A Bayesian Approach to Modeling Operational Risk When Data is Scarce.

K. Petter Svensson

February 26, 2015

Abstract

The goal of this thesis is to investigate whether it is possible to construct an advanced measurement approach (AMA) model for operational risk when the number of internal data points are very scarce. An AMA model should combine internal data, external data, scenario data, and business environment and internal control factors to give a one year VaR estimate with 99.9 % confidence of operational risk. Out of the methods of combining the different data sources suggested in the literature, only the Bayesian inference approach is suitable due to the small amount of data available. In order to not be restricted to suitable conjugate-pairs, a numerical approach to evaluating the posterior distributions is undertaken, and three different severity distributions are tried out. The distributions tried are the Weibull; the generalized Champernowne, which is suggested by the literature due to its tail behavior; and the g-and-h, which is suggested by the literature due to both its versatility and tail behavior. The conclusion of this thesis is that it is possible to construct an AMA model with Poisson loss frequencies using Bayesian inference to combine the different data sources. However, the data material was too scarce to draw any reliable conclusions about the severity distribution.

Key words: AMA, Bayesian inference, Basel II, g-and-h distribution, generalized Champernowne distribution, loss distribution approach, operational risk

Contents

Acknowledgments

This thesis would not have been possible without the support of Jonas Björk and Stèphanie Axelsson at Klarna, who formulated the initial problem statement and have been a great support along the way. Furthermore, I would like to thank Dragi Anevski from the Department of Mathematical Statistics at the Faculty of Engineering at Lund University for his great support and encouragement throughout the process of writing this thesis. Finally, I would like to give a special thanks to Juliana Crim for her help with proofreading, and to my family and close friends who have put up with my stressed behavior the last couple of months.

Acronyms

ABC approximate Bayesian computation.

AIC Akaike's information criterion.

AM adaptive Metropolis.

AMA advanced measurement approach.

BCBS Basel Committee on Banking Supervision.

BIA basic indicator approach.

BIC Bayesian information criterion.

CML conditional maximum likelihood.

GCD generalized Champernowne distribution.

IID independent and identically distributed.

LDA loss distribution approach.

MCMC Markov chain Monte-Carlo.

ML maximum likelihood.

PDF probability distribution function.

SA standardized approach.

UHS upper half spread.

List of Figures

List of Tables

1 Introduction

1.1 Background

The need to manage operational risk is as old as the existence of banks, since the possibility of internal and external fraud and processing errors is an integral part of the business. These risks were, up until the last decade, handled by insurance policies and audits. However the Basel Committee on Banking Supervision (BCBS) started to formalize the management of operational risk in 1998 (Schevchenko, 2011). This work led to the inclusion of operational risk in the regulatory framework Basel II that was introduced in 2004.

Under Basel I, banks and financial institutes governed by the regulating body were required to allocate capital to be used in case of large losses due to market and credit events. In the Basel II rules, operational risk was recognized as the third source of large losses. Rules concerning the allocation of regulatory capital to cover for operational losses were also put forth. As a part of formalizing the management of operational risk, BCBS defined operational risk as:

> The risk of direct or indirect loss resulting from inadequate or failed internal processes, people, and systems, or from external events. Including legal risk, but excluding strategic and reputational risk.

BCBS also categorized operational losses according to two dimensions, business line of occurrence and type of event. Both of these dimensions are specified on multiple levels to allow for different levels of granularity in the analysis and measurement of operational losses. Furthermore, the Basel II rules define three methods of calculating the amount of regulatory capital needed. The least sophisticated is the basic indicator approach (BIA), the next is the standardized approach (SA), and the most sophisticated method is the advanced measurement approach (AMA).

The two lower methods rely on a fairly rough assumption that the amount of yearly operational losses is proportional to the yearly gross income, while AMA requires statistical estimation of incurred losses, external losses, scenarios, and the use of internal control factors to form the capital estimate. This means that BIA and SA are merely heuristics and are quite conservative in their estimation while AMA has the potential of giving accurate estimations and highlighting problems in the company's current risk management processes. The incentive for banks to develop these advanced models is mainly twofold. First, it opens up the possibility of reduced levels of regulatory capital, leading to lower expenses for capital cost, and secondly, the model can be a great tool for the company's risk management team to mitigate possible losses.

Since the introduction of Basel II, a number of European banks have developed and implemented their own AMA models. In Sweden, one or two of the larger banks have done this. Even though it is mainly larger banks that have gotten their AMA models approved by the regulatory authorities, the incentives make the possibility of using AMA attractive to smaller companies as well.

1.2 Problem Statement

The aim of this thesis is to investigate if it is possible to construct an AMA model for measuring operational risk for a smaller company. The company examined is a small financial institution that is subject to the Basel II rules and has experienced a very small amount of operational losses, making data very scarce.

2 Regulatory Framework

2.1 Business Lines and Event-Types

The Basel II regulations categorizes operational losses in two dimensions, business line and event types. Each of these two dimensions are defined on multiple levels depending on the number of losses at hand and the need for granularity. There exist seven main event types and eight main business lines, which are presented in tables 1 and 2 below. When categorizing losses, it is common to talk about risk cells, which are defined as the combination of the two dimensions to form a 54 cell matrix for the lowest level of granularity.

Table 1: *The Basel II Event-Type Categories*

Event-Type Category (Level 1)	**Categories (Level 2)**
Internal Fraud	Unauthorized Activity Theft and Fraud
External Fraud	Theft and Fraud Systems Security
Employment Practices and Workplace Safety	Employee Relations Safe Environment Diversity and Discrimination
Clients, Products, and Business Practices	Suitability, Disclosure, and Fiduciary Improper Business or Market Practices Product Flaws Selection, Sponsorship, and Exposure Advisory Activities
Damage to Physical Assets	Disasters and other events
Business Disruption and System Failures	Systems
Execution, Delivery, and Process Management	Transaction Capture, Execution, and Maintenance Monitoring and Reporting Customer Intake and Documentation Customer Client Account Management Trade Counterparties Vendors and Suppliers

Table 2: *The Basel II Business Lines*

Business Unit	**Business Line Level 1**	**Business Line Level 2**
Investment Banking	Corporate Finance	Corporate Finance
		Municipal/Government Finance
		Merchant Banking
		Advisory Services
	Trading and Sales	Sales
		Market Making
		Proprietary Positions
		Treasury
Banking	Retail Banking	Retail Banking
		Private Banking
		Card Services
	Commercial Banking	Commercial Banking
	Payment and Settlement	External Clients
	Agency Services	Custody
		Corporate Agency
		Corporate Trust
Others	Asset Management	Discretionary Fund Management
		Non-Discretionary Fund Management
	Retail Brokerage	Retail Brokerage

2.2 Three Methods of Measuring Operational Risk

In Basel II, BCBS recognizes three different methods of calculating the regulatory capital needed for operational risk. The simplest method is called the basic indicator approach (BIA); the expansion that is a bit more granular is called the standardized approach (SA); and the most advanced and only actual statistical method is called the advanced measurement approach (AMA).

2.2.1 The Basic Indicator Approach

Under BIA, the total capital requirement K_{BIA} is defined as

$$K_{BIA} = \frac{1}{n}\sum_{i=1}^{3} GI_i \cdot \alpha, \tag{2.1}$$

where GI_i is the annual gross income, set to zero if negative, i years back, α is a percentage set by the Basel Committee, currently at 15% and n is the number of the past three years with a positive gross income. Hence, calculating regulatory capital according to BIA means taking 15 percent of the average positive gross income over the past three years.

2.2.2 The Standardized Approach

Table 3: *The β values currently used in SA*

Index	Business Line	Percentage
1	Corporate Finance	18%
2	Trading and Sales	18%
3	Retail Banking	12%
4	Commercial Banking	15%
5	Payment and Settlement	18%
6	Agency Services	15%
7	Asset Management	12%
8	Retail Brokerage	12%

The standardized approach (SA) is very similar to BIA but accounts for differences in risk among different business lines. The total capital requirement K_{SA} is defined as

$$K_{SA} = \frac{1}{3}\sum_{i=1}^{3}\sum_{j=1}^{8} GI_{i,j} \cdot \beta_j, \tag{2.2}$$

where $GI_{i,j}$ is the annual gross income, set to zero if negative, i years back for business line j, and β_j is a percentage dependent on the business line set by the Basel Committee with current percentages shown in table 3.

2.2.3 The Advanced Measurement Approach

AMA is a loosely specified quantitative approach to calculating regulatory capital. The Basel II rules require the use of the following four data elements:

(i) Internal Data

(ii) External Data

(iii) Scenario Analysis

(iv) Business environment and internal control factors

The first three are usually combined in some fashion, and number four is used to scale the other data sources (Schevchenko, 2011). Furthermore, the Basel II rules require the use of VaR as the risk measure with a time horizon of one year and at a confidence level of 99.9 %. Also, there are additional qualitative requirements governing the risk management procedures of the company that lie outside of the scope of this thesis.

2.3 Modeling Losses - Loss Distribution Approach

The Loss Distribution Approach (LDA) is a popular statistical approach for calculating VaR levels. It falls within the AMA category and has its roots in actuarial science. To apply the LDA, two drivers of the loss distribution are identified, the yearly frequency of which events occur and the severity of a loss given that it occurs. These two drivers are modeled separately and are assumed to be independent. Once both distributions are fitted, they are convoluted to give the loss distribution. The annual loss for year t can then be expressed as

$$L_t = \sum_{i=1}^{N_t} X_i^t, \tag{2.3}$$

where N_t is the stochastic number of loss events during year t and $\boldsymbol{X}_t = (X_1^t, ..., X_{N_t}^t)$ is a vector of independent random losses during year t.

2.4 Measuring Risk - VaR and Expected Shortfall

Value at risk is a measure widely used to quantify risk and is specified for a certain time horizon. This is the measurement used in the Basel II rules. The VaR of the random variable $L \sim F_L(l)$ for a certain confidence level α is defined as

$$\mathrm{VaR}_\alpha(L) = F_L^{-1}(\alpha) \tag{2.4}$$

More intuitively VaR_α is the value such that the probability of a loss exceeding this over the specified time horizon is $1 - \alpha$ %. The most commonly used values for α are 95 % and 99 %.

3 Theoretical Framework

3.1 Distributions

3.1.1 Gamma Distribution

The stochastic variable X is said to follow a gamma distribution if its probability density is given by

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

and its cumulative distribution by

$$F(x; \alpha, \beta) = \frac{1}{\Gamma(\alpha)} \gamma(\alpha, \frac{x}{\beta}).$$

The gamma distribution is commonly used in Bayesian statistics since it is a conjugate prior for a number of different distributions (Schevchenko, 2011). There will be more on this further on.

3.1.2 Weibull Distribution

The Weibull distribution for a stochastic variable X has the probability density

$$f(x; \alpha, \beta) = \alpha\beta x^{-\beta x^{\alpha}}$$

and the cumulative distribution

$$F(x; \alpha, \beta) = 1 - e^{-\beta x^{\alpha}}$$

with $x > 0$, the shape parameter $\alpha > 0$, and the scale parameter $\beta > 0$. The Weibull distribution is a generalization of the exponential distribution where the added parameter allows for greater flexibility and a heavier tail. For $\alpha = 1$, it reduces to the exponential distribution and for shape parameters $\alpha < 1$, the distribution is heavy tailed (Chernobai et al., 2007).

3.1.3 Generalized Champernowne Distribution

The original Champernowne distribution was proposed by D.G. Champernowne in 1936 in the context of describing income distribution. More recently, the generalized Champernowne distribution (GCD) has been proposed and has become popular in risk estimation due to its appealing properties. The distribution behaves similar to a log-normal distribution for small values of x and converges to a Pareto distribution in the tail (Bolancé et al., 2012). This means that the distribution can fit the entire support of a loss distribution instead of having to use different distributions for different loss sizes, posing the problem of choosing the cutoffs for different distributions. The generalized Champernowne distribution's probability density is defined by

$$f(x;\alpha,M,c)=\frac{\alpha(x+c)^{\alpha-1}((M+c)^{\alpha}-c^{\alpha})}{((x+c)^{\alpha}+(M+c)^{\alpha}-2c^{\alpha})^2},$$

and its cumulative distribution by

$$F(x;\alpha,M,C)=\frac{(x+c)^{\alpha}-c^{\alpha}}{(x+c)^{\alpha}+(M+c)^{\alpha}-2c^{\alpha}},$$

where $x\geq 0, M>0, \alpha>0$, and $c\geq 0$.

3.1.4 The g-and-h Distribution

The g-and-h distribution is a quantile distribution, meaning that it is defined by its quantile function. Distributions defined in this way usually lack an analytical expression for the probability density and the cumulative distribution. The formulation does however offer a great deal of flexibility, and the g-and-h distribution has been shown to be suitable for modeling operational losses (Dutta and Perry, 2007). The g-and-h distribution is defined in the following way: Let $Z\sim N(0,1)$ be a standard normal random variable. A random variable X is said to follow a g-and-h distribution with parameters $A,B,g,h\in\mathbb{R}^+$ if X can be expressed as

$$X=A+B(e^{gZ}-1)\frac{e^{hZ^2/2}}{g}$$

Where A is the location parameter, B the scale parameter, g the skewness parameter and h the elongation parameter, i.e. it controls the distribution's tail length (Hoaglin, 1985).

3.2 Parameter Estimation

3.2.1 Maximum Likelihood Estimation

For a sample $x_1,...,x_n$ that is independent and identically distributed (IID) the joint density function is $f(x_1,...,x_n|\theta)=f(x_1|\theta)\cdot ...\cdot f(x_n|\theta)$. This is used to define a function for the parameter θ given the observed data $(x_1,...,x_n)$, the function is called the likelihood function and is defined as

$$\mathcal{L}(\theta|x)=\prod_{i=1}^{n}f(x_i|\theta).$$

A summation is usually easier to work with than a multiplication. The log-likelihood is therefore usually used instead and is defined as

$$\ell(\theta|x)=\log\left(\prod_{i=1}^{n}f(x_i|\theta)\right)=\sum_{i=1}^{n}\log\left(f(x_i|\theta)\right).$$

The parameter θ can now be estimated by maximizing the likelihood or log-likelihood function, i.e.

$$\hat{\theta}_{mle} = \arg\max_{\theta} \mathcal{L}(\theta) = \arg\max_{\theta} \ell(\theta).$$

The intuition is that θ is estimated with the value that makes the estimated parametric model fit best with the observed data.

