convergence case respectively.

**Example 4.1 (Pareto)** Let \( X_1, \ldots, X_d \) iid \( \sim \text{Pareto}(x_0, \xi) \), parameterized by \( F(x) = (x/x_0)^{-1/\xi} \) for \( x > x_0 > 0 \) and some \( \xi > 0 \). Note that the second-order parameter \( \rho \) controlling the convergence rate in (2.5) may be understood as \( \rho = -\infty \), i.e. \( F \in 2RV_{-1/\xi,-\infty} \), and hence the exact Pareto is a loss model of fast convergence. According to Theorem 3.1 of Paper D, we obtain for \( d \geq 1 \) and as \( \alpha \to 1 \),

\[
1 - \delta(\alpha) = \begin{cases} 
\frac{d \xi^{-1} + \frac{d-1}{d} (1-\alpha)^{\xi}}{1-\xi} + o \left( (1 - \alpha)^{\xi} \right), & \xi < 1, \\
1 - \frac{d}{d-1} (1-\alpha) \log (1 - \alpha) + o ((1 - \alpha) \log (1 - \alpha)), & \xi = 1, \\
\frac{d \xi^{-1} - \xi d \xi^{-1} - 1 d \xi^{-1} \Gamma(1-1/\xi)}{d \xi^{-1} \Gamma(1-1/\xi)} (1 - \alpha) + o(1 - \alpha), & \xi > 1.
\end{cases}
\]

Note that \( 1 - \delta(\alpha) \) is independent of the left endpoint \( x_0 \) of the Pareto\((x_0, \xi)\) distribution.

The results of Example 4.1 may be used to approximate diversification benefit curves as in Figure 4.1. This is illustrated in Figure 4.2 where we compare

![Figure 4.2](image-url)

**Figure 4.2:** Empirical diversification benefits (full, based on \( n = 10^7 \) simulations) together with first-order approximation \( 1 - \delta_1(\alpha) \equiv \sqrt{2}/2 \approx 0.71 \) and second-order approximation \( 1 - \delta_2 \) (dashed) for two iid Burr \((\tau = 0.25, \kappa = 8)\), Pareto \((x_0 = 1, \xi = 0.5)\) and g-and-h \((g = 2, h = 0.5)\) rvs.
the first-order approximation $1 - \delta_1(\alpha) = d^{\xi - 1}$ and the second-order approximation resulting from (4.2), given by $1 - \delta_2(\alpha) = d^{\xi - 1} + A(\alpha)K$, for the case $d = 2$ iid rvs from an exact Pareto, a Burr and a g-and-h distribution. Note that the first order approximation is the same for all three models as they have the same tail index $\xi = 0.5$.

**Remark:** Unlike Figure 4.2 might suggest, $1 - \delta(\alpha)$ need not necessarily approach its ultimate value $d^{\xi - 1}$ from above as $\alpha \to 1$ since $A(\cdot)$ and $K$ in (4.2) may be positive or negative; see Paper D, Section 4, for a discussion of this issue.

Summing up, Figure 4.2 confirms that the behavior of diversification benefits $\delta(\alpha)$ at levels of $\alpha$ close to 1 (typically of interest for applications) may indeed be very sensitive to small changes of $\alpha$. Moreover, for the g-and-h model of Figure 4.2 for instance, the regime switch from sub- to superadditivity takes place at the extreme level of $\alpha \approx 99.95\%$. Altogether, it seems that the second-order approximation $1 - \delta_2$ is able to capture this behavior better than $1 - \delta_1$ and in this sense provides a further step towards a better understanding of the concept of diversification.

### 4.1 Estimation—A Brief Outlook

In Paper D we discuss probabilistic properties of diversification benefits. The question of statistical estimation of diversification benefits $\delta$ is not addressed. Estimation of diversification benefits using the second-order approximation $1 - \delta_2(\cdot)$ is difficult in general, as one needs to estimate the function $A(\cdot)$ as well as the constant $K$. Prior to that, one would need to estimate the second-order regular variation parameter $\rho$ in order to determine which of the two asymptotic regimes applies and, based on that, which of the two resulting second-order approximations one should choose; see Paper D, Theorem 3.1.

In some cases however, for instance for the exact Pareto distribution as underlying model $F$, the quantities $A(\cdot)$ and $K$ simplify considerably and can be calculated explicitly; see Example 4.1. Therefore, for an unknown underlying model $F$ one possibility might be to use penultimate approximations to approximate its tail quantile function $U(t) = e^{\phi(\log t)}$ by a series of exact Pareto distributions, whose tail-indices $\xi = \xi(\alpha)$ depend on the quantile level $\alpha$. Note that in that case $\rho < -(1 \wedge \xi(\alpha))$ and hence a corresponding
Analysis of Diversification Benefits

second-order approximation for the diversification benefit may be given by

\[
1 - \delta_2(\alpha) = \begin{cases} 
    d \xi(\alpha)^{-1} + \frac{d-1}{d} \frac{(1-\alpha)\xi(\alpha)}{1-\xi(\alpha)}, & \xi(\alpha) < 1, \\
    1 - \frac{d}{d} (1 - \alpha) \log (1 - \alpha), & \xi(\alpha) = 1, \\
    d \xi(\alpha)^{-1} - \xi(\alpha) d \xi(\alpha)^{-1} \frac{d-1}{d} \frac{\Gamma(1-1/\xi(\alpha))}{2\Gamma(1-2/\xi(\alpha))} (1 - \alpha), & \xi(\alpha) > 1,
\end{cases}
\]

where \( \xi(\alpha) = \varphi'(\log t) \) with \( t = 1/(1 - \alpha) \) is the local tail index of the underlying model \( U \). Hence, the estimation of \( \delta_2 \) at a given quantile level \( \alpha \) requires the estimation of the local tail index \( \varphi'(\log t) \) at the level \( t = 1/(1 - \alpha) \). Estimators for the latter have been introduced in Chapter 3 of this review.

4.2 Concluding Remarks

One of the key messages from Paper D is a word of warning against an imprudent use of the concept of diversification. Indeed, already in the simplest case of iid random variables \((S_i)_{1 \leq i \leq d}\), understanding the mechanisms of diversification is not an easy task. Adding dependence to \((S_1, \ldots, S_d)\) will quickly increase the complexity of the problem and it is not clear to what extent the problem will still be analytically tractable; see for instance Embrechts et al. [18] and references therein.

On a more general note, we tried to write all four accompanying Papers A–D in such a way that some parts may be of interest to the readership interested in some of the more methodological aspects underlying certain EVT-based methods while other parts may help (warn) the more applied risk management end-user interested in possible applications of such methods. The light in which these papers shall be seen may aptly be summed by quoting Christoffersen et al. [12]:

“Thus, we believe that best-practice applications of EVT to financial risk management will benefit from awareness of its limitations, as well as the strengths [...] Our point is simply that we should not ask more of the theory than it can deliver.”
Bibliography


The Quantitative Modeling of Operational Risk: Between g-and-h and EVT.

The Quantitative Modeling of Operational Risk: Between g-and-h and EVT

Matthias Degen Paul Embrechts Dominik D. Lambrigger
Department of Mathematics, ETH Zurich, 8092 Zurich, Switzerland

Abstract

Operational risk has become an important risk component in the banking and insurance world. The availability of (few) reasonable data sets has given some authors the opportunity to analyze operational risk data and to propose different models for quantification. As proposed in Dutta and Perry [12], the parametric g-and-h distribution has recently emerged as an interesting candidate.

In our paper, we discuss some fundamental properties of the g-and-h distribution and their link to extreme value theory (EVT). We show that for the g-and-h distribution, convergence of the excess distribution to the generalized Pareto distribution (GPD) is extremely slow and therefore quantile estimation using EVT may lead to inaccurate results if data are well modeled by a g-and-h distribution. We further discuss the subadditivity property of Value-at-Risk (VaR) for g-and-h random variables and show that for reasonable $g$ and $h$ parameter values, superadditivity may appear when estimating high quantiles. Finally, we look at the g-and-h distribution in the one-claim-causes-ruin paradigm.

Keywords: Extreme Value Theory, g-and-h Distribution, Hill Estimator, LDA, Operational Risk, Peaks Over Threshold, Second Order Regular Variation, Subadditivity, Value-at-Risk.

1 Introduction

Since the early discussion around Basel II and Solvency 2, the pros and cons of a quantitative (Pillar I) approach to operational risk have been widely
put forward. Some papers, like Danielsson et al. [7], have early on warned against an over optimistic view that tools from market (and to some extent credit) risk management can easily be transported to the Basel II framework for operational risk. Also, the actuarial community working on Solvency II so far defied a precise definition, and as a consequence a detailed quantitative capital measurement for operational risk. The situation in the banking world is very different indeed, not only did Basel II settle on a precise definition, “The risk of loss resulting from inadequate or failed internal processes, people and systems or from external events. Including legal risk, but excluding strategic and reputational risk.”, also concrete suggestions for risk capital calculation have been made. These include the basic-indicator approach (BIA), the standardized approach (SA) and the loss distribution approach (LDA). BIA and SA are easy to calculate pure volume based measures. In the LDA however, banks are basically given full methodological freedom for the calculation of regulatory capital. The main reason being that for this new, and especially from a statistical data point of view, poorly understood risk class, regulators hope that modeling freedom would yield a healthy competition among the quant groups of various financial institutions. Whereas this point of view is no doubt a laudable one, the imposed boundary conditions make a practical implementation more than difficult. Some of these constraints are the use of the risk measure (VaR), the level (99.9%) and the “holding” period (1 year). Of these, the extremely high quantile (corresponding to a 1 in 1000 year event estimation) is no doubt the most critical one. Beyond these, banks are required to augment internal data modeling with external data and expert opinion. An approach that allows for combining these sources of information is for instance discussed in Lambrigger et al. [17]. The fact that current data—especially at the individual bank level—are far from being of high quality or abundant, makes a reliable LDA for the moment questionable.

By now, numerous papers, reports, software, textbooks have been written on the subject. For our purposes, as textbooks we would like to mention McNeil et al. [21] and Panjer [27]. Both books stress the relevance of actuarial methodology towards a successful LDA; it is no coincidence that in McNeil et al. [21], Chapter 10 carries the title “Operational Risk and Insurance Analytics”. Another recent actuarial text that at some point will no doubt leave its footprint on the LDA platform is Bühlmann and Gisler [6].

For the present paper, two fundamental papers, which are center stage to the whole LDA controversy, are Moscadelli [25] and Dutta and Perry [12]. Both are very competently written papers championing different analytic approaches to the capital charge problem. Whereas Moscadelli [25] is strongly based on EVT, Dutta and Perry [12] introduce as a benchmark model the parametric g-and-h distribution. Moscadelli [25] concludes that, based on the

46
2002 Loss Data Collection Exercise (LDCE) of the Basel Committee, EVT yields reasonable capital estimates when data are pooled at Business Line (BL) level. A considerable broader range for BL $\beta$-coefficients in the SA beyond the Basel II (12-18)% range is arrived at. The overall $\alpha = 15\%$ coefficient in the BIA is corroborated. The information coming through from individual banks with respect to the use of EVT is mixed. As explained in Nešlehová et al. [26], the statistical properties of the data are no doubt a main fact underlying this diffuse image. When it comes to high quantile estimation (and 99.9% is very high) EVT emerges as a very natural key methodological player; more on this later in the paper.

In Dutta and Perry [12] the authors conclude that “Many different techniques being tested by researchers are centered around EVT. In many of those cases we observe that attempts are made to fit a distribution or apply a method without understanding the characteristics of the loss data or the limitation of the models”. And further, “In our experiment we found that EVT often did not work for data where we observed many large losses”. Based on the 2004 LDCE, Dutta and Perry [12] suggest the g-and-h distribution as a viable option. They also stress that the 2002 LDCE data were pooled across many banks outside of the US. The quality of this data was better in the sense of comprehensiveness because it included banks all around the world. Compared to the 2004 LDCE data, the 2002 LDCE time series were shorter and many non-US banks did not suffer large losses.

As already stated above, we consider both Moscadelli [25] and Dutta and Perry [12] as very well written. The latter paper also introduces a fundamental, more qualitative yardstick against which any capital charge model ought to be tested:

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

In our paper, we will mainly look carefully at the g-and-h approach and compare and contrast its properties with EVT based methodology. As academics we do not possess real operational risk data so that our comments may be
“academic” in nature; we do however hope that the various results discussed will contribute positively towards the quest for a reliable (in the sense of 1.-5. above) capital charge for operational risk. Based on the empirical findings of Dutta and Perry [12] that (1) operational risk data seem to be modeled appropriately by the g-and-h; and that (2) the EVT based Peaks Over Threshold (POT) approach does not seem to model the data well, we like to bridge these findings with theory.

We expect the reader to have studied Moscadelli [25] and Dutta and Perry [12] in detail. A basic textbook for EVT in the context of insurance and finance is Embrechts et al. [13]; see also Chapter 7 in McNeil et al. [21]. Before we start our discussion, we find it worthwhile to put the record straight on EVT: papers like Diebold et al. [10] and Dutta and Perry [12] highlight weaknesses of EVT when it comes to some real applications, especially in finance. In Embrechts et al. [13] these points were already stressed very explicitly. Like any statistical method, EVT (for instance in its Peaks Over Threshold (POT) or Hill estimator variant) only promises to deliver when a very precise set of conditions is satisfied. We strongly advice the reader to revisit Embrechts et al. [13] and look carefully at the following examples: Figure 4.1.13, Figure 5.5.4 and Figure 6.4.11. Nešlehová et al. [26] yields further warnings when EVT is applied blindly to operational risk data. We very much hope that some of these underlying issues will become more clear when we progress through the paper.

The paper is organized as follows. In Section 2 we recall the definition of the g-and-h distribution and discuss some fundamental first order regular variation properties. In Section 3 we focus on second order regular variation as well as on the (slow) rate of convergence of a relevant subclass of the g-and-h excess distribution functions to the corresponding GPD. Further we analyze the impact of these results on quantile estimation via the POT method. Subadditivity properties of VaR for g-and-h distributed random variables (rvs) are discussed in Section 4. In Section 5 we highlight the link between g-and-h and the one-claim-causes-ruin phenomenon. We conclude in Section 6.

2 The g-and-h distribution

2.1 The basic definition

Throughout this paper, rvs are denoted by capital letters $X_1, X_2, \ldots$ and assumed to be defined on a common probability space $(Ω, \mathcal{F}, P)$. These rvs will typically represent one-period risk factors in a quantitative risk management context. The next definition is basic to the analysis in Dutta and Perry [12].
Definition 2.1 Let $Z \sim N(0,1)$ be a standard normal rv. A rv $X$ is said to have a g-and-h distribution with parameters $a, b, g, h \in \mathbb{R}$, if $X$ satisfies

$$X = a + b \frac{e^{gZ} - 1}{g} e^{hZ^2/2},$$

with the obvious interpretation for $g = 0$. We write $X \sim$ g-and-h, or when $X$ has distribution function (df) $F$, $F \sim$ g-and-h.

Instead of $g$ and $h$ being constants, a more flexible choice of parameters may be achieved by considering $g$ and $h$ to be polynomials including higher orders of $Z^2$. In Dutta and Perry [12], such a polynomial choice was necessary for some banks and business lines. For our paper, we restrict our attention to the basic case where $g$ and $h$ are constants. The parameters $g$ and $h$ govern the skewness and the heavy-tailedness of the distribution, respectively; see Hoaglin et al. [16].

In the case $h = 0$, equation (1) reduces to $X = a + b \frac{e^{gZ} - 1}{g}$, which is referred to as the g-distribution. The g-distribution thus corresponds to a scaled lognormal distribution. In the case $g = 0$, equation (1) is interpreted as $X = a + bZ e^{hZ^2/2}$, which is referred to as the h-distribution. The case $g = h = 0$ corresponds to the normal case. The linear transformation parameters $a$ and $b$ are of minor importance for our purposes. Unless otherwise stated we restrict our attention to the g-and-h distribution with parameters $a = 0$ and $b = 1$. Furthermore we assume $g, h > 0$. Parameters of the g-and-h distributions used in Dutta and Perry [12] to model operational risk (at enterprise level) are within the following ranges: $g \in [1.79, 2.30]$ and $h \in [0.10, 0.35]$.

Remark: Since the function $k(x) = \frac{e^{gx} - 1}{g} e^{hx^2/2}$ for $h > 0$ is strictly increasing, the df $F$ of a g-and-h rv $X$ can be written as

$$F(x) = \Phi(k^{-1}(x)),$$

where $\Phi$ denotes the standard normal df. This representation immediately yields an easy procedure to calculate quantiles and hence the Value-at-Risk of a g-and-h rv $X$,

$$\text{VaR}_\alpha(X) = F^{-1}(\alpha) = k(\Phi^{-1}(\alpha)), \quad 0 < \alpha < 1.$$

In the next section we derive some properties of the g-and-h distribution which are important for understanding its estimation properties of high quantiles.
2.2 Tail properties and regular variation

In questions on high quantile estimation, the statistical properties of the estimators used very much depend on the tail behavior of the underlying model. The g-and-h distribution is very flexible in that respect. There are numerous graphical techniques for revealing tail behavior of dfs. We restrict our attention to mean excess plots (me-plots) and log-log density plots. In Figure 1 we show a me-plot for a g-and-h distribution with parameter values typical in the context of operational risk. Besides the thick line corresponding to the theoretical mean excess function, we plot 12 empirical mean excess functions based on $n = 10^5$ simulated g-and-h data. The upward sloping behavior of the me-plots indicates heavy-tailedness as typically present in the class of subexponential dfs $S$ (see Embrechts et al. [13], Figure 6.2.4), linear behavior corresponding to Pareto (power) tails. In the latter case, the resulting log-log-density plot shows a downward sloping linear behavior; see Figure 2 for a typical example. Figure 1 also highlights the well-known problem when interpreting me-plots, i.e. a very high variability of the extreme observations made visible through the simulated me-plots from the same underlying model. Both figures give insight into the asymptotic heavy-tailedness of the g-and-h. We now make this property analytically precise.

![Mean Excess Plot](image)

Figure 1: Theoretical mean excess function (thick line) together with 12 empirical mean excess plots of the g-and-h distribution.

A standard theory for describing heavy-tailed behavior of statistical models is Karamata’s theory of regular variation. For a detailed treatment of the theory, see Bingham et al. [4]. Embrechts et al. [13] contains a summary
useful for our purposes. Recall that a measurable function $L : \mathbb{R} \to (0, \infty)$ is slowly varying (denoted $L \in SV$) if for $t > 0$:

$$\lim_{x \to \infty} \frac{L(tx)}{L(x)} = 1.$$  

A function $f$ is called regularly varying (at $\infty$) with index $\alpha \in \mathbb{R}$ if $f(x) = x^\alpha L(x)$ and is denoted by $f \in RV_\alpha$; note that $RV_0 = SV$. The following proposition is an immediate consequence of Karamata’s Theorem; see Embrechts et al. [13], Theorem A3.6. It provides an easy tool for checking regular variation. In the context of EVT, the result is known as von Mises condition for the Fréchet df; see Embrechts et al. [13], Corollary 3.3.8. Throughout we denote $F = 1 - F$.

**Proposition 2.1** Let $F$ be an absolutely continuous df with density $f$ satisfying

$$\lim_{x \to \infty} \frac{xf(x)}{F(x)} = \alpha > 0,$$

then $F \in RV_{-\alpha}$.

Note the slight abuse of notation, in the sense that we should restrict $RV$ to non-negative rvs. Through tail-equivalence (see Embrechts et al. [13], Definition 3.3.3) we can easily bypass this issue.

We proceed by showing that the g-and-h distribution is indeed regularly varying (at $\infty$) with index $-1/h$ (still assuming $h > 0$). Assume $X \sim g$-and-h, then

$$F(x) = \Phi(k^{-1}(x)), \quad f(x) = \frac{\varphi(k^{-1}(x))}{k'(k^{-1}(x))},$$
where $\varphi$ denotes the density of a standard normal rv. Using $u(1-\Phi(u))/\varphi(u) \to 1$, as $u \to \infty$, we have that
\[
\lim_{x \to \infty} \frac{xf(x)}{F(x)} = \lim_{x \to \infty} \frac{x\varphi(k^{-1}(x))}{(1-\Phi(k^{-1}(x)))k'(k^{-1}(x))}
\]
\[
= \lim_{u \to \infty} \frac{k(u)\varphi(u)}{(1-\Phi(u))k'(u)}
\]
\[
= \lim_{u \to \infty} \frac{\varphi(u)(e^{gu} - 1)}{(1-\Phi(u))(ge^{gu} + hu(e^{gu} - 1))}
\]
\[
= \frac{1}{h}
\]
and hence by Proposition 2.1 $F \in RV_{-1/h}$.

**Remark:** In a similar way, one shows that also the $h$-distribution ($h > 0$) is regularly varying with the same index. This was already mentioned in Morgenthaler and Tukey [24]. The $g$-distribution ($g > 0$) however is—as a scaled lognormal distribution—subexponential but not regularly varying. At this point the reader is advised to have a look at Section 1.3.2 and Appendix A 3.2 in Embrechts et al. [13], or Section 5 later in the paper.

In summary, we have the following result.

**Theorem 2.1**

Suppose $F \sim g$-and-$h$ with $g,h > 0$, then $F \in RV_{-1/h}$. For $h = 0$ and $g > 0$, we have $F \in S \backslash RV$, where $S$ denotes the class of subexponential dfs.

Hence, if $X \sim g$-and-$h$ with $h > 0$ we have by definition of regular variation
\[
F(x) = x^{-1/h}L(x)
\]
for some slowly varying function $L$. A key message from our paper is that the precise behavior of $L$ may profoundly affect the statistical properties of EVT-based high quantile estimations. This point was very clearly stressed in Embrechts et al. [13]; see Figure 4.1.13 and Example 4.1.12. End-users in risk management and financial applications seem largely to have missed out on this message. We will show how absolutely crucial this point is. The quality of high quantile estimation for power tail data very much depends on the second order behavior of the underlying (mostly unknown) slowly varying function $L$; for further insight on this, see Degen and Embrechts [9].

Below we derive an explicit asymptotic formula for the slowly varying function $L$ in the case of the g-and-h distribution. For $g,h > 0$ we have
\[
L(x) = F(x)x^{1/h} = (1 - \Phi(k^{-1}(x)))x^{1/h},
\]
and hence
\[
L(k(x)) = (1 - \Phi(x))(k(x))^{1/h} = (1 - \Phi(x)) \left( \frac{e^{g x} - 1}{g} \right)^{1/h} e^{x^2/2} = \frac{1}{\sqrt{2\pi}x} \left( \frac{e^{g x} - 1}{g} \right)^{1/h} \left( 1 + O \left( \frac{1}{x^2} \right) \right),
\]
leading to
\[
L(x) = \frac{1}{\sqrt{2\pi} g^{1/h}} \left( \frac{e^{g k^{-1}(x)} - 1}{k^{-1}(x)} \right)^{1/h} \left( 1 + O \left( \frac{1}{(k^{-1}(x))^2} \right) \right), \quad x \to \infty.
\]

In order to find an asymptotic estimate for \(k^{-1}\), define
\[
\tilde{k}(x) = \frac{1}{g} e^{h^2 x + g x} \sim k(x), \quad x \to \infty,
\]
with inverse function
\[
\tilde{k}^{-1}(x) = -\frac{g}{h} + \frac{1}{h} \sqrt{g^2 + 2h \log(gx)}, \quad x > 0.
\]

Here and throughout the paper, \(f(x) \sim g(x), \ x \to a\) means that \(\lim_{x \to a} \frac{f(x)}{g(x)} = 1\).

Note that \(\tilde{k}^{-1}(x) \sim k^{-1}(x)\) for \(x \to \infty\). Altogether we obtain:

**Theorem 2.2**

Let \(F \sim g\) and \(h\) with \(g, h > 0\). Then \(\overline{F}(x) = x^{-1/h} L(x)\), with \(L \in SV\), where for \(x \to \infty\),
\[
L(x) = \frac{1}{\sqrt{2\pi} g^{1/h}} \left[ \exp \left( g \left( -\frac{g}{h} + \frac{1}{h} \sqrt{g^2 + 2h \log(gx)} \right) \right) - 1 \right]^{1/h} \left( 1 + O \left( \frac{1}{\log x} \right) \right).
\]

**Proof:** Define
\[
\tilde{L}(x) = \frac{1}{\sqrt{2\pi} g^{1/h}} \left( \frac{e^{g \tilde{k}^{-1}(x)} - 1}{\tilde{k}^{-1}(x)} \right)^{1/h}.
\]
Note that $u := k^{-1}(x)$ is a strictly increasing function for $g, h > 0$. Hence,

$$\frac{L(x)}{\tilde{L}(x)} = \frac{\sqrt{2\pi} g^{1/h} \tilde{k}^{-1}(x) (1 - \Phi(k^{-1}(x))) x^{1/h}}{(e^{g k^{-1}(x)} - 1)^{1/h}}$$

$$= \frac{\sqrt{2\pi} g^{1/h} \tilde{k}^{-1}(k(u))(1 - \Phi(u))(k(u))^{1/h}}{(e^{g \tilde{k}^{-1}(k(u))} - 1)^{1/h}}$$

$$= \left(\frac{e^{g u} - 1}{e^{g \tilde{k}^{-1}(k(u))} - 1}\right)^{1/h} \tilde{k}^{-1}(k(u))(1 - \Phi(u)) \varphi(u)$$

$$= 1 + O\left(\frac{1}{u^2}\right)$$

$$= 1 + O\left(\frac{1}{\log x}\right), \ x \to \infty$$

which completes the proof. \qed

REMARKS:

• The slowly varying function $L$ in the above theorem is (modulo constants) essentially of the form $\exp\left(\sqrt{\log x}/\sqrt{\log x}\right)$. This will turn out to be a particularly difficult type of slowly varying function in the context of EVT.

• In this context, many authors consider $U(x) = F^{-1}(1 - 1/x)$ instead of $\tilde{F}$; see Section 3.2. This would make some proofs easier, but from a pedagogical point of view not always more intuitive. \qed

In the next section we will study the second order behavior of $L$ more carefully. For this we will first make a link to EVT and discuss how the properties of $L$ may influence the statistical estimation of high quantiles based on EVT.