3.2.2 Pseudo Maximum Likelihood Estimation for GCD

Bolancé et al. (2012) suggests a modified way of estimating the parameters of a generalized Champernowne distribution. The reason for doing this is that numerical methods used to calculate the maximum likelihood estimate of the model distribution parameters do not always converge. The authors suggest estimating M with the empirical median of the sample, since $F(M)_{GCD} = 0.5$, and then estimating α and c with ML which ensures that the likelihood function is concave and has a maximum. The log-likelihood function to be maximized numerically is:

$$\begin{aligned}\ell(\alpha, c) =& n \log \alpha + n \log((M+c)^{\alpha} - c^{\alpha}) + (\alpha - 1) \sum_{i=1}^{n} \log(x_i + c) \\ &- 2 \sum_{i=1}^{n} \log((x_i + c)^{\alpha} + (M+c)^{\alpha} - 2c^{\alpha})\end{aligned}$$

3.2.3 Conditional Maximum Likelihood Estimation for GCD

Buch-Kromann (2009) suggests a conditional maximum likelihood method that puts larger emphasis on the tail estimation to try and correct for sparsity in observations of extreme events. This is done by defining a modified PDF for x that lies above some known threshold t:

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

with $\alpha > 0, M > 0$ and $c \geq 0$. Note that for $t = 0$, this is the same as the ordinary PDF of the generalized Champernowne distribution.

$$\begin{aligned}\ell_t(\boldsymbol{\theta}) =& n_t \log \alpha + (\alpha - 1) \sum_{i=1}^{n_t} \log(x_i + c) - 2 \sum_{i=1}^{n_t} \log\left((x_i + c)^{\alpha} + (M+c)^{\alpha} - 2c^{\alpha}\right) \\ &+ n_t \log\left((t+c)^{\alpha} + (M+c)^{\alpha} - 2c^{\alpha}\right)\end{aligned} \tag{3.1}$$

The estimation is done in two steps:

1. M_1 and α_1 are estimated by setting $c_1 = 0$ and maximizing (3.1) for a given t. This ensures a good fit in the tail portion of the parametric distribution.

2. Set $\alpha = \alpha_1$ and $M = \left(\frac{\tau}{\alpha_1} + c^{\alpha_1}\right)^{1/\alpha_1} - c$ where $\tau = \alpha_1 M_1^{\alpha_1}$ and estimate c_2 by maximizing (3.1) for $t = 0$.

The final estimates are obtained as $\hat{\alpha} = \alpha_1$, $\hat{c} = c_2$ and $\hat{M} = \left(\frac{\tau}{\hat{\alpha}} + \hat{c}^{\hat{\alpha}}\right)^{1/\hat{\alpha}} - \hat{c}$.

3.2.4 Quantile Estimation for the g-and-h Distribution

Dutta and Perry (2007) suggest the following procedure for estimating the parameters of a g-and-h distribution:
Estimate A as the empirical median. Define X_p and Z_p as the percentiles of the g-and-h and standard normal distribution, then

$$g_p = -\left(\frac{1}{Z_p}\right) \ln\left(\frac{X_{1-p} - X_{0.5}}{X_{0.5} - X_p}\right).$$

Choose a number of percentile points p, Hoaglin (1985) suggests using $\log_2$ of the sample size, g is then estimated as the median of $\{g_p\}$.

The upper half spread (UHS) is defined as

$$UHS = \frac{g(X_{1-p} - X_{0.5})}{e^{-gZ_p} - 1},$$

and it can be shown that the following holds for operational risk

$$\ln UHS = \ln B + hZ_p^2/2.$$

Therefore, B and h can be estimated by a linear regression model thus concluding the estimation methodology.

3.2.5 Numerical Maximal Likelihood Estimation of the g-and-h Distribution

Rayner and MacGillivray (2002) propose a methodology for evaluating the likelihood of the generalized g-and-h distribution, similar to the g-and-h distribution used in this paper. An adaptation of this methodology is presented below.

A quantile function is defined as the inverse of the CDF, i.e. the quantile function is given by

$$x = Q(u|\boldsymbol{\theta}) = F^{-1}(u|\boldsymbol{\theta}).$$

Furthermore, the likelihood function can be expressed as

$$\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{x}) = \prod_{i=1}^{n} f(x_i|\boldsymbol{\theta}) = \prod_{i=1}^{n} \frac{\partial}{\partial x} Q^{-1}(x_i|\boldsymbol{\theta}) = \left(\prod_{i=1}^{n} Q'\left(Q^{-1}(x_i|\boldsymbol{\theta})|\boldsymbol{\theta}\right)\right)^{-1},$$

where

$$Q'(u|\boldsymbol{\theta}) = \frac{\partial Q}{\partial u} = \frac{\partial Q}{\partial z} \cdot \frac{dz}{du}$$

and since $z = \Phi^{-1}(u)$, then

$$\frac{dz}{du} = \sqrt{2\pi} e^{z^2/2}$$

and

$$\frac{\partial Q(z|\boldsymbol{\theta})}{\partial z} = \frac{\partial}{\partial z}\left(A + B(e^{gz} - 1)\frac{e^{hz^2/2}}{g}\right) = B\left(e^{gz}(g + hz) - hz\right)\frac{e^{hz^2/2}}{g}.$$

Combining this gives

$$Q'(u|\boldsymbol{\theta}) = \sqrt{2\pi} B\left(e^{gz}(g + hz) - hz\right)\frac{e^{(h+1)z^2/2}}{g} \tag{3.2}$$

The log-likelihood for the data vector $\boldsymbol{x} = (x_1, ..., x_n)$ given the parameters $\boldsymbol{\theta}$ can now be obtained as follows:

1. Numerically solve $x_i = Q(z_i|\boldsymbol{\theta})$ for each x_i to get $z_i = Q^{-1}(x_i|\boldsymbol{\theta})$
2. For each z_i, calculate $Q'(u_i|\boldsymbol{\theta})$ according to (3.2)
3. Sum the negative logarithm of each term obtained in step 2 to get the log-likelihood

$$\ell(\boldsymbol{\theta}|\boldsymbol{x}) = -\sum_{i=1}^{n} \log\left(Q'\left(Q^{-1}(x_i|\boldsymbol{\theta})|\boldsymbol{\theta}\right)\right).$$

This can now be used in conjunction with some numerical maximization scheme to obtain a maximum likelihood estimate. This approach is, however, extremely computationally heavy due to the need to numerically obtain $\boldsymbol{z}$ in each iteration.

3.3 Goodness of Fit

3.3.1 Information Criterion

Information criteria are measures of fit based on the likelihood of a fitted model and some penalty for number of used parameters. There exist a number of information criteria. The two most commonly used are the Akaike information criterion (AIC) and the Bayesian information criterion (BIC), defined as

$$\mathrm{AIC} = 2k - \ell(\boldsymbol{\theta}) \tag{3.3}$$

$$\mathrm{BIC} = k \cdot \log(n) - \ell(\boldsymbol{\theta}) \tag{3.4}$$

where k is the number of parameters in the model and n is the number of data points in the fitted sample. Among a number of alternative models, the one with the smallest value of an information criterion is favored by that information criterion.

3.3.2 The QQ-Plot

The quantile quantile plot is a scatter plot in which the quantiles of two probability distributions are plotted against each other. The quantiles can either be empirical or calculated from some fitted parametric distribution. If the shape of the two distributions only differ by a linear transformation, the scatter plot will form a straight line with a slope depending on scale difference and intercept depending on location difference of the two distributions. To help evaluate how well the shapes align, a straight line can be drawn through the 25th and 75th percentiles which is done in the plots used in this paper.

3.4 Parametric Bootstrapping

Statistical Bootstraps is a wide category of methods aimed at measuring the accuracy of sample estimates. This is done by resampling from an approximate distribution and using some measure of accuracy to determine the uncertainty in the original estimate. A common approximate distribution to sample from is the empirical distribution of the observed data and methods involving this falls within the non-parametric bootstrap category. Another category is the parametric bootstrap method in which a parametric model is fitted and then used as the approximate distribution. Given a random sample $\boldsymbol{x} = (x_1, x_2, ..., x_N)$, the following is done to create a parametric bootstrap estimate:

1. Using some goodness of fit measure, choose a suitable parametric distribution $F(x; \boldsymbol{\theta})$.
2. Estimate the parameter vector $\hat{\boldsymbol{\theta}}$ using ML or some other estimation methodology.
3. Simulate M data sets $\boldsymbol{x}_m^* = (x_{m,1}^*, x_{m,2}^*, ..., x_{m,N}^*)$ from the estimated distribution $F(x; \hat{\boldsymbol{\theta}})$.
4. Estimate $\hat{\boldsymbol{\theta}}_m^*$ for each generated data set $\boldsymbol{x}_m^*$
5. Calculate measure of accuracy of the original estimates from the set of $\hat{\boldsymbol{\Theta}}^* = (\hat{\boldsymbol{\theta}}_1^*, \hat{\boldsymbol{\theta}}_2^*, ..., \hat{\boldsymbol{\theta}}_M^*)$

3.5 Bayesian Inference

Bayesian inference uses Bayes' theorem in order to draw statistical inference about observed data. The main difference to the basic frequentist inference is the use of a prior hypothesis about data and updating the probability of this using acquired data. Bayes' theorem is in its simplest form is expressed as

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)} \tag{3.5}$$

3.5.1 Definitions

Shevchenko and Peters (2013) define Bayesian inference in the following way. Consider a random vector of data $\boldsymbol{X} = (X_1, X_2, ..., X_N)$ whose joint density, for a vector of parameters $\boldsymbol{\Theta} = (\theta_1, \theta_2, ..., \theta_d)$, is $h(\boldsymbol{x}|\boldsymbol{\theta})$. In the Bayesian case, both the data and the parameters are considered random or uncertain. This means that the joint density is

$$h(\boldsymbol{x}, \boldsymbol{\theta}) = h(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta}|\boldsymbol{x})h(\boldsymbol{x}) \tag{3.6}$$

Where

- $h(\boldsymbol{x}, \boldsymbol{\theta})$ is the joint density of observed data and parameters.
- $h(\boldsymbol{x}|\boldsymbol{\theta})$ is the density of observations given parameters. This is the same as the likelihood function of $\boldsymbol{\theta}$, i.e. $\mathcal{L}_{\boldsymbol{x}}(\boldsymbol{\theta}) = h(\boldsymbol{x}|\boldsymbol{\theta})$
- $\pi(\boldsymbol{\theta})$ is the probability density of the parameters, commonly referred to as the prior density. $\pi(\boldsymbol{\theta})$ usually depend on a set of hyper-parameters, but these are left out here to keep the notation uncluttered.
- $\pi(\boldsymbol{\theta}|\boldsymbol{x})$ is the density of parameters given data $\boldsymbol{x}$, usually referred to as the posterior density.
- $h(\boldsymbol{x})$ is the marginal density of $\boldsymbol{X}$ that can be derived as $h(\boldsymbol{x}) = \int h(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$.

Bayes' theorem (3.5) can be used to express the posterior density as

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{\pi(\boldsymbol{\theta})h(\boldsymbol{x}|\boldsymbol{\theta})}{h(\boldsymbol{x})} \tag{3.7}$$

For the purpose of inferring parameter values, the scale of (3.7) is of lesser importance, leading to the simplified expression

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}) \propto \pi(\boldsymbol{\theta})h(\boldsymbol{x}|\boldsymbol{\theta}). \tag{3.8}$$

3.5.2 Conjugate Prior

There are cases in which the posterior distribution can be calculated analytically, which makes the usage of Bayesian inference a lot less cumbersome. This is the case for certain choices of priors and distributions that result in a posterior of the same type as the prior, formally named conjugate priors.

Definition 1. *Let F denote a class of density functions $f(x|\theta)$, indexed by θ. A class U of prior densities $\pi(\theta)$ is said to be a conjugate family for F and $F - U$ is called a conjugate pair, if the posterior density $\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta)d\theta}$ is in the class U for all $f \in F$ and $\pi \in U$.*

3.6 Markov Chain Monte-Carlo Methods

Markov chain Monte-Carlo methods (MCMC) are used to evaluate complicated and possibly high dimensional integrals. This can be used to evaluate the posterior density if analytical methods are unsuccessful. The method works by ”walking” around on the multidimensional surface to be integrated creating a chain of steps. For each iteration, the algorithm searches in the vicinity of of the previous step to find a new step on the probability surface that has a high enough contribution to the integral. There exist a large number of MCMC algorithms, all constructed in such a way that the Markov chains they give rise to have the integrand as their equilibrium distribution. A common algorithm in use is the Metropolis, which exists in a number of different adaptations each with its own specific advantages.

3.6.1 Metropolis Algorithm

1. Initialize $\boldsymbol{\theta}^0$
2. For $l = 1, ..., L$
   (a) Generate a proposal $\boldsymbol{\theta}^*$ from the proposal density $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{l-1})$.
   (b) Draw U from $\text{unif}(0, 1)$
   (c) If
   $$U < \min\left\{1, \frac{\pi(\boldsymbol{\theta}^*|\boldsymbol{x})}{\pi(\boldsymbol{\theta}^{l-1}|\boldsymbol{x})}\right\},$$
   set $\boldsymbol{\theta}^l = \boldsymbol{\theta}^*$, otherwise set $\boldsymbol{\theta}^l = \boldsymbol{\theta}^{l-1}$
3. Drop the first L_b samples of the Markov chain.

The Metropolis algorithm uses a symmetrical proposal distribution q to propose new points to add to the chain. To decide whether to accept or reject proposed points, a function $f(\boldsymbol{\theta})$ is used that is proportional to the target distribution $P(\boldsymbol{\theta})$. For each proposed point $\boldsymbol{\theta}^*$, the acceptance ratio $\alpha = f(\boldsymbol{\theta}^*)/f(\boldsymbol{\theta}^{l-1})$ is calculated, and the proposed point is accepted with the probability α. Note that if $f(\boldsymbol{\theta}^*) > f(\boldsymbol{\theta}^{l-1})$, the point is always accepted. This means that the chain will contain a higher amount of steps from areas with higher probability density and a few steps from areas with low probability density. When the number of steps grows large, the Markov chain becomes a better approximation of the target function $P(\boldsymbol{\theta})$. Values in the start of the generated chain are usually dropped to ensure that the chain represents a steady state approximation of the target distribution, the dropped steps are usually called the burn in period.

3.6.2 Adaptive Metropolis Algorithm

The Metropolis algorithm offers no guidance as to which proposal distribution to use, other than that it should be symmetric. Furthermore, there is a tradeoff between low and high dispersion of drawn proposals. A too low variance may lead to the chain getting stuck in areas of lower density while a too high variance may cause a very slow convergence rate. Haario et al. (2001) suggests the usage of an adaptive Metropolis (AM) algorithm for tackling these two problems. The algorithm uses a Gaussian proposal distribution and adjusts the variance in each step. This methodology leads to a chain that is non-Markovian, but the algorithm maintains the correct ergodicity properties if the target density is bounded from above and has a bounded support (Haario et al., 2001).

The proposal distribution $q(\cdot|\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_{l-1})$ used in the AM algorithm is a multivariate Gaussian distribution with the current point $\boldsymbol{\theta}^{l-1}$ as the mean and covariance $\Sigma_l = s_d \operatorname{cov}(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_{l-1}) + s_d \zeta I$. In the covariance expression, s_d is a parameter that only depends on the dimension of the target distribution, ζ is a small constant, and the second term of the expression ensures that Σ_l never becomes singular. The algorithm needs to be initialized by using a fixed covariance matrix for a number of iterations before applying the adaptive covariance matrix. This leads to formally stating the used covariance matrix as

$$\Sigma_l = \begin{cases} \Sigma_0, & l \leq l_0, \\ s_d \operatorname{cov}(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_{l-1}) + s_d \zeta I, & l > l_0. \end{cases} \tag{3.9}$$

Where Σ_0 is stated using best prior knowledge and l_0 is the length of the initialization period.

3.6.3 Approximate Bayesian Computation

When the likelihood of the loss data is intractable or unknown in closed form, as is the case with the g-and-h distribution, the approximate Bayesian computation (ABC) can be used to evaluate the posterior distribution. For the observed data set $\boldsymbol{x}$, a specific parametric distribution is chosen with the only requirement that data can easily be simulated from it using arbitrary parameters. A possible parameter vector $\boldsymbol{\theta}^*$ is then drawn from a proposal distribution and a data set $\boldsymbol{x}^*$ is simulated using these. The proposed parameter vector $\boldsymbol{\theta}^*$ is then accepted with some probability P if the distance between the summary statistic $S(\cdot)$ of the random sample $\boldsymbol{x}$ and the simulated sample $\boldsymbol{x}^*$ are smaller than ε. For a sufficient summary statistics, as $\epsilon \to 0$, for an appropriate distance function $\rho(\cdot,\cdot)$, $\pi_{ABC}(\boldsymbol{\theta}|\boldsymbol{x},\epsilon) \to \pi(\boldsymbol{\theta}|\boldsymbol{x})$ (Schevchenko, 2011). This methodology can be coupled with some MCMC method, and in the algorithm below, the ABC is coupled with the Metropolis algorithm.

1. Initialize $\boldsymbol{\theta}^0$
2. For $l = 1, ..., L$
   (a) Generate a proposal $\boldsymbol{\theta}^*$ from the proposal density $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{l-1})$.
   (b) Simulate a dataset $\boldsymbol{x}^*$ from the model with parameters $\boldsymbol{\theta}^*$.
   (c) Draw U from $\mathrm{unif}(0,1)$
   (d) If
   $$U < \min\left\{1, \frac{\pi(\boldsymbol{\theta}^*)}{\pi(\boldsymbol{\theta}^{l-1})} 1_{\{\rho(S(\boldsymbol{x}),S(\boldsymbol{x}^*))\leq\varepsilon\}}\right\},$$
   set $\boldsymbol{\theta}^l = \boldsymbol{\theta}^*$, otherwise set $\boldsymbol{\theta}^l = \boldsymbol{\theta}^{l-1}$
3. Drop the first L_b samples of the Markov-chain.

A sufficient summary statistic is a function of $\boldsymbol{x}$ which summarizes all available sample information about $\boldsymbol{\theta}$ (Schevchenko, 2011). $S(\cdot)$ can, for example, be defined as the following vectors:

- $S(\cdot) = (\hat{\mu}, \hat{\sigma}, \hat{\gamma}_1, \hat{\gamma}_2)$ where $\hat{\mu}$ is the empirical mean, $\hat{\sigma}$ the empirical variance, $\hat{\gamma}_1$ the empirical skewness, and $\hat{\gamma}_2$ the empirical kurtosis.
- $S(\cdot) = (x_{(1)}, x_{(2)}, ..., x_{(N)})$, i.e. the sorted vector $\boldsymbol{x}$.

4 Data Material

4.1 Loss Severity Data

4.1.1 Internal Loss Severity Data

The internal data comes from the company's internal database where managers in the organization should report all incidents that may lead to an operational loss. Incidents with an actual loss are initially used, and there is no lower reporting limit. The losses are classified according to Basel II event type category level 2, but due to scarcity in data, only level 1 event types are used.

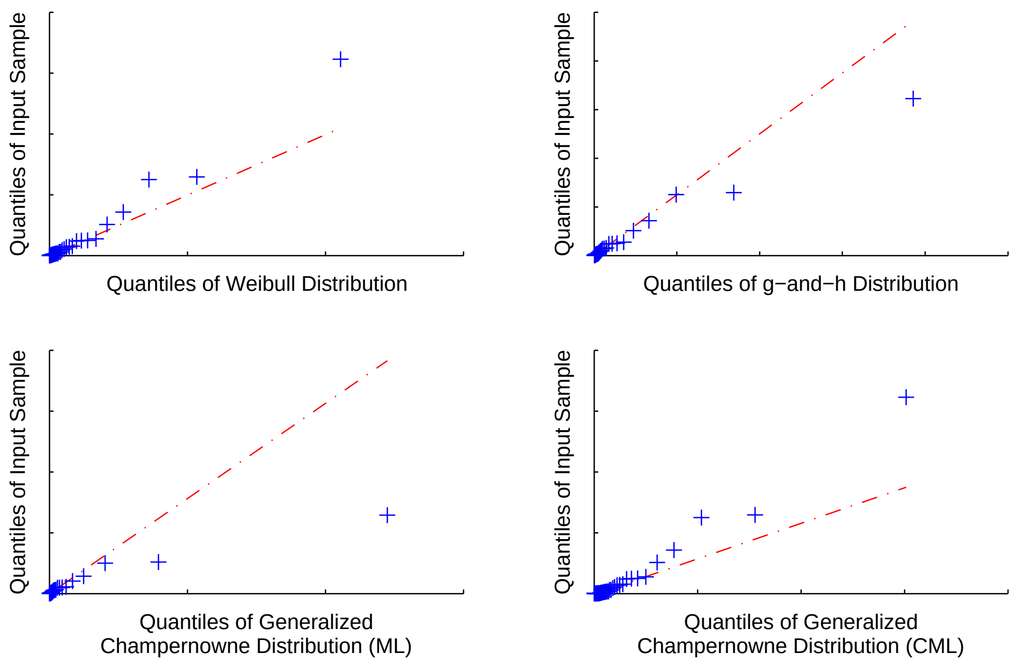

Figure 1: *QQ-plots of all internal loss severity data*

Table 4: *AIC and BIC for each distribution fitted to the internal data*

	Weibull	g-and-h	GCD	GCD CML
AIC	653.81	688.48	654.19	682.28
BIC	657.71	696.29	660.05	688.13

According to the AIC and BIC in table 4, the distribution that fits the data sample best overall is the Weibull distribution, but the generalized Champernowne are not too far off. Furthermore, figure 1 shows that the tail behavior of the observed data is best described by the g-and-h distribution or the Weibull distribution. Other distributions were also examined but left out. For example, the log-normal did have a slightly better AIC and BIC, but a terrible tail fit and was therefore excluded.

4.1.2 External Loss Severity Data

The external data used in this case study stems from a commercial data pool and consists of approximately 155,000 losses that took place in the retail banking business line. The data includes loss severity in Euro, date of occurrence, level 2 event types, level 2 business lines, a five level measure of institution size, and rough geographic location of each loss. The normal thing to do would be to try and scale all loss severities based on factors assumed to drive the number and severity of losses. Examples of such factors are geographic location of incidents, yearly revenue of bank or business line, company asset values, number of employees, et cetera (Petersson and Svensson, 2013). However, the data at hand is only a sample from the commercial database and lacks all such parameters that could be used for scaling, forcing the use of data where the geographic area and company size matches the company in the case study, leaving approximately 4,100 observations. In this sample, there seems to be a lower reporting threshold, since the lowest value is around 1,200 Euros.

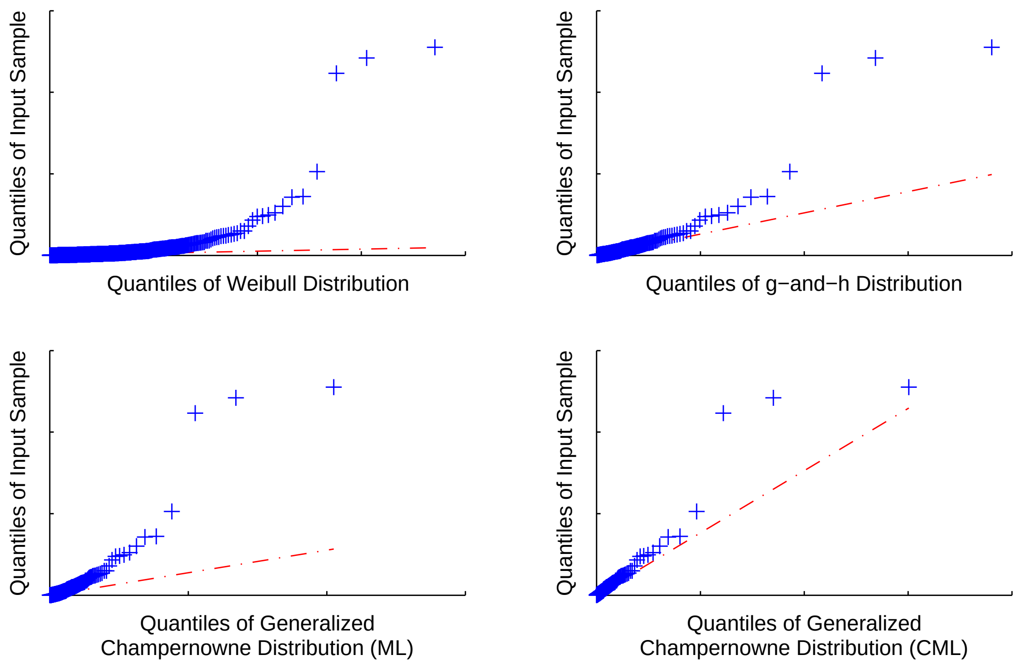

Figure 2: *QQ-plot of all external loss severity data*

Table 5: AIC and BIC for each distribution fitted to the external data

	Weibull	g-and-h	GCD	GCD CML
AIC	53445	51583	52187	52417
BIC	53458	51608	52206	52436

According to table 5, the best overall fit for the aggregated external severity data is the g-and-h distribution. When each event type is examined individually in table A.1, five event types is best modeled by the the g-and-h distribution and two by the generalized Champernowne distribution. Figure 2 shows that the tail is best represented by the g-and-h distribution or the generalized Champernowne distribution estimated with CML. QQ-plots for each separate business line are shown in figures A.1-A.7, found in appendix A. They also supports the use of either the generalized Champernowne distribution estimated with CML or the g-and-h distribution.

Table 6: Number of external observations per event type.

ID	Event Type	Observations
1	Internal Fraud	216
2	External Fraud	2065
3	Employment Practices and Workplace Safety	120
4	Clients, Products, and Business Practices	401
5	Damage to Physical Assets	107
6	Business Disruption and System Failures	127
7	Execution, Delivery, and Process Management	1066

Finally, the distribution of loss events among different event types are shown in table 6, and it is clear that data sufficiency will hinder us from drawing sharp conclusions about some of the event types.

4.1.3 Scenario Loss Severity Data

The scenario data is generated by the company's risk management team in cooperation with department managers and consists of a large number of what-if scenarios describing risks in the daily operations. Each scenario is classified in accordance with the event types specified in the Basel II rules and has a loss severity estimated by expert judgment.

Table 7: *AIC and BIC for each distribution fitted to the scenario data*

	Weibull	**g-and-h**	**GCD**	**GCD CML**
AIC	3360.9	3535.9	3382.1	3406.4
BIC	3367.7	3549.4	3392.3	3416.5

According to table 7 the best overall fit among the evaluated distributions is offered by the Weibull distribution. Furthermore, the QQ-plots in figure 3 suggests that the Weibull distribution fits the tail best as well and that the second-best option is the generalized Champernowne distribution estimated with CML.

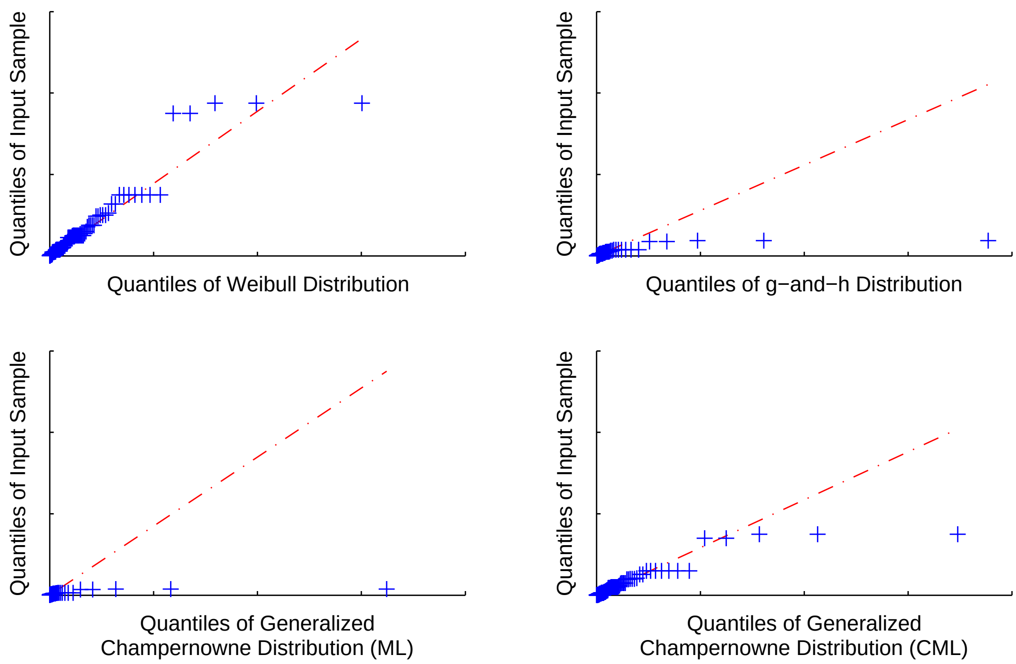

Figure 3: *QQ-plot of scenario data*

4.1.4 Comparison Between Data Sets

The Basel II rules force the use of external and scenario data as a proxy for internal data. In this section, a graphical justification for this methodology is given.

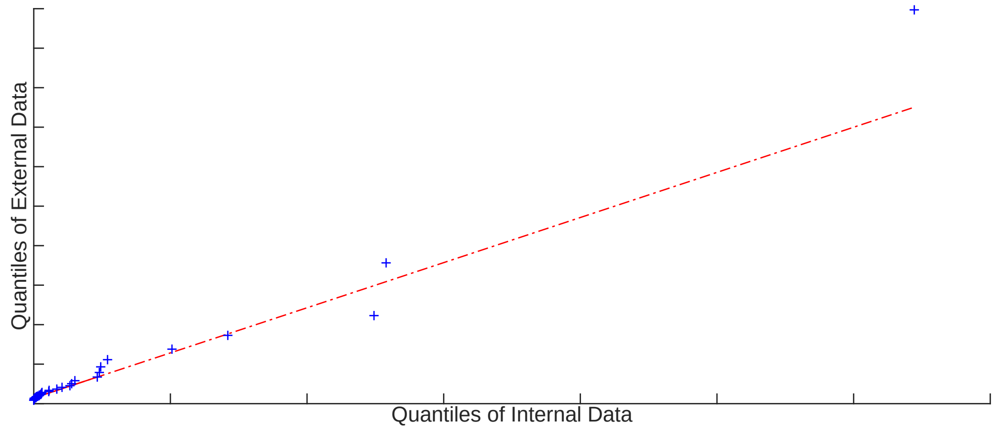

Figure 4: *QQ-plot comparing the empirical quantiles of internal versus external data*

Figure 4 visualizes how well the distribution of the internal and external data aligns. It shows that the tail of the external data is slightly thicker than that of the internal data. It is, however, worth noting that the internal data is comprised by less than 100 points, making it hard to draw any strong conclusions about how the tails of these heavy-tailed data sets compare.