3 Second order regular variation

3.1 The Pickands-Balkema-de Haan Theorem

We assume the reader to be familiar with univariate EVT. The notation used in this section is taken from Embrechts et al. [13]. For a g-and-h rv $X$ (with $g, h > 0$) it was shown in the previous section that $F \in \text{MDA}(H_\xi)$, i.e. belongs to the maximum domain of attraction of an extreme value distribution with index $\xi = h > 0$. The Pickands-Balkema-de Haan Theorem, Theorem 3.4.13(b) in Embrechts et al. [13], implies that for $F \in \text{MDA}(H_\xi)$, $\xi \in \mathbb{R}$,
there exists a positive measurable function $\beta(\cdot)$, such that

$$\lim_{u \uparrow x_0} \sup_{x \in (0, x_0 - u)} |F_u(x) - G_{\xi, \beta(u)}(x)| = 0.$$ 

We denote the upper support point of $F$ by $x_0$. In the case of a g-and-h distribution, $x_0 = \infty$. By the above theorem, the excess df $F_u$, defined by $F_u(x) = P(X - u \leq x | X > u)$, is well approximated by the df of a GPD, $G_{\xi, \beta(u)}$, for high threshold values $u$. This first-order convergence result stands at the heart of EVT and its numerous applications. For practical purposes however second-order properties of $F$ are of considerable importance for the performance of parameter estimates or the estimation of high quantiles. We are in particular interested in the rate of convergence of $F_u$ towards $G_{\xi, \beta(u)}$, i.e. in how fast does

$$d(u) := \sup_{x \in (0, x_0 - u)} |F_u(x) - G_{\xi, \beta(u)}(x)|$$

converge to 0 for $u \to x_0$. For this, define

$$V(t) := (1 - F)^{-1}(e^{-t})$$
$$A(t) := \frac{V''(\log t)}{V'(\log t)} - \xi,$$

for some $F \in \text{MDA}(H_\xi)$. The following proposition (see Raoult and Worms [28], Corollary 1) gives insight into the behavior of the rate of convergence to 0 of $d(u)$ in cases including, for example, the g-and-h distribution with $\xi = h > 0$.

**Proposition 3.1** Let $F \in \text{MDA}(H_\xi)$ be a df which is twice differentiable and let $\xi > -1$. If the following conditions are satisfied:

i) $\lim_{t \to \infty} A(t) = 0$,

ii) $A(\cdot)$ is of constant sign near $\infty$,

iii) there exists $\rho \leq 0$ such that $|A| \in \text{RV}_\rho$,

then, for $u \to x_0$,

$$d(u) := \sup_{x \in (0, x_0 - u)} |F_u(x) - G_{\xi, V'(V^{-1}(u))}(x)| = O(A(e^{V^{-1}(u)})).$$

The parameter $\rho$ is called the second order regular variation parameter. Recall that for a g-and-h distribution $F(x) = \Phi(k^{-1}(x))$ and hence $F^{-1}(x) =$
\(k(\Phi^{-1}(1 - x))\). In this case the function \(V\) defined above is given by \(V(t) = k(\Phi^{-1}(1 - e^{-t}))\).

Moreover,

\[ V'(\log t) = \frac{k'(\nu(t))}{t\varphi(\nu(t))} \]

and

\[ V''(\log t) = \frac{k''(\nu(t)) - tk'(\nu(t))\left(\varphi(\nu(t)) + \varphi'(\nu(t))/t\varphi(\nu(t))\right)}{(t\varphi(\nu(t))^2}, \]

where \(\nu(t) := \Phi^{-1}(1 - \frac{1}{t})\). One easily checks conditions i) and ii) above. In addition, using Lemma 2 of Raoult and Worms [28], it can be shown that \(|A| \in RV_{\rho}\) with second order parameter \(\rho = 0\). By definition of \(V\) we have

\[ A\left(e^{V^{-1}(u)}\right) = \frac{V''(\log e^{V^{-1}(u)})}{V'(\log e^{V^{-1}(u)})} - h = \frac{V''(\log 1/F(u))}{V'(\log 1/F(u))} - h = \frac{k''(k^{-1}(u))F(u)}{k'(k^{-1}(u))F(u)} + \frac{k^{-1}(u)F(u)}{\varphi(k^{-1}(u))} - 1 - h. \]

**Lemma 3.1** For \(X \sim g\text{-and-}h\) with \(g,h > 0\), the following asymptotic relation holds:

\[ A\left(e^{V^{-1}(k(x))}\right) \sim \frac{g}{x}, \quad x \to \infty. \]

**Proof:** Using the expansion\(^1\)

\[ \frac{x\Phi(x)}{\varphi(x)} = 1 + O\left(\frac{1}{x^2}\right), \quad x \to \infty, \]

and the fact that

\[ \frac{k''(x)}{k'(x)} = hx + g + o(1), \quad x \to \infty, \]

we have

\[ \lim_{x \to \infty} \frac{A\left(e^{V^{-1}(k(x))}\right)}{g/x} = \frac{1}{g} \lim_{x \to \infty} \left(\frac{x\Phi(x)}{\varphi(x)} \left(\frac{k''(x)}{k'(x)} + x\right) - x(h + 1)\right) = 1. \]

\(^1\)This proof is revised as the proof in the originally published version contained a typo.
By Proposition 3.1 and since $k^{-1}(\cdot)$ is increasing (still assuming $g, h > 0$), the rate of convergence of the excess df of a g-and-h distributed rv towards the GPD $G_{\xi, \beta(u)}$ with $\xi = h$ and $\beta(u) = V'(V^{-1}(u))$ is given by

$$d(u) = O\left(\frac{1}{k^{-1}(u)}\right) = O\left(\frac{1}{\sqrt{\log u}}\right), \quad u \to \infty.$$ 

At this point we would like to stress that $d(u) = O\left(\frac{1}{\sqrt{\log u}}\right)$ does not imply that the rate of convergence is independent of the parameters $g$ and $h$. Not a detailed derivation of this fact, but rather a heuristic argument is provided by the following:

$$\frac{\log L(x)}{\log x} \sim \sqrt{\frac{g}{h^{3/2}}} \frac{1}{\sqrt{\log x}} = O\left(\frac{1}{\sqrt{\log u}}\right), \quad x \to \infty.$$ 

Clearly the value $\frac{g}{h^{3/2}}$ affects the rate of convergence of $\log L(x)/\log x$ as $x \to \infty$. For our purposes however, this is not important.

In Table 1 we have summarized the rates of convergence in the GPD approximation as a function of the underlying df. For both the exponential as well as the exact Pareto, $d(u) = 0$. For dfs like the double exponential parent, normal, Student $t$ and Weibull, convergence is at a reasonably fast rate. Already for the very popular lognormal and loggamma dfs, convergence is very slow. This situation deteriorates further for the g-and-h where the convergence is extremely slow. Note that one can always construct dfs with arbitrary slow convergence of the excess df towards the GPD; see Resnick [29], Exercise 2.4.7. This result is in a violent contrast to the rate of convergence in the Central Limit Theorem which, for finite variance rvs, is always $n^{-1/2}$.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Parameters</th>
<th>$F(x)$</th>
<th>$\rho$</th>
<th>$d(u)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential($\lambda$)</td>
<td>$\lambda &gt; 0$</td>
<td>$e^{-\lambda x}$</td>
<td>$-\infty$</td>
<td>0</td>
</tr>
<tr>
<td>Pareto($\alpha$)</td>
<td>$\alpha &gt; 0$</td>
<td>$x^{-\alpha}$</td>
<td>$-\infty$</td>
<td>0</td>
</tr>
<tr>
<td>Double exp. parent</td>
<td>$\nu &gt; 0$</td>
<td>$t\nu(x)^{1/\nu}$</td>
<td>$-2/\nu$</td>
<td>$O\left(\frac{1}{\nu}\right)$</td>
</tr>
<tr>
<td>Student $t$</td>
<td>$\mu \in \mathbb{R}, \sigma &gt; 0$</td>
<td>$\Phi(x)$</td>
<td>0</td>
<td>$O\left(\Phi\left(\frac{1}{\nu}\right)\right)$</td>
</tr>
<tr>
<td>Normal(0,1)</td>
<td></td>
<td>$\Phi(x)$</td>
<td>0</td>
<td>$O\left(\Phi\left(\frac{1}{\nu}\right)\right)$</td>
</tr>
<tr>
<td>Weibull($\tau, c$)</td>
<td>$\tau \in \mathbb{R}_+ \setminus {1}, c &gt; 0$</td>
<td>$e^{-(cx)^{r}}$</td>
<td>0</td>
<td>$O\left(\frac{1}{\log u}\right)$</td>
</tr>
<tr>
<td>Lognormal($\mu, \sigma$)</td>
<td>$\mu \in \mathbb{R}, \sigma &gt; 0$</td>
<td>$\Phi\left(\log x - \frac{\mu}{\sigma}\right)$</td>
<td>0</td>
<td>$O\left(\frac{1}{\log u}\right)$</td>
</tr>
<tr>
<td>Loggamma($\gamma, \alpha$)</td>
<td>$\alpha &gt; 0, \gamma \neq 1$</td>
<td>$\Gamma(\alpha, \gamma(x))$</td>
<td>0</td>
<td>$O\left(\frac{1}{\log u}\right)$</td>
</tr>
<tr>
<td>g-and-h</td>
<td>$g, h &gt; 0$</td>
<td>$\Phi(k^{-1}(x))$</td>
<td>0</td>
<td>$O\left(\frac{1}{\sqrt{\log u}}\right)$</td>
</tr>
</tbody>
</table>

Table 1: Rate of convergence to the GPD for different distributions, as a function of the threshold $u$. 

\[ t\nu(x) \sim c(\nu)x (1 + x^2/\nu)^{-(\nu+1)/2} \]

\[ \Gamma_{\alpha, \gamma}(x) \sim c(\alpha, \gamma)x^{-\alpha}(\log x)^{\gamma-1} \]
From a theoretical point of view this already yields a first important result: if data are well modeled by a g-and-h distribution with $g, h > 0$, then high quantile estimation for such data based on the POT method will typically converge very slowly. In the next section we will look at this issue in somewhat more detail.

It is often stated by some authors that they have “solved” the (critical) optimal choice of threshold problem in the POT or Hill method. On several occasions we have stressed that this problem has no general solution; optimality can only be obtained under some precise second order properties on the underlying slowly varying function $L$ (we concentrate on the Fréchet case). It is precisely this $L$ (let alone its second order properties) which is impossible to infer from statistical data. Hence, the choice of a reasonable threshold (we avoid using the word “optimal”) remains the Achilles heel of any high quantile estimation procedure based on EVT. For a more pedagogic and entertaining presentation of the underlying issues, see Embrechts and Nešlehová [14].

3.2 Threshold choice

There exists a huge literature on the optimal threshold selection problem in EVT; see for instance Beirlant et al. [2] for a review. Within a capital charge calculation problem, the choice of threshold $u$ above which EVT fits well the tail of the underlying df may significantly influence the value estimated. We stress the word “may”; indeed in some cases the quantile estimate is rather insensitive with respect to the choice of $u$, in other cases it is very sensitive. This stresses the fact that for the modeling of extremes, great care as to the underlying model and data properties has to be taken. The analysis below is indicative of the underlying issues and definitely warrants a much broader discussion. We have included it to warn the reader for some of the difficulties in using automatic procedures for determining so-called “optimal” tail regions for the estimation of high quantiles. We restrict our attention to g-and-h dfs and estimate quantiles using the Hill estimator. The conclusions obtained also hold for the MLE based POT method.

We assume that $X_1, X_2, \ldots, X_n$ are iid realizations from a continuous df $F$ with $F \in RV_{-1/\xi}$, i.e. $F(x) = x^{-1/\xi}L(x)$, $L \in SV$.

**Definition 3.1** The Hill estimator is defined by

$$H_{k,n} := \frac{1}{k} \sum_{j=1}^{k} \log X_{n-j+1,n} - \log X_{n-k,n} \quad (1 < k < n),$$

where $X_{1,n} \leq \ldots \leq X_{n,n}$ are the order statistics of $X_1, \ldots, X_n$. 

58
Consider the quantile function $U(x) := F^{-1}(1 - 1/x)$. Since $F \in RV_{-1/\xi}$, we have $U(x) = x^{\xi}l(x)$, for some slowly varying function $l$; see for instance Beirlant et al. [2]. If there exist $\rho \leq 0$ and a positive function $b$ with $b(x) \to 0$ for $x \to \infty$, such that for all $t \geq 0$,

$$
\log \frac{l(tx)}{l(x)} \sim b(x)k_p(t), \quad x \to \infty,
$$

with

$$
k_p(t) = \begin{cases} \frac{\rho - 1}{\rho} & \rho < 0, \\
\log t & \rho = 0,
\end{cases}
$$

then the asymptotic mean square error (AMSE) of the Hill estimator satisfies

$$
AMSE_{H_{k,n}} := (\text{ABias}_{H_{k,n}})^2 + \text{AVar}_{H_{k,n}} = \left( b\frac{(n + 1)/(k + 1)}{1 - \rho} \right)^2 + \frac{\xi^2}{k}, \tag{2}
$$

see for instance Matthys and Beirlant [19]. Applying this result to the regularly varying g-and-h df $F$ with index $1/\xi = 1/h \in (0, \infty)$, we get

$$
l(x) = \frac{1}{g(2\pi)^{h/2}} \frac{e^{\Phi^{-1}(1-1/x)} - 1}{\Phi^{-1}(1-1/x)^h}
= \frac{1}{g(2\pi)^{h/2}} \frac{e^{\Phi^{-1}(1-1/x)}}{\Phi^{-1}(1-1/x)^h} (1 - e^{-g\Phi^{-1}(1-1/x)}).
$$

Using the following approximation for the quantile function of the normal,

$$
\Phi^{-1}\left(1 - \frac{1}{x}\right) \sim \sqrt{\log \frac{x^2}{2\pi}} - \log \log \frac{x^2}{2\pi}, \quad x \to \infty,
$$
(see e.g. Dominici [11], Proposition 21) we arrive at

$$
\log \frac{l(tx)}{l(x)} = g \left( \Phi^{-1}(1 - 1/(tx)) - \Phi^{-1}(1 - 1/x) \right)
+ \frac{h}{\Phi^{-1}(1-1/(tx))} \log \frac{1 - e^{-g\Phi^{-1}(1-1/(tx))}}{1 - e^{-g\Phi^{-1}(1-1/x)}}
= \left[ g \frac{(2 \log x)^{1/2}}{2 \log x} - \frac{h}{(2 \log x)^{3/2}} + o\left( \frac{1}{(\log x)^{3/2}} \right) \right] \log t,
$$

as $x \to \infty$. Hence, in particular, $\rho = 0$ for the g-and-h distribution with $g, h > 0$. Note that, given the precise model assumptions above, formula (2) yields an obvious approach to estimating the optimal sample fraction to
calculate the Hill estimator.
In practice however, one usually does not have any information about the second order properties of the underlying df. Thus for each \( k, 1 < k < n, b(\cdot), \rho \) and \( h \) have to be estimated form the data by estimators \( \hat{b}(\cdot), \hat{\rho} \) and \( \hat{h} \), which are for example the maximum likelihood or the least squares estimators; see Beirlant et al. [2]. One then chooses \( k \) in the following way:

\[
k_{\text{opt}} = \arg\min_{k \in \mathbb{N}_+} \left[ \left( \frac{\hat{b}((n+1)/(k+1))}{1-\hat{\rho}} \right)^2 + \frac{\hat{h}^2}{k} \right].
\]

We will now apply this procedure to simulated g-and-h data. For each pair of parameter values \( g \) and \( h \) (see Table 2 below) we simulate a hundred samples of 2000 observations from a g-and-h distribution. For each of the 100 samples we compute the Hill estimator \( \hat{h}_{\text{Hill}}^{Hill} \) of \( h \) using \( k_{\text{opt}} \) number of upper order statistics of the 2000 observations of that sample. For every cell in Table 2 we thus get 100 estimates \( \left( \hat{h}_{\text{Hill},m}^{Hill}, m \right)_{1 \leq m \leq 100} \) of \( h \). To analyze the performance of the Hill estimator \( \hat{h}_{\text{Hill}}^{Hill} \) we calculate the standardized root mean square error (SRMSE), which for a single cell is given by

\[
\frac{1}{h} \sqrt{\frac{1}{100} \sum_{m=1}^{100} \left( \hat{h}_{\text{Hill},m}^{Hill} - h \right)^2}.
\]

The SRMSE of the Hill estimator \( \hat{h}_{\text{Hill}}^{Hill} \) is summarized in Table 2.

<table>
<thead>
<tr>
<th>( g \backslash h )</th>
<th>0.1</th>
<th>0.2</th>
<th>0.5</th>
<th>0.7</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>142</td>
<td>82</td>
<td>33</td>
<td>23</td>
<td>18</td>
<td>11</td>
</tr>
<tr>
<td>0.2</td>
<td>165</td>
<td>97</td>
<td>42</td>
<td>32</td>
<td>25</td>
<td>20</td>
</tr>
<tr>
<td>0.5</td>
<td>224</td>
<td>132</td>
<td>49</td>
<td>38</td>
<td>27</td>
<td>19</td>
</tr>
<tr>
<td>0.7</td>
<td>307</td>
<td>170</td>
<td>63</td>
<td>44</td>
<td>29</td>
<td>20</td>
</tr>
<tr>
<td>1</td>
<td>369</td>
<td>218</td>
<td>86</td>
<td>58</td>
<td>36</td>
<td>26</td>
</tr>
<tr>
<td>2</td>
<td>696</td>
<td>385</td>
<td>151</td>
<td>108</td>
<td>74</td>
<td>31</td>
</tr>
<tr>
<td>3</td>
<td>1097</td>
<td>613</td>
<td>243</td>
<td>163</td>
<td>115</td>
<td>54</td>
</tr>
</tbody>
</table>

Table 2: SRMSE (in %) of the Hill estimator \( \hat{h}_{\text{Hill}}^{Hill} \) of \( h \) for g-and-h data for different parameter values of \( g \) and \( h \).

From Table 2 we may deduce a characteristic pattern which essentially remains the same for other threshold selection procedures. We confirmed this by implementing the optimal threshold selection method proposed by Guillou and Hall [15] and by applying an ad-hoc selection method, using a fixed
percentage of exceedances of 5%. Further, we applied a method based on a logarithmic regression model provided by Beirlant et al. [3], where the authors try to handle the case $\varrho = 0$. They analyze slowly varying functions of the following form,

$$L(x) = C (\log(x))^\beta (1 + o(1)),$$

with $C, \beta > 0$.

If data come from a loggamma distribution, for which the slowly varying function fulfills (3), numerical calculations show rather good results when compared to a Hill estimator without bias reduction. However, for g-and-h data with $g$ and $h$ in a typical operational risk range, even given the extra knowledge about the second order parameter, the corresponding SRMSEs are in a similar range as for the other estimators.

An identical study was performed for the MLE estimates $\hat{h}^{MLE}$ of $h$, yielding very similar results to the case of $\hat{h}^{Hill}_{\text{opt}}$. Therefore, whether using Hill or MLE to estimate $h$, the key message we infer from Table 2 is that EVT-based tail index estimation leads to highly inaccurate results. Moreover, the larger the ratio $g/h$, the larger the SRMSE. In particular, for parameter values reported in Dutta and Perry [12], which are in a range around $g = 2, h = 0.2$, the SRMSE is close to 400%. The numbers reported in Table 2 are somewhat counterintuitive. Indeed in papers like McNeil and Saladin [22] and Dutta and Perry [12] it is stated that heavier tailed models require higher thresholds and likewise a larger sample size to achieve a similar error bound. Table 2 on the other hand indicates, that for fixed $g$, the SRMSE decreases for increasingly heavier tails.

The poor performance of EVT-based tail index estimation, especially for parameter values with a large ratio $g/h$, e.g. $g = 2$ and $h = 0.2$, is further confirmed by a Hill plot; see Figure 3 (right panel). On the other hand, we expect a “good” Hill plot for $g/h$ small, e.g. $g = 0.1$ and $h = 1$, which is confirmed by Figure 3 (left panel).

In the left panel, the Hill plot is rather flat over a large range of threshold values yielding an accurate estimate of the true value $h = 1$. In the right panel however, the Hill plot is absolutely misleading. Though being temptingly flat, an estimation of the shape parameter $\xi = h$ based on the Hill plot would in that case lead to a completely wrong estimate of $\hat{h}^{Hill} \approx 0.7$, whereas the true value is $h = 0.2$. One can easily come up with finite mean g-and-h examples (i.e. $h < 1$) leading to infinite mean EVT estimates ($\hat{h}^{Hill} > 1$). Such an example can be constructed by choosing the skewness parameter sufficiently high. We exemplify this issue in Figure 4, where we present a Hill plot for $n = 10^6$ realizations of a g-and-h rv with parameter values $g = 4$ and $h = 0.2$ (finite mean). Again the Hill plot shows a relatively flat behavior, suggesting a value of $\hat{h}^{Hill} \approx 1.2$, indicating an infinite mean.
In summary, given that data are well modeled by a g-and-h distribution where \(g/h\) is large, as is the case with the data reported by Dutta and Perry [12], an EVT based estimation of the tail index \(h\) unavoidably leads to highly inaccurate estimates. Consequently in such cases high quantile estimation using standard EVT methodology becomes highly sensitive to specific numerical estimation procedures. We emphasize this further in the next section.

Figure 3: Hill plot for \(g = 0.1, h = 1\) and \(n = 10^6\) (left panel) and \(g = 2, h = 0.2\) and \(n = 10^6\) (right panel).

Figure 4: Hill plot for \(g = 4, h = 0.2\) and \(n = 10^6\).
3.3 Quantile estimation

To confirm our findings of the previous section we performed a quantile estimation study along the lines of McNeil and Saladin [22], [23]. Instead of applying sophisticated optimal threshold selection procedures we likewise concentrated on an ad-hoc method by taking into account only a certain percentage of the highest data points; see McNeil and Saladin [22] for details. We generated g-and-h data and calculated the POT estimator of the 99% and the 99.9% quantiles for different values of \( g \) and \( h \). We compared our results \((h = 0.2 \text{ and } h = 1)\) with the findings of McNeil and Saladin [22] to conclude that the performance of the POT estimator for the g-and-h distribution is much worse—in terms of high standardized bias and SRMSE—than for any of the distributions used in that paper.

From a methodological point of view, Makarov [18] is also relevant in this respect. In that paper, the author shows that uniform relative quantile convergence in the Pickands-Balkema-de Haan Theorem necessarily needs a slowly varying function \( L \) which is asymptotically constant. Clearly, \( L \) in the g-and-h case is far from being constant; a more detailed discussion on this is to be found in Degen and Embrechts [9].

All the results shown so far point to the fact that the slowly varying function \( L \) for the g-and-h distribution for \( g,h > 0 \) renders high quantile estimation based on EVT methodology difficult: for g-and-h type data, all EVT based procedures show extremely slow convergence and hence for small to medium size data samples, these estimators may be highly inaccurate.

In order to better understand the relative merits of EVT and g-and-h, we now turn to estimating quantiles in cases where EVT is known to do well and see how g-and-h based estimation compares. In the Tables 3 and 4 we give the estimated quantiles for two empirical data sets; the daily S&P data from 1960 to 1993 and the Danish fire insurance data from 1980 to 1990, as discussed in Embrechts et al. [13].

<table>
<thead>
<tr>
<th>Quantile</th>
<th>Empirical</th>
<th>POT</th>
<th>g-and-h</th>
</tr>
</thead>
<tbody>
<tr>
<td>90%</td>
<td>0.93</td>
<td>1.10</td>
<td>0.92</td>
</tr>
<tr>
<td>95%</td>
<td>1.30</td>
<td>1.34</td>
<td>1.29</td>
</tr>
<tr>
<td>99%</td>
<td>2.14</td>
<td>2.13</td>
<td>2.23</td>
</tr>
<tr>
<td>99.9%</td>
<td>4.10</td>
<td>4.30</td>
<td>3.98</td>
</tr>
</tbody>
</table>

Table 3: Quantile estimation of S&P-data with \( n = 8414 \) data points. In the case of the POT-Method we fix the threshold at \( u = 1.45 \).
We compare empirical quantile estimation, POT quantile estimation and the g-and-h method. For the latter we fit a g-and-h distribution to the data, where we allow for location and scale parameters to be different from \( a = 0, b = 1 \). The parameters \( a, b, g, h \) are estimated using Tukey’s percentiles. Using the language of Hoaglin et al. [16], we take approximately \( \log_2(n) \) letter values, where \( n \) is the number of available data points, with the full spread (FS) for the S&P data and with the upper half spread (UHS) for the Danish fire insurance data; see for instance Hoaglin et al. [16] and Dutta and Perry [12], Appendix C. The quantile is then given by \( a + bk(\Phi^{-1}(\alpha)) \).

<table>
<thead>
<tr>
<th></th>
<th>Empirical</th>
<th>POT</th>
<th>g-and-h</th>
</tr>
</thead>
<tbody>
<tr>
<td>90%</td>
<td>5.54</td>
<td>5.64</td>
<td>5.72</td>
</tr>
<tr>
<td>95%</td>
<td>9.97</td>
<td>9.30</td>
<td>9.43</td>
</tr>
<tr>
<td>99%</td>
<td>26.04</td>
<td>27.51</td>
<td>27.32</td>
</tr>
<tr>
<td>99.9%</td>
<td>131.55</td>
<td>121.17</td>
<td>101.51</td>
</tr>
</tbody>
</table>

Table 4: Quantile estimation of Danish fire insurance data with \( n = 2167 \) data points. In the case of the POT-Method we fix the threshold to \( u = 5 \).

We conclude that for the 95% and 99% levels all methods yield rather similar results, whereas for very high quantiles, the results differ substantially. Of course for the S&P data a more dynamic modeling, as for instance given in McNeil and Frey [20] including careful backtesting, would be useful. In the case of the Danish data backtesting to find the better fitting procedure is not really available. Once more, these results are in no way conclusive. We have included them to highlight some issues and hopefully encourage further research.

As a final comparison we test the three quantile estimation methods mentioned above by means of two selected examples in line with McNeil et al. [21], Section 7.2.5. We will distinguish between a “soft” and a “hard” problem. With regards to the “soft” problem, we generate 1000 realizations of a standard normal rv and estimate the 95%-quantile, whereas for the “hard” problem we generate 1000 realizations of a \( t_3 \) rv and estimate the 99.9%-quantile. So in the “soft” problem we estimate a quantile well within the range of light-tailed data. For the “hard” problem we estimate a quantile at the edge of heavy-tailed data. In both problems our estimations are based on the empirical, the POT and the g-and-h method by means of the procedure mentioned above. In the case of the g-and-h method the full spread is used to estimate the parameters \( a, b, g, h \). In Figure 5 we plot the SRMSE as a function of the chosen threshold of the GPD.
In the soft case, where the quantile is estimated at a moderate level, g-and-h fits well and its SRMSE is smaller than the SRMSE obtained by the POT method. This is not surprising, as the normal distribution perfectly fits into the g-and-h framework. In the hard case, the g-and-h method as well as the POT method clearly outperform the empirical estimator.

4 Subadditivity of VaR

As stated above, we can give an explicit formula for the Value-at-Risk in the case of a g-and-h rv:

\[ \text{VaR}_\alpha(X) = k(\Phi^{-1}(\alpha)), \quad 0 < \alpha < 1, \]

with

\[ k(x) = \frac{e^{g x} - 1}{g} e^{h x^2 / 2}. \]

In Dutta and Perry [12] the authors state: “We have not mathematically verified the subadditivity property for g-and-h, but in all cases we have observed empirically that enterprise level capital is less than or equal to the sum of the capitals from business lines or event types”. Of course, a mathematical discussion of subadditivity would involve multivariate modeling; we will return to this issue in a future publication.