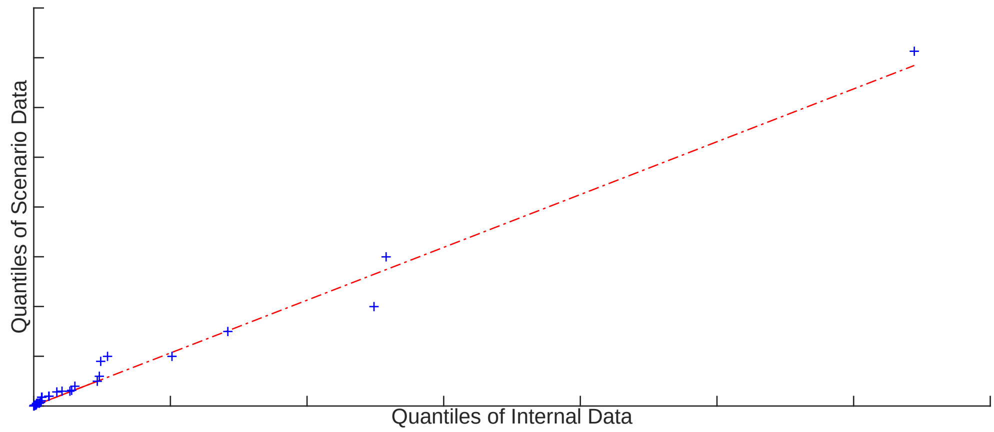

Figure 5: *QQ-plot comparing the empirical quantiles of internal versus scenario data*

According to figure 5, the internal and scenario data align quite well in terms of distribution shape.

4.1.5 Choice of Parametric Loss Severity Model

As seen in sections 4.1.1-4.1.3, there is no obvious choice of parametric model. Internal and scenario data suggest the usage of the Weibull distribution, while the external data suggests the use of either the g-and-h or the generalized Champernowne distribution. Due to these differences among the data sets, all three distributions will be tried out in the proposed model below.

4.1.6 Shifting Loss Data

In the external data set, there is a lower reporting threshold, meaning that all loss severities that have occurred below this level are missing from the data set. One way of dealing with this kind of truncated data set is to try and fit a truncated distribution to it. Another method is to assume that there are no losses below the threshold and to therefore subtract the threshold from the data set and, when sampling from the distribution, adding the threshold back to each sample (Ergashev et al., 2014). There is a discrepancy between the three data sets since the internal data set contains observations below the threshold but the external and the scenario sets do not. In this paper, this is handled by excluding the lower internal observations of the internal data. This approach can be justified as a method of shutting out noise from the sample, since it is fair to assume that managers in the organization are more hesitant on reporting small events, resulting in a large reporting bias for smaller loss sizes. After truncating the internal data, all three sets are shifted and the support of the data is now $\mathbb{R}^+$.

4.2 Loss Frequency Data

Figure 6 shows the distribution of the number of daily events for one event type in the external data used. Popular choices of parametric distributions to model the loss frequency are the Poisson distribution and the negative binomial distribution, both fitted in the figure. It is shown that the negative binomial distribution offers a somewhat better fit over the Poisson distribution, but the latter is favored in the modeling later on, since there exists a suitable conjugate pair and since it seems to be the best practice in the industry.

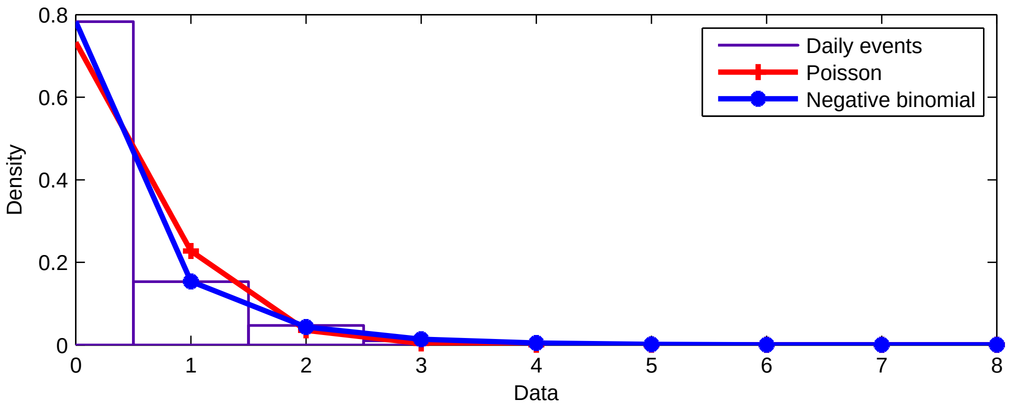

Figure 6: *Probability mass function of daily events for one of the event types in the external data with fitted negative binomial and poisson distributions*

Figure 7 shows the same as its counterpart above, but since only three years-worth of data exists, the plot is a lot more ambiguous. It is shown that there are no instances where more than one event take place on one day. In this figure, there is no visible difference between the fitted Poisson and negative binomial distribution.

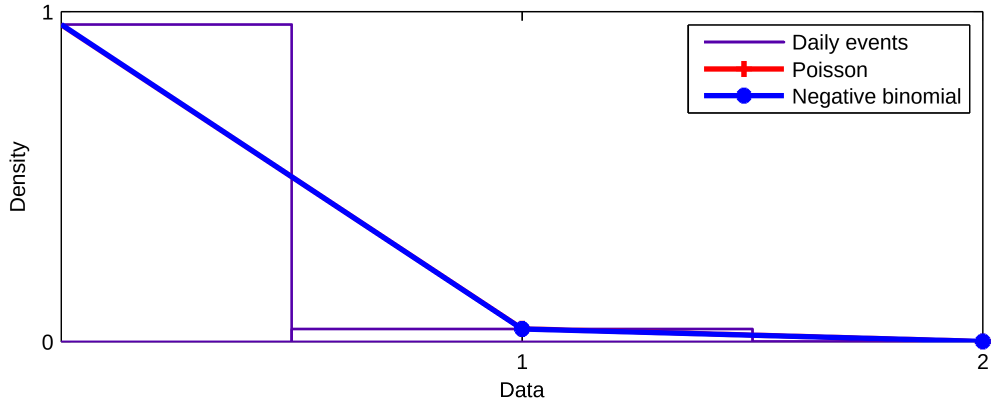

Figure 7: *Probability mass function of daily events in the internal data with fitted negative binomial and poisson distributions*

Finally, in figure 8, the QQ-plot of a fitted Poisson distribution is shown, and once again, only three years-worth of observations are available, but they align quite well with the fitted distribution.

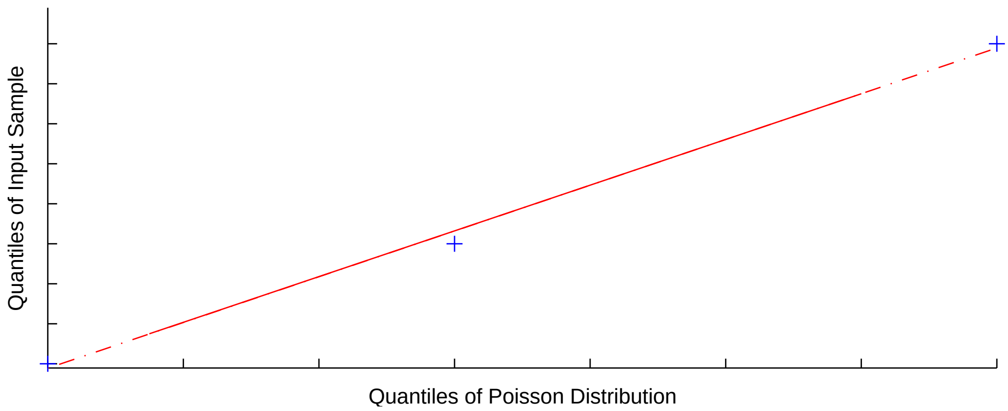

Figure 8: *QQ-plot of the yearly internal loss frequency*

The scenario data was elicited through a workshop in which relevant managers at the company came to a consensus regarding the expected number of events and the variance of this estimate for each event type.

5 Procedure

Schevchenko (2011) proposes the use of Bayesian inference to combine different data sources in order to estimate reliable frequency and severity distributions. In the case of severities, external data is used to estimate an empirical, less subjective prior. The posterior is then formed by adding the likelihood of internal and scenario data given the prior distribution of model parameters. This is a modification of the author's initial approach in which only two data sets were combined, and the third added using some other methodology. One example is forming the prior distribution using scenario analysis with external events as input, then updating using internal events to form the posterior distribution. The developed approach with all three sources was chosen since it fits best with the examined company's current risk management methodology. In the frequency case, the external data cannot be used to form the prior since it would give an estimate of the whole data pool's loss frequency. Instead, scenario data is used to form a prior, and the likelihood of internal observations given this prior is then used to form the posterior distribution. Once both the severity and frequency distributions are fitted, these are combined in equation (2.3) to form the distribution of yearly losses for each risk cell. Finally, Monte Carlo simulations are carried out to get the value-at-risk for each risk cell.

5.1 Modeling the Loss Frequency Distribution

Since the data presented in section 4.2 suggests that a Poisson distribution describes the loss frequencies well, and because a suitable conjugate pair exists, a Poisson model is used. The conjugate pair has a gamma distribution as prior and the Poisson model as the likelihood resulting in a gamma distributed posterior. The parameters of the prior gamma distribution are estimated from the elicited expert opinions on expected yearly frequency and variance of estimate. Then the analytical expression of the posterior is derived as below using the fact that the gamma and Poisson distributions form a conjugate pair. Finally, the new expected value of the frequency distribution is used in the VaR_α simulations.

Let the prior distribution of Λ be defined by the gamma distribution

$$\pi(\lambda) = \frac{(\lambda/\beta)^{\alpha-1}}{\Gamma(\alpha)\beta}\exp(-\lambda/\beta), \quad \lambda > 0, \alpha > 0, \beta > 0.$$

And the data be Poisson distributed given $\Lambda = \lambda$

$$h(\boldsymbol{n}|\lambda) = \prod_{i=1}^{T} e^{-\lambda}\frac{\lambda^{n_i}}{n_i!}.$$

Inserting this into 3.7 results in

$$\pi(\lambda|\boldsymbol{n}) \propto \frac{(\lambda/\beta)^{\alpha-1}}{\Gamma(\alpha)\beta}\exp(-\lambda/\beta)\prod_{i=1}^{T} e^{-\lambda}\frac{\lambda^{n_i}}{n_i!} \propto \lambda^{\alpha_T-1}exp(-\lambda/\beta_T).$$

Which again is gamma distribution with parameters

$$\alpha_T = \alpha + \sum_{i=1}^{T} n_i, \quad \beta_T = \frac{\beta}{1+\beta T}.$$

With the expected value given the observed data expressed as

$$E[\lambda|\boldsymbol{N} = \boldsymbol{n}] = \alpha_T\beta_T$$

The prior and resulting posterior distribution for one of the event types is shown in figure 9 below.

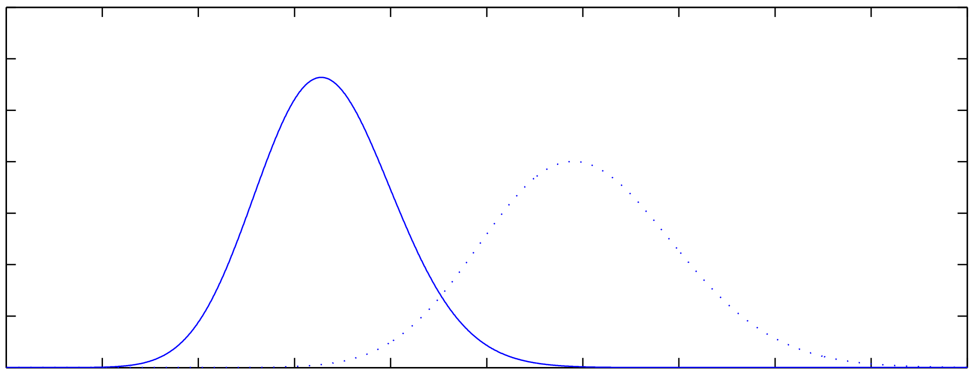

Figure 9: *The prior frequency distribution in dotted and the posterior in solid.*

5.2 Modeling the Loss Severity Distribution

For a choice of parametric distribution, the company's loss distribution is described by the parameter vector $\boldsymbol{\theta}$, commonly referred to as the company's risk profile. In the Bayesian context, the company's risk profile is treated as a realization from a random stochastic vector $\boldsymbol{\Theta}$ indicating that the true value of $\boldsymbol{\theta}$ is uncertain (Lambrigger et al., 2007). Now assume that $\boldsymbol{\Theta}$ is a stochastic vector describing the market's risk profile, and each company i in the market has their own realization, $\boldsymbol{\theta}_i$, of the market's risk profile. This means that a prior distribution can be formed by estimating the distribution of a set $\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ..., \boldsymbol{\theta}_N\}$ of risk profile realizations. In the ideal case, a large sample size from a large number of companies would be available, but the losses in the available data set cannot be accredited to specific companies. Instead, a parametric bootstrap procedure is used in order to construct a prior distribution $\pi(\boldsymbol{\theta})$ from the available data.

Judging from the QQ-plots and IC values in section 4.1.1, the internal data is well-modeled by the Weibull distribution. Section 4.1.2, on the other hand, suggests the use of either the generalized Champernowne or g-and-h distribution for the external data. Each distribution poses its own specific challenges, with Weibull being the least complicated to implement and the g-and-h distribution the most complicated. This is the case since no suitable conjugate pair exists for any of the distributions, but an analytical likelihood function exists for both the Weibull and the generalized Champernowne distributions. This calls for the use of MCMC to evaluate the posterior distribution and, due to the third parameter, the generalized Champernowne distribution poses more of a challenge than the Weibull distribution. For the g-and-h distribution, no analytical likelihood function exists. Therefore, one option is to use ABC coupled with MCMC to evaluate the posterior, adding another difficulty to the implementation. Another option is to use the numerical likelihood function described in section 3.2.5 together with MCMC. One way of choosing which model to use would be to just rely on Occam's Razor and go with the model that requires the least amount of assumptions, in this case, the Weibull option. However, since the external data favors the g-and-h and the generalized Champernowne distributions, an attempt on each distribution is made with varying results. For each attempt, a prior is formed by applying a parametric bootstrap with $M = 10,000$ for each event type individually. The generalized Champernowne and g-and-h distributions need some further work than the steps suggested in section 3.4, and this is described below.