In order to statistically investigate the subadditivity property for the g-and-h distribution, we perform a simulation study. Let \( X_1, X_2 \) be iid g-and-h rvs with parameters \( g = 2.4 \) and \( h = 0.2 \). We estimate (by simulation of \( n = 10^7 \) realizations) the diversification benefit \( \delta_{g,h}(\alpha) = \text{VaR}_\alpha(X_1) + \)
VaR_\alpha(X_2) - VaR_\alpha(X_1 + X_2), where of course \delta_{g,h}(\alpha) will be non-negative if and only if subadditivity occurs. Our results are displayed in Figure 6. For the above realistic choice of parameters, superadditivity holds for \alpha smaller than a certain level \tilde{\alpha} \approx 99.4\%. The fact that subadditivity, i.e. \text{VaR}_\alpha(X_1 + X_2) \leq \text{VaR}_\alpha(X_1) + \text{VaR}_\alpha(X_2), holds for \alpha sufficiently large is well known; see Proposition 4.1 below. That superadditivity enters for typical operational risk parameters at levels below some \tilde{\alpha} may be somewhat surprising. The latter may be important in the discussion around the scaling of risk measures. Indeed, risk managers realize that estimating \text{VaR}_\alpha at a level \alpha \geq 99\%, say, is statistically difficult. It has been suggested to estimate \text{VaR}_\alpha deeper down in the data, \alpha = 90\%, say, and then scale up to 99.9\%. The change from super- to subadditivity over this range should be of concern.

Note that one can even construct finite-mean examples (choosing the skewness parameter \( g \) large enough) for levels \tilde{\alpha} = 99.9\% and higher, such that subadditivity of Value-at-Risk fails for all \alpha < \tilde{\alpha}. This should be viewed in contrast to the following proposition by Danielsson et al. [8]. See also that paper for a definition of bivariate regular variation.

**Proposition 4.1** Suppose that the non-degenerate vector \((X_1, X_2)\) is regularly varying with extreme value index \( \xi < 1 \). Then \text{VaR}_\alpha is subadditive for \alpha sufficiently large. \( \square \)

Figure 6 exemplifies the subadditivity of VaR only in the very upper tail region. The reader should thus be warned that Proposition 4.1 is an asymp-
totic statement and does not guarantee subadditivity for a broad range of high quantiles. Furthermore, note that for $\xi = h > 1$ subadditivity typically fails. The reason being that for $h > 1$ one deals with infinite mean models; see Nešlehová et al. [26] for more details on this.

For practitioners it will be of prime importance to know for which choices of $g$ and $h$ values one can expect subadditivity. As shown in Figure 6, this depends on the level $\alpha$. We restrict ourself to the $\alpha$-values 99% and 99.9%, relevant for practice. Assume that the operational risk data of two business lines of a bank are well modeled by iid $g$-and-$h$ rvs with parameter values $g \in [1.85, 2.30]$, $h \in [0.15, 0.35]$. Note that these values roughly correspond to the parameters estimated by Dutta and Perry [12] at enterprise level. It would be of interest to figure out if aggregation at business line level leads to diversification in the sense of subadditivity of VaR. For this purpose we consider two iid $g$-and-$h$ rvs with $g$ and $h$ values within the above mentioned ranges. In Figure 7 we display a contour plot of $\delta_{g,h}(\alpha)$ for a fixed $\alpha$, together with the rectangle containing the parameter values of interest.

The number attached to each contour line gives the value of $\delta_{g,h}(\alpha)$ and the lines indicate levels of equal magnitude of diversification benefit. The 0-value corresponds to models where VaR$_\alpha$ is additive, VaR$_\alpha(X_1 + X_2) = \text{VaR}_\alpha(X_1) + \text{VaR}_\alpha(X_2)$. The positive values (bottom left hand corner) correspond to models yielding subadditivity. The top right hand corner, corresponding to negative values for $\delta_{g,h}(\alpha)$, leads to superadditivity for the corresponding parameter values. Note that for $\alpha = 99.9\%$, the entire parameter rectangle lies within the region of subadditivity; see right panel of Figure 7. It is though important to realize that with only relatively small changes
in the underlying $g$ and $h$ parameters, one may end up in the superadditivity region. The situation becomes more dramatic at lower quantiles. The left panel of Figure 7 corresponds to $\alpha = 99\%$ (which is still relatively high!). There the superadditivity region extends and a substantial fraction of our parameter rectangle lies therein.

The above statements were made under the iid assumption. In the example below we allow for dependence. For this we link the marginal $g$-and-$h$ distributions with the same parameters as in Figure 6 by a Gauss-copula; see McNeil et al. [21], p. 191. In Figure 8 we plot $\delta_{g,h}(\alpha)$ for three different correlation parameters $\rho = 0, 0.5$ and 0.7. This figure should be compared with Figure 6.

![Figure 8](image)

Figure 8: Plot of $\delta_{g,h}(\alpha)$ as a function of $\alpha$ for $g = 2.4, h = 0.2, n = 10^7$ and Gauss-copula with correlation $\rho = 0, 0.5$ and 0.7. Note that $\rho = 0$ corresponds to the independence case in Figure 6.

It appears that in a range below 95%, $|\delta_{g,h}(\alpha)|$ becomes smaller when the correlation parameter increases. This is not surprising because VaR is additive under comonotonic dependence, i.e. for risks with maximal correlation; see McNeil et al. [21], Theorem 5.25. As a consequence $\delta_{g,h}(\alpha)$ would be tending to 0 for $\rho \to 1$. The effect of dependence can clearly be seen for large values of $\alpha$. Based on our simulation study, it appears that with increasing correlation $\rho$, the range of superadditivity extends to even higher values of $\alpha$. Hence the stronger the dependence the higher the level $\alpha$ has to be in order to achieve a subadditive model. Formulated differently, for strong dependence ($\rho$ large), most levels $\alpha$ chosen in practice will lie within the range of superadditivity. We have worked out these results also for other dependence structures, like
the t- and the Gumbel-copula. For these cases we also elaborated contour plots as in Figure 7. The results do not differ significantly from Figure 7 and thus we refrain from displaying these plots here.

The situation in any LDA is of course in general much more complicated than in our simple example above. Practitioners and risk managers should therefore interpret our statements rather from a methodological and pedagogical point of view. It seems that diversification of operational risk can go the wrong way due to the skewness and heavy-tailedness of this type of data.

5 The one-claim-causes-ruin paradigm

In several publications on operational risk it is stated that often relatively few claims cause the major part of the total operational risk loss. Papers highlighting this phenomenon in an operational risk context are Nešlehová et al. [26] and Böcker and Klüppelberg [5]. Though these publications contain the relevant results, for matter of completeness we reformulate the main conclusions in terms of the g-and-h distribution. We concentrate on the iid case, changes incorporating dependence between the different loss rvs along the lines of Böcker and Klüppelberg [5] can easily be made.

Let \( X_1, \ldots, X_d \) be iid g-and-h rvs and \( S_d = \sum_{i=1}^{d} X_i \) the total loss. Recall that for \( g > 0, h \geq 0 \) the g-and-h distribution is subexponential, i.e.

\[
P[S_d > x] \sim P[\max_{1 \leq i \leq d} X_i > x], \quad x \to \infty.
\]

The above relation expresses the fact that for subexponential distributions, the tail distribution of the total loss \( S_d \) is determined by the tail distribution of the maximum loss. We are in the so-called “one-claim-causes-ruin” regime; see Embrechts et al. [13], Section 8.3, or Asmussen [1].

More generally, consider \((X_t)_{t \geq 0}\) a sequence of iid g-and-h rvs, independent of a counting process \((N_t)_{t \geq 0}\) and \( S_t = \sum_{i=1}^{N_t} X_i \). Hence we have

\[
G_t(x) := P[S_t \leq x] = \sum_{n=0}^{\infty} P[N_t = n] F^{\ast n}(x),
\]

where \( F^{\ast n} \) denotes the n-th convolution of \( F \). Furthermore, by Theorem 1.3.9 of Embrechts et al. [13], if there exists \( \epsilon > 0 \) such that

\[
\sum_{n=0}^{\infty} (1 + \epsilon)^n P[N_t = n] < \infty, \quad (4)
\]

then the tail df of \( S_t \) satisfies

\[
P[S_t > x] \sim E[N_t] F(x), \quad x \to \infty.
\]
Note that condition (4) is for instance satisfied in the Poisson, Binomial and Negative Binomial case. The above representation implies
\[ G_t^{-1}(\alpha) \sim F^{-1} \left( \frac{1 - \alpha}{E[N_t]} \right), \quad \alpha \to 1, \]
and hence for \( F \sim g\text{-and-}h \) with \( g > 0, h \geq 0 \),
\[ \text{VaR}_\alpha(S_t) \sim k \left( \Phi^{-1} \left( 1 - \frac{1 - \alpha}{E[N_t]} \right) \right), \quad \alpha \to 1. \]
Though these results yield explicit analytic approximations for \( \text{VaR}_\alpha \), \( \alpha \) large, their practical importance is questionable.

6 Conclusion

In this paper we have highlighted some of the modeling issues for g-and-h severity distributions within an LDA for operational risk. There seems to be a discrepancy in practice between results which strongly favor EVT methodology (Moscadelli [25]) and g-and-h methodology (Dutta and Perry [12]). Our main results are as follows. First, the g-and-h class of dfs yields an overall very slow rate of convergence in applications using EVT based techniques. This is mainly due to the second order behavior of the slowly varying function underlying the g-and-h for \( h > 0 \). As a consequence, setting an optimal threshold for an EVT based POT approach becomes very difficult and hence quantile (risk capital) estimates may become unreliable. Second, the issue of sub- or superadditivity of g-and-h based VaR estimation very much depends on the parameter values \( g \) and \( h \). It is shown that, both for iid as well as for dependent data, small changes in the underlying parameters may lead VaR to switch regime (super to sub or vice versa). Finally, since the class of g-and-h distributions is subexponential (for \( g > 0, h \geq 0 \)), this class of dfs also yields the one-claim-causes-ruin phenomenon.

Several of the above results (observations) were based on simulation studies. We do however believe that the messages delivered in our paper may already have considerable relevance for practical application of the LDA for operational risk. In future publications we shall come back to some of these issues in a more analytic form. In particular, we are working on QRM relevant properties of multivariate g-and-h models.

7 Acknowledgments

The authors would like to thank Valérie Chavez-Demoulin, Kabir Dutta, Johanna Nešlehová and Jason Perry for several discussions related to the quan-
quantitative modeling of operational risk and for the numerous useful comments which led to a substantial improvement of the paper. They also would like to thank two anonymous referees for their comments.

Bibliography


EVT-Based Estimation of Risk Capital and Convergence of High Quantiles.

*Advances in Applied Probability, 40(3)*, 696-715.
EVT-Based Estimation of Risk Capital
and Convergence of High Quantiles

Matthias Degen Paul Embrechts
Department of Mathematics, ETH Zurich, 8092 Zurich, Switzerland

Abstract

We discuss some issues regarding the accuracy of a quantile-based estimation of risk capital. In this context, Extreme Value Theory (EVT) emerges naturally. The paper sheds some further light on the ongoing discussion concerning the use of a semi-parametric approach like EVT and the use of specific parametric models such as the g-and-h. In particular, the paper discusses problems and pitfalls evolving from such parametric models when using EVT and highlights the importance of the underlying second-order tail behavior.

Keywords: Extreme Value Theory; g-and-h Distribution; Operational Risk; Peaks Over Threshold; Penultimate Approximation; Second-Order Regular Variation; Slow Variation; Value-at-Risk

1 Introduction

Over recent years, we have witnessed a growing interest in the theory and applications of EVT. For instance, textbooks such as de Haan and Ferreira [16], Balkema and Embrechts [2] or Resnick [26] discuss new methodological developments together with specific applications to such fields as environmental statistics, telecommunication, insurance and finance. In applying EVT one still runs into theoretical issues which need further study. In this paper we present such a problem, discuss some partial solutions and indicate where more research is needed.

Our starting point is a problem from the regulatory framework (so-called
Basel II) of banking and finance. The reader interested in this more applied background is referred to Chapter 1 in McNeil et al. [23]. For the purposes of the present paper, we concentrate on the quantitative modeling of Operational Risk (OR). The latter is defined as the risk of losses resulting from inadequate or failed internal processes, people and systems or from external events. This definition includes legal risk, but excludes strategic and reputational risk. OR can be viewed as complementary to the widely studied risk classes Market Risk (MR) and Credit Risk (CR). Without going into a full description of OR data, it suffices to know that risk capital for OR has to be calculated (statistically estimated) using the concept of Value-at-Risk (VaR) at the extreme level of 99.9% and for losses aggregated over a 1-year period; see Section 3 for a definition of VaR. Because of this, early on EVT was recognized as a canonical tool (see Moscadelli [24]) but also criticized for the possible instability of its output; see Dutta and Perry [11]. Degen et al. [10] highlights some of the main problems in applying EVT to OR data and moreover compares and contrasts EVT with the alternative g-and-h approach as championed by Dutta and Perry [11]. One of the main conclusions of these earlier analyses of OR data is that in standard loss severity models used in OR practice, it is the asymptotic behavior of the tail-distribution (in particular the associated slowly varying function in a Pareto-type model) that may cause problems. In the present paper we highlight these issues in more detail. In particular we show that for a simple distribution such as the g-and-h, EVT-based estimation for ranges relevant for practice may give answers which differ significantly from the asymptotics.

In Section 2 we give basic notation, model assumptions and review some standard facts from EVT. Section 3 discusses rates of convergence for quantiles and penultimate approximations. Section 4 looks more carefully into the second-order tail behavior under specific model assumptions for EVT applications to OR. We show that the slowly varying function underlying the g-and-h model has second-order properties which may give rise to misleading conclusions when such data are analyzed using standard EVT methodology. Section 5 concludes and gives hints for further research.

2 Univariate EVT - background and notation

We assume the reader to be familiar with univariate EVT, as presented for instance in Embrechts et al. [12]. Below we review some basic facts. Throughout we assume that our loss data \(X\) are modeled by a continuous df \(F(x) = P(X \leq x)\) and standardly write \(F = 1 - F\).

We use the notation MDA(\(H_\xi\)) for Maximum Domain of Attraction of a generalized extreme value df \(H_\xi\); see Embrechts et al. [12] for details.
Throughout the paper we restrict our attention to the case $\xi > 0$. Then $F \in \text{MDA}(H_\xi)$ is equivalent to $\overline{F} \in RV_{-1/\xi}$; see for instance Embrechts et al. [12], Theorem 3.3.7. In terms of its tail quantile function $U(x) = F^{-}(1 - 1/x)$, this is equivalent to $U \in RV_\xi$. We standardly use the notation $\overline{F} \in RV_{-1/\xi}$ for $\overline{F}(x) = x^{-1/\xi}L(x)$, where $L$ is some slowly varying function in the sense of Karamata, i.e. for all $x > 0$,

$$\lim_{t \to \infty} \frac{L(tx)}{L(t)} = 1. \tag{1}$$

We write $L_F$ and $L_U$ for the slowly varying functions associated with $F$ and $U$ respectively. $F$ and $U$ are always assumed to be continuous and sufficiently smooth where needed.

Often it turns out to be more convenient to work on a log-log scale for $\overline{F}$ and $U$. As such, for $\overline{F} \in RV_{-1/\xi}$ with density $f$ we will write

$$\overline{F}(x) = e^{-\Psi(s)} = e^{-s/\xi + \Psi_{L_F}(s)}, \quad s = \log x,$$

where $\Psi$ and $\Psi_{L_F}$ denote the log-log transform of $\overline{F}$ and $L_F$ respectively; see Appendix A for details. Similarly we define $U(t) = e^{\phi(r)} = e^{\xi r + \phi_{L_U}(r)}$, with $r = \log t$. Note that $L$ varies slowly if $\Psi_L$ vanishes at infinity.

For $F \in \text{MDA}(H_\xi)$ the result below yields a natural approximation for the excess df $F_u$, defined by $F_u(x) = P(X - u \leq x | X > u)$, in terms of the generalized Pareto distribution (GPD) function $G_{\xi,\beta}(x) = 1 - (1 + \xi x/\beta)^{-1/\xi}$ where $1 + \xi x/\beta > 0$.

**Proposition 2.1 (Pickands-Balkema-de Haan)** For $\xi \in \mathbb{R}$ the following statements are equivalent:

(i) $F \in \text{MDA}(H_\xi)$,

(ii) There exists a strictly positive measurable function $\beta$ such that

$$\lim_{u \to x_F} \sup_{x \in (0,x_F-u)} |F_u(x) - G_{\xi,\beta(u)}(x)| = 0. \tag{2}$$

The defining property of the scale function $\beta$ is given by the asymptotic relationship $\beta(u) \sim u\xi$, as $u \to \infty$. Moreover, if (2) holds in the case $\xi > 0$ for some function $\beta > 0$, then it also holds for $\beta(u) = \xi u$; see de Haan and Ferreira [16], Theorem 1.2.5.

**Remark:** A key property (and hence data assumption) underlying EVT estimation based on Proposition 2.1 is that of stability. The class of GPDs is stable under successive conditioning over increasing thresholds, and this
with the same parameter \( \xi \); see Embrechts et al. [12], Theorem 3.4.13 (c) and Remark 5), p.166. All EVT-based estimation procedures are based on this tail-stability property. It may typically hold for environmental data; whether or not it is tenable for socio-economic data is an issue of current debate. □

As a consequence of Proposition 2.1, for \( F \in \text{MDA}(H_\xi) \) a natural approximation of the tail \( \overline{F}(x) \), for \( x \geq u \) and \( u \) sufficiently large, is provided by \( \overline{F}(u)\overline{G}_{\xi,\beta(u)}(x-u) \). For \( \xi > 0 \) we may without loss of generality take \( \beta(u) = \xi u \) and hence consider the approximation

\[
\overline{K}(u,x) := \begin{cases} 
\overline{F}(x), & x < u, \\
\overline{F}(u)\overline{G}_{\xi,\xi u}(x-u) = c(u)x^{-1/\xi}, & x \geq u,
\end{cases}
\]

where \( c(u) = u^{1/\xi}\overline{F}(u) = L_F(u) \).

For practical purposes, in order to appreciate the goodness of the tail approximation (3) for \( x \geq u \), it is important to quantify the rate at which \( \overline{K} \) converges to \( \overline{F} \), i.e. to determine the rate of convergence in (2)—or equivalently in (1).

In an OR context, rates of convergence for loss severity distributions are discussed in Degen et al. [10]. In the latter paper, the authors focus on the so-called g-and-h distribution recently proposed by Dutta and Perry [11] to model operational risk losses. Recall that a random variable (rv) \( X \) is said to have a g-and-h distribution, if \( X \) satisfies

\[
X = a + b \frac{e^{gZ} - 1}{g} e^{hZ^2/2}, \quad a, g, h \in \mathbb{R}, b > 0,
\]

where \( Z \sim N(0,1) \). The linear transformation parameters \( a \) and \( b \) are of minor importance for our analysis. Unless stated otherwise we therefore restrict our attention to the standard case \( a = 0 \) and \( b = 1 \). This class of dfs was introduced in Tukey [28] and studied from a statistical point of view for instance in Hoaglin et al. [18] and Martinez and Iglewicz [22].

Degen et al. [10] show that for \( g, h > 0 \) (typical for OR data) the g-and-h distribution tail is regularly varying with index \(-1/h\), i.e. \( \overline{F}(x) = x^{-1/h}L_F(x) \). The corresponding slowly varying \( L_F \) is, modulo constants, asymptotically of the form \( \exp(\sqrt{\log x}/\sqrt{\log x}) \) (see also (7) below) and turns out to be a particularly difficult function to handle from a statistical data analysis point of view. Indeed, below we show that the behavior of \( L_F \) in ranges relevant for practice is very different from its limit behavior, which may cause EVT-based estimation methods of \( h \) to be very inaccurate over such ranges.
3 Convergence of quantiles

In quantitative risk management, often risk capital charges are based on estimates of high quantiles (Value-at-Risk) of underlying profit-and-loss distributions; see Chapters 1 and 2 in McNeil et al. [23] for details.

**Definition 3.1** The generalized inverse of a df $F$,

$$F^-(q) = \inf \{ x \in \mathbb{R} : F(x) \geq q \}, \quad 0 < q < 1,$$

is called the quantile function of $F$. In a financial risk management context, for given $q$, $F^-(q)$ is referred to as the $q100\%$ Value-at-Risk, denoted by $\text{VaR}_q(F)$.

The tail approximation (3) suggests estimating VaR of an (unknown) underlying df $F \in MDA(H_\xi)$, i.e. estimating $F^-(q)$ for some level $q \in (0,1)$, typically close to 1, by its approximating counterpart $K^-(u,q)$. By the properties of inverse functions (see for instance Resnick [25], Proposition 0.1) the quantiles $K^-(u,q)$ converge pointwise to the true value $F^-(q)$ as $u \to \infty$, but the convergence need not be uniform.

**Definition 3.2** We say that for some df $F \in MDA(H_\xi)$, uniform relative quantile (URQ) convergence holds if

$$\lim_{u \to \infty} \sup_{q \in (0,1)} \left| \frac{K^-(u,q)}{F^-(q)} - 1 \right| = 0;$$

see also Makarov [21], who gives a necessary condition for URQ convergence.

According to Makarov [21], failure of uniform relative quantile (URQ) convergence may lead to unrealistic risk capital estimates. Another possible reason for the discrepancy between an EVT-based methodology and certain parametric approaches for high quantile estimation is provided by the fact that the excess dfs of many loss rvs used in practice (e.g. log-normal, log-gamma, g-and-h) show very slow rates of convergence to the GPD. At the model level, this is due to the second-order behavior of the underlying slowly varying functions. Consequently, tail index estimation and also quantile (i.e. risk capital) estimation using EVT-based methodology improperly may yield inaccurate results; see Degen et al. [10]. Below we combine both lines of reasoning and embed URQ convergence in the theory of second-order regular variation.
3.1 Rates of convergence for quantiles

Assume $F \in RV_{-1/\xi}$ for some $\xi > 0$, or equivalently, $U \in RV_{\xi}$, in terms of its tail quantile function $U(t) = F^-(1 - 1/t)$. In order to assess the goodness of the tail approximation (3), the rate at which $K^-(u, q)$ tends to the true quantile $F^-(q)$ as $u \to \infty$ has to be specified. We are thus interested in the rate at which

$$\frac{U(tx)}{U(t)} - x^\xi$$

(4)
tends to 0 as $t \to \infty$.

In the sequel we focus on the rate of convergence in (4) within the framework of the theory of second-order regular variation, as presented for instance in de Haan and Stadtmüller [17] or de Haan and Ferreira [16], Section 2.3 and Appendix B.

$U$ is said to be of second-order regular variation, written $U \in 2RV_{\xi, \rho}$, $\xi > 0, \rho \leq 0$, if for some positive or negative function $A$ with $\lim_{t \to \infty} A(t) = 0$,

$$\lim_{t \to \infty} \frac{U(tx)}{U(t)} - x^\xi A(t) =: H_{\xi, \rho}(x), \quad x > 0,$$

(5)
exists, for some $H_{\xi, \rho}$ which is non-trivial. In that case we necessarily have $H_{\xi, \rho}(x) = x^{\xi/p} - 1$ for $x > 0$. Note that for $\xi > 0$, $U \in 2RV_{\xi, \rho}$ is equivalent to $F \in 2RV_{-1/\xi, \rho/\xi}$.

In order to find sufficient conditions for (5) to hold, consider the following intuitive reasoning in terms of log-log transforms. Let $U \in RV_{\xi}$ and assume that $U'$ exists. This ensures that we may write $U$ as

$$U(t) = e^{\varphi'(\log t)}, \quad \varphi(r) = \int_1^r \frac{\varepsilon(y)}{y} dy + c,$$

where $\varepsilon(y) = y U'(y)/U(y)$ and $c = \log U(1)$; see Appendix A. The log-log plot of $U$ depicts the graph of $\varphi$. With $s = \log x$ and $r = \log t$ we may write (4) in terms of log-log transforms as

$$\frac{U(tx)}{x^\xi U(t)} - 1 = e^{\varphi(r+s) - \varphi(r) - \xi s} - 1 \sim \varphi(r + s) - \varphi(r) - \xi s, \quad r = \log t \to \infty.$$

The expression $\varphi(r + s) - \varphi(r) - \xi s$ may be approximated by $(\varphi'(r) - \xi)s$ and therefore, the convergence rate of $\frac{U(tx)}{U(t)} - x^\xi$ to 0 is of the same order as the rate at which $\varphi'_{L_1}(\log t) = \varphi'(\log t) - \xi = \frac{U'(t)}{U(t)} - \xi$ tends to 0 as $t \to \infty$. This motivates the next result.

**Theorem 3.1**

Suppose $U(t) = e^{\varphi'(\log t)}$ is twice differentiable and $A(t) = \varphi'(\log t) - \xi$. If for some $\xi > 0$ and some $\rho \leq 0$
\[ i) \lim_{t \to \infty} \phi'(t) = \xi, \]

\[ ii) \phi'(t) - \xi \text{ is of constant sign near infinity, and} \]

\[ iii) \lim_{t \to \infty} \frac{\phi''(t)}{\phi'(t) - \xi} = \rho, \]

then, for \( x > 0 \),

\[ \lim_{t \to \infty} \frac{U(tx) - x^\xi}{A(t)} = H_{\xi, \rho}(x), \]

with

\[
H_{\xi, \rho}(x) = \begin{cases} 
  x^\xi \frac{x^\rho - 1}{\rho}, & \rho < 0 \\
  x^\xi \log x, & \rho = 0.
\end{cases}
\]

**Proof:** Recall the definition of \( A(t) = \phi'(\log t) - \xi \) and observe that

\[
\lim_{t \to \infty} \frac{U(tx)}{U(t)} - x^\xi = H_{\xi, \rho}(x) \iff \lim_{r \to \infty} \frac{\phi(r + s) - \phi(r) - \xi s}{\phi'(r) - \xi} = e^{-\xi s} H_{\xi, \rho}(e^s),
\]

where \( s = \log x \) and \( r = \log t \).

By assumption \( iii) \) we have \( \lim_{t \to \infty} t A'(t)/A(t) = \rho \) which, by the Representation Theorem for regularly varying functions (see for instance Bingham et al. [4], Theorem 1.3.1), guarantees \( |A| \in RV_\rho \). In particular \( A(tx)/A(t) \to x^\rho \) locally uniformly as \( t \to \infty \), and we obtain for every \( s \in \mathbb{R} \) as \( r \to \infty \)

\[
\frac{\phi(r + s) - \phi(r) - \xi s}{\phi'(r) - \xi} = \int_0^s (\phi'(r + u) - \xi) \, du \quad \to \quad \int_0^s e^{\rho u} \, du = \begin{cases} 
  e^{\rho s} - 1, & \rho < 0 \\
  s, & \rho = 0.
\end{cases}
\]

This finishes the proof. \( \square \)

Note that in smooth cases, the rate of convergence is thus uniquely determined by the underlying slowly varying function \( L_U \), i.e. \( A(t) = \phi'_{L_U}(\log t) = t L_U'(t)/L_U(t) \).

**Remarks (URQ Convergence):**

i) Define \( \Psi(s) = -\log \bar{F}(e^s) \) and \( \Psi^u(s) = -\log \bar{K}(u, e^s) \) for \( s \in [\log u, \infty) \) (and \( \Psi \equiv \Psi^u \) on \( (-\infty, \log u) \) by definition of \( \bar{K} \)). With this notation, URQ convergence holds for \( F \) if and only if \( \|\Psi^-(s) - (\Psi^u)^-(s)\|_\infty \to 0 \) on \( (-\infty, 0) \) as \( u \to \infty \).

It is not difficult to see that URQ convergence holds for \( F \) if and only if the log-log transform \( \Psi_{L_F} \) of \( L_F \) satisfies \( \Psi_{L_F}(s) \to c_F \) as \( s \to \infty \), for some constant \( c_F \in (0, \infty) \).
For $U$ satisfying the assumptions of Theorem 3.1 with $\rho < 0$ we have that $\varphi'(\log x) - \xi \sim Cx^\rho, x \to \infty$, for some constant $C > 0$, and hence $\varphi_{L_U}(t) = \varphi(t) - t\xi$ (or equivalently $\Psi_{L_U}(s) = s/\xi - \Psi(s)$) tends to a finite limit. In particular, a strictly negative second-order parameter $\rho$ implies URQ convergence for $U$ (or equivalently $F$). In the case $\rho = 0$, $L_U$ belongs to the so-called de Haan class $\Pi$, which is a subclass of the class of the slowly varying functions; see de Haan and Ferreira, Theorem B.3.6. For $L_U \in \Pi$, $\lim_{x \to \infty} L_U(x) =: L_U(\infty)$ exists, but $L_U(\infty)$ may be infinite; see de Haan and Ferreira [16], Corollary B.2.13.

For $F \in RV_{-1/\xi}$, the above theorem gives the rate at which $K^-$ tends to the true quantile $F^-$, or equivalently, the rate at which the corresponding properly scaled excess quantiles $F_{u}^-$ converge to $G_{\xi,1}^-$. For $A(t) = tU''(t)/U(t) - \xi$ satisfying the conditions of Theorem 3.1, we obtain for every $x > 1$

$$\lim_{u \to \infty} \frac{F_{u}^-(1-1/x)}{A\left(1/F(u)\right)} = \frac{1}{\xi} H_{\xi,\rho}(x),$$

where—for $F$ sufficiently smooth—the corresponding convergence rate satisfies

$$A\left(1/F(u)\right) = \frac{F(u)}{uf(u)} - \xi \sim \frac{\xi^2 uL_F(u)}{L_F(u)}, \quad u \to \infty.$$ 

Whereas for distributions with $\rho < 0$ this convergence is rather fast, it may be very slow in the case $\rho = 0$. The log-gamma for instance is well-known for its slow convergence properties with $A(1/F(u)) = O\left(1/\log u\right)$. The situation for the g-and-h is even worse, with $A(1/F(u)) = O\left(1/\sqrt{\log u}\right)$.

Summing up, in terms of the second-order parameter $\rho$, the tail $F$ (or equivalently $U$) will be "better" behaved if $\rho < 0$ than in the case $\rho = 0$. In the former case the convergence in (5) is not "too slow", in the sense that the rate function $|A|$ is regularly varying with index $\rho < 0$, i.e. if the influence of the nuisance term $L_U$, or equivalently $L_F$, vanishes fast enough in terms of $L_U$ (or $L_F$) behaving "nicely" and tending to some positive constant (thus implying URQ convergence for $F$). In the case $\rho = 0$, the rate function $|A|$ is slowly varying and hence the excess quantiles typically converge very slowly. However, in certain cases the slow convergence rate may be improved using the concept of penultimate approximation.
3.2 Penultimate approximations

So far we have been concerned with the ultimate approximation, i.e. for every $x > 0$ and for large values of $t$,

$$U(tx) \approx x^\xi U(t).$$

One method to improve the rate of convergence in the above approximation goes back to the seminal work of Fisher and Tippett [13]. More recent accounts on this are found for instance in Cohen [7], Gomes and de Haan [14] or Worms [29].

The basic idea behind penultimate approximations is to vary the shape parameter $\xi$ as a function of the threshold $t$, i.e. to consider

$$U(tx) \approx x^{\xi(t)} U(t),$$

with $\xi(t) \to \xi$ for $t \to \infty$, where one hopes to improve the convergence rate by choosing $\xi(.)$ in an appropriate way.

In order to illustrate how to find a feasible function $\xi(.)$, consider the following. In terms of log-log transforms, the ultimate approximations uses that, for large values of $r$,

$$\varphi(r + s) \approx \varphi(r) + \xi s,$$

where $r = \log t$ and $s = \log x$, i.e. $\varphi$ is approximated linearly by a straight line with slope $\xi$. A better approximation might be achieved if, for large values of $r$, $\varphi$ is still approximated linearly, but now by its tangent line in the respective points $r = \log t$, i.e. by a straight line with slope $\varphi'$, leading to

$$\varphi(r + s) \approx \varphi(r) + \varphi'(r)s.$$

Thus, a reasonable choice of a threshold-dependent shape parameter is $\xi(t) = \varphi' \log t = t U'(t)/U(t)$.

At this point it is worth noting that there is a close connection between the theory of penultimate approximations and the theory of second-order regular variation. Suppose $U$ satisfies the conditions given in Theorem 3.1. In that case we obtain, for large values of $t$,

$$U(tx) \approx U(t) \left( x^\xi + A(t) H_{\xi,\rho}(x) \right),$$

or equivalently, for large values of $r = \log t$,

$$\varphi(r + s) \approx \varphi(r) + \xi s + (\varphi'(r) - \xi) e^{-\xi s} H_{\xi,\rho}(e^s),$$

which, by definition of $H_{\xi,\rho}$ for $\rho = 0$, is the same as $\varphi(r + s) \approx \varphi(r) + \varphi'(r)s$. As a consequence, in the case $\rho = 0$, the theory of second-order regular
variation yields (asymptotically) the same approximation as the penultimate theory. Note however, that we may easily construct examples where the second-order theory does not apply but the penultimate does.

**Example 3.1** Consider $U(t) = e^{\varphi(\log t)}$, with $\varphi(r) = \xi r + \varphi_{Lu}(r) = \xi r + \sin \sqrt{r}$. In this case, $\varphi_{Lu}$ changes sign infinitely often as we move out, and thus Theorem 3.1 does not apply. A penultimate approximation $\varphi(r + s) = \varphi(r) + \varphi'(r)s$ may nevertheless be considered. In particular, the approximation error $e_r(s) = \varphi(r + s) - \varphi(r) - \varphi'(r)s$ is of the order $O(s^2)$ for $s \to 0$, whereas the error in the ultimate case $\varphi(r + s) - \varphi(r) - \xi s$ is of the order $O(s)$.

The above example shows that, although no second-order improvement exists in that particular case, the penultimate approximation may locally still lead to an improvement over the ultimate approximation. In addition, we shall show below that in the case $\rho = 0$, the rate of convergence in the penultimate approximation may indeed (asymptotically) improve compared to the ultimate approximation.

Intuitively it is clear that the rate at which $\frac{U(tx)}{U(t)} - x \xi(t)$ tends to 0 is of the same order as the rate at which the linear approximations (i.e. the tangent lines in points $t$) approach the straight line with slope $\xi$. Hence, the speed at which the penultimate convergence rate $a(.)$ tends to 0 is of the same order as the speed at which the slope $\varphi'$ tends to its ultimate value $\xi$, which is measured by $\varphi''$. So as a candidate for the penultimate convergence rate we choose $a(t) = \varphi''(\log t) = t \xi'(t)$. Condition iii) of Theorem 3.1, $\lim_{t \to \infty} \frac{\varphi''(t)}{\varphi'(t)} = \rho$ for some $\rho \leq 0$, implies that for this choice of $a$, the convergence rate may asymptotically only be improved in cases where $\rho = 0$ and we have $a(t) = o(A(t))$ for $t \to \infty$. Indeed, under the conditions discussed above and under the additional condition

$$\frac{\varphi''(x)}{\varphi'(x)} \to 0 \text{ or equivalently } \frac{t \xi''(t)}{\xi'(t)} \to -1, \quad x = \log t \to \infty,$$

the (improved) penultimate rate of convergence may be given as follows.

**Theorem 3.2**

Let $U$ satisfy the conditions of Theorem 3.1 with $\rho = 0$ and define $\xi(t) = tU'(t)/U(t) = A(t) + \xi$. If $\lim_{t \to \infty} \frac{t \xi''(t)}{\xi'(t)} = -1$, and $\xi'$ is of constant sign near infinity, then, for all $x > 0$,

$$\lim_{t \to \infty} \frac{U(tx) - x \xi(t)}{t A'(t)} = J_\xi(x),$$

where

$$J_\xi(x) = \frac{1}{2} x^2 \log^2 x.$$
Proof: Following the lines of the proof of Theorem 3.1, note first that for 
$s = \log x$ and $r = \log t$,

$$
\lim_{t \to \infty} \frac{U(tx) - x^\xi(t)}{tA'(t)} = J_\xi(x) \iff \lim_{r \to \infty} \frac{\varphi(r + s) - \varphi(r) - \varphi'(r)s}{\varphi''(r)} = \frac{1}{2}s^2.
$$

For every $s \in \mathbb{R}$ we have

$$
\frac{\varphi(r + s) - \varphi(r) - \varphi'(r)s}{\varphi''(r)} = \int_0^s \int_0^y \frac{\varphi''(r + z)}{\varphi''(r)} dz dy.
$$

Moreover, by assumption $t\xi''(t)/\xi'(t) \to -1$ (or equivalently $\varphi'''(x)/\varphi''(x) \to 0$), which implies that $|\varphi''| \in RV_0$, which guarantees $\varphi''(r + s)/\varphi''(r) \to 1$ locally uniformly for $r \to \infty$, and hence for $r \to \infty$,

$$
\int_0^s \int_0^y \frac{\varphi''(r + z)}{\varphi''(r)} dz dy \to \int_0^s \int_0^y 1 dz dy = \frac{1}{2}s^2,
$$

which finishes the proof.

Remarks:

i) We want to stress again that is important to distinguish between the second-order theory which is an asymptotic theory (i.e. is concerned with the limit behavior) while the penultimate theory is a local theory. Only for the special case $\rho = 0$ and under certain additional conditions do the second-order theory and the penultimate theory yield asymptotically the same approximation.

ii) From the proof of Theorem 3.2 it is clear that, although the original rate is improved asymptotically, i.e. $tA'(t) = o(A(t))$ for $t \to \infty$, the improvement is not spectacular as the new rate $tA'(t)$ is again slowly varying. Nevertheless, locally the improvements may be considerable, as we show in the next paragraph.

3.3 Implications for practice

To illustrate the above results, we compare the (theoretical) relative approximation error $e(u) := \left| \frac{K(u,1-1/x)}{F-(1-1/2)} - 1 \right|$ for the 99.9% quantiles (confidence level required under Basel II) as a function of the threshold $u$, in the ultimate and penultimate approximation for certain frequently OR-loss severity models. Besides the well-known Burr and log-gamma distribution (see for

87
instance Embrechts et al. [12], Chapter 1), we consider the g-and-h, the modified Champernowne and the generalized Beta distribution.

The modified Champernowne was recently proposed by Buch-Larssen et al. [5] in an OR context. Its tail df is given by

$$
F(x) = \frac{(M + c)^\alpha - c^\alpha}{(x + c)^\alpha + (M + c)^\alpha - 2c^\alpha}, \quad \alpha, M, x > 0, c \geq 0,
$$
hence $F \in RV^{-\alpha}$. The density of a generalized Beta (GB2) distribution is given by

$$
f(x) \propto \frac{x^{ap-1}}{(1 + (\bar{x})^a)^{p+q}}, \quad a, b, p, q, x > 0,
$$
so that $f \in RV^{-aq-1}$; see for instance Dutta and Perry [11] for its use in OR.

In Figure 1 we show the approximation errors $e(.)$ in % for the g-and-h and the log-gamma distribution (left panel) and for the Burr, the modified Champernowne and the GB2 distribution (right panel). To enable a qualitative comparison across different distributions we take the thresholds as quantile levels $q$ and scale the horizontal axis by the 99.9% quantile.

In order to compare quantitatively and to check how the GPD-approximation for high quantiles performs, we fix a relative error level of $e^{99.9\%}(u) = 5\%$, say, and compute the excess probabilities over the corresponding $u$ levels. In practice, in order to estimate a 99.9% quantile using the POT-method, a certain amount $N_u$ of data exceeding this threshold $u$ is needed (we take $u$ such that $e^{99.9\%}(u) = 5\%$), so as to come up with reasonable estimates.
For illustrative purposes we choose $N_u$ to be 100. From this we may infer the number $n$ of data points we would expect to have to generate, in order to have $N_u = 100$ excesses over the threshold $u$ for a given relative error $e^{99.9\%}(u) = 5\%$; see Table 1 for the results.

Table 1: Expected number of data points $n$ needed to get $N_u = 100$ exceedances over a fixed threshold $u$ for the distributions of Figure 1 with the respectively specified parameter values.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>$n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>g-and-h ($u_1$)</td>
<td>585</td>
</tr>
<tr>
<td>g-and-h ($p_1$)</td>
<td>221</td>
</tr>
<tr>
<td>Log-gamma ($u_2$)</td>
<td>15.6</td>
</tr>
<tr>
<td>Log-gamma ($p_2$)</td>
<td>8.87</td>
</tr>
<tr>
<td>Burr ($u_3$)</td>
<td>2.55</td>
</tr>
<tr>
<td>Burr ($p_3$)</td>
<td>2.13</td>
</tr>
<tr>
<td>Modified Champernowne ($u_4$)</td>
<td>2.41</td>
</tr>
<tr>
<td>Modified Champernowne ($p_4$)</td>
<td>2.15</td>
</tr>
<tr>
<td>Generalized Beta ($u_5$)</td>
<td>1.28</td>
</tr>
<tr>
<td>Generalized Beta ($p_5$)</td>
<td>1.68</td>
</tr>
</tbody>
</table>

From the five loss models considered above, the Burr, the modified Champernowne and the GB2 satisfy $\rho < 0$, whereas the log-gamma and the g-and-h distribution have second-order parameter $\rho = 0$. In the latter case, $n$ increases vastly to about 47780 and 91540 for the log-gamma and the g-and-h, respectively, reflecting the slow convergence properties of these distributions.

Using penultimate approximations above a reasonably high threshold, the number of data needed to achieve the same level of accuracy may be lowered significantly in all cases but for the GB2. Note however that in the GB2 case the convergence is rather fast and the approximation error $e(u)$ is negligible for high threshold values $u$ anyway. Though the theory of penultimate approximations seems to be very promising towards the (theoretical) improvement of high-quantile estimation accuracy, its practical relevance may be limited since the slope $\varphi'$ has to be estimated from the data. More statistical work would be highly useful here.

**Remark:** The above examples are of course idealized since we assume the underlying distributions $F$ to be known and hence also the corresponding tail index $\xi$ of the GPD $G_\xi$. In practice we will encounter an additional error source due to estimation errors of the parameters.

From an applied risk management point of view our analysis admits the following conclusion. The closer the second-order parameter $\rho$ is to zero, the
slower EVT-based estimation techniques converge. Thus, if data seem to be modeled well by a df $F$ with second-order parameter $\rho = 0$, the amount of data needed in order to come up with reasonable results may be prohibitively large. In an operational risk context this is also one of the main reasons why banks have to combine internal loss-data with external data and expert opinion, leading to further important statistical issues; see Lambrigger et al. [20]. In addition, for distributions with "bad" second-order behavior (i.e. $\rho = 0$) the exact shape of the associated slowly varying function $L_F$ may make things worse and give rise to misleading conclusions about the underlying data. As we already saw, a prime example of this situation, important for practice, is provided by the g-and-h slowly varying function $(g, h > 0)$, which is analyzed in more detail in the next section.

4 A slowly varying function with a sting in its tail

According to Dutta and Perry [11], the typical parameter values of a g-and-h df $F$ used to model OR loss severities are in a range around $g \in (1.7, 2.3)$ and $h \in (0.1, 0.4)$. In the sequel we therefore always assume $g$ and $h$ to be strictly positive, hence $F \in RV_{-1/h}$. Adopting the notation of Degen et al. [10], we consider $X \sim g$-and-$h$ with df $F(x) = \Phi(k^{-1}(x))$, where $k(x) = \frac{e^{gx} - 1}{g} e^{hx^2/2}$ and $\Phi$ denotes the standard normal df. In Figure 2 we plot $F$ on a log-log scale for OR-typical parameter values. As $F \in RV_{-1/h}$, a straight line with slope $-1/h$ is to be expected as we move out in the right tail.

![Figure 2: Log-log plot of the tail of a g-and-h distribution with $g = 2$ and $h = 0.2$ and straight line (dotted) as a reference line with slope $-1.2$.](image-url)
According to Figure 2, the log-log plot is almost linear over a large region of practical interest (for quantile levels of 90% up to 99.99%). Therefore, over such ranges, the g-and-h tail behavior is close to an exact Pareto and thus the influence of the slowly varying part seems to be minimal. Figure 2 is fallacious however, as the slope of a linear approximation \( g(.) \) implies a tail index of around 0.8 whereas the theoretical tail index was chosen to be \( h = 0.2 \). The consequences for statistical estimation of this may be better understood by the concept of what we will call local heavy-tailedness.

### 4.1 Local heavy-tailedness

Consider \( \overline{F}(x) = x^{-1/\xi}L_F(x) \in RV_{-1/\xi}, \xi > 0 \), where \( F \) has a density \( f \). Recall that we then may write \( \overline{F} \) as

\[
\overline{F}(x) = e^{-\Psi(s)} = e^{-s/\xi + \Psi_{L_F}(s)}, \quad s = \log x,
\]

with \( \Psi \) and \( \Psi_{L_F} \) denoting the log-log transforms of \( \overline{F} \) and \( L_F \) respectively; see Appendix A. As a graph of \( -\Psi \) corresponds to the log-log plot of \( \overline{F} \), the total amount of heavy-tailedness at a point \( x \) is measured by \( \Psi'(\log x) \). It consists of the ultimate heavy-tailedness of the model (tail index \( \xi \)) and an additional source of local heavy-tailedness due to \( L_F \). The local heavy-tailedness is measured by the slope of \( \log L_F \) which is given by

\[
\Psi'_{L_F}(s) = \frac{1}{\xi} - \Psi'(s) = \frac{1}{\xi} - \frac{xf(x)}{\overline{F}(x)}, \quad s = \log x.
\]

Clearly, depending on the underlying model \( F \), the amount of local heavy-tailedness due to the shape of its associated slowly varying function \( L_F \) may be significant. As indicated above, this is particularly evident in the case of the g-and-h slowly varying function, for which it turns out that the behavior of \( \overline{F} \) (or \( L_F \)) in ranges relevant for risk management applications is very different from its ultimate asymptotic behavior. Neglecting this issue for data of (or close to) g-and-h type may lead to problems when applying standard EVT methodology. Whereas this issue is well known from a theoretical point of view within the EVT community (see for instance Resnick [25], Exercise 2.4.7), it is somewhat surprising that it manifests itself so clearly in a fairly straightforward, and increasingly used, parametric model such as the g-and-h.

To get a feeling for the behavior of the g-and-h slowly varying function \( L_F \) we show in Figure 3 a log-log plot of \( L_F \) (left panel) with corresponding slope (right panel) for \( g \) and \( h \) parameter values typical for OR (the function \( k \) is inverted numerically using a Newton algorithm with an error tolerance level of \( 10^{-13} \)).
Figure 3: Log-log plot of the g-and-h slowly varying function \( L_F \) with \( g = 2 \) and \( h = 0.2 \) (left panel) with corresponding slope (right panel) as defined in (6).

The almost linear log-log plot suggests that the function \( L_F \) behaves approximately like a power function \( x^{1/\eta} \) for some \( \eta > 0 \) and with \( 1/\eta \) given in the slope plot. The asymptotic behavior of the log-log transform of \( \bar{F} \) is given by

\[
-\Psi(s) = -\frac{1}{h}s + \frac{\sqrt{2}g}{h^{3/2}}\sqrt{s} - \frac{1}{2}\log s - c + O\left(\frac{1}{\sqrt{s}}\right), \quad s = \log x \to \infty, \quad (7)
\]

with \( c = \frac{1}{2}\log \frac{4x}{h} + \frac{g^2}{h^2} + \frac{\log g}{h} \); see Appendix A, equation (11).

The deviation from exact power-law decay (i.e. the deviation from linearity in a log-log plot) is due to the slowly varying part. The amount of local heavy-tailedness is measured by the slope \( \Psi'_{L_F} \) of \( \log L_F \), which behaves like

\[
\Psi'_{L_F}(s) = \frac{g}{\sqrt{2}h^{3/2}}\sqrt{s} - \frac{1}{2s} + O\left(\frac{1}{s^{3/2}}\right), \quad s = \log x \to \infty. \quad (8)
\]

Therefore, the rate at which the influence of \( L_F \) vanishes, i.e. the rate at which \( \Psi'_{L_F} \) tends to 0, is of the order \( O\left(1/\sqrt{\log x}\right) \), \( x \to \infty \).

Equation (8) gives us a first impression of how unpleasant the g-and-h slowly varying function might be. Indeed, its slow convergence properties together with its (power-like) behavior in ranges relevant for OR practice may lead to serious difficulties in the statistical estimation of extremes based on EVT, given that data follow such a model. At this point we wish to stress that this is not a weak point of EVT but should rather be viewed as a warning against "gormless guessing" of a parametric model, and this in the words of Richard Smith and Jonathan Tawn; see Embrechts et al. [12], Preface p.VII.

In the next paragraph we study EVT estimation within a g-and-h model, i.e. for g-and-h generated data.
4.2 Tail index estimation for g-and-h data

In Degen et al. [10], the problem of the estimation of the tail index $h$ within a g-and-h model was pointed out using the Hill estimator. To emphasize that EVT-based tail index estimation for g-and-h data may be problematic whatever method one uses, we shall work below with the increasingly popular POT-MLE method for which the statistical basis was laid in the fundamental papers of Davison [8], Smith [27] and Davison and Smith [9]. For further background reading, see Embrechts et al. [12]. We additionally implemented other tail-index estimators such as the moment estimator, a bias-reduced MLE or an estimator based on an exponential regression model as discussed in Beirlant et al. [3]. As might be expected from the discussion in the previous paragraphs, all these estimators led to similar conclusions and we therefore refrain from showing those results.

In Figure 4 we simulated $n = 10^4$ observations from a g-and-h model $(g = 2, h = 0.2)$ and plot the POT-MLE of the tail index $h$ as a function of the number of exceedances used, together with the 95% confidence bounds.

![MLE-POT estimates of tail index](image)

**Figure 4:** Theoretical tail index (straight solid line) and MLE-POT estimates of $\xi = h$ for g-and-h data with $g = 2$ and $h = 0.2$, based on $n = 10^4$ observations.

At first glance, Figure 4 suggests that the df of the underlying data follows nearly perfectly an exact Pareto law, or at least converges rather fast towards an exact Pareto law. Indeed, the deviation from exact power law behavior seems to vanish quickly as the MLE behaves stably and is flat over a large region of thresholds (as was of course to be expected from Figure 2). As a consequence one would accept an estimate of the tail index of around $\hat{h} \approx 0.85$
EVT-based estimation thus significantly overestimates the true parameter $h = 0.2$, suggesting a rather heavy-tailed model ($\hat{\alpha} = 1/h \approx 1.2$, i.e. infinite variance), whereas the data are simulated from a model with $1/h = 5$, which has finite moments up to order five. The reason behind this lies in the behavior of the underlying slowly varying function over moderate ranges. More precisely, for g-and-h data ($g = 2$, $h = 0.2$) with for practice reasonable sample sizes, a number of $k = 500$, say, largest order statistics is taken into account for the estimation of $h$. For this realistic choice of parameter values of $g$ and $h$, the values of such order statistics will typically range from around 10 to around $10^3$. Over such ranges the slope of $\Psi_{LF}$ is nearly constant and hence $L_F$ grows approximately like a power $x^{1/\eta}$ with (averaged) $1/\eta \approx 3.8$; see Figure 5. Therefore, in ranges relevant for practical applications, the "regularly varying"-like slowly varying part $L_F$ adds a significant amount of local heavy-tailedness to the model. Together with the asymptotic tail decay of $F \in RV_{-5}$, the local power-like growth of $L_F$ leads to a regular variation index of around $-1.2$, i.e. to a "local" tail index $h_{loc}$ of $F$ of around 0.83.

![Figure 5: Slope of the log-log plot of the g-and-h slowly varying function ($g = 2, h = 0.2$) in a range relevant for operational risk.](image)

**Remark:** The notion of regular variation is an asymptotic concept, hence our emphasis on the distinction between local (i.e. finite ranges relevant for practice) and asymptotic tail behavior. Due to its slow convergence properties, the (asymptotic) tail index for the g-and-h is significantly overestimated by standard (i.e. ultimate) EVT estimation methods. On the other hand, after a penultimate correction accounting for the local heavy-tailedness due to the behavior of the slowly varying part, the behavior of $F$ over ranges relevant for practice is captured well by EVT methods (the MLE in Figure 4 is rather stable and flat). This is due to the extremely slow decay of $\Psi'_{LF}$. 
which is close to constant over large ranges. Over such ranges, the g-and-h therefore behaves approximately like an exact Pareto, for which EVT-based estimation methods are known to perform well.

Furthermore, as can be seen from (8), the accuracy of tail index estimation depends crucially on the ratio $g/h^{3/2}$ of the g-and-h parameters. For parameter values of $g$ and $h$ relevant for operational risk applications (see Dutta and Perry [11], p.43), $g/h^{3/2}$ is relatively large and thus the slope of $\Psi_{L_F}$—the share of additional local heavy-tailedness—is large.

For cases with $g/h^{3/2}$ small, for instance $g = 0.1$ and $h = 0.5$, $\Psi_{L_F}'$ is less than 0.2 over a large range, which includes a major part of largest order statistics of reasonable sample sizes, and thus the influence of $L_F$ is minor, i.e. the difference between local and asymptotic behavior is marginal. Consequently the accuracy of EVT-based tail index estimation increases considerably and the estimate captures the structure of the data correctly; see for example Figure 6.

![MLE-POT estimates of tail index](image)

Figure 6: Theoretical tail index (straight solid line) and MLE-POT estimates of $\xi = h$ for g-and-h data with $g = 0.1$ and $h = 0.5$, based on $n = 10^4$ observations.