5.2.1 The Generalized Champernowne Prior

The initial parametric estimate could be estimated with ML or CML. Both estimating methodologies are tried out. When the ML was used, the resulting bootstrap sample $\hat{c}^*$ needed some further corrections in order for a parametric model to be fitted. By taking the logarithm of the c parameter, a bimodal distribution like the one in figure 10 was obtained. Using the left part of this distribution would be equivalent to assuming that c is more or less zero, while using the right part would allow c to obtain higher values. Since we are trying to model the prior as a multinormal distribution, the bootstrap is split into two models, one with the prior distribution of c close to zero and one with a prior distribution of c that allows for higher values. Furthermore, the bootstrapped α parameter tended to have some high outlier values for the event types where the sample size was small, which were left out before fitting the prior distributions. This procedure left two multivariate prior distributions, one with high probability of c values close to zero and one where higher values of c are more likely. Finally, the usage of the CML to make the initial estimate of the parametric model resulted in erratic behavior of the bootstrap estimate and were therefore abandoned.

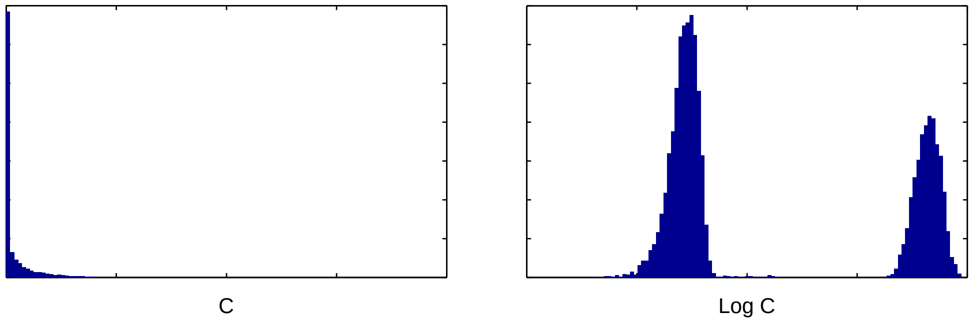

Figure 10: *Distribution of c and log c from bootstrap.*

5.2.2 The g-and-h Prior

In the g-and-h case, the quantile estimation approach is used to fit the distribution to the data in each event type. Then M samples of the same size as the original set is drawn and fitted with the same approach giving a bootstrap sample $\hat{\mathbf{\Theta}}^*$ of the market risk profile. The prior distribution is then well modeled by a multivariate normal distribution without any further transformations or other corrections. The prior distribution of the four parameters for one event type is shown in figure 11. The distribution is shifted to mask confidential data.

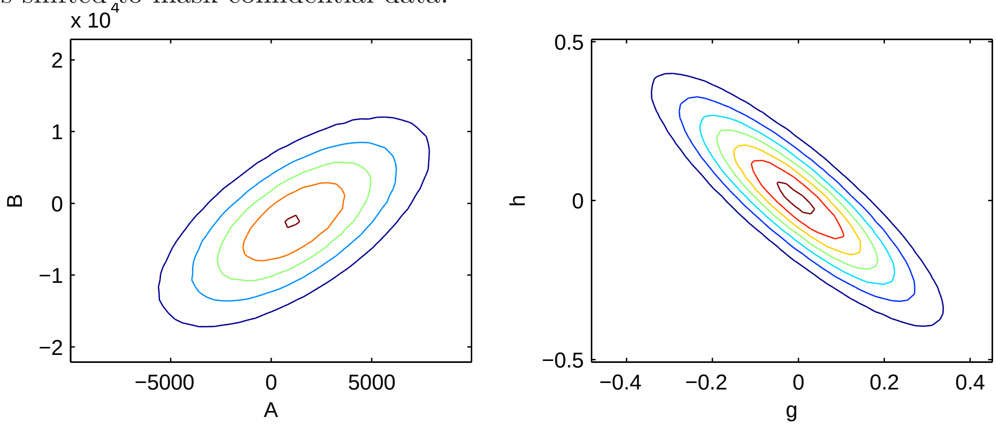

Figure 11: *The prior distribution for one of the event types with shifted to origin for masking purposes.*

5.2.3 Combining Internal Data and Scenario Analysis

Schevchenko (2011) suggests forming $h(\boldsymbol{x}, \boldsymbol{\delta}|\boldsymbol{\theta})$ by assuming that the internal data $\boldsymbol{x}$ and scenario data $\boldsymbol{\delta}$ are independent given $\boldsymbol{\Theta} = \boldsymbol{\theta}$, i.e.

$$h(\boldsymbol{x}, \boldsymbol{\delta}|\boldsymbol{\theta}) = h_1(\boldsymbol{x}|\boldsymbol{\theta})h_2(\boldsymbol{\delta}|\boldsymbol{\theta}), \tag{5.1}$$

where each conditional density can be expressed as likelihood functions

$$h_1(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_{k=1}^{K} f_1(x_k|\boldsymbol{\theta}) = \mathcal{L}(\boldsymbol{x}|\boldsymbol{\theta}) \tag{5.2}$$

$$h_2(\boldsymbol{\delta}|\boldsymbol{\theta}) = \prod_{m=1}^{M} f_2(\delta_m|\boldsymbol{\theta}) = \mathcal{L}(\boldsymbol{\delta}|\boldsymbol{\theta}). \tag{5.3}$$

This works well for the Weibull and generalized Champernowne distributions but not for the g-and-h distribution, due to the lack of analytical expression for the likelihood function. Instead, the ABC method or the numerical likelihood approach can be used to evaluate the posterior. When ABC is used, equation (5.1) expresses that scenario data and internal data should be combined into one set of actual and hypothetical losses before applying the ABC.

5.2.4 Forming and Evaluating the Posterior

The posterior is now formed in order to combine the internal, external and scenario data into one parameter estimation of the chosen model. The following is the expression for the posterior density:

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}, \boldsymbol{\delta}) \propto \pi(\boldsymbol{\theta})\mathcal{L}(\boldsymbol{x}|\boldsymbol{\theta})\mathcal{L}(\boldsymbol{\delta}|\boldsymbol{\theta}). \tag{5.4}$$

Compared to (3.7), the denominator is left out here since only the shape of the posterior distribution is needed and not the scale. The most efficient way of evaluating (5.4) is via some numerical integration method, and to ensure quick convergence, the adaptive Metropolis algorithm is used. Once the posterior distribution is approximated, the median in each dimension is used as the point estimate of the parameter vector.

5.2.5 Approximating the Posterior Distribution for the Weibull Model

The Weibull model has a prior that is well-behaved, and a known likelihood function. It is therefore quite straightforward to apply a MCMC algorithm to approximate the posterior distribution. In this paper, the adaptive

Metropolis algorithm is used since it offers a proposal distribution that has been shown to be very efficient and is easily tuned. This means that a multivariate Gaussian distribution is used as the proposal distribution, i.e:

$$q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{l-1}) = \mathcal{N}(\boldsymbol{\theta}^{l-1}, \Sigma_l)$$

where Σ_l is updated in each iteration to the covariance of the previous steps as described in section 3.6.2. As for the scaling factor, Haario et al. (2001) suggest using $s_d = 2.4^2/d = 2.88$ in (3.9) since it has been shown to be somewhat optimal for a Gaussian target distribution. As the initial Σ_0, the covariance matrix of the prior distribution is used to give reasonable initial dispersion of steps. Furthermore, the stability part is left out, i.e. $\zeta = 0$ since the algorithm ran fine without it. The algorithm is run with a burn-in period of 1,000 iterations and then an initialization period $l_0 = 200$ iterations for the adaptive updating of Σ_l. The overall iteration length were set to 100,000 iterations to ensure a high resolution of the posterior distribution, and the resulting acceptance rates were just shy of 40 %.

5.2.6 Approximating the Posterior Distribution for the Generalized Champernowne Model

There exist two prior distributions for each event type for the generalized Champernowne model due to the bimodal property discovered in the bootstrap process. For each event type and for each of these two models, the posterior distribution is evaluated using the adaptive Metropolis algorithm. This is done in the same way as for the Weibull model but with $s_d = 2.4^2/d = 1.92$. Otherwise, the settings were the same, and the resulting acceptance rates were approximately 36 %.

5.2.7 Approximating the Posterior Distribution for the g-and-h Model

Since the g-and-h distribution does not have an analytical expression of the likelihood function, other approaches are needed.

First off, the approximate Bayesian calculation method coupled with the adaptive metropolis algorithm is used to evaluate the posterior distribution. The adaptive algorithm is applied in the same way as described above in section 5.2.5, and ABC is coupled in the way described in section 3.6.3.

Allingham et al. (2009) suggests using a summary statistic $S(\boldsymbol{x})$ that is quantile-based and accounts for all available quantiles, defined as $S(x) = \text{sort}(\boldsymbol{x})$. Furthermore, they suggest the usage of the Euclidean norm as the distance function defined by

$$\rho(S(\boldsymbol{x}), S(\boldsymbol{x}')) = \sqrt{\sum_{i=1}^{N} (S(x_i) - S(x_i'))^2},$$

and both of these suggestions are implemented.

With these functions defined, the tuning of the ABC algorithm is quite straightforward. First, the algorithm is run for a large number of iterations, in this implementation 10,000, for an ε high enough to accept almost all values of ρ. This gives an acceptance rate of about 37 % percent which would be in the desired range for an ordinary Metropolis algorithm and is therefore accepted. The histogram of the resulting values of ρ are shown in figure 12, an ε value in the far left part of the distribution is desirable, and therefore, this is set to the 5th percentile as seen in the figure. After this is done, the tuning is completed and the ABC algorithm is executed with 1,000,000 iterations per event type to evaluate the posterior distribution with sufficient accuracy and the resulting acceptance rates are around 4 %. The median for each parameter is then calculated as the point estimate of the parameter vector for each posterior distribution.

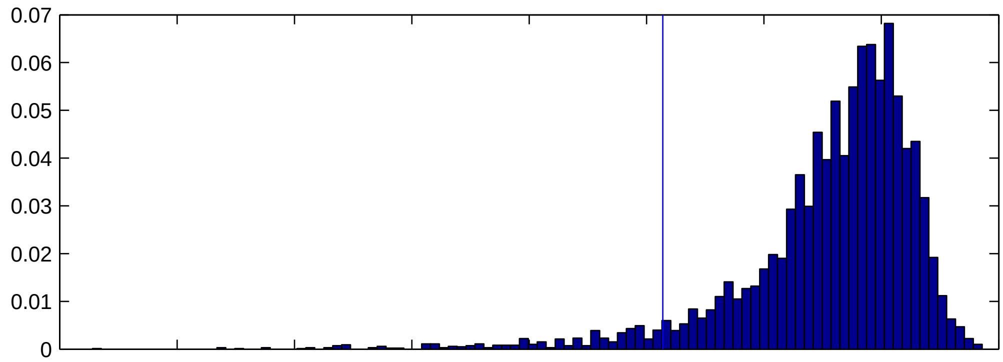

Figure 12: *Distribution of accepted ρ values from initial run with the 5th percentile marked.*

The second option is to use the numerical likelihood approach from section 3.2.5 and approximate the posterior with the adaptive Metropolis algorithm in the same way as for the Weibull and the generalized Champernowne model. The same settings are used but in this case, $s_d = 2.4^2/d = 1.44$. This approach is feasible since there are very few internal and scenario events, so that even if the likelihood calculation is very computationally heavy, it can be done in a reasonable amount of time. The resulting acceptance rates are approximately 29 %.

5.3 Combining Frequency and Severity Models to Form the Loss Distribution

At this point, there exist fitted frequency distributions and three different fitted severity distributions for all of the seven examined event types. These severity and frequency distribution pairs are now combined according to (2.3), and Monte Carlo simulations are performed for each risk cell to estimate the loss distribution. The number of years simulated is set to 1,000,000 to get more than a few estimates worse than the one in a thousand years level sought after. The simulations for each event type are carried out according to

1. Calculate posterior medians of $\pi(\boldsymbol{\theta}|\boldsymbol{x}, \boldsymbol{\delta})$ and mean of $\pi(\lambda|\boldsymbol{x})$.
2. For $y = 1, ..., Y$
   (a) Draw number of events N from Pois(λ_{mean})
   (b) Draw N losses from g-and-h($\boldsymbol{\theta}_{median}$)
   (c) Sum losses to get the aggregated yearly loss L and add to vector of yearly losses $\boldsymbol{L}$
3. Calculate the VaR as the 99.9th percentile of $\boldsymbol{L}$

Once VaR estimates for each included risk cell are available, these are simply added together to form the overall capital requirement estimate. The reason for doing this is since no estimation of correlation has been done due to scarcity in data, the Basel II rules state that perfect correlation should be assumed between the risk cells, i.e. the worst outcome in each cell will happen simultaneously.

6 Results and Conclusion

Since the data material used is too small to make an accurate choice of which parametric model to use, and since further testing of different alternatives needs to be performed with access to better data, results from all distributions are presented and analyzed with respect to each other and other factors. In appendix B, fitted prior distributions are shown for the Weibull, generalized Champernowne with c close to zero, generalized Champernowne with positive c and the g-and-h distribution. These are the priors that are then used with the methodology described in section 5 to evaluate five posterior distributions: the Weibull, the two generalized Champernowne models, the g-and-h estimated with ABC, and the g-and-h estimated with the numerical likelihood function. The posterior distributions are left out due to confidentiality reasons but an example of a masked posterior distribution for one event type of the g-and-h ABC model is shown in figure 13. The figure also shows how the point estimate of the severity parameters has changed from the prior to the posterior distribution.

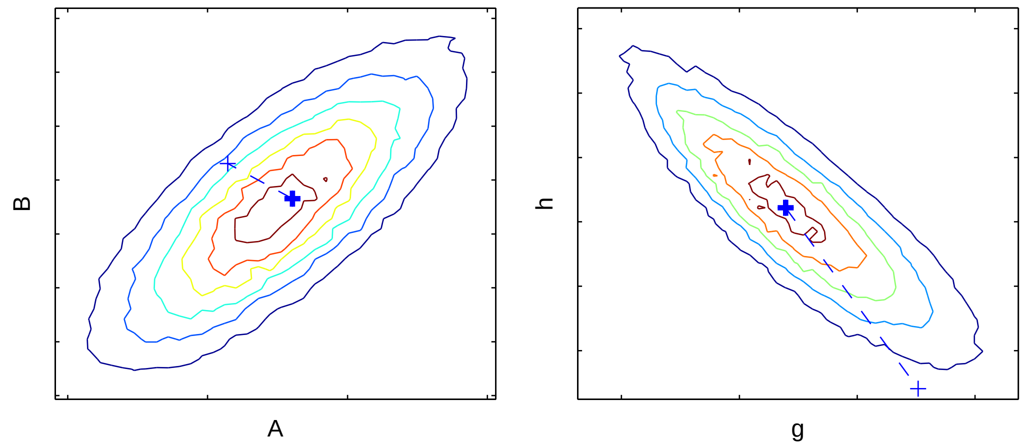

Figure 13: *Contour plot of the joint posterior distribution for the parameter pairs A,B and g,h for the g-and-h ABC model for one of the event types. The median of the prior distribution value is marked with a thin cross and the median of the posterior distribution with a thick cross.*

The point estimates of the posterior distributions for the frequency model and the five different severity models are used to simulate yearly losses to get the VaR levels. The combined VaR levels for each of the five implemented models are shown in figure 14, where it is obvious that the estimates diverge greatly. The Weibull model gives VaR estimates that are the lowest, and at the 99.9 percentile sought after, it gets rather close to the estimate given by the standardized approach that the investigated company currently employs.