### 4.3 Risk capital estimation for operational risk

Based on the 2004 loss data collection exercise (LDCE) for operational risk data, Dutta and Perry [11] note that in the estimation of OR risk capital (1-year 99.9% VaR) there seems to be a serious discrepancy between an
EVT-based approach and the g-and-h approach. They find that EVT often yields unreasonably high risk capital estimates whereas the g-and-h approach would not. Fitting a g-and-h model to the 2004 LDCE data (aggregated at enterprise level), Dutta and Perry [11] find $\xi = h \in (0.1, 0.4)$ for different banks. By contrast, applying EVT methodology to the 2002 LDCE data (aggregated at business line level), Moscadelli [24] comes up with infinite-mean models (i.e. $\xi > 1$) for 6 out of 8 business lines. Based on this, resulting risk capital estimates may be expected to differ widely for the two approaches. See also Jobst [19] on this issue; in that paper it is claimed that for typical OR data the g-and-h approach is superior only from a confidence level of 99.99% onwards.

As we do not possess the data underlying these analyses, our findings below, based on simulated data, may be academic in nature. However, even though EVT-based techniques may, for simulated g-and-h data, lead to completely wrong estimates of the (asymptotic) tail index $h$ (see Figure 4), this does not need to carry over to estimates of relevant risk measures such as VaR or return periods. Indeed, based on $n = 10^4$ simulated g-and-h data ($g = 2$, $h = 0.2$), the MLE-POT estimates of the 99.9% quantile seem to be rather accurate (though the 95% confidence band is quite broad); see Figure 7.

![MLE-POT estimates of 99.9% quantile](image)

Figure 7: Theoretical 99.9% g-and-h quantile (straight solid line) and MLE-POT estimates for $g = 2$ and $h = 0.2$, based on $n = 10^4$ observations.

The seeming incompatibility of the two Figures 4 and 7 can be explained by taking into account the effect of local heavy-tailedness caused by the slowly varying part. For $g = 2$ and $h = 0.2$, the theoretical 99.9% g-and-h quantile
is approximately 626. The amount of local heavy-tailedness in this point is $\Psi'_{L_F}(626) \approx 3.7$. Moreover, recall that over large ranges the g-and-h behaves approximately like an exact Pareto (with tail index $\neq h$), and thus the situation is as if we were to estimate the 99.9% quantile of a distribution which seems to be modeled well by an exact Pareto distribution with parameter $\alpha = -5 + 3.7 = -1.3$ (i.e. with tail index $\xi \approx 0.77$). This quantile is then estimated using $k = 500$, say, upper order statistics, whose values will—for a sample of $n = 10^4$ observations—typically range from around 10 to around $10^3$. Due to the almost constant slope of $\Psi_{L_F}$ ($\approx 3.8$ averaged) over this range, the 500 order statistics thus seem to be modeled well by an exact Pareto distribution with parameter $\alpha = -1.2$ (i.e. $\xi \approx 0.83$). This may explain the good performance of the MLE-POT estimates of the 99.9% quantile in Figure 7.

Clearly, these quantile estimates are likely to get worse in cases where one is estimating far out-of-sample quantiles. Especially in an OR context, estimating at a level of 99.9% is a serious issue as today’s OR loss databases are rather sparse. But still, as the influence of $L_F$ changes only very slowly over large ranges (we are in the $\rho = 0$ case), the size of estimation errors is not nearly as large as in the case for tail-index estimation. In this sense, the $\rho = 0$ case for high quantile estimation based on EVT is not necessarily as troublesome as can be expected from the well-known poor performance of tail-index estimation in such a case. A more in-depth study of this phenomenon would however be highly desirable and have important consequences for QRM-practice.

## 5 Conclusion

In this paper we highlight some issues regarding a quantile-based estimation of risk capital motivated by the Basel II regulatory framework for Operational Risk. From a theoretical point of view, EVT-based estimation methodologies of high quantiles arise very naturally. Our main results are as follows.

First, according to Makarov [21], failure of URQ convergence may lead to inaccurate risk capital estimates. We complement these findings by showing that for sufficiently smooth $F \in RV_{-1/\xi}$, $\xi > 0$, the asymptotic behavior of the associated slowly varying function $L_F$ determines whether or not URQ convergence holds. This then allows one to embed URQ convergence in the framework of second-order regular variation for quantiles: $L_F(x) \rightarrow c \in (0, \infty)$ implies a second-order parameter $\rho < 0$, whereas $L_F(x) \rightarrow \infty$ (or 0) implies $\rho = 0$. In the latter case, the slow convergence properties together with the possibly delusive behavior of $L_F$ may cause serious problems when applying standard EVT methodology.
Second, we stress the fact that, when using EVT methodology, the second-order behavior of the underlying distribution, which (in smooth cases) is fully governed by its associated slowly varying function, is crucial. If data are well modeled by a distribution with “bad” second-order behavior, i.e. with second-order parameter $\rho = 0$, EVT-based estimation techniques will typically converge slowly. As a consequence, the amount of data needed in order for EVT to deliver reasonable results may be unrealistically high and largely incompatible with today’s situation for OR data bases. The idea of penultimate approximations seems very promising in this respect. So far this concept has been of a more theoretical nature and further applied research would be desirable.

Third, the g-and-h distribution of Dutta and Perry [11] corresponds to a class of loss dfs for which the slowly varying function $L_F (g, h > 0)$ is particularly difficult to handle. Due to its slow convergence properties ($\rho = 0$), its behavior in ranges relevant for OR practice is very different from its ultimate asymptotic behavior. For broad ranges of the underlying loss values, the slowly varying function $L_F$ behaves like a regularly varying function, putting locally some extra weight to the tail $F$. As a consequence, standard EVT-based tail-index estimation (asymptotic behavior matters) may result in completely wrong estimates. However, this poor performance need not carry over to high-quantile estimation (finite range behavior matters).

For risk management applications in general and operational risk in particular, a key property to look for is the second-order behavior of the underlying loss severity models. Models encountered in practice often correspond to the case $\rho = 0$. Especially in the latter case, more research on the statistical estimation of high quantiles using EVT is needed; see for instance Gomes and Pestana [15] and references therein for some ideas.

Finally: as already discussed in Remark 2.1, EVT assumes certain tail-stability properties of the underlying loss-data. These may or may not hold. In various fields of application outside of finance and economics these properties seem tenable, and hence EVT has established itself as a most useful statistical modeling tool. Within financial risk management, discussions on stability are still ongoing and may lie at the basis of critical statements on the use of EVT; see for instance Christoffersen et al. [6], Dutta and Perry [11] and Jobst [19]. In their discussion on the forecasting of extreme events, the authors of the former paper state "Thus, we believe that best-practice applications of EVT to financial risk management will benefit from awareness of its limitations, as well as the strengths. When the smoke clears, the contribution of EVT remains basic and useful: it helps us to draw smooth curves through the extreme tails of empirical survival functions in a way that is consistent with powerful theory. Our point is simply that we should not
ask more of the theory than it can deliver."

We very much hope that our paper has helped in lifting a bit of the smoke screen and will challenge EVT experts to consider more in detail some of the statistical challenges related to the modeling of extremes in financial risk management.

6 Acknowledgements

The authors would like to thank Guus Balkema for his many valuable comments and suggestions that helped considerably to improve on previous versions of this paper and Johanna Nešlehová for early discussions related to the project of uniform quantile convergence. An anonymous referee made several useful remarks.

Bibliography


A Appendix

By our standing assumption $F$ and $F^{-}$ are throughout assumed to be continuous. We shall furthermore always assume sufficient smoothness for $F$ and $F^{-}$ where necessary. Consider $F(x) = x^{-1/\xi}L_F(x) \in RV_{-1/\xi}$, $\xi > 0$, and denote by $f$ the density of $F$. This is equivalent to assuming $F$ to be normalized regularly varying with index $-1/\xi$ (see Bingham et al. (1987), Section 1.3) and ensures that we may write $F$ as

$$
F(x) = e^{-\Psi(\log x)}, \quad \Psi(s) = \int_1^e \frac{\eta(y)}{y} dy - c,
$$

(9)

with $\eta(y) = yf(y)/F(y)$ and $c = \log F(1)$. So the graph of $-\Psi$ corresponds to the log-log plot of $F$ and we have $\Psi'(s) = xf(x)/F(x)$, where $s = \log x$. As $F \in RV_{-1/\xi}$, the slope $\Psi'(s)$ converges to $1/\xi$ as $s \to \infty$. The log-log transform of the associated SV function $L_F$ is given by $\Psi_{L_F}(s) = s/\xi - \Psi(s)$. Its slope $\Psi_{L_F}'$ tends to 0 and measures the speed at which the influence of the slowly varying nuisance part vanishes.

Analogously, for the tail quantile function $U(t) = F^{-}(1 - 1/t)$, we obtain $U(t) = e^{\Psi^{-}(\log t)}$, where $\Psi^{-}$ denotes the generalized inverse of $\Psi$ (see Definition 3.1), and we write

$$
U(t) = e^{\phi(\log t)}, \quad \phi(r) = \Psi^{-}(r) = \int_1^e \frac{\varepsilon(y)}{y} dy + c,
$$

(10)

with $\varepsilon(y) = yU''(y)/U(y)$ and $c = \log U(1)$.

As $\phi(\Psi(r)) = r$, simple calculus shows that the sufficient first- and second-order conditions in the ultimate and penultimate approximation (see Theorems 3.1 and 3.2) may be equivalently expressed in terms of $F$ or $U$.

Asymptotics for the g-and-h distribution

A random variable $X$ is said to have a $g$-and-$h$ distribution if $X$ satisfies

$$
X = a + bk(Z) = a + b\frac{e^{gZ} - 1}{g}e^{hZ^2/2}, \quad a, g, h \in \mathbb{R}, b > 0,
$$

where $Z \sim N(0,1)$. We concentrate on the case $a = 0$ and $b = 1$. Degen et al. [10] show that for $g, h > 0$, the g-and-h distribution is regularly varying with index $-1/h$, i.e. $F(x) = \Phi(k^{-1}(x)) = x^{-1/h}L_F(x)$. Since $F$ is differentiable, we may write

$$
F(x) = \Phi(k^{-1}(x)) = e^{-\Psi(\log x)},
$$

102
where, as above, $\Psi$ denotes the log-log transform of $F$.

By definition of $k$ we have $y := \log k(s) = hs^2/2 + gs - \log g + O(e^{-gs})$, $s \to \infty$. We rewrite this as $s^2 = 2y/h - 2gs/h + 2\log g/h + O(e^{-gs})$, $s \to \infty$, and hence

$$s = k^{-1}(e^y) = \sqrt{\frac{2y}{h}} - g + \frac{1}{\sqrt{y}} \left( \frac{g^2}{(2h)^{3/2}} + \frac{\log g}{\sqrt{2h}} \right) + O\left( \frac{1}{y^{3/2}} \right), \quad y \to \infty.$$  

Recall the standard asymptotic expansion for the normal distribution tail given by

$$\Phi(x) = e^{-x^2/2}/(\sqrt{2\pi}x) \left( 1 - 1/x^2 + O(1/x^4) \right), \quad x \to \infty;$$

see Abramowitz and Stegun [1], p. 932. Hence we obtain the following asymptotics for the g-and-h tail $F$:

$$-\Psi(t) = \log \Phi(k^{-1}(e^t))$$

$$= -\frac{1}{2} \left( k^{-1}(e^t) \right)^2 - \log k^{-1}(e^t) - \log 2\pi + O\left( \frac{1}{(k^{-1}(e^t))^2} \right),$$

as $t \to \infty$, which then leads to

$$-\Psi(t) = -\frac{1}{h} t + \frac{\sqrt{2g}}{h^{3/2}} \sqrt{t} - \frac{1}{2} \log t - c + O\left( \frac{1}{\sqrt{t}} \right), \quad t \to \infty,$$

with $c = \frac{1}{2} \log \frac{4\pi}{h} + \frac{g^2}{h^2} + \frac{\log g}{h}$.

**Asymptotics for the log-gamma distribution**

A random variable $X$ follows a log-gamma distribution with parameters $\alpha$ and $\beta$, if its density satisfies

$$f(x) = \frac{\alpha^\beta}{\Gamma(\beta)} (\log x)^{\beta-1} x^{-\alpha-1}, \quad \alpha, \beta > 0, x > 1,$$

where $\Gamma$ denotes the Gamma function. The corresponding tail is given by

$$F(x) = \frac{\Gamma(\beta, \alpha \log x)}{\Gamma(\beta)} = e^{-\Psi(\log x)}, \quad x > 1,$$

where

$$\Gamma(\beta, \alpha \log x) = \int_{\alpha \log x}^\infty t^{\beta-1} e^{-t} dt, \quad x > 1,$$

denotes the upper incomplete Gamma function. Therefore we obtain $-\Psi(t) = \log \Gamma(\beta, \alpha t) - \log \Gamma(\beta)$. For $\beta = 1$ we are in the Pareto case with $-\Psi(t) = -\alpha t$.

For $\beta \in \mathbb{R}_+ \setminus \{1\}$, using the standard asymptotic expansion for the upper incomplete Gamma function given in Abramowitz and Stegun [1], p. 263, we obtain

$$-\Psi(t) = -\alpha t + (\beta - 1) \log t + c + O\left( \frac{1}{t} \right), \quad t \to \infty,$$

where $c = (\beta - 1) \log \alpha - \log \Gamma(\beta)$.

103
Matthias Degen and Paul Embrechts (2009).

Scaling of High-Quantile Estimators.

Submitted.
Scaling of High-Quantile Estimators

Matthias Degen       Paul Embrechts
DEPARTMENT OF MATHEMATICS, ETH ZURICH, 8092 ZURICH, SWITZERLAND

Abstract

Enhanced by the global financial crisis, the discussion about an accurate estimation of regulatory (risk) capital a financial institution needs to hold in order to safeguard against unexpected losses has become highly relevant again. The presence of heavy tails in combination with small sample sizes turns estimation at such extreme quantile levels into an inherently difficult statistical issue. We discuss some of the problems and pitfalls that may arise. In particular, based on the framework of second-order extended regular variation, we compare different high-quantile estimators and propose methods for the improvement of standard methods by focussing on the concept of penultimate approximations.

Keywords: Extreme Value Theory; Peaks Over Threshold; Penultimate Approximation; Power Normalization; Second-Order Extended Regular Variation

1 Introduction

It is fair to say that the global financial system is going through a deep crisis. Whereas for some time a regulatory framework was put into place to avoid systemic risk, the current problems highlight the total insufficiency of this (so-called) Basel framework. Warnings for this were voiced early on; see for instance Danielsson et al. [7]. Also the weaknesses of Value-at-Risk (VaR), the risk measure required by the Basel framework, were discussed over and over again; see for instance Nešlehová et al. [18] and references therein. Nevertheless, it has turned out to be extremely difficult to convince regulators
to "think again". As a consequence, and mainly spurred on by the subprime crisis, statisticians are increasingly called upon to single out research themes with considerable practical usefulness. A key example of this is the long-term joint project between the Office of the Comptroller of the Currency (OCC) and the National Institute of Statistical Sciences (NISS) on the topic of "Financial Risk Modeling and Banking Regulation". The current paper is motivated by this research program.

Our starting point is the discussion about the estimation of regulatory (risk) capital a financial institution needs to hold in order to safeguard against unexpected losses. Without going into a full description of financial data—be it Market Risk (MR), Credit Risk (CR) or Operational Risk (OR)—it suffices to know that, according to the current regulatory standards in the banking industry (Basel II framework), risk capital has to be calculated (statistically estimated) using the concept of VaR at very high levels of confidence (for MR usually 99% at a 10-day horizon, for CR and OR 99.9%, for economic capital even 99.97%, all three of them at a 1-year horizon). The credit crisis prompted the introduction of an extra 99.9%, 1-year capital charge for MR, the so-called Incremental Risk Charge; see Basel Committee [3]. Because of the extreme quantile levels required, early on extreme value theory (EVT) was recognized as a potentially useful tool. However, and this often from practice, critical voices have been raised against an imprudent use of standard EVT. In the context of quantitative risk management (QRM), the use of EVT-based high-quantile estimators may indeed be a delicate issue and warrants careful further study.

The aim of our paper is twofold. In a first and more theoretical part, we analyze different choices of normalization and their influence on the rate of convergence in certain limit laws underlying EVT. More precisely, we compare linear and power norming for high-risk scenarios and quantiles which leads to techniques that are not part of the standard EVT toolkit. In particular we propose the use of so-called penultimate approximations to estimate extreme quantiles. The idea of penultimate approximations goes back to Fisher and Tippett [10], its potential for practical applications however seems to have received little attention so far; see Degen and Embrechts [8] for some references.

In a second part, concrete applications of the methodology developed in the first part are discussed. We compare the performance of different high-quantile estimators. One method increasingly championed in practice estimates quantiles at a lower level (e.g. 99%) and then scales up to the desired higher level (e.g. 99.9%) according to some scaling procedure to be specified. In this context, the usefulness of penultimate approximations in situations of very heavy tails together with small sample sizes (typical for
2 Univariate EVT

We assume the reader to be familiar with univariate EVT, as presented for instance in Embrechts et al. [9] or in de Haan and Ferreira [13]. Throughout we assume that our loss data \( X > 0 \) are modeled by a continuous distribution function (df) \( F \) with upper end-point \( x_F \leq \infty \) and standardly write \( F^{-} = 1 - F \). The corresponding tail quantile function is denoted by \( U(t) = F^{-} \left( 1 - 1/t \right) \), where \( F^{-} \) denotes the (generalized) inverse of \( F \). To avoid confusion we will—where necessary—denote the df and the tail quantile function of a random variable (rv) \( X \) by \( F_X \) and \( U_X \), respectively.

As our focus is on the application of EVT-based methods to quantitative risk management, we prefer to work within the framework of exceedances (Peaks Over Threshold (POT) method) rather than within the classical framework of block-maxima. The two concepts however are closely linked as the next result shows; see de Haan and Ferreira [13], Theorem 1.1.6.

Proposition 2.1 For \( \xi \in \mathbb{R} \) the following are equivalent.

i) There exist constants \( a_n > 0 \) and \( b_n \in \mathbb{R} \) such that
\[
\lim_{n \to \infty} F^n( a_n x + b_n ) = H_\xi(x) = \exp \left\{ - (1 + \xi x)^{-1/\xi} \right\},
\]
for all \( x \) with \( 1 + \xi x > 0 \).

ii) There exists a measurable function \( a(.) > 0 \) such that for \( x > 0 \),
\[
\lim_{t \to \infty} \frac{U(tx) - U(t)}{a(t)} = D_\xi(x) = \frac{x^\xi - 1}{\xi}.
\]

iii) There exists a measurable function \( f(.) > 0 \) such that
\[
\lim_{t \to x_F} \frac{F(t + xf(t))}{F(t)} = (1 + \xi x)^{-1/\xi},
\]
for all \( x \) for which \( 1 + \xi x > 0 \).

Moreover, (1) holds with \( b_n = U(n) \) and \( a_n = a(n) \). Also, (3) holds with \( f(t) = a \left( 1/F(t) \right) \).

DEFINITION 2.1 A df \( F \) satisfying (1) is said to belong to the linear maximum (l-max) domain of attraction of the extreme value distribution \( H_\xi \) and
we write $F \in D_{l}^{\text{max}}(H_\xi)$. For necessary and sufficient conditions for distributions $F$ to belong to $D_{l}^{\text{max}}(H_\xi)$ we refer to de Haan and Ferreira [13], Chapter 1.

Domain of attraction conditions have been formulated directly in terms of regular variation of $F$ at $x_F \leq \infty$ for the cases $\xi > 0$ and $\xi < 0$, but not for the case $\xi = 0$; see Gnedenko [11]. The novelty of Proposition 2.1 (originally due to de Haan [12]) is that it treats the domain of attraction conditions for the three cases in a unified way by making use of the more general concept of extended regular variation (ERV) for $U$. Recall that a function $U$ is said to be of extended regular variation with index $\xi \in \mathbb{R}$ and with auxiliary function $a(.)$ if it satisfies (2); see de Haan and Ferreira [13], Appendix B.2. In that case we write $U \in \text{ERV}_\xi(a)$.

**Remark:** Even within the unified framework of ERV, the case $\xi = 0$ is still somewhat special. Acting as limiting cases, the right hand sides in (2) and (3) are interpreted as $\log x$ and $e^{-x}$ respectively. In that case, $U$ and $1/F$ are said to be of $\Pi$-variation and $\Gamma$-variation, respectively, and we write $U \in \Pi(a)$ (or $U \in \text{ERV}_0$) and $1/F \in \Gamma(f)$.

From a theoretical point of view, this full generality of the framework of extended regular variation is certainly to be appreciated. For applications to QRM however, a framework treating $\xi \geq 0$ but not $\xi < 0$ in an as simple as possible way is to be preferred. This is done below basically by working with $\log U$ instead of $U$.

## 3 Asymptotic properties of normalized high-risk scenarios and quantiles

For a positive rv $X \sim F$ we introduce the notation of $X^t$, which is defined as the rv $X$, conditioned to exceed the threshold $t > 0$. Within QRM, $X^t$ is often referred to as a high-risk scenario; see also Balkema and Embrechts [1] for this terminology.

With this notation, Proposition 2.1 iii) states that high-risk scenarios, linearly normalized, converge weakly to a non-degenerate limit, i.e. for $\xi \in \mathbb{R}$ and $x > 0$,

$$
P \left( \frac{X^t - t}{f(t)} > x \right) = \frac{\overline{F}(t + xf(t))}{\overline{F}(t)} \to -\log H_\xi(x) = (1 + \xi x)^{-1/\xi}, \quad t \to x_F,
$$

(4)
for some measurable function \( f(.) > 0 \). In that case we shall say that \( F \) belongs to the \textit{linear POT} (l-POT) \textit{domain of attraction} of \( H_\xi \) and write \( F \in D^\text{POT}_l(H_\xi) \).

While the limit behavior of random variables (exceedances as well as block-maxima) under linear normalizations is well understood and frequently used in applications, the theory under non-linear normalizations has been studied less. Pantcheva [19] and Mohan and Ravi [16] developed a theory of power norming within the block-maxima framework.

We shall adopt this idea of non-linear norming and study the limit behavior of power normalized high-risk scenarios. Inspired by Barakat et al. [2], who compare the convergence rates under linear and power normalization within the block-maxima setting, we study the first- and second-order asymptotic behavior of power-normalized high-risk scenarios and quantiles.

**Definition 3.1** We say that a df \( F \) belongs to the \textit{power POT} (p-POT) \textit{domain of attraction} of some non-degenerate df \( K \) and write \( F \in D^\text{POT}_p(K) \), if there exists a measurable function \( g(.) > 0 \) such that the (power) normalized high-risk scenario \( \left( X^t/t \right)^{1/g(t)} \) converges weakly to \( K \), in the sense that

\[
P \left( \left( X^t/t \right)^{1/g(t)} > x \right) \to \overline{K}(x), \quad t \to x_F, \tag{5}
\]

for every continuity point \( x > 0 \) of \( K \).

Introducing logarithms proves useful at this point, as it provides a link between the two concepts of linear and power norming for high-risk scenarios. In particular we have the following result about the relation between the respective domains of attraction \( D^\text{POT}_l \) and \( D^\text{POT}_p \).

**Proposition 3.1** For \( X > 0 \) with df \( F_X \) and for \( \xi \in \mathbb{R} \) the following holds:

\[
i) \quad F_{\log X} \in D^\text{POT}_l(H_\xi) \iff F_X \in D^\text{POT}_p(K_\xi), \\
ii) \quad F_X \in D^\text{POT}_l(H_\xi) \implies F_X \in D^\text{POT}_p(K_\xi^{-}),
\]

where \( \overline{K}_\xi(x) = -\log H_\xi(\log x) \) for \( x > 0 \) and \( \xi^{-} = \xi \wedge 0 \).

**Proof:** i) Let \( \xi \in \mathbb{R} \) and \( x > 0 \). Setting \( Y = \log X \), the corresponding high-risk scenario satisfies \( Y^s = \log (X^t) \) for \( s = \log t \) and thus it immediately follows that

\[
\lim_{s \to x_F} P \left( \frac{Y^s - s}{f(s)} > x \right) = -\log H_\xi(x) \iff \lim_{t \to x_F} P \left( \left( \frac{X^t}{t} \right)^{1/g(t)} > x \right) = \overline{K}_\xi(x),
\]

where \( f(s) = g(t) \) and with \( s = \log t \).
Degen and Embrechts (2009)

\[ P \left( \frac{X_t}{f(t)} > x \right) \rightarrow (1 + \xi x)^{-1/\xi}, \quad t \rightarrow x_F. \]

We make use of the fact that the convergence above is uniformly in \( t \). Moreover, define \( \lambda_t(x) = \frac{t x^{g(t)}}{f(t)} \) and observe that if \( \lim_{t \rightarrow x_F} \lambda_t(x) =: \lambda_\infty(x) \) exists, we have for every \( x > 0 \) and as \( t \rightarrow x_F \)

\[ P \left( \left( \frac{X_t}{t} \right)^{1/g(t)} > x \right) = \frac{F(tx^{g(t)})}{F(t)} = \frac{F(t + \lambda_t(x)f(t))}{F(t)} \rightarrow (1 + \xi \lambda_\infty(x))^{-1/\xi}. \]

Now, set \( g(t) = f(t)/t \) so that \( \lambda_t(x) = \frac{x f(t)/t - 1}{f(t)/t} \) for \( x > 0 \).

**\( \xi > 0 \):** In this case \( g(t) \rightarrow \xi \), as \( t \rightarrow x_F \); see de Haan and Ferreira, Theorem 1.2.5. Therefore, the limit \( \lambda_\infty \) exists, is finite and we have \( \lim_{t \rightarrow x_F} \lambda_t(x) = (x^\xi - 1)/\xi \).

**\( \xi < 0 \):** Note first that \( x_F < \infty \). Moreover we have \( f(t)/(x_F - t) \rightarrow -\xi \) as \( t \rightarrow x_F \) (see de Haan and Ferreira [13], Theorem 1.2.5.) and hence \( g(t) \rightarrow 0 \) for \( t \rightarrow x_F \). Therefore we obtain \( \lim_{t \rightarrow x_F} \lambda_t(x) = \log x \).

**\( \xi = 0 \):** For \( \xi = 0 \), the right endpoint \( x_F \) may be finite or infinite. Moreover, \( f(.) \) is asymptotically equivalent to a function \( \tilde{f}(.) \), whose derivative vanishes at \( x_F \). For the case \( x_F = \infty \) we thus have

\[ \frac{\tilde{f}(t) - \tilde{f}(t_0)}{t} = \frac{1}{t} \int_{t_0}^t \tilde{f}(s)ds \rightarrow 0, \quad t \rightarrow \infty. \]

Therefore \( \tilde{f}(t)/t \rightarrow 0 \) as \( t \rightarrow x_F \) (and hence also \( g(t) \rightarrow 0 \)), which in turn implies \( \lim_{t \rightarrow x_F} \lambda_t(x) = \log x \).

In the case \( x_F < \infty \), \( \tilde{f}(t) \rightarrow 0 \) as \( t \rightarrow x_F \) (and hence also \( g(t) \rightarrow 0 \)); see de Haan and Ferreira [13], Theorem 1.2.6. Therefore we obtain \( \lim_{t \rightarrow x_F} \lambda_t(x) = \log x \).

Altogether, \( F \in D^\text{POT}_l(H_\xi) \) with \( \xi \in \mathbb{R} \) implies that for every \( x > 0 \) and as \( t \rightarrow x_F \),

\[ P \left( \left( \frac{X_t}{t} \right)^{1/g(t)} > x \right) \rightarrow \begin{cases} 
(1 + \xi \lambda_\infty(x))^{-1/\xi} & \xi > 0, \\
\lambda_\infty(x) & \xi = 0,
\end{cases} \]

i.e. \( F \in D^\text{POT}_p(K_{\xi_-}) \), where \( \xi_- = \xi \wedge 0 \). This finishes the proof. \[\square\]
For later purposes we shall reformulate Proposition 3.1 in terms of quantiles. Due to the convergence properties of inverse functions (see Resnick [20], Proposition 0.1) this is immediate and we have the following result.

**Corollary 3.1** For \( X > 0 \) with tail quantile function \( U_X \) and \( \xi \in \mathbb{R} \) the following holds:

i) \( \log U_{\log X} \in \text{ERV}_\xi(a) \iff \log U_X \in \text{ERV}_\xi(a) \),

ii) \( U_X \in \text{ERV}_\xi(a) \implies \log U_{\log X} \in \text{ERV}_\xi(a) \),

where \( \xi_- = \xi \wedge 0 \) and \( b(t) = a(t)/U(t) \) for some measurable function \( a(.) > 0 \).

**Remarks:**

1) According to Assertion ii) of Corollary 3.1, convergence of linearly normalized quantiles \( U(tx) \), i.e.

\[
\frac{U(tx) - U(t)}{a(t)} \to D_\xi(x) = \frac{x^\xi - 1}{\xi}, \quad t \to \infty,
\]

for some \( x > 0 \), implies convergence of power normalized quantiles, i.e.

\[
\left( \frac{U(tx)}{U(t)} \right)^{1/b(t)} \to \exp \{ D_{\xi_-}(x) \}, \quad t \to \infty.
\]

In the case of main interest for QRM applications, i.e. for \( \xi \geq 0 \), this rewrites as \( U \in \text{ERV}_\xi(a) \implies \log U \in \Pi(b) \).

2) The respective converse implications in ii) of Proposition 3.1 and Corollary 3.1 do not hold; \( D_p^{\text{POT}} \) attracts in fact more distributions than \( D_t^{\text{POT}} \). Consider for example \( F_X(x) = (\log x)^{-1} \) with \( x > e \), hence \( F_X \notin D_t^{\text{POT}} \) but \( F_X \in D_p^{\text{POT}} \). □

Note that for \( F \in D_p^{\text{POT}}(K) \), the possible limit laws \( K \) are unique up to what we might call p-types (in the POT setting), where we call two dfs \( K_1 \) and \( K_2 \) of the same p-type if \( K_1(x) = K_2(x^p) \) for some \( p > 0 \).

**Proposition 3.2 (Convergence to p-types)** Let \( X \sim F \) be a positive rv and assume \( K_1 \) and \( K_2 \) are two non-degenerate distribution functions.

i) If there exist measurable functions \( g_1(.) > 0 \) and \( g_2(.) > 0 \), such that for \( x > 0 \)

\[
\frac{F(x^{g_1(t)})}{F(t)} \to K_1(x), \quad \frac{F(x^{g_2(t)})}{F(t)} \to K_2(x), \quad t \to x_F,
\]

113
ii) If (9) holds, then either of the two relations in (8) implies the other and (10) holds.

**Proof:** ii) Assume that (9) holds and that \( \frac{F(tx^{g_1(t)})}{F(t)} \to K_1(x) \) as \( t \to x_F \). From the theory of ERV it clear that the existence of a non-degenerate limit \( K \) implies that necessarily \( K(x) = 1 - (1 + \xi \log x)^{-1/\xi} \).

Since the limit laws \( K \) are continuous, uniform convergence holds and we obtain

\[
\frac{F(tx^{g_2(t)})}{F(t)} = \frac{F(t(x^{g_2(t)/g_1(t)}))}{F(t)} \to K_1(x^p), \quad t \to x_F.
\]

i) Assume that the two relations in (8) hold and set \( V(t) = F^{-r}(1 - t) \) and \( W_i(t) = K_i^{-r}(1 - t) \) for \( 0 < t < 1 \) and \( i = 1, 2 \). As \( K_1 \) and \( K_2 \) are non-degenerate, we may find points \( x_1, x_2 \) such that \( W_1(x_1) > W_1(x_2) \) and \( W_2(x_1) > W_2(x_2) \). Due to the convergence properties of generalized inverse functions, we have

\[
\lim_{t \to x_F} \left( \frac{V(F(t)x_i)}{t} \right)^{1/g_j(t)} = W_j(x_i), \quad i, j \in \{1, 2\}.
\]

Taking logarithms we find

\[
\frac{1}{g_j(t)} \log \frac{V(F(t)x_1)}{V(F(t)x_2)} \to \log \frac{W_j(x_1)}{W_j(x_2)} > 0, \quad t \to x_F, \quad j \in \{1, 2\}.
\]

From this we obtain

\[
\lim_{t \to x_F} \frac{g_2(t)}{g_1(t)} = \frac{\log W_1(x_1)}{\log W_1(x_2)} =: p > 0,
\]

which finishes the proof.

In order to appreciate the approximations under linear and power norming, (6) and (7) respectively, we need to quantify and compare the goodness of these approximations. More precisely, we are interested in a comparison of the convergence rates for normalized quantiles in (6) and (7)—or equivalently
of the rates for corresponding normalized high-risk scenarios in (4) and (5). It turns out that a judicious choice of the power normalization $b(.)$, respectively $g(.)$, may improve the convergence rate over linear norming. This may be important for applications, as we might hope to improve the accuracy of standard EVT-based high-quantile estimators when using power norming.

3.1 Second-order asymptotics of normalized quantiles

In the sequel we prefer to work with quantiles $U$ rather than distribution tails $F$. However, any statement formulated in the $U$–framework may equivalently be expressed in the $F$–framework. Moreover, while we worked in full generality (i.e. $\xi \in \mathbb{R}$) so far, we shall henceforth restrict ourselves to the case $\xi \geq 0$, most of interest for QRM applications. Similar results for the case $\xi < 0$ may be worked out.

In order to avoid unnecessary technicalities and to exclude pathological cases we shall throughout assume sufficient smoothness for $U$. For our purposes, the following representation for $U$ turns out to be convenient to work with:

$$U(t) = e^{\varphi(\log t)}, \quad \varphi(t) = \int_1^t \frac{ds}{u(s)} + c,$$

where $u(s) = U(s)/U'(s)$ and $c = \log U(1)$. Furthermore we shall assume that

(A1) the von Mises condition holds, i.e. $tU''(t)/U'(t) \to \xi - 1$, for some $\xi \geq 0$; see de Haan and Ferreira [13] for details.

Assumption (A1) is equivalent to assuming $\varphi' \to \xi \geq 0$ together with $\varphi''/\varphi' \to 0$. It reflects the fact that the log-log plot $\varphi$ of $U$ is assumed to behave "nicely" in the sense of being ultimately linear, i.e. with converging slope $\varphi'$ and vanishing convexity $\varphi''$. Moreover, (A1) is sufficient to guarantee $U \in ERV_\xi(a)$, i.e. for $x > 0$

$$\frac{U(tx) - U(t)}{a(t)} \to D_\xi(x) = \frac{x^\xi - 1}{\xi}, \quad t \to \infty, \quad (11)$$

for some measurable function $a(.) > 0$. Note that from the theory of extended regular variation it is clear that $a(t)/U(t) \to \xi \geq 0$. By Corollary 3.1, this in turn implies $\log U \in \Pi(b)$ and hence

$$\left(\frac{U(tx)}{U(t)}\right)^{1/b(t)} \to x, \quad t \to \infty, \quad (12)$$
where \( b(t) = a(t)/U(t) > 0 \) and such that \( b(t) \to \xi \).

Our interest is in second-order results for (11) and (12), i.e., we want to consider functions \( A \) and \( B \) with 
\[
\lim_{t \to \infty} A(t) = \lim_{t \to \infty} B(t) = 0,
\]
which for \( \xi \geq 0 \) and \( x > 0 \) satisfy
\[
\frac{U(tx) - U(t)}{a(t)} - D_\xi(x) \to S(x), \quad t \to \infty, \tag{13}
\]
and
\[
\frac{D_\xi \left( \left( \frac{U(tx)}{U(t)} \right)^{1/b(t)} \right) - D_\xi(x)}{B(t)} \to T(x), \quad t \to \infty, \tag{14}
\]
with \( b(t) = a(t)/U(t) \) and for some non-trivial limits \( S(.) \) and \( T(.) \). Again, the limiting case \( D_0(.) \) is interpreted as \( \log(.) \).

Clearly, the precise conditions for (13) and (14) to hold will depend on the choice of the normalization \( a(.) \), respectively \( b(.) \). Different choices of normalization may lead to different asymptotics in the respective second-order relations. We discuss the following three cases.

**Case I: \( b_1(t) \equiv \xi \) for \( \xi > 0 \)**

For this choice of power normalization, the limit relations (13) and (14) coincide and we have the following second-order asymptotics.

**Proposition 3.3** Suppose \( U(t) = e^{\varphi(\log t)} \) is twice differentiable and let \( A_1(t) = ta'(t)/a(t) - \xi \) with \( a(t) = \xi U(t) \) for some \( \xi > 0 \). Assume that (A1) holds, that \( \varphi' \) is ultimately monotone, and that

(A2) \( \lim_{t \to \infty} \varphi''(t)/\varphi'(t) - \xi = \rho, \) for some \( \rho \leq 0 \).

Then, for \( x > 0 \),
\[
\lim_{t \to \infty} \frac{U(tx)}{A_1(t)} - x^\xi = H_{\xi,\rho}(x) = x^\xi D_\rho(x),
\]
where \( D_\rho(x) = x^{\rho-1}/\rho \).

**Proof:** See Degen and Embrechts [8], Theorem 3.1.

**Case II: \( b_2(t) = tU''(t)/U(t) \) for \( \xi \geq 0 \)**

An intuitive reasoning behind this choice of normalization may be given by noting that \( b(t) = tU''(t)/U(t) = \varphi'(\log t) \) is the slope of the log-log plot \( \varphi \) of
U. In the sequel we will therefore refer to \( b(t) = \varphi'(\log t) \) as the local slope (or local tail index) of the log-log plot of \( U \) at points \( t \) (as opposed to the ultimate slope \( \xi = \varphi'(\infty) \) typically considered in standard EVT); see also Degen and Embrechts [8]. We obtain the following second-order asymptotics for linear and power normalized quantiles.

**Proposition 3.4** Suppose \( U(t) = e^{\varphi(\log t)} \) is twice differentiable and let \( A_2(t) = ta'(t)/a(t) - \xi \) with \( a(t) = tU'(t) \). Assume that (A1) holds for some \( \xi \geq 0 \), that \( A_2 \) is ultimately monotone, and that \( |A_2| \in RV_\rho \), for some \( \rho \leq 0 \).

Then, for \( x > 0 \),

\[
\lim_{t \to \infty} \frac{U(tx) - U(t)}{a(t)} - D_\xi(x) = S_{\xi,\rho}(x),
\]

where

\[
S_{\xi,\rho}(x) = \int_1^x \int_1^s y^{\rho - 1} dy ds = \begin{cases} 
\frac{1}{\rho} (D_{\xi+\rho}(x) - D_\xi(x)), & \rho < 0, \\
\frac{1}{\xi} (x^\xi \log x - D_\xi(x)), & \rho = 0, \xi > 0, \\
\frac{1}{\log x^2}, & \rho = \xi = 0.
\end{cases}
\]

**Proof:** See de Haan and Ferreira [13], Theorem 2.3.12.

**Proposition 3.5** Suppose \( U(t) = e^{\varphi(\log t)} \) is three times differentiable and let \( B_2(t) = tb'(t)/b(t) \) with \( b(t) = tU'(t)/U(t) \). Assume that (A1) holds for some \( \xi \geq 0 \), that \( \varphi'' \) is ultimately monotone, and that

\( (A3) \lim_{t \to \infty} \varphi''(t) = \rho \), for some \( \rho \leq 0 \).

Then, for \( x > 0 \),

\[
\lim_{t \to \infty} \frac{D_\xi \left( \left( \frac{U(tx)}{U(t)} \right)^{1/b(t)} \right) - D_\xi(x)}{B_2(t)} = T_{\xi,\rho}(x),
\]

where

\[
T_{\xi,\rho}(x) = \begin{cases} 
\frac{x^\xi}{\rho} (D_\rho(x) - \log x), & \rho < 0, \\
\frac{x^\xi}{\log x^2}, & \rho = 0.
\end{cases}
\]

**Proof:** We rewrite (14) for \( t \to \infty \) as

\[
D_\xi \left( \left( \frac{U(tx)}{U(t)} \right)^{1/b(t)} \right) - D_\xi(x) \sim x^\xi \log U(tx) - \log U(t) - b(t) \log x
\]

\[
= x^\xi \int_1^x \frac{b(ts)}{b(t)B(t)} \frac{1}{s} ds.
\]
For the above integral to converge to a non-trivial limit, it is sufficient to require \( b \in ERV_\rho \) for some \( \rho \leq 0 \) and with auxiliary function \( b(t)B(t) \). In that case, it is clear from the theory of extended regular variation that, in the case \( \rho < 0 \), we may take the auxiliary function to satisfy \( b(t)B(t) = tb'(t) \), if \( b' \) is ultimately monotone.

For the special case \( \rho = 0 \) note that, when \( b' \) is ultimately monotone, \( b' \in RV_{-1} \) if and only if \( b \in \Pi(c) \). In this case one may choose \( c(t) = tb'(t) \). This follows from the Monotone Density Theorem for \( \Pi - \)Variation; see Bingham et al. [5], Theorem 3.6.8.

Altogether, \((A3)\) being equivalent to \( \frac{tb''(t)}{b'(t)} \to \rho - 1 \), together with the assumption of \( \varphi'' \) (i.e. \( b' \)) being ultimately monotone ensures that \( b \in ERV_\rho(c) \) for some \( \rho \leq 0 \) and such that we may choose \( c(t) = tb'(t) \). By the Uniform Convergence Theorem for \( ERV\) (see Bingham et al. [5], Theorem 3.1.7a), the convergence

\[
\lim_{t \to \infty} \frac{b(ts) - b(t)}{tb'(t)} = \begin{cases} \frac{x^{\rho - 1}}{\rho}, & \rho < 0, \\ \log x, & \rho = 0. \end{cases}
\]

holds locally uniformly on \((0, \infty)\) which finishes the proof. \(\square\)

**Case III:** \( b_3(t) = \log U(t) - \frac{1}{t} \int_{t_0}^{t} \log U(s)ds \) for some \( t_0 > 0 \) and for \( \xi \geq 0 \)

Compared with the previous two cases, this choice of \( b(.) \) does not seem to be very intuitive. We shall therefore briefly give a heuristic argument about its raison d’être in the literature. Recall from Case II that in smooth cases we may choose \( b(.) \) as the local slope of the log-log plot of \( U \), i.e. \( b(t) = \varphi'(\log t) \). However, if one does not want to a priori assume differentiability, \( \varphi' \) need not exist. In that case, by Karamata’s Theorem, \( \varphi(\log t) \) is of the same order as its average \( \varphi'(\log t) := \frac{1}{t} \int_{t_0}^{t} \varphi(\log s)ds \), for some \( 0 < t_0 < t \), i.e. \( \frac{\varphi(\log t)}{\varphi(\log t) - \varphi'(\log t)} \to 1 \) as \( t \to \infty \). However, unlike \( \varphi, \varphi' \) is always differentiable and hence—similar to the (smooth) Case II—one may choose \( b(t) = \varphi'(\log t) = \varphi(\log t) - \frac{1}{t} \int_{t_0}^{t} \varphi(\log s)ds \) with \( 0 < t_0 < t \). Following the terminology in Case II, we might refer to \( \varphi' \) as a kind of local ”pseudo” slope of \( \varphi \) (or local ”pseudo” tail index).

**Proposition 3.6** Suppose \( U(t) = e^{\varphi(\log t)} \) is twice differentiable and let \( B_3(t) = tb'(t)/b(t) \) with \( b(t) = \log U(t) - \frac{1}{t} \int_{t_0}^{t} \log U(s)ds \) for some \( t_0 > 0 \). Assume that \((A1)\) holds for some \( \xi \geq 0 \), that \( b' \) is ultimately monotone, and that

\[(A4) \lim_{t \to \infty} \frac{\varphi''(\log t)}{(\varphi'(\log t) - b(t))} - 1 = \tau, \text{ for some } \tau \leq 0.\]
Then, for \( x > 0 \),
\[
\lim_{t \to \infty} \frac{D_\xi \left( \frac{U(tx)}{U(t)} \right)^{1/b(t)} - D_\xi(x)}{B_3(t)} = T_{\xi,\tau}(x) = x^\xi \left( D_\tau(x) + S_{\xi,\tau}(x) \right).
\]

**Proof:** Assumption (A4) rewrites as \( tb''(t)/b'(t) \to \tau - 1 \), as \( t \to \infty \), and, together with \( b' \) being ultimately monotone, ensures that \( b \in ERV_\tau(c) \) with \( c(t) = tb'(t) \). For \( b(t) = \log U(t) - \frac{1}{\tau} \int_{t_0}^t \log U(s)ds \) and for some \( x > t_0 \) we obtain by partial integration
\[
\int_{t_0}^x \frac{b(t)}{t} dt = \int_{t_0}^x \frac{\log U(t)}{t} dt - \int_{t_0}^x \frac{1}{t^2} \int_{t_0}^t \log U(s)ds dt
\]
\[
= \frac{1}{x} \int_{t_0}^x \log U(s)ds = \log U(x) - b(x),
\]
so that
\[
\log U(x) = b(x) + \int_{t_0}^x \frac{b(t)}{t} dt.
\]
Therefore, by the Uniform Convergence Theorem for ERV (see Bingham et al. [5], Theorem 3.1.7a), we obtain for \( t \to \infty \),
\[
D_\xi \left( \frac{U(tx)}{U(t)} \right)^{1/b(t)} - D_\xi(x) \sim x^\xi \frac{\log U(tx) - \log U(t) - b(t) \log x}{b(t)B_3(t)}
\]
\[
= x^\xi \left( \frac{b(tx) - b(t)}{tb'(t)} + \int_1^x \frac{b(ts) - b(t)}{tb'(t)} \frac{1}{s} ds \right)
\]
\[
\to x^\xi \left( \frac{x^\tau - 1}{\tau} + S_{\xi,\tau}(x) \right),
\]
where \( S_{\xi,\tau} \) is as in Proposition 3.4.

Concerning a second-order result under linear norming in the Case III, we draw on the work of Vanroelen [22]. The author relates different second-order relations for cases where the normalization \( a(.) \) is replaced by \( \tilde{a}(.) \) with \( a(t) \sim \tilde{a}(t) \), as \( t \to \infty \). We have the following result.

**Proposition 3.7** Suppose that the assumptions of Proposition 3.4 hold, as well as (A3) for some \( \rho \leq 0 \) and (A4) with \( \tau \neq -1 \). Define \( \tilde{a}(t) = U(t) \left( \log U(t) - 1/t \int_{t_0}^t \log U(s)ds \right) \) for some \( t_0 > 0 \). Then, for \( x > 0 \),
\[
\lim_{t \to \infty} \frac{U(tx) - U(t)}{\tilde{a}(t)} - D_\xi(x) = S_{\xi,\rho,\tau}(x),
\]
where $A_3(.) = A_2(.)$ and $A_2(.)$ is as in Proposition 3.4 and with

$$S_{ξ,ρ,τ}(x) =
\begin{cases}
\frac{1}{ρ}(D_{ξ,ρ}(x) - (1 + ρτ) D_ξ(x)), & ρ < 0, \\
\frac{1}{ξ}(x^ξ \log x + (1 + ξ τ) D_ξ(x)), & ρ = 0, ξ > 0, \\
\frac{(\log x)^2}{2} - τ \log x, & ρ = ξ = 0.
\end{cases}$$

**Proof:** In terms of $ϕ$ and its derivatives we have $B_2(t) = ϕ''(r)/ϕ'(r)$ and $A_2(t) = B_2(t) + A_1(t) = ϕ''(r)/ϕ'(r) + ϕ'(r) - ξ$, where $r = \log t$ and with $A_i$ and $B_i$ as defined above. Moreover, $B_3(t) = ϕ'(\log t)/b_3(t) - 1$, so that Assumption (A4) may be rewritten as $B_2(t)/B_3(t) \to 1$ for $t \to \infty$ and with $τ \leq 0$.

The proof now follows from Proposition 1.3.1 of Vanroelen [22] by remarking that Assumption (A4) with $τ ≠ -1$ implies Condition (1.17) of Vanroelen [22]. Indeed, with $a(t) = tU'(t)$ as in Proposition 3.4, we have

$$\lim_{t \to \infty} \frac{1}{A_2(t)} \left( \frac{\tilde{a}(t)}{a(t)} - 1 \right) = \lim_{t \to \infty} \frac{B_3(t)}{A_2(t)} = \lambda,$$

for some $λ \in \mathbb{R}$ ($λ = 0$ in the case $ρ = 0$).

### 3.2 Comparison of the convergence rates under linear and power norming

It is not surprising that in the case of a constant power normalization (Case I), no improvement may be achieved when using power norming for quantiles $U(tx)$ instead of linear norming. However, the need for the case of non-constant power normalizations. Indeed, a comparison of the respective second-order limit relations in the Cases II and III shows that there are situations in which the convergence rate for power normalized quantiles is faster than that for a linear normalization is applied.

**Case II:** $b_2(t) = tU'(t)/U(t)$

Recall from the proof of Proposition 3.7 that we have $B_3(t)/A_2(t) = 1 - A_1(t)/A_2(t)$ and $A_2(t)/A_1(t) = \frac{ϕ''(r)}{ϕ'(r)(ϕ'(r) - ξ)} + 1$, with $r = \log t$ and $A_i$ and $B_i$ as defined above.

$ξ > 0$: In this case, $A_2(t)/A_1(t) \to ρ/ξ + 1$ as $t \to \infty$ and, consequently, the rates under linear and power norming, $A_2(t)$ and $B_2(t)$ respectively, are both of the same order $O(A_1(t))$, $t \to \infty$, as long as $ρ < 0$ (and $-ξ ≠ ρ$). For $ρ = 0$ however, $B_2(t) = o(A_2(t))$, $t \to \infty$, i.e the convergence rate
is (asymptotically) faster when using power norming instead of linear normalization.

$\xi = 0$: Assumption (A3) implies $\varphi''/\varphi' \to \rho \leq 0$, so that under (A1) a non-degenerate second-order relation (14) is only possible for $\rho = 0$. In this case, the convergence rate $B_2(\cdot)$ for power normalized quantiles may (asymptotically) still be improved over the rate for linearly normalized quantiles under the additional requirement that $\varphi''/(\varphi')^2 \to 0$.

Moreover, note that for $\rho = 0$ (and $\xi > 0$) the rate under a non-constant power normalization $b_2(t) = \varphi'(\log t)$ (local slope) is also faster than with a constant normalization $b(t) \equiv \xi = \varphi'(\infty)$ (ultimate slope). This in particular motivates the consideration of so-called penultimate approximations and their application to QRM.

**Remark (Penultimate Approximation):** From an asymptotic point of view, in the case $\xi = 0$ it matters whether we consider $U(tx)/U(t)$ together with a power normalization $1/b_2(t) = 1/\varphi'(\log t)$ and convergence to a constant limit $x$ as in relation (12), or convergence of $U(tx)/U(t)$ to the non-constant, threshold-dependent limit $x^{b_2(t)}$. While in the former case the convergence rate is $\varphi''/\varphi'$ (see Proposition 3.5), the rate in the latter case is $\varphi''$; see Degen and Embrechts [8], Theorem 3.2 (which also holds for $\xi = 0$ as one easily verifies). Clearly this difference is merely of theoretical interest, since, with QRM applications in mind, both procedures give rise to the same penultimate approximation $U(tx) \approx x^{b_2(t)}U(t)$ for $x > 1$ and $t$ large. For an introduction to the concept of local slopes and penultimate approximations as well as references for the latter, see Degen and Embrechts [8].

**Case III:** $b_3(t) = \log U(t) - 1/t \int_{t_0}^{t} \log U(s) ds$

Assumption (A4) may be rewritten as $B_2(t)/B_3(t) - 1 \to \tau \leq 0$ for $t \to \infty$; see proof of Proposition 3.7. For $\tau \neq -1$, this implies that Condition (15) holds, i.e. $B_3(t)/A_3(t) \to \lambda$ for some $\lambda \in \mathbb{R}$. For the special case $\tau = -1$ which may lead to $\lambda = \infty$ in (15), we refer to Vanroelen [22] for further reading. As a consequence, for $\tau \neq -1$ and due to the second-order asymptotics in Case II, $B_3(\cdot)$ and $A_3(\cdot)$ are of the same order if and only if the second-order parameter satisfies $\rho < 0$. Similar to Case II, for $\rho = 0$, the rate under power norming may be improved over the rate under linear norming, i.e $B_3(t) = o(A_3(t))$, $t \to \infty$. Also, the rate under a non-constant normalization $b_3(\cdot)$ is faster than with constant norming $b(t) \equiv \xi$ (for $\xi > 0$). With applications in mind, this motivates the consideration of a penultimate approximation given by $U(tx) \approx x^{b_3(t)}U(t)$ for $x > 1$ and $t$ large.
From the above we may draw the following conclusions about an EVT-based approximation of high quantiles under linear and power normalization. There are indeed situations where the convergence rate for quantiles with non-constant power normalization improves both the rates for linear normalization as well as constant power normalization. This in particular necessitates the second-order parameter $\rho$ associated with the underlying model to satisfy $\rho = 0$ which pertains to many loss models important for QRM-practice. Hence we may hope for an improvement of the accuracy of classical EVT-based high-quantile estimators by the use of penultimate approximations.

In cases where the rate may be improved under power norming, the improvement does not seem to be spectacular at first glance as the improved rate is again slowly varying and hence may still be arbitrarily slow. However, the above second-order statements are about the asymptotic behavior of quantiles $F^{-}(\alpha)$ only, i.e. as the confidence level $\alpha$ tends to 1. Of greater interest for QRM practice is the local behavior as one usually considers a fixed level $\alpha = 99.9\%$, say. Having said that, slow convergence (i.e. $\rho = 0$) does not at all need to be an impediment. In ranges relevant for practice the improvement in the estimation of high quantiles may well be considerable; see Degen and Embrechts [8], Figure 1.

4 Implications for quantitative risk management

We discuss the relevance of power norming, or more precisely of the corresponding penultimate approximations, for practical applications to QRM. In particular we study the EVT-based estimation of high quantiles together with possible fallacies it may bring with it. We hope that for the EVT-community, our discussion will lead to further relevant research—especially for the important case $\rho = 0$.