This gives reason to suspect that the other models might not work well given the input data since the industry standard in Europe is that an AMA model produces a capital requirement that is a couple of percentage points lower than that of the basic indicator approach and standardized approach (Basel_Committee, 2009).

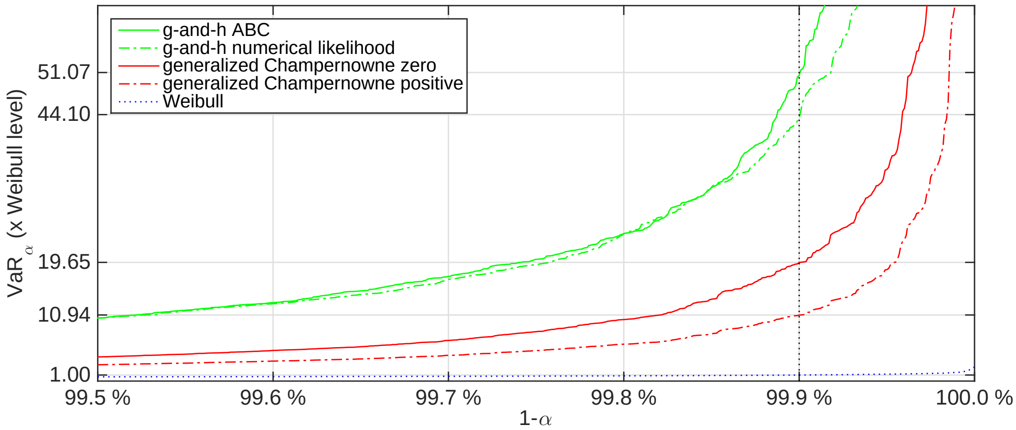

Figure 14: *Tail of VaR combined levels for each fitted model*

When examining the individual event types, it becomes obvious that one of them is contributing the majority of the capital to the generalized Champernowne and the g-and-h estimates. This might be due to over-sensitivity in these models that have a a higher number of parameters and are more closely related to extreme value theory. To illustrate the effect, figure 15 shows the VaR levels without the specific event type. In this figure, the difference compared to the Weibull model are a lot smaller, which points towards instability due to the small amount internal and scenario data.

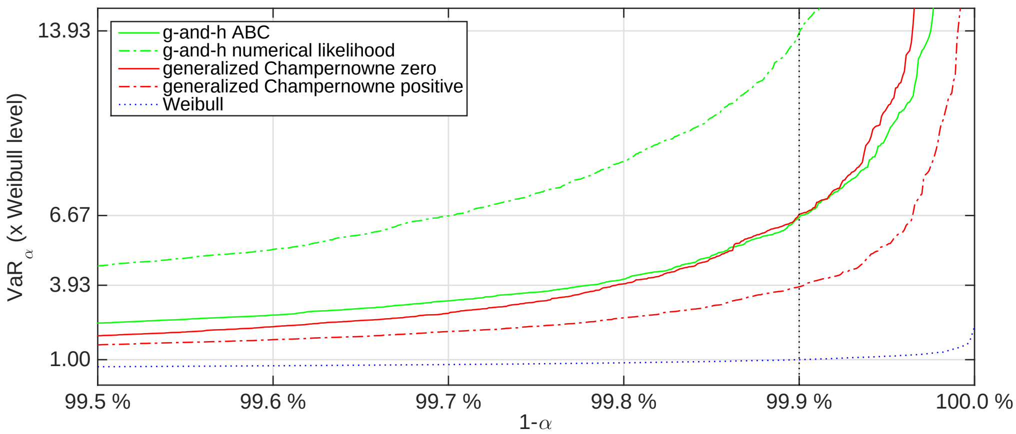

Figure 15: *Tail of VaR combined levels for each fitted model excluding one event type*

Furthermore, when using the median of the prior distribution as parameter values for the three severity distributions, the capital estimates lie a bit closer together. This is shown in figure 16, and in this case, both of the Champernowne-based estimates lie within a few percent of the current estimate based on the standardized approach.

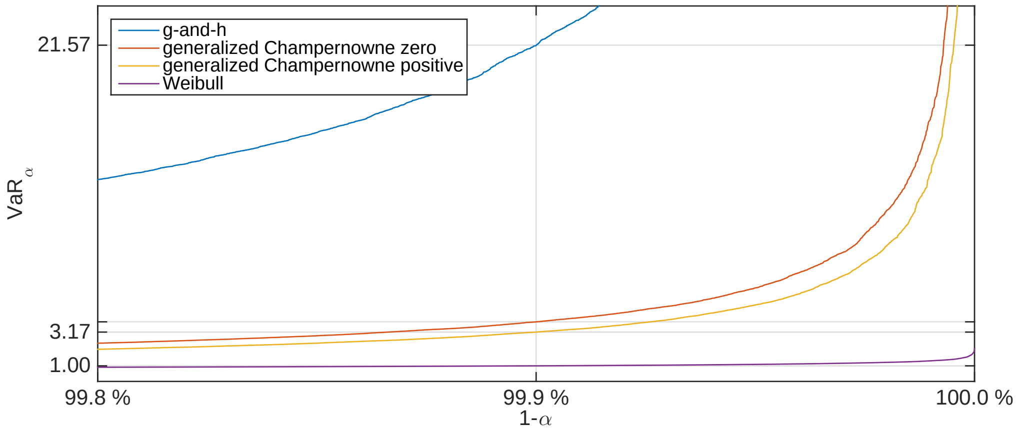

Figure 16: *Tail of VaR combined levels for the four prior severity distributions*

Due to scarcity in data, it is reasonable to suspect that most of the prior distributions are quite unreliable. In figure 17, the tail of the estimated VaR levels for event type 2 is shown, and it is clear that the spread among the different models is far more reasonable. This might indicate that, given better input data, the usage of the g-and-h or the generalized Champernowne distributions are viable options.

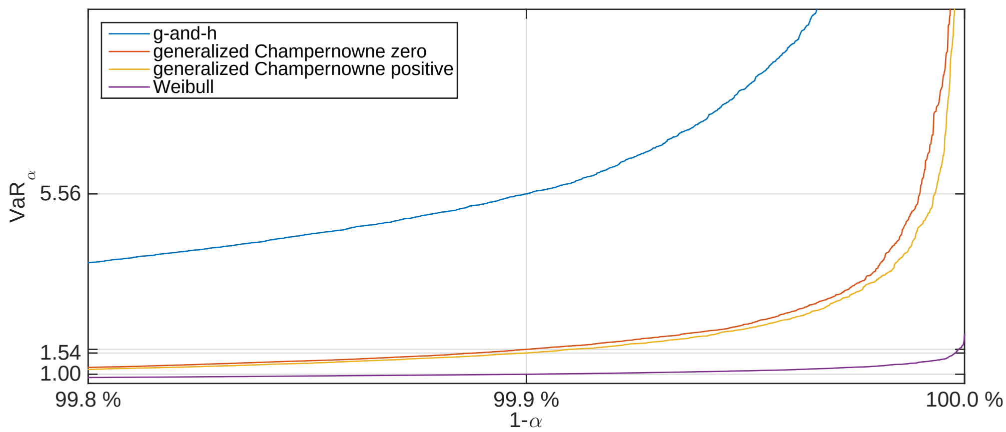

Figure 17: *Tail of VaR levels for event type 2 for the four prior severity distributions*

6.1 The Weibull Model

Since the Weibull model turned out to come quite close to the level given by the standardized approach, and because it responded much more consistently to different stability tests, this section is devoted to analyzing the behavior of the proposed Bayesian inference model implemented for the Weibull distribution. To give some insights into how different event types contribute to the estimated capital requirement, the tail portion of the estimated VaR levels for each event type are shown in figure 18. The sum of the 99.9th percentile of all seven event types gives a total estimated capital requirement that is roughly 70 % of the amount required by the standardized approach that the company currently uses.

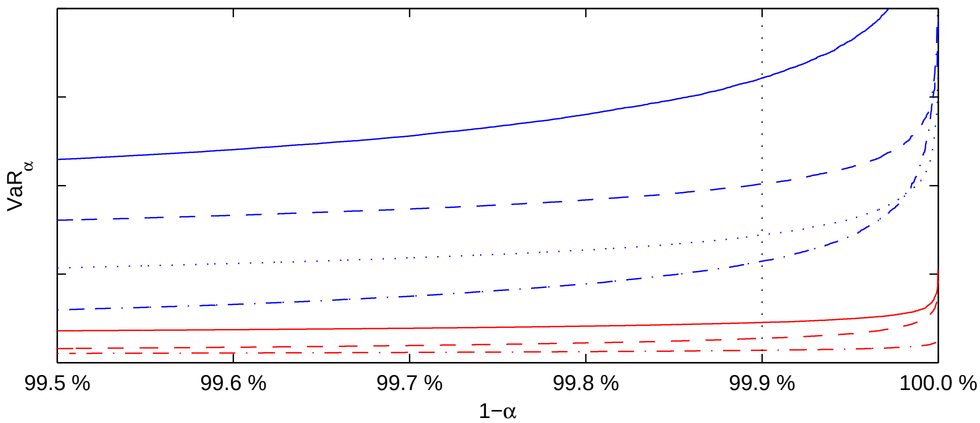

Figure 18: *The simulated yearly VaR for each event type in the Weibull model.*

6.1.1 Impact of External Events

To analyze the stability of the output from the Weibull model, a ten-multiplier stress test is performed for both the internal and the external data. The aim of the test is to see how the estimated capital requirement reacts to new events in different data sources. Since this test is more about seeing how the capital requirements change rather than the actual level, and because there is a shortage of data, each data source is first pooled so that all event types are combined into one. Then new observations are added with values that are multiples of the median of the set they are added to. For each test, the capital requirement is then estimated and compared to the original estimate to see what impact new events would have on the model output.

Table 8: *Ten-multiplier stress test of external losses*

Number of losses added	Size of loss	Change VaR (%)
1	1M	0.47
1	10M	0.42
1	100M	0.58
1	1,000M	2.61
1	10,000M	12.10
1	100,000M	42.80
10	1M	0.22
10	10M	-0.04
10	100M	4.17
10	1,000M	23.24
10	10,000M	92.91
10	100,000M	296.53

For the external data, the current maximum loss is about 900M, and table 8 shows that for one added event, there is not much change until the event added is 10 times as large as the current largest event. When 10 events are added instead, a similar response is seen but heavily amplified for larger events.

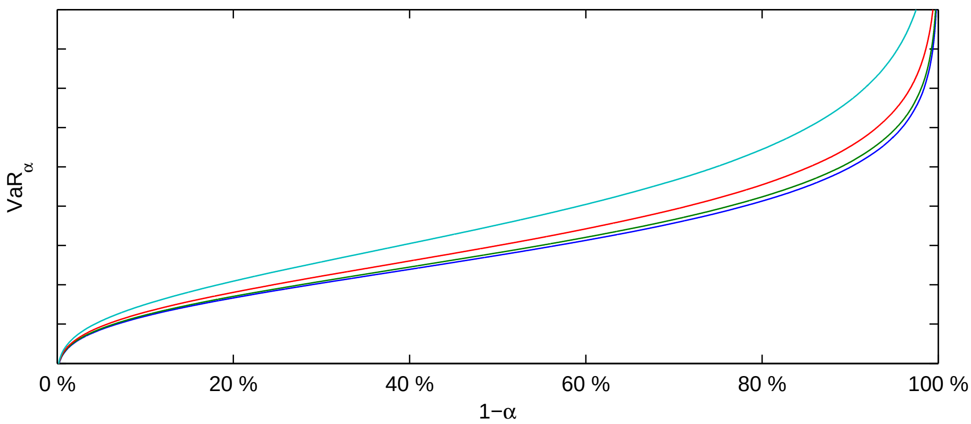

Figure 19: *VaR levels for stress test of the Weibull model with 1 external event added*

Figure 19 shows the VaR levels for the ten-multiplier stress test with one added event. The lowest amount is when no event is added, and the highest is when an event of size 100,000M is added.

Figure 20 shows the tail of the same stress test with corresponding capital levels shown indexed so that the baseline has index 100.

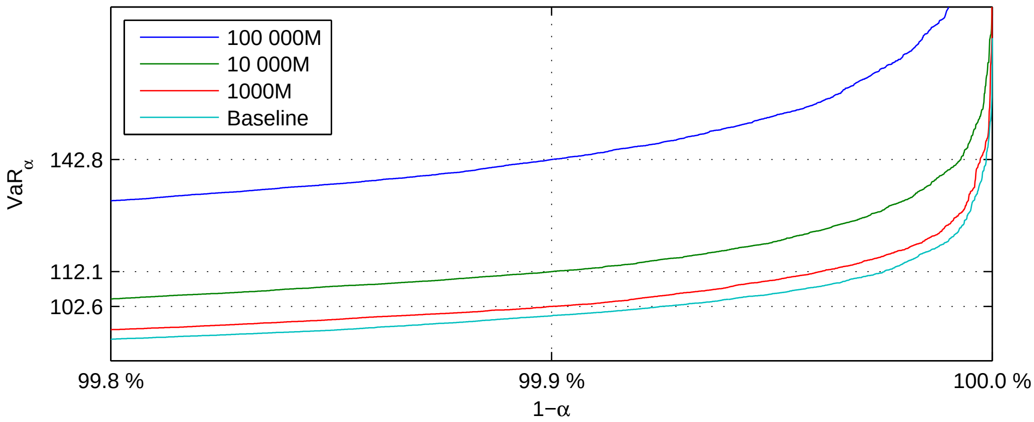

Figure 20: *Tail of VaR levels for stress test with 1 external event*

Finally, figure 21 shows the tail of the VaR levels for the stress test with 10 events added. This concludes the results from the stress test of the external data, and the overall conclusion is that the model responds to external events in a moderate way so that losses experienced in the industry will affect the regulatory capital, but for larger changes, extreme losses are required. This is very intuitive and contributes towards regarding this model as a viable option for modeling operational risk when internal data is scarce.