Recall the Basel II regulatory guidelines for CR and OR according to which risk capital has to be calculated using VaR (i.e. quantiles) at the high level of 99.9%. While this is standard for CR, where a variety of by now well-established models has been tried and tested, it is not so for OR. The latter has only been incorporated in the Basel II framework relatively recently, so that the resulting lack of historical data makes the estimation of a 99.9% quantile a daunting and inherently difficult task. Estimation methods include for instance the use of empirical quantiles as well as the fitting of some specific parametric loss models. For the latter method one is usually left with a reasonably good fit in the body but not in the tails of the data.

Due to the nature of the problem, the use of EVT has emerged naturally;
see for instance Moscadelli [17], where the application of the popular Peaks Over Threshold (POT) method to OR is discussed. Having an extreme quantile level in mind, level scaling inherent to standard EVT (estimate at lower levels, 99% say, and then scale up to the desired higher levels such as 99.9%) provides a potential alternative. In either case however, accurate estimation of the tail index $\xi$ is challenging, so that, in the end some constructive scepticism concerning the wiseness to base risk capital on high-level quantiles of some (profit and) loss df, even when using standard EVT methods, is still called for; see for instance Danielsson et al. [7] and Nešlehová et al. [18].

The second-order results on power norming suggest that moving away from the tail index $\xi$—the indicator of the ultimate heavy-tailedness of the loss model—and focusing instead on the local tail index $b(t) = \varphi'(\log t)$, or on its pseudo equivalent $\tilde{b}(t) = \tilde{\varphi}'(\log t)$, might prove useful at this point. In particular it motivates the consideration and comparison of estimation methods for high quantiles based on

i) standard EVT, and

ii) "advanced" EVT.

As for i), we incorporate two methods belonging to the standard EVT toolkit. Recall from the asymptotics for quantiles under linear norming (see relation (11)) that we may consider $U(tx) \approx U(t) + a(t) x^{\xi - 1}$ and, due to regular variation of $U$, also $U(tx) \approx x^{\xi} U(t)$ for $x > 1$ and large values of $t$. This suggests the following scaling properties of high-quantile estimators. For some quantile levels $\tilde{\alpha}, \alpha \in (0, 1)$ with $\tilde{\alpha} < \alpha$,

$$\hat{\text{VaR}}_{\alpha} = \hat{\text{VaR}}_{\tilde{\alpha}} + \hat{a}(t) \frac{x^{\xi} - 1}{\xi}, \quad (16)$$

and similarly

$$\hat{\text{VaR}}_{\alpha} = x^{\xi} \hat{\text{VaR}}_{\tilde{\alpha}}, \quad (17)$$

with $x = (1 - \tilde{\alpha})/(1 - \alpha) > 1$ and some estimates of $\xi$, $a(t)$ and $\text{VaR}_{\tilde{\alpha}}$ at the lower level $\tilde{\alpha}$.

Relation (16) is better known as the POT-estimator of $\text{VaR}_\alpha$. Indeed, setting $u = \hat{\text{VaR}}_{\tilde{\alpha}}$, and using Proposition 2.1, we arrive at a natural estimator

$$\hat{\text{VaR}}_{\alpha} = u + \hat{f}(u) \left( \frac{N_u}{n(1-\alpha)} \right)^\frac{\hat{\xi}}{\xi} - 1, \quad (18)$$

for some estimates $\hat{\xi}$ and $\hat{f}(u)$ of $\xi$ and of $f(u)$. Here $\frac{N_u}{n}$ is an estimate of $\hat{F}(u)$, where $N_u$ denotes the number of exceedances over the threshold $u$ (set
by the user) of a total number of \( n \) data points; see for instance Embrechts et al. [9], Chapter 6.5.

In the simulation study below, (18) and (17) are referred to as the Standard EVT I and II methods, respectively. The tail index \( \xi \) and (threshold-dependent) scale parameter \( f(u) \) are estimated using the POT-MLE method with an ad-hoc threshold choice of 10\% of the upper order statistics. Compared to the POT-MLE, the performance of other implemented tail index estimators such as the Hill, the method of moments, and the exponential regression model (see for instance Beirlant et al. [4]) did not show significant differences.

Method ii) makes use of penultimate approximations. Based on relation (12), with a non-constant power normalization \( b(. \) ), we suggest the following scaling procedure for high-quantile estimators. For quantile levels \( \tilde{\alpha}, \alpha \in (0,1) \) with \( \tilde{\alpha} < \alpha \),

\[
\text{VaR}_\alpha = x^{\tilde{b}(t)} \text{VaR}_{\tilde{\alpha}},
\]

with \( t = 1/(1 - \tilde{\alpha}), \ x = (1 - \tilde{\alpha})/(1 - \alpha) > 1 \) and some estimates of \( b(t) \) and \( \text{VaR}_{\tilde{\alpha}} \). For the simulation study, we incorporate the two choices \( b(t) = \varphi'(\log t) \), the local slope, as well as \( b(t) = \tilde{\varphi}'(\log t) \), the local "pseudo" slope, and will refer to these methods as the Advanced EVT I and II methods, respectively.

The advanced EVT methods, are included in the simulation study in order to outline the potential of penultimate approximations for practical applications. For the aim of this paper, we do not elaborate on the respective estimation procedures for \( \varphi' \) and \( \tilde{\varphi}' \). In both cases, the estimates are based on a prior local regression procedure for the log-data. This is done with the 'locfit' function (with a tricube weight function and smoothing parameter of 3/4) provided in S-Plus (see Loader [15], Chapter 3 and Section 6.1). The integral appearing in \( \tilde{\varphi}' \) is approximated by a composite trapezoidal rule. Finally, the (lower) quantile \( \text{VaR}_{\tilde{\alpha}} \) for (17) and (19) is estimated by the empirical quantile.

**Remark (Local Tail Index):** The two scaling procedures (17) and (19) use the idea of a linear extrapolation of the log-log plot \( \varphi \) of \( U \), but with slopes \( \varphi' \) at different quantile levels. While the penultimate approximation (19) requires the estimation of the local tail index \( \varphi'(\log t) \) (or of \( \tilde{\varphi}'(\log t) \)) at a specified levels \( t \), the ultimate approximation (17)—in theory—makes use of estimates of the ultimate tail index \( \varphi'(\infty) = \xi \).

In practice, given a sample of size a thousand, say, one will use a number of largest order statistics (above a certain threshold \( t_0 \)) to estimate \( \xi \) in (17). It is clear that this yields an estimate of \( \varphi'(\log u) \) at some (unknown) level \( u > t_0 \) rather than of \( \xi = \varphi'(\infty) \). One of the differences between (17) and
(19) thus is, that in the former case the level \( u \) is random (\( u \) depends on the underlying data), while the latter case uses estimates of the slope \( \varphi'(\log t) \) at predefined levels \( t = 1/(1 - \tilde{\alpha}) \), set by the user. \( \Box \)

### 4.1 Simulation study

The simulation study is based on sample data from six frequently used OR loss models, such as the loggamma, the lognormal, the g-and-h, the Pareto, the Burr and the generalized Beta distribution of the second kind (GB2). For convenience we recall the definition of a g-and-h rv \( X \) which is obtained from a standard normal rv \( Z \) through

\[
X = a + b e^{gZ} - 1 \quad \frac{g}{e^{hZ^2}/2},
\]

with parameters \( a, g, h \in \mathbb{R} \) and \( b \neq 0 \). Note that in the case \( h = 0 \) one obtains a (shifted) lognormal rv. For the Pareto df we use the parameterization \( F(x) = (x/x_0)^{-1/\xi} \), for \( x > x_0 > 0 \) and some \( \xi > 0 \). The GB2 is parameterized as in Kleiber and Kotz [14], p. 184, while the remaining three loss models are as in Embrechts et al. [9], p. 35.

For Table 1 we simulate 200 samples of 1000 observations from each of the six loss models. For each of the 200 data sets we compare bias and the standardized root mean square error (SRMSE) of the four above-mentioned EVT-based estimation methods for VaR at level 99.9%. Several simulations with different choices of (for risk management practice relevant) parameter values were performed, all of them showing a similar pattern concerning the performance of the different estimation methods; see Table 1.

**Remark:** Despite its inconsistency with the well-known stylized facts of OR data (power-tail, i.e. \( \xi > 0 \)), the lognormal distribution (semi heavy-tailed, i.e. \( \xi = 0 \)) is widely used in OR practice as a loss severity model. We include it in our simulation study primarily to question its omnipresence by highlighting some of the problems its use may bring with it. \( \Box \)

As mentioned above, estimation at very high quantile levels by means of fitting a parametric loss model may be hard to justify. For illustrative purposes we nevertheless perform a simulation for the six resulting parametric high-quantile estimators, based on the same data sample. An excerpt of these (expectedly) disappointing results is given in Table 2. Here, the model parameters are estimated using MLE, except for the g-and-h distribution, for which there is no agreed standard estimation method so far. For that case
Table 1: Bias and SRMSE (in %) of four EVT-based estimators for VaR at the 99.9% level based on 200 datasets of 1000 observations of six different loss models.

<table>
<thead>
<tr>
<th>Loss model</th>
<th>Bias</th>
<th>SRMSE</th>
<th>Bias</th>
<th>SRMSE</th>
<th>Bias</th>
<th>SRMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loggamma ((\alpha = 1.75, \beta = 2))</td>
<td>8.41</td>
<td>52.88</td>
<td>5.20</td>
<td>32.93</td>
<td>9.65</td>
<td>57.63</td>
</tr>
<tr>
<td>Lognormal ((\mu = 3.5, \sigma = 1.25))</td>
<td>5.26</td>
<td>56.53</td>
<td>-8.88</td>
<td>39.24</td>
<td>4.97</td>
<td>62.62</td>
</tr>
<tr>
<td>g-and-h ((a = b = 3, g = 0.8, h = 0.4))</td>
<td>7.60</td>
<td>36.84</td>
<td>42.44</td>
<td>53.21</td>
<td>9.53</td>
<td>43.46</td>
</tr>
<tr>
<td>Std. EVT I (POT)</td>
<td>8.41</td>
<td>52.88</td>
<td>5.20</td>
<td>32.93</td>
<td>9.65</td>
<td>57.63</td>
</tr>
<tr>
<td>Std. EVT II ((\alpha = 0.99))</td>
<td>13.99</td>
<td>72.48</td>
<td>6.10</td>
<td>62.20</td>
<td>0.21</td>
<td>51.65</td>
</tr>
<tr>
<td>Adv. EVT I ((\alpha = 0.99))</td>
<td>-9.53</td>
<td>28.29</td>
<td>1.98</td>
<td>41.34</td>
<td>-5.10</td>
<td>29.94</td>
</tr>
<tr>
<td>Adv. EVT II ((\alpha = 0.99))</td>
<td>2.66</td>
<td>41.95</td>
<td>3.60</td>
<td>39.80</td>
<td>-1.69</td>
<td>32.35</td>
</tr>
<tr>
<td>Pareto ((x_0 = 1.2, \xi = 0.75))</td>
<td>13.73</td>
<td>62.73</td>
<td>7.79</td>
<td>54.12</td>
<td>1.20</td>
<td>45.80</td>
</tr>
<tr>
<td>Burr ((\alpha = 1, \kappa = 2, \tau = 1.5))</td>
<td>188.78</td>
<td>200.70</td>
<td>6.10</td>
<td>62.20</td>
<td>0.21</td>
<td>51.65</td>
</tr>
<tr>
<td>GB2 ((a = b = 2, p = 1.5, q = 0.75))</td>
<td>-89.77</td>
<td>89.81</td>
<td>1.26</td>
<td>32.09</td>
<td>-2.00</td>
<td>25.36</td>
</tr>
</tbody>
</table>

A comparison of the results in the Tables 1 and 2 clearly shows that the estimation of high quantiles based on fitting parametric models may indeed be problematic. The model uncertainty involved may be considerable (large fluctuation of the estimation errors). Moreover, from a QRM regulatory point of view, a large negative bias (i.e. underestimation of risk capital) is to be avoided. Not surprisingly, the lognormal parametric model underestimates risk capital charges considerably. While intolerable from a sound regulatory perspective this at the same time may explain the "attractiveness" of its use for a financial institution.

On the other hand, given the high level of 99.9%, the performance of all four EVT-based methods is promising; see Table 1. A comparison within the
EVT-based methods does not yield a clear ranking. However, the advanced EVT methods seem to work at least as well as the standard EVT methods, in particular exhibiting smaller SRMSE. This finding is not by accident. Recall that the estimation of $\varphi'$ and $\tilde{\varphi}'$ in the advanced EVT I and II methods is based on a local regression procedure (i.e. smoothing) of the log-data. As a consequence, the estimates are more robust, which leads to smaller SRMSE-values. For smaller sample sizes we expect this behavior to become even more pronounced.

To confirm the above findings on EVT-based high-quantile estimators, we perform a second, similar study and estimate quantiles at the even more extreme level of 99.97%, relevant for the calculation of so-called economic capital; see for instance Crouhy et al. [6], Chapter 15. Owing to Remark 4.2 we leave out the lognormal data sample. We again simulate 200 samples of 1000, 500 and 250 observations in Table 3.

Table 3: Bias and SRMSE (in %) of four EVT-based estimators for VaR at the 99.97% level based on 200 datasets of 1000, 500 and 250 observations.

<table>
<thead>
<tr>
<th></th>
<th>$n = 1000, \hat{\alpha} = 0.99$</th>
<th>$n = 500, \hat{\alpha} = 0.98$</th>
<th>$n = 250, \hat{\alpha} = 0.96$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loggamma ($\alpha = 1.25, \beta = 1.25$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Std. EVT I (POT)</td>
<td>39.47</td>
<td>159.44</td>
<td>81.57</td>
</tr>
<tr>
<td>Std. EVT II</td>
<td>38.19</td>
<td>160.53</td>
<td>82.15</td>
</tr>
<tr>
<td>Adv. EVT I</td>
<td>-2.99</td>
<td>46.88</td>
<td>-3.93</td>
</tr>
<tr>
<td>Adv. EVT II</td>
<td>7.49</td>
<td>68.89</td>
<td>1.94</td>
</tr>
<tr>
<td>g-and-h ($a = b = 1.5, g = 0.8, h = 0.6$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Std. EVT I (POT)</td>
<td>43.06</td>
<td>149.69</td>
<td>80.63</td>
</tr>
<tr>
<td>Std. EVT II</td>
<td>39.94</td>
<td>163.40</td>
<td>84.14</td>
</tr>
<tr>
<td>Adv. EVT I</td>
<td>7.76</td>
<td>60.52</td>
<td>16.76</td>
</tr>
<tr>
<td>Adv. EVT II</td>
<td>17.52</td>
<td>83.57</td>
<td>18.38</td>
</tr>
<tr>
<td>Pareto ($x_0 = 1, \xi = 0.85$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Std. EVT I (POT)</td>
<td>33.31</td>
<td>120.47</td>
<td>105.22</td>
</tr>
<tr>
<td>Std. EVT II</td>
<td>35.14</td>
<td>135.80</td>
<td>118.95</td>
</tr>
<tr>
<td>Adv. EVT I</td>
<td>-16.29</td>
<td>35.67</td>
<td>-29.95</td>
</tr>
<tr>
<td>Adv. EVT II</td>
<td>5.46</td>
<td>63.49</td>
<td>-8.24</td>
</tr>
<tr>
<td>Burr ($\alpha = 1, \kappa = 1.5, \tau = 1.25$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Std. EVT I (POT)</td>
<td>29.94</td>
<td>159.70</td>
<td>68.72</td>
</tr>
<tr>
<td>Std. EVT II</td>
<td>27.77</td>
<td>166.73</td>
<td>68.98</td>
</tr>
<tr>
<td>Adv. EVT I</td>
<td>5.29</td>
<td>69.86</td>
<td>24.87</td>
</tr>
<tr>
<td>Adv. EVT II</td>
<td>9.26</td>
<td>75.01</td>
<td>16.09</td>
</tr>
<tr>
<td>GB2 ($a = 1, b = 2, p = 1.5, q = 1.25$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Std. EVT I (POT)</td>
<td>12.93</td>
<td>88.16</td>
<td>104.19</td>
</tr>
<tr>
<td>Std. EVT II</td>
<td>11.63</td>
<td>93.63</td>
<td>108.70</td>
</tr>
<tr>
<td>Adv. EVT I</td>
<td>6.58</td>
<td>58.63</td>
<td>29.20</td>
</tr>
<tr>
<td>Adv. EVT II</td>
<td>12.96</td>
<td>59.20</td>
<td>24.79</td>
</tr>
</tbody>
</table>

From Table 3 we may draw the following conclusions. Most importantly,
the potential of an advanced EVT approach to estimate extreme quantiles in the presence of very heavy tails and small sample sizes is clearly revealed. The performance of the advanced EVT I and II methods is by far superior compared to that of the two standard EVT approaches. This confirms that the idea of using penultimate approximations instead of ultimate approximations may indeed be promising in certain situations relevant for practice (and not only from a second-order asymptotic viewpoint). While the estimation errors of the two advanced EVT methods remain comparably moderate even for small sample sizes, standard EVT-based methods hit the wall. The estimation errors explode for decreasing sample sizes so that the usefulness of these methods seems questionable in such situations. From a QRM perspective this means that relying on high-quantile estimates based on these conventional methods may be questionable.

5 Conclusion

In this paper we consider EVT-based high-quantile estimators and discuss scaling properties and their influence on the estimation accuracy at very high quantile levels. The scarcity of data together with the heavy-tailedness present in the data (especially for OR), turns high-quantile estimation into an inherently difficult statistical task. The nature of the problem calls for EVT in some or other form. The application of methods from the standard EVT toolkit in such applied situations is however not without problems. Our main results are as follows.

First, from a methodological perspective, it is de Haan’s framework of $\Pi$-variation that is most useful for our purposes, as it allows for a unified treatment of the for QRM important cases $\xi > 0$ and $\xi = 0$. Inherent to $\Pi$-variation is the notion of power norming (as opposed to the standardly used linear norming) of quantiles and high-risk scenarios. The use of different normalizations leads to different second-order asymptotics. It turns out that, in certain cases relevant for practice, judicious choices of a (non-constant) power normalization—instead of a linear or a constant power normalization—may improve the rate of convergence in the respective limit results.

Second, the theory of second-order extended regular variation provides a methodological basis for the derivation of new high-quantile estimators. The application of different normalizations in the respective second-order relations translates into different scaling properties of the resulting high-quantile estimators. Our findings motivate the derivation of new estimation procedures for high quantiles by means of penultimate approximations. In particular we propose two ”advanced” EVT methods which are based on the estimation of the local (pseudo) slope $\varphi'$ (and $\tilde{\varphi}'$) of the log-log plot $\varphi$ of the underlying
loss model \( U(t) = e^{\varphi(\log t)} \). The methods proposed are intended to complement, rather than to replace, methods from the standard EVT toolkit. Their applications may be useful in situations in which the reliability of standard methods seems questionable.

Third, by means of a simulation study we show that, in the presence of heavy tails together with data scarcity, reliable estimation at very high quantile levels, such as the 99.9% or 99.97%, is a tough call. While our study highlights the limitations of standard EVT approaches in such cases, at the same time it reveals the potential of more advanced EVT methods.

Further statistical research on advanced EVT approaches to estimate high quantiles, together with a more in-depth study of their benefits as well as limitations for practical applications would be desirable.

6 Acknowledgements

The authors would like to thank Guus Balkema for useful discussions related to the paper.

Bibliography


Matthias Degen, Dominik D. Lambrigger and Johan Segers (2009).

Risk Concentration and Diversification: Second-Order Properties.

Submitted.
Abstract

The quantification of diversification benefits due to risk aggregation plays a prominent role in the (regulatory) capital management of large firms within the financial industry. However, the complexity of today’s risk landscape makes a quantifiable reduction of risk concentration a challenging task. In the present paper we discuss some of the issues that may arise. The theory of second-order regular variation and second-order subexponentiality provides the ideal methodological framework to derive second-order approximations for the risk concentration and the diversification benefit.

Keywords: Diversification; second-order regular variation; second-order subexponentiality; subadditivity; Value-at-Risk

JEL Classification: C14

Subject Category and Insurance Brand Category: IE43

1 Introduction

Diversification is one of the most popular techniques to mitigate exposure to risk in all areas of banking and (re-)insurance. Due to the increasing complexity of financial and insurance products, a quantitative analysis of portfolio diversification (reduction of risk concentration) has become a demanding task.
Many authors have warned against an imprudent application of diversification concepts, especially when the underlying risk factors show a heavy-tailed pattern; see for instance [10], p. 269, [19], [9] or [6]. More recently, [16] and [15] have discussed diversification benefits linking heavy-tailed distributions to specific economic models.

In the present paper we discuss risk concentration and diversification within the new regulatory framework for banks (Basel II) and insurance companies (Solvency II), where diversification within and among market, credit, operational and insurance risk plays a prominent role. More specifically, we study second-order properties of diversification benefits for independent and identically distributed (iid) risks under the risk measure Value-at-Risk (VaR).

For a risky position $X$ with distribution function $F$, the Value-at-Risk at the level $\alpha$ is defined by

$$\text{VaR}_\alpha(X) = F^{-1}(\alpha), \quad 0 < \alpha < 1,$$

where $F^{-1}(\alpha)$ denotes the generalized inverse of $F$. Under the Basel II/Solvency II framework, $\text{VaR}_\alpha(X)$ essentially corresponds to the regulatory risk capital a financial institution needs to hold in order to be allowed to carry the risky position $X$ on its books. Note that the level $\alpha$ is given by the respective regulatory authority and is typically close to 1.

Throughout we assume that our potential future losses $X_1, \ldots, X_n$, $n \geq 2$, are non-negative iid random variables with continuous distribution function $F$. We write $\overline{F} = 1 - F$ for the tail of $F$. We assume that $\overline{F} \in RV_{-1/\xi}$, i.e. for every $x > 0$,

$$\frac{\overline{F}(tx)}{\overline{F}(t)} \to x^{-1/\xi}, \quad t \to \infty,$$

so that

$$C(\alpha) = \frac{\text{VaR}_\alpha\left(\sum_{k=1}^n X_k\right)}{\sum_{k=1}^n \text{VaR}_\alpha(X_k)} \to n^{\xi - 1}, \quad \alpha \to 1. \quad (1)$$

We will refer to $C(\alpha)$ as the risk concentration (at level $\alpha$) and to $1 - C(\alpha)$ as the diversification benefit. Note that due to the non-coherence of VaR, diversification benefits may be positive or negative. A risk concentration value $C(\alpha)$ greater than 1 means non-diversification (i.e. superadditivity of $\text{VaR}_\alpha$) at the level $\alpha$. In such cases, aggregation of risks would even lead to an increase of regulatory risk capital.

Given the high $\alpha$-levels typically of interest for risk management practice, analyzing risk concentration by means of its empirical counterpart will in general not yield much insight. One is therefore advised to consider (suitable) analytic approximations of $C(\cdot)$. In the simple case of $n$ regularly varying iid losses, relation (1) gives rise to a first-order approximation $C_1(\alpha) \equiv n^{\xi - 1}$ of $C(\alpha)$ for $\alpha$ close to 1. The asymptotic result (1) has been generalized to situations where the vector $\mathbf{X} = (X_1, \ldots, X_n) \geq \mathbf{0}$ is multivariate regularly
varying with index $-1/\xi$ and with identically distributed margins; see [2], Proposition 2.1 and [8], Theorem 4.1.

The main issue discussed in this paper is that the convergence in (1) may be arbitrarily slow. As a consequence, in risk management practice, where we are interested in $C(\alpha)$ at some fixed level of $\alpha$ typically ranging from 95% to 99.97%, the first-order approximation $C_1(\alpha)$ may be too crude as the following example illustrates.

**Example 1.1** Consider a financial institution holding two risks $X_1$ and $X_2$. Assume that these positions are modeled by $X_1, X_2 \overset{iid}{\sim} F$ with $F \in RV_{-2}$. In that case the risk concentration satisfies $C(\alpha) \to 2^{\xi-1} = 1/\sqrt{2} \approx 0.71$, for $\alpha \to 1$. Therefore, when aggregating $X_1$ and $X_2$, a diversification benefit (reduction of regulatory risk capital) of about 29% would seem reasonable at first sight (for high levels of $\alpha$). Figure 1 however shows that such an argumentation needs careful rethinking. 

![Figure 1: Empirical risk concentration (based on $10^7$ simulations) together with a first-order approximation $C_1 = 1/\sqrt{2} \approx 0.71$ for two iid random variables from a Burr ($\tau = 0.25, \kappa = 8$), a Pareto ($\xi = 0.5$) and a g-and-h ($a = 0, b = 1, g = 2, h = 0.5$) distribution; see Remark 2.1 and Examples 4.1 and 4.2 for the parameterization used in the respective models.](image)

In what follows, we suggest the reader to keep Figure 1 in mind as a warning against a careless use of asymptotics to justify diversification benefits. Most
importantly, the behavior of the risk concentration $C(\alpha)$ at levels of $\alpha$ close to 1 (typically of interest for applications) may be very sensitive, i.e. small changes of $\alpha$ may lead to large changes of $C(\alpha)$. In economic terms this means that while we may well expect diversification benefits of considerable size at a certain level $\alpha$, this may change rather drastically into non-diversification once we move away only little from that level.

Altogether this motivates the consideration of a second-order approximation for the risk concentration $C$. Concerning methodology, we draw on the theories of second-order regular variation and second-order subexponentiality. Our main result, Theorem 3.1, gives the precise asymptotic behavior of the approximation error $C(\alpha) - n^{\xi-1}$ as $\alpha \to 1$. As it turns out, asymptotically two situations may arise. Without going into details at this point, it is either the asymptotic behavior in the second-order regular variation part or the one in the second-order subexponential part that dominates in the limit.

For the more applied risk management end-user, the main message is that, not only for infinite mean but also for finite mean models used in financial and insurance risk management, aggregating risks may somewhat surprisingly result in a negative diversification benefit (under VaR).

The paper is organized as follows. Section 2 recalls basic definitions and results on second-order regular variation and second-order subexponentiality. In Section 3 we present our main result on the second-order behavior of the risk concentration $C(\alpha)$. In Section 4 we apply this result to different distribution functions relevant for practice.

2 Preliminaries

The tail quantile function associated with the distribution function $F$ is denoted by $U_F(t) = (1/F)^-(t) = F^-(1 - 1/t)$, $t > 1$. Where clear from the context, we omit the subscript and write $U$ instead of $U_F$. Recall that $U \in RV_{-1/\xi}$ for some $\xi > 0$ is equivalent to $U \in RV_\xi$. In this case we write $U(t) = t^\xi L_U(t)$, where $L_U \in RV_0$ denotes the slowly varying function associated with $U$.