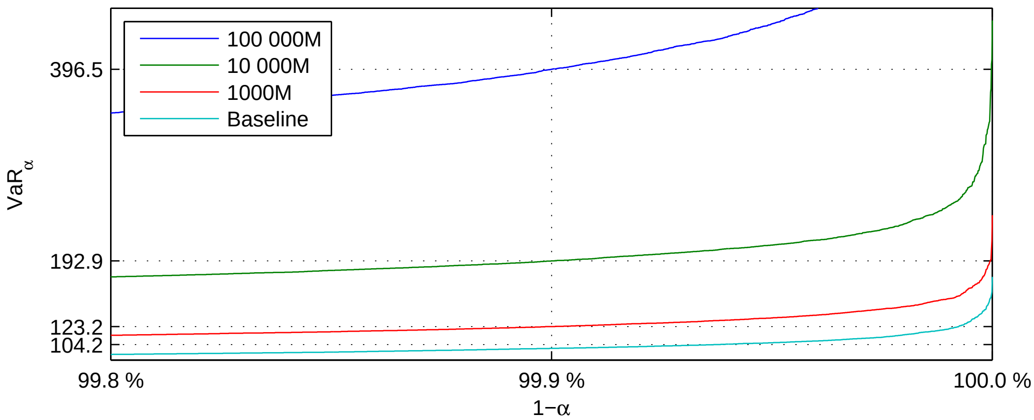

Figure 21: *Tail of VaR levels for stress test with 10 external events*

6.1.2 Impact of Internal Events

The same type of ten-multiplier stress test is performed for the internal data set but only for one added event, since there are so few internal events to start with. For the internal data, the maximum event is approximately 300M, and the current maximum event is only a fraction of the current required regulatory capital. A 10,000M event exceeds the current regulatory level slightly, so for this level, the test shows what would happen to the estimate if an event of the same magnitude as the current regulatory capital were to happen.

Table 9: *Ten-multiplier stress test of internal losses*

Number of losses added	Size of loss	Change VaR (%)
1	1M	0.20
1	10M	0.49
1	100M	0.72
1	1,000M	2.07
1	10,000M	11.20

Figure 22 gives a graphical representation of the stress test. This differs from the external test by changing less due to internal events. This is somewhat counterintuitive and might need further investigation. It might be a result of scarcity in the internal data leading to the majority of the estimated effects coming from the external data.

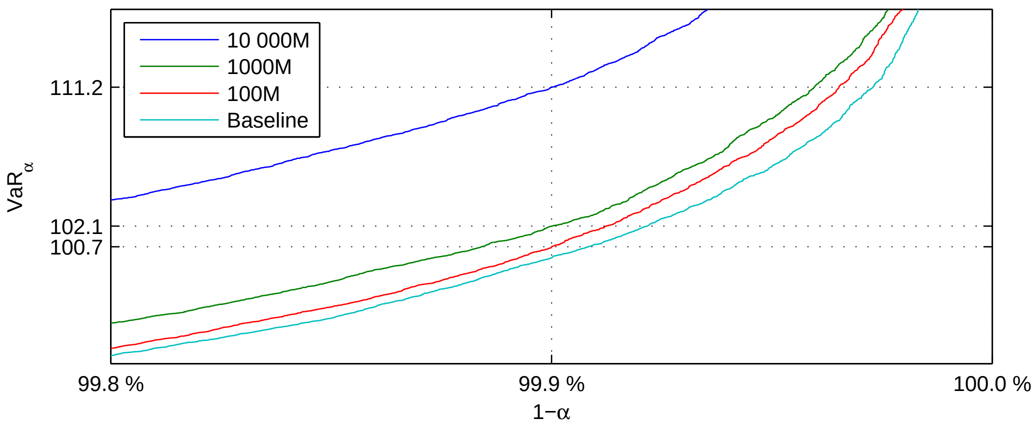

Figure 22: *Tail of VaR levels for stress test with 1 internal event*

6.2 Conclusion and Final Thoughts

The goal of this thesis was to investigate if it was possible to construct an AMA model when the number of internal events are extremely scarce. One viable methodology that fit well with the investigated company's current risk management framework was identified. Then, since the initial findings were inconclusive as to which parametric model to use for the severity distribution, the methodology was tried out with a number of different models and estimation techniques for these models. It is interesting that the output from one of the tested models came so close to the level of capital requirement given by the standardized approach. Furthermore, the stress testing showed that the proposed model reacted in moderation to new extreme events but with the appropriate change in estimated capital requirement.

Therefore, the main result of this thesis is that it is plausible to construct an AMA model using Bayesian inference, and that the Weibull severities and Poisson frequencies gave estimates that came close to those given by heuristics. To get a more robust model, it is, however, crucial to enhance the quality of the input data, gather the full 5 years of internal data, develop the scenario analysis so that it is better suited for the proposed model, and to properly identify which distribution gives realistic capital estimates and, at the same time, is the best fit for the internal data when more of it is available. It is also crucial to acquire external data that is richer in quantity and has parameters that enables scaling of external data to better align with the internal losses observed. Until this is done, it's impossible to say more than that it seems like this model is reasonable.

As a final remark, it should be noted that when access to proper external data is established, the model version where the prior severity distributions are formed using scenario analysis based on knowledge of the company and external events should be evaluated as well. This might be a better implementation for this type of company but has not been tested in this thesis due to the lack of suitable external data and the massive undertaking this would amount to for the company's risk control team.

References

Allingham, D., King, R. A., and Mengersen, K. L. (2009). Bayesian estimation of quantile distributions. *Statistics and Computing*, 19(2):189–201.

Basel_Committee (2009). Results from the 2008 loss data collection exercise for operational risk. Technical report, Bank of International Settlements.

Bolancé, C., Guillén, M., Gustafsson, J., and Nielsen, J. P. (2012). *Quantitative Operational Risk Models*. CRC Press.

Buch-Kromann, T. (2009). Comparison of tail performance of the champernowne transformed kernel density estimator, the generalized pareto distribution and the g-and-h distribution. *The Journal of Operational Risk*, 4(2):43–67.

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

Cruz, M. G. (2002). *Modeling, Measuring and hedging operational risk*. Chicester: John Wiley & Sons.

Dutta, K. and Perry, J. (2007). A tale of tails: An empirical analysis of loss distribution models for estimating operational risk capital. *Federal Reserve Bank of Boston, Working paper.*

Ergashev, B., Pavlikov, K., Uryasev, S., and Sekeris, E. (2014). Estimation of truncated data samples in operational risk modeling. April 2014 Draft.

Haario, H., Saksman, E., and Tamminen, J. (2001). An adaptive metropolis algorithm. *Bernoulli*, 7(2):223–242.

Hoaglin, D. C. (1985). Summarizing shape numerically: The g-and-h distributions. In Hoaglin, D. C., Mosteller, F., and Tukey, J. W., editors, *Exploring Data Tables, Trends, and Shapes*, pages 461–511. Hoboken: John Wiley & Sons.

Lambrigger, D. D., Shevchenko, P. V., and Wüthrich, M. V. (2007). The quantification of operational risk using internal data, relevant external data and expert opinions. *The Journal of Operational Risk*, 2(3):3–27.

Lambrigger, D. D., Shevchenko, P. V., and Wüthrich, M. V. (2008). Give credit where credit is due: Operational risk goes bayesian.

Petersson, J. and Svensson, M. (2013). Operativ risk: Skalning av extern förlustdata för kapitalkravsberäkning. Master's thesis, KTH.

Rayner, G. D. and MacGillivray, H. L. (2002). Numerical maximum likelihood estimation for the g-and-k and generalized g-and-h distributions. *Statistics and Computing*, 12(1):57–75.

Schevchenko, P. V. (2011). *Modelling Operational Risk Using Bayesian Inference*. Springer Berlin Heidelberg.

Shevchenko, P. V. and Peters, G. W. (2013). Loss distribution approach for operational risk capital modelling under basel ii: Combining different data sources for risk estimation. *The Journal of Governance and Regulation*, 2(3):33–57.

Appendices

A Goodness of Fit for the External Data by Event Type

Table A.1: *BIC for each fitted distribution fitted to the event types of the external data*

ET	Weibull	g-and-h	GCD	GCD CML
1	2982.5	2929.0	2931.4	2937.5
2	26428	25371	25686	25916
3	1540.4	1532.3	1530.6	1530.9
4	5284.5	5102.1	5133.6	5309.1
5	1372.0	1337.7	1352.8	1369.2
6	1670.4	1653.9	1649.3	1655.1
7	14122	13743	13882	13891