The distribution function of the sum $X_1 + \cdots + X_n$ is denoted by $G$ and due to the iid assumption we have $G(x) = F^*(x)$, the $n$-fold convolution of $F$. Since $F$ is regularly varying, $F$ is subexponential and hence

$$\frac{\overline{G}(x)}{\overline{F}(x)} \to n, \quad x \to \infty;$$

see for instance [7], Corollary 1.3.2. In terms of quantiles, setting $G^-(\alpha) = x,$
we obtain
\[
\frac{G^-(\alpha)}{F^-(\alpha)} = \frac{U_F\left(1/F(x)\right)}{U_F\left(1/G(x)\right)} = \frac{U_F\left(1/F(x)\right)}{U_F\left(F(x)/G(x)\cdot 1/F(x)\right)} \to \frac{n}{\xi}, \quad \alpha \to 1, \quad (2)
\]
due to the Uniform Convergence Theorem for regularly varying functions; see for instance [4], Theorem 1.5.2. This implies
\[
C(\alpha) = \frac{\text{VaR}_\alpha\left(\sum_{k=1}^n X_k\right)}{\sum_{k=1}^n \text{VaR}_\alpha(X_k)} \to \frac{n}{\xi-1}, \quad \alpha \to 1.
\]
In order to analyze the convergence rate of $C(\alpha)$ to $n^{\xi-1}$ as $\alpha \to 1$, the derivation in (2) suggests to study the second-order behavior in the two limit relations
\[
\frac{U(ts)}{U(t)} \to s^{\xi}, \quad t \to \infty, \quad (3)
\]
\[
\frac{G(x)}{F(x)} \to n, \quad x \to \infty. \quad (4)
\]
Rate of convergence results for (3) are well-established within the framework of second-order regular variation; see for instance [12], Section 2.3 and Appendix B.3 for an introduction. Rate of convergence results for (4) may be obtained using the framework of second-order subexponentiality; see for instance [17, 18] and [3]. Below we review these two concepts.

**Second-order regular variation**

**Definition 2.1 (Second-order regular variation)** A function $U \in RV_{\xi}$ with $\xi > 0$ is said to be of second-order regular variation with parameter $\rho \leq 0$, if there exists a function $a(\cdot)$ with $\lim_{t \to \infty} a(t) = 0$ such that
\[
\lim_{t \to \infty} \frac{U(ts)}{U(t)} - \frac{s^{\xi}}{a(t)} = H_{\xi,\rho}(s) = s^{\rho} \frac{s^\rho - 1}{\rho}, \quad (5)
\]
with the obvious interpretation for $\rho = 0$. In this case we write $U \in 2RV_{\xi,\rho}(\alpha)$ and refer to $a(\cdot)$ as the auxiliary function of $U$. \qed

Note that $U \in 2RV_{\xi,\rho}$ is equivalent to $F \in 2RV_{-1/\xi,\rho/\xi}$ for $\xi > 0, \rho \leq 0$. It is well known that if a non-trivial limit $H_{\xi,\rho}$ in (5) exists which is not a multiple of $s^{\xi}$, then it is necessarily of the form stated. Furthermore, the
auxiliary function satisfies $|a| \in RV_\rho$; see for instance [13], Theorem 1. The second-order parameter $\rho$ thus governs the rate of convergence in (3), i.e. the smaller $|\rho|$, the slower the convergence.

A broad and frequently used subclass of models satisfying (5) is given by the so-called Hall class; see also [14].

**Definition 2.2 (Hall Class)** A distribution function $F$ is said to belong to the Hall class if its quantile function $U$ admits the asymptotic representation

$$U(t) = c t^\xi (1 + d t^\rho + o(t^\rho))$$

as $t \to \infty$, for some $c > 0$, $d \in \mathbb{R} \setminus \{0\}$, and first- and second-order parameters $\xi > 0$ and $\rho < 0$.

In terms of tail functions this means that we consider models of the form

$$F(x) = \left(\frac{x}{c}\right)^{-1/\xi} \left(1 + \frac{d}{\xi} \left(\frac{x}{c}\right)^{\rho/\xi} + o\left(x^{\rho/\xi}\right)\right), \quad x \to \infty.$$

Note that the tail quantile function of a loss model in the Hall class obviously satisfies $U \in 2RV_{\xi,\rho}(a)$ with $a(t) \sim d t^\rho$ as $t \to \infty$. [Throughout the paper we mean by $f_1(t) \sim f_2(t)$ for $t \to t_0$ that $f_1(t)/f_2(t) \to 1$ as $t \to t_0$.] Conversely, loss models that are second-order regularly varying with $\rho < 0$ and with auxiliary function $a(t) \sim d t^\rho$ for $t \to \infty$ are members of the Hall class. This follows from the Representation Theorem for extended regularly varying functions; see [4], Theorem 3.6.6.

**Remark:** For the standard Pareto model $U(t) = t^\xi$, the convergence in (3) is immediate. We interpret this case as $U \in 2RV_{\xi,-\infty}$.

Heavy-tailed models $U \in 2RV_{\xi,\rho}$ not belonging to the Hall class include for instance the loggamma or the $g$-and-$h$ distribution ($\rho = 0$ in both cases).

**Second-order subexponentiality**

Second-order subexponentiality results by [3] (Theorems 2.2 and 2.5) are summarized in the following proposition; see also [17, 18] for similar results.

**Proposition 2.1** Assume that $F$ is differentiable with $F(0) = 0$. If for some $\xi > 0$, $\frac{G(x)}{f(x)} \to \xi$ as $x \to \infty$, then for $n \geq 2$ and with $G(x) = F^{n_\ast}(x)$,

$$\lim_{x \to \infty} \frac{G(x)}{F(x)} = J_\xi(n) = n(n - 1)c_\xi,$$
with
\[ c_\xi = \begin{cases} \frac{1}{\xi}, & \text{if } \xi \leq 1, \\ \frac{(1 - \xi)\Gamma^2(1 - 1/\xi)}{2\Gamma(1 - 2/\xi)}, & \text{if } \xi > 1, \end{cases} \]

and
\[ b(x) = \begin{cases} \frac{\mu_F}{x}, & \text{if } \xi \leq 1, \mu_F < \infty, \\ \frac{\mu_F(x)}{x}, & \text{if } \xi = 1, \mu_F = \infty, \\ \frac{F(x)}{(\xi - 1)}, & \text{if } \xi > 1, \end{cases} \]

where \( \mu_F(x) = \int_0^x t \, dF(t) \), \( \mu_F = \lim_{x \to \infty} \mu_F(x) \) and where \( \Gamma \) denotes the gamma function.

![Figure 2: \( c_\xi \) in Proposition 2.1 as a function of \( \xi \).](image)

**Remarks:**

i) In defining \( c_\xi \) for \( 1 < \xi < 2 \) we make use of the analytic continuation of the gamma function \( \Gamma \) to \( \mathbb{C} \{0, -1, -2, \ldots\} \); see [1], Formula 6.1.2.

ii) Note that \( c_\xi \) is strictly decreasing in \( \xi \) and thus \( c_\xi = 0 \) if and only if \( \xi = 2 \); see also Figure 2. In that case, Proposition 2.1 does not yield a (proper) second-order result for convolutions in that particular case. To the best of our knowledge, second-order asymptotics of the above form for \( \xi = 2 \) are not available in the literature (except for special cases, such as stable laws).
iii) In the case $\xi = 1$ and $\mu_F = \infty$, we have $f = F' \in RV_{-2}$. Karamata’s Theorem implies that $\mu_F(x) = \int_0^x t f(t)dt$ is slowly varying; see for example [12], Theorem B.1.5.

iv) By Proposition 2.1, the asymptotic behavior of the function $b$ is fully specified and we have $b \in RV_{-(1/\xi)}$ with $\xi > 0$.

3 Main result

Combining the concepts of second-order regular variation and second-order subexponentiality, we obtain a second-order result for the risk concentration $C$. It may be viewed as a partial quantile analogue of Theorem 3.2 in [11]. Recall the notations in the previous sections, in particular in Proposition 2.1.

**Theorem 3.1**

Let $X_1, \ldots, X_n \overset{iid}{\sim} F$ be positive random variables and let $U = (1/F)^-$ be such that $tU'(t)/U(t) \to \xi > 0$. Suppose that $U \in 2RV_{\xi,\rho}(a)$ for some $\rho \leq 0$ and with auxiliary function $a(\cdot)$ of ultimately constant sign. If $\rho \neq -(1/\xi)$, then, for fixed $n \geq 2$ and as $\alpha \to 1$,

$$C(\alpha) = \frac{\text{VaR}_\alpha(\sum_{k=1}^n X_k)}{\sum_{k=1}^n \text{VaR}_\alpha(X_k)} = n^{\xi-1} + K_{\xi,\rho}(n)A(\alpha) + o(A(\alpha))$$

where $A(\alpha)$ and $K_{\xi,\rho}(n)$ are given as follows:

(i) case $\rho < -(1/\xi)$:

$$A(\alpha) = b(F^{-}(\alpha)) = \begin{cases} \mu_F / F^{-}(\alpha), & \text{if } \xi \leq 1, \rho < -\xi, \mu_F < \infty, \\ \mu_F(F^{-}(\alpha))/F^{-}(\alpha), & \text{if } \xi = 1, \rho < -1, \mu_F = \infty, \\ (1 - \alpha)/(\xi - 1), & \text{if } \xi > 1, \rho < -1; \end{cases}$$

$$K_{\xi,\rho}(n) = \begin{cases} (n-1)/n, & \text{if } \xi \leq 1, \rho < -\xi, \\ n^{\xi-2}(n-1)\xi c_\xi, & \text{if } \xi > 1, \rho < -1; \end{cases}$$

(ii) case $\rho > -(1/\xi)$:

$$A(\alpha) = a(1/(1 - \alpha)),$$

$$K_{\xi,\rho}(n) = n^{\xi-1} \frac{n^\rho - 1}{\rho}.$$  

**Proof:** See Appendix.  

□
As an approximation to the risk concentration $C(\alpha)$, Theorem 3.1 suggests to consider a second-order approximation $C_2(\alpha) = n^{\xi-1} + K_{\xi,\rho}(n)A(\alpha)$, for $\alpha < 1$.

According to Theorem 3.1 two situations arise. Note that

$$A(\alpha) = \begin{cases} b(F^-(\alpha)), & \text{if } \rho < -(1 \wedge \xi), \\ a(1/(1 - \alpha)), & \text{if } \rho > -(1 \wedge \xi), \end{cases}$$

where $b \circ F^- \in RV_{-(1\wedge\xi)}$ and $|a| \in RV_{\rho}$. Now if $\rho < -(1 \wedge \xi)$, then $b(F^-(\alpha))$ vanishes faster than $a(1/(1 - \alpha))$ as $\alpha \to 1$. This motivates the following terminology: a loss model is said to be of fast convergence if $U \in 2RV_{\xi,\rho}$ with first- and second-order parameters satisfying $\rho < -(1 \wedge \xi)$, and of slow convergence, if $\rho > -(1 \wedge \xi)$; see Figure 3.

A refinement of the proof of Theorem 3.1 also allows to treat the boundary case $\rho = -(1 \wedge \xi)$. This case seems to be of less relevance for practical applications though.

**Addendum**

Let $U \in 2RV_{\xi,\rho}(a)$ satisfy the conditions of Theorem 3.1. Suppose, in addition, that for some $q \in \mathbb{R} \setminus \{0\}$

$$\frac{b(F^-(\alpha))}{a(1/(1 - \alpha))} \to q, \quad \alpha \to 1,$$

where $b(\cdot)$ is as in Proposition 2.1. Then, for $n \geq 2$ and $\alpha \to 1$,
with $H_{\xi,\rho}(\cdot)$ and $J_\xi(\cdot)$ as in Definition 2.1 and Proposition 2.1 respectively.

**Remark:** Assume that $U \in 2RV_{\xi,\rho}(a)$. If the associated slowly varying function $L_U$ is differentiable with ultimately monotone derivative $L'_U$, then the auxiliary function $a(\cdot)$ can be chosen as

$$a(t) = \frac{t U'(t)}{U(t)} - \xi.$$ 

A proof is given in the Appendix.

### 4 Examples

In this section we consider the situation of Theorem 3.1 for different loss models. For notational convenience we focus on the case $\rho \neq -(1 \wedge \xi)$. For the Hall class, Theorem 3.1 specializes as follows.

**Corollary 4.1** Let $U$ belong to the Hall class (i.e. $U(t) = c t^\xi (1 + d t^\rho + o(t^\rho))$ as $t \to \infty$, for some $c > 0$, $d \in \mathbb{R} \setminus \{0\}$, $\xi > 0$, and $\rho < 0$) and satisfy the assumptions of Theorem 3.1. Then, the function $A(\cdot)$ in Theorem 3.1 can be chosen as

$$A(\alpha) = \begin{cases} \frac{\mu_F}{c} (1-\alpha)^{\xi}, & \text{if } \xi \leq 1, \rho < -\xi, \mu_F < \infty, \\ -(1-\alpha) \log (1-\alpha), & \text{if } \xi = 1, \rho < -1, \mu_F = \infty, \\ (1-\alpha)/(\xi-1), & \text{if } \xi > 1, \rho < -1, \\ d\rho(1-\alpha)^{-\rho}, & \text{if } \rho > -(1 \wedge \xi). \end{cases}$$

**Example 4.1 (Burr)** Let $X_1, \ldots, X_n \sim_{\text{id}} \text{Burr}(\tau, \kappa)$, with tail function $F(x) = (1 + x^\tau)^{-\kappa}$ for some $\tau, \kappa > 0$. In terms of its tail quantile function this writes as

$$U(t) = (t^{1/\kappa} - 1)^{1/\tau} = t^{1/(\tau\kappa)} \left(1 - \frac{1}{\tau} t^{-1/\kappa} + o(t^{-1/\kappa})\right), \quad t \to \infty,$$

so that $U$ belongs to the Hall class with parameters $c = 1$, $d = -1/\tau$, $\xi = 1/(\tau\kappa)$, and $\rho = -1/\kappa$. By (6) the tail quantile function $U$ is given in an...
explicit form as well as through an asymptotic expansion, so that we may use either Theorem 3.1 or Corollary 4.1 to derive a second-order result for $C$. Using the former we obtain in the case of fast convergence, i.e. for $\kappa \wedge (1/\tau) < 1$,

$$C(\alpha) = \begin{cases} \frac{n^{1/(\tau\kappa)-1}}{\tau} + \left(\frac{n-1}{n}\kappa B(\kappa - \frac{1}{\tau}, 1 + \frac{1}{\tau}) + o(1)\right)(1 - \alpha)^{1/(\tau\kappa)}, & \text{if } \tau \kappa > 1, \\ 1 - \left(\frac{n-1}{n} + o(1)\right)(1 - \alpha) \log (1 - \alpha), & \text{if } \tau \kappa = 1, \\ \frac{n^{1/(\tau\kappa)-1} - \frac{n-1}{\tau\kappa} n^{1/(\tau\kappa)-2} \Gamma^2(1-\tau\kappa)}{2 \Gamma(1-2\tau\kappa)} (1 - \alpha) + o(1 - \alpha), & \text{if } \tau \kappa < 1, \end{cases}$$

as $\alpha \to 1$ and where $B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$ denotes the Beta function.

In case of slow convergence, i.e. if $\kappa \wedge (1/\tau) > 1$, Theorem 3.1 together with Remark 3.1 suggests to consider the expansion

$$C(\alpha) = \frac{n^{1/(\tau\kappa)-1}}{\tau} + \frac{1}{\tau} n^{1/(\tau\kappa)-1}(1 - n^{-1/\kappa})(1 - \alpha)^{1/\kappa} + o((1 - \alpha)^{1/\kappa}), \ \alpha \to 1.$$  

\[\square\]

Figure 4: Empirical risk concentration (full, based on $10^7$ simulations) together with first-order approximation $C_1 \equiv 1/\sqrt{2} \approx 0.71$ (left panel) and $C_1 \equiv 2^{1/4} \approx 1.19$ (right panel) and second-order approximation $C_2$ (dashed) for two iid Burr($\tau, \kappa$) and Pareto($1/\xi$) random variables for a finite mean case (left panel) and an infinite mean case (right panel). Note the different scales on the vertical axis.

The gain of a second-order approximation $C_2$ over a first-order approximation $C_1$ is illustrated in Figure 4 for the case of a fast converging Burr and an exact Pareto model.

For practical purposes, it is essential to know whether $C(\alpha)$ approaches its ultimate value $n^{\xi-1}$ from above or from below as $\alpha$ tends to 1. For a loss model $U \in 2RV_{\xi, \rho}$ satisfying the assumptions of Theorem 3.1 with $\rho < -(1 \wedge \xi)$ (fast
convergence case), the derivative of $C_2$ satisfies

$$\lim_{\alpha \to 1} C_2' (\alpha) = \begin{cases} -\infty, & \text{if } \xi < 1, \text{ or } \xi = 1, \mu_F = \infty, \\
\frac{n^{-1} \mu_F}{c}, & \text{if } \xi = 1, \mu_F < \infty, \\
n^{\xi-2}(n-1)\xi \frac{\Gamma^{2(1-1/\xi)}}{2 \pi (1-2/\xi)}, & \text{if } \xi > 1,
\end{cases}$$

for some $c = \lim_{t \to \infty} L_U (t) \in (0, \infty)$. Moreover, we have $C_2' (1) > 0$ if and only if $\xi > 2$, i.e. it is only in very heavy-tailed cases that $C(\alpha)$ increases to $n^{\xi-1}$ as $\alpha \to 1$.

In the case of slowly converging loss models, the situation is more involved. For $U \in 2RV_{\xi, \rho}$ with $\rho > -(1 \wedge \xi)$, the ultimate behavior of $C_2'$ depends on the exact form of $L_U$ so that in general no precise statement can be made without extra assumptions. Still, for distributions in the Hall class, $C(\alpha)$ will approach its limit $n^{\xi-1}$ from above (below) if $d$ is negative (positive).

Within the class of slowly converging loss models, the case $\rho = 0$ deserves special attention, as in this case the decay of $|a|$ in (5) may be arbitrarily slow. This is due to the behavior of the associated slowly varying function $L_U$ which, in the case $\rho = 0$, may indeed be rather misleading; see for instance [5], Section 4. A prime example for this is provided for instance by Tukey’s g-and-h distribution.

**Example 4.2 (g-and-h)** A random variable $X$ is said to follow Tukey’s g-and-h distribution with parameters $a, b, g, h \in \mathbb{R}$, if $X$ satisfies

$$X = a + b \frac{e^{gZ} - 1}{g} e^{hZ^2/2},$$

where $Z \sim \mathcal{N}(0, 1)$ and with the obvious interpretation for $g = 0$. Note that in principle such random variables need not be positive. In financial risk management practice relevant cases the parameters typically satisfy $b, g, h > 0$, so that we may bypass this issue using the notion of right tail dominance; see [3].

Suppose $X_1, \ldots, X_n$ are iid g-and-h random variables with $a = 0$, $b = 1$ and $g, h > 0$. Then one shows that $U \in 2RV_{\xi, \rho}$ with $\xi = h$, $\rho = 0$ and that

$$a \left( \frac{1}{1-\alpha} \right) = \frac{U'}{(1-\alpha)U'} - \xi = \frac{g}{\Phi^{-1}(\alpha)} + O \left( \frac{1}{(\Phi^{-1}(\alpha))^2} \right), \quad \alpha \to 1,$$

where we used the standard asymptotic expansion for the tail of the normal distribution given by $\Phi(x) = e^{-x^2/2}/(\sqrt{2\pi}x) (1 + O(1/x^2))$, $x \to \infty$; see for
instance [1], p. 932. Therefore, we obtain the following second-order asymptotics for the risk concentration:

\[ C(\alpha) = n^{h-1} + n^{h-1} \log(n) \frac{g}{\Phi^{-1}(\alpha)} + o \left( \frac{1}{\Phi^{-1}(\alpha)} \right), \quad \alpha \to 1. \]

Depending on the parameter values of \( g \) and \( h \), \( C(\cdot) \) may be growing extremely fast when moving away from \( \alpha = 1 \); see Figure 5. In that figure we compare a g-and-h with a Burr model of slow convergence and with a standard Pareto model. Note that we choose the tail index as \( \xi = 0.5 \) in each model.

Figure 5: Empirical risk concentration (full, based on \( n = 10^7 \) simulations) together with the first-order approximation \( C_1(\alpha) \equiv 1/\sqrt{2} \approx 0.71 \) and the second-order approximation \( C_2 \) (dashed) for two iid Burr (\( \tau = 0.25, \kappa = 8 \)), Pareto (\( \xi = 0.5 \)) and g-and-h (\( g = 2, h = 0.5 \)) random variables.

Figure 5 allows us to draw the following conclusions. Even at high levels of \( \alpha < 1 \), the diversification benefit promised by first-order theory may vanish rather quickly and may even get negative. For the g-and-h model of Figure 5, the regime switch from sub- to superadditivity takes place at the extreme level of \( \alpha \approx 99.95\% \). The second-order approximation \( C_2 \) is able to capture this behavior better than \( C_1 \).
Appendix

PROOF (THEOREM 3.1): For $\alpha < 1$, define $x = G^{-}(\alpha) = (F^{\alpha})^{-}(\alpha)$. Note that the convergence in the definition of second-order regular variation holds locally uniformly on $(0, \infty)$; see [12], Remark B.3.8. Therefore, replacing $t$ by $1/G(x)$ and $s$ by $G(x)/F(x)$ in (5), $U \in 2RV_{\xi, \rho}$ implies

$$
\lim_{x \to \infty} \frac{U(1/G(x))}{nU(1/G(x))} = \frac{1}{n} \left( \frac{G(x)}{F(x)} \right)^{\xi} = n^{\xi-1} \frac{n^{\rho} - 1}{\rho} = n^{-1} H_{\xi, \rho}(n),
$$

with the obvious interpretation for $\rho = 0$. From Proposition 2.1 we then get

$$
C(\alpha) = \frac{1}{n} \left( \frac{G(x)}{F(x)} \right)^{\xi} + n^{-1} H_{\xi, \rho}(n) a \left( \frac{1}{1 - \alpha} \right) + o \left( a \left( \frac{1}{1 - \alpha} \right) \right)
$$

$$
= n^{\xi-1} [1 + \xi n^{-1} J_{\xi} n b(G^{\alpha}(\alpha))] + n^{-1} H_{\xi, \rho}(n) a \left( \frac{1}{1 - \alpha} \right)
$$

$$
+ o [b(G^{-}(\alpha))] + o \left( a \left( \frac{1}{1 - \alpha} \right) \right),
$$

for $\alpha \to 1$ and where we have used the expansion $(1 + y)^{\xi} = 1 + \xi y + o(y)$ as $y \to 0$. Note that, due to regular variation of $b$, we have

$$
b(G^{-}(\alpha)) \sim \left( \frac{G^{-}(\alpha)}{F^{-}(\alpha)} \right)^{-(1 + 1/\xi)} \to n^{-(\xi + 1)}, \quad \alpha \to 1.
$$

Define $\tilde{A}(\alpha) = b(F^{-}(\alpha)) + a \left( \frac{1}{1 - \alpha} \right)$. Note that $b \circ F^{-} \in RV_{-(1 + \xi)}$ and $|a| \in RV_{\rho}$. Due to the regular variation properties of $b$ and $a$ and since $\rho \neq -(1 + \xi)$ this implies that

$$
\frac{C(\alpha) - n^{\xi-1}}{A(\alpha)} = \xi n^{\xi-2} n^{-(\xi + 1)} J_{\xi} n b(F^{-}(\alpha)) \frac{b(F^{-}(\alpha))}{A(\alpha)} + n^{-1} H_{\xi, \rho}(n) a \left( \frac{1}{1 - \alpha} \right) + o(1),
$$

as $\alpha \to 1$, which yields the result. \hfill \Box

PROOF (REMARK 3.1): $U \in 2RV_{\xi, \rho}(a)$ (with $\rho \leq 0 < \xi$) with auxiliary function $a(\cdot)$ implies $U \in RV_{\xi}$ and we write $U(t) = t^{2} L(t)$ for some slowly varying function $L$. With this notation and for $s > 0$,

$$
\lim_{t \to \infty} \frac{U(ts)}{U(t)} - s^{\xi} = s^{\xi} \frac{s^{\rho} - 1}{\rho} \iff \lim_{t \to \infty} \frac{L(ts) - L(t)}{a(t)L(t)} = \frac{s^{\rho} - 1}{\rho}.
$$

Hence $L \in ERV_{\xi}(B)$, i.e. $L$ is extended regularly varying with index $\rho \leq 0$ and auxiliary function $B(t) = a(t)L(t)$. For an introduction to ERV we refer to [12], Appendix B.2.
Case $\rho = 0$: We write $L(t) = L(t_0) + \int_{t_0}^{t} L'(s)ds$. The ultimate monotonicity of $L'$ guarantees $L' \in RV_{-1}$ by the Monotone Density Theorem for II-variation; see [4], Corollary 3.6.9. In that case, $tL'(t)$ is an auxiliary function, hence necessarily $B(t) \sim tL'(t), t \to \infty$; see [12], Remark B.2.6.

Case $\rho < 0$: In that case, the limit $\lim_{t \to \infty} L(t) = L(\infty)$ exists and is finite. Set $f(t) = L(\infty) - L(t) = \int_{t}^{\infty} L'(s)ds$. Then $\lim_{t \to \infty} \frac{f(t)}{B(t)} = -1/\rho$ and $f(t) \in RV_{\rho}$, by Theorem B.2.2 in [12]. Ultimate monotonicity of $L'$ implies that $\frac{L'(t)}{f(t)} \to -\rho$ by Proposition B.1.9 11) of [12] and hence $B(t) \sim tL'(t)$ as $t \to \infty$.

Altogether, we thus obtain $a(t) = \frac{B(t)}{L(t)} \sim \frac{tL'(t)}{L(t)} = \frac{tU'(t)}{U(t)} - \xi$ as $t \to \infty$. □

Acknowledgments

The authors would like to thank Paul Embrechts and Mario V. Wüthrich for several discussions related to this paper and the numerous useful comments. The first two authors gratefully acknowledge the financial support by RiskLab Switzerland. The research of the third author is supported by IAP research network grant nr. P6/03 of the Belgian government (Belgian Science Policy) and by contract nr. 07/12/002 of the Projet d’Actions de Recherche Concertées of the Communauté française de Belgique, granted by the Académie universitaire Louvain.

Bibliography


Curriculum Vitae

PERSONAL DATA

Name                     Matthias Degen
Date of Birth            June 21, 1981
Nationality             Swiss

EDUCATION

Ph.D. studies in mathematics, ETH Zurich.
Topic: Extreme value theory.
Supervisor: Prof. Dr. Paul Embrechts.

Undergraduate Studies    10/2001 – 03/2006
Diploma with distinction in mathematics, extra diploma in insurance mathematics, ETH Zurich.
Major: Financial and insurance mathematics.

Matura at Kantonsschule Trogen AR.
Major: Mathematics and Physics.

EMPLOYMENT

ETH Zurich               09/2003 – 03/2006
Teaching assistant.

ETH Zurich               07/2006 – 12/2009
Research and teaching assistant at the Department of Mathematics at ETH Zurich.