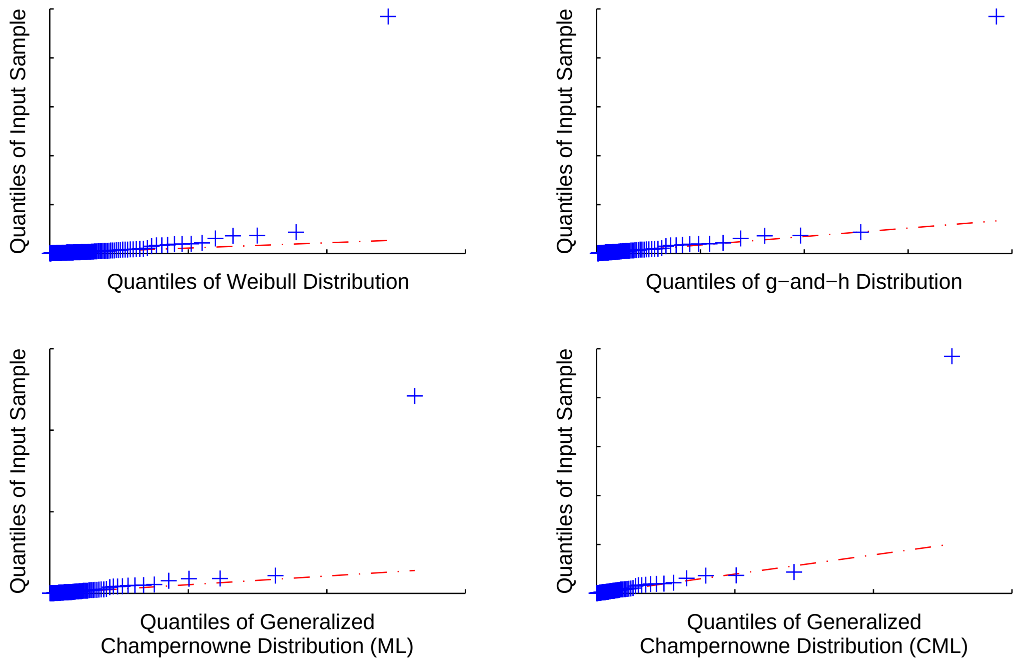

Figure A.1: *QQ-Plot of internal fraud*

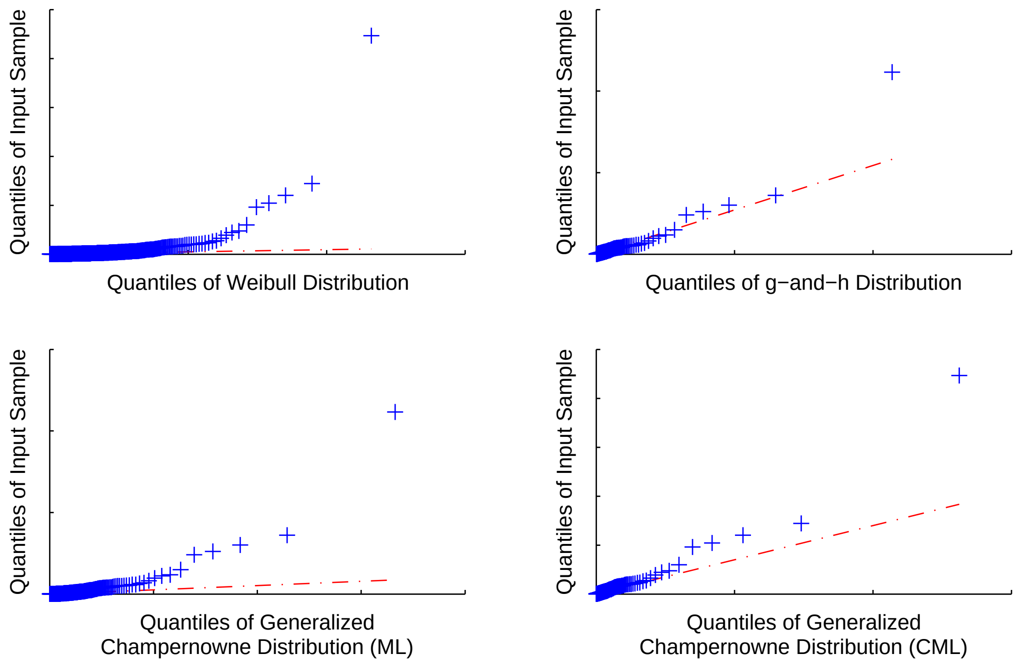

Figure A.2: *QQ-plot of external fraud*

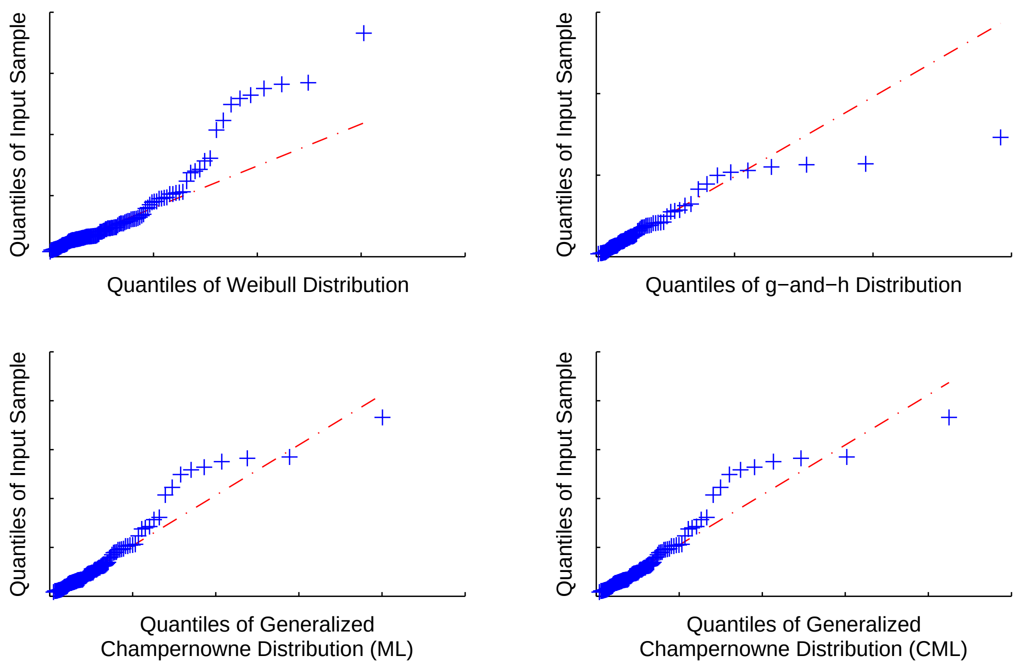

Figure A.3: *QQ-plot of employment practices and workplace safety*

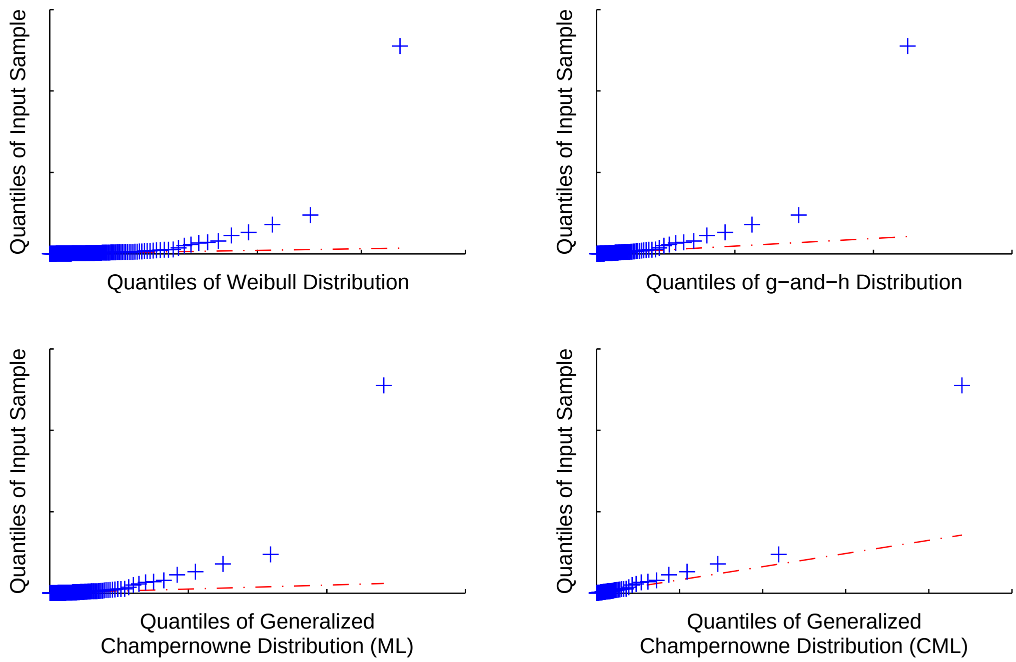

Figure A.4: *QQ-plot of clients, products and business practices*

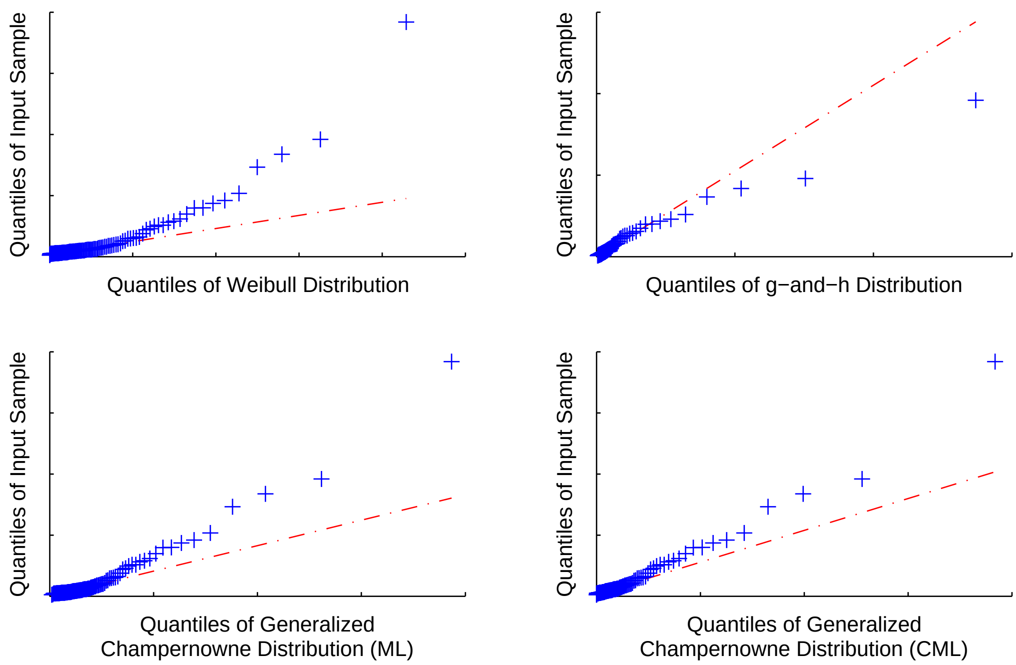

Figure A.5: *QQ-plot of damage to physical assets*

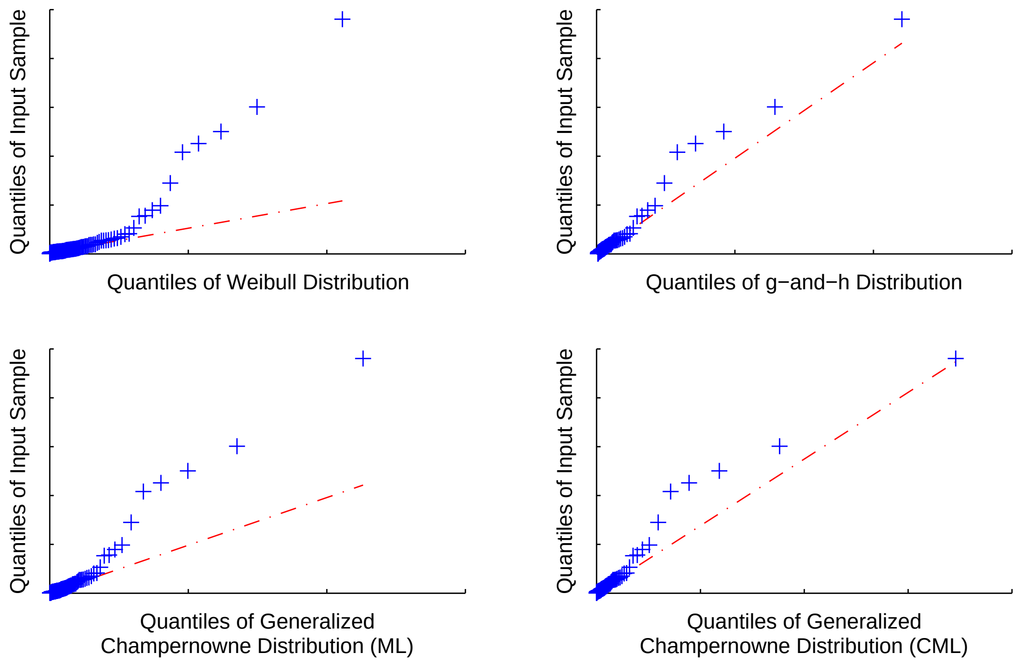

Figure A.6: *QQ-plot of business disruption and system failures*

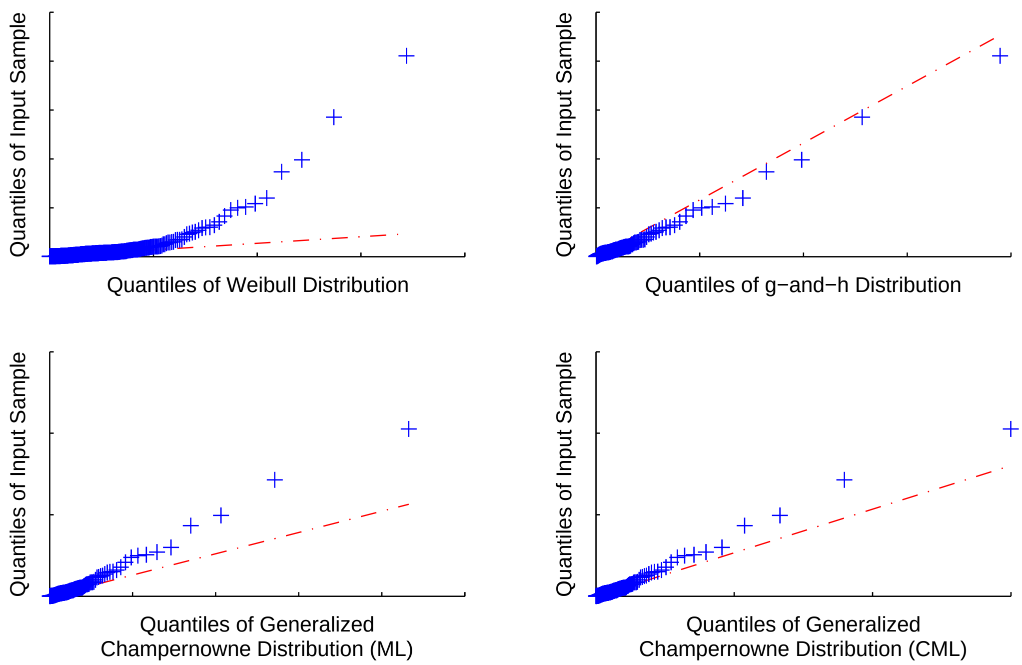

Figure A.7: *QQ-plot of execution, delivery, and process management*

B Prior Parameter Distributions

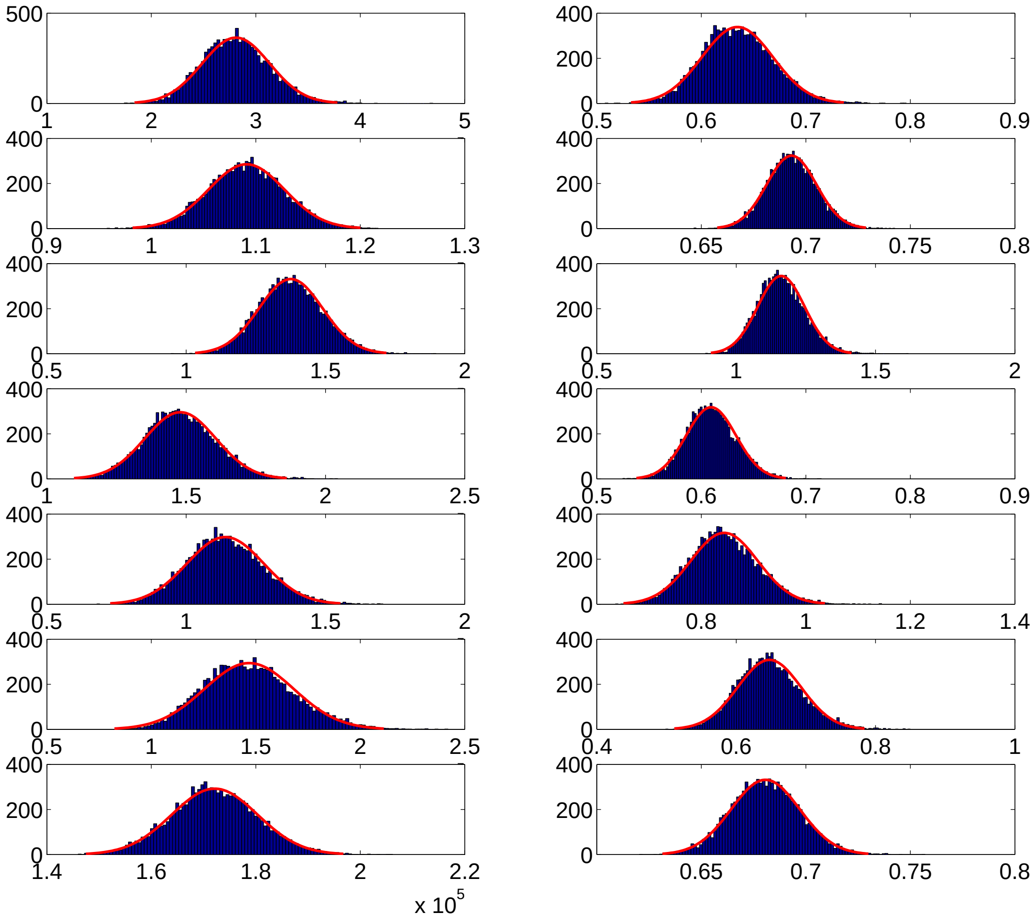

Figure A.8: *Prior distribution of the Weibull parameters*

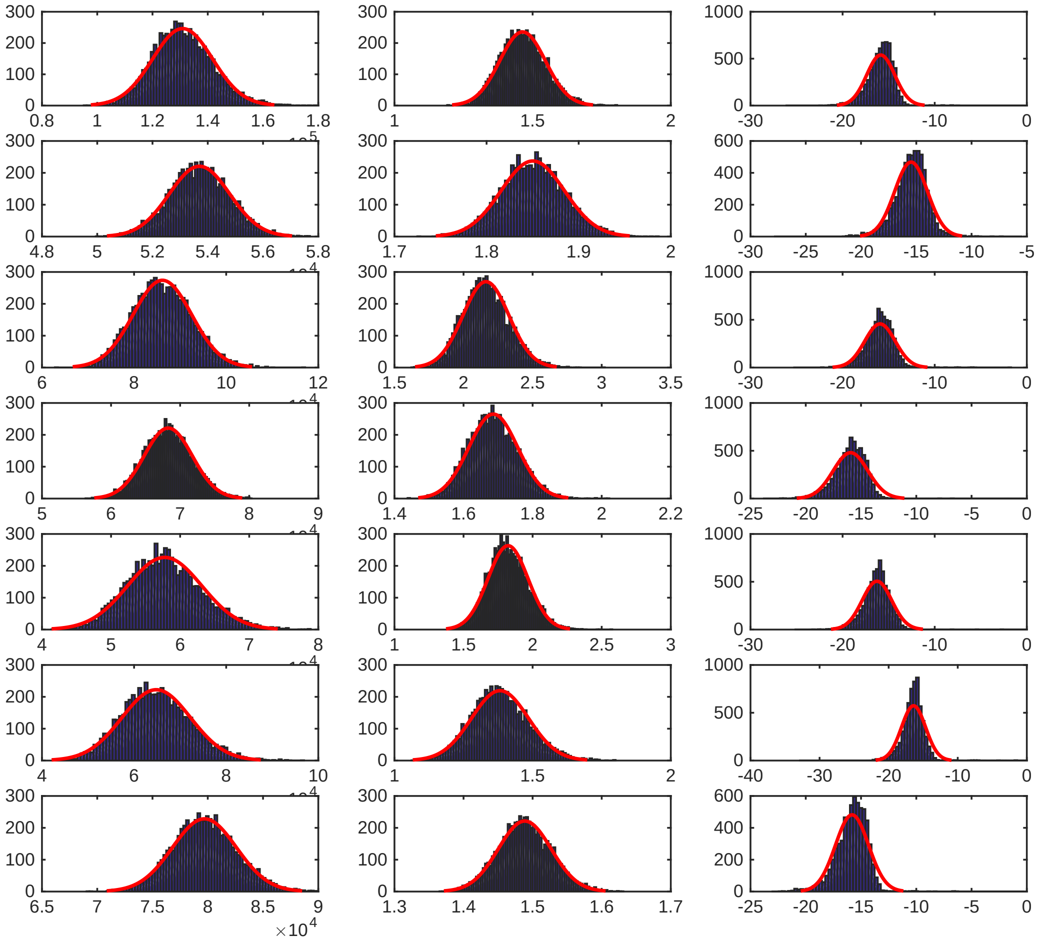

Figure A.9: *Prior distribution of the generalized Champernowne parameters with c close to zero*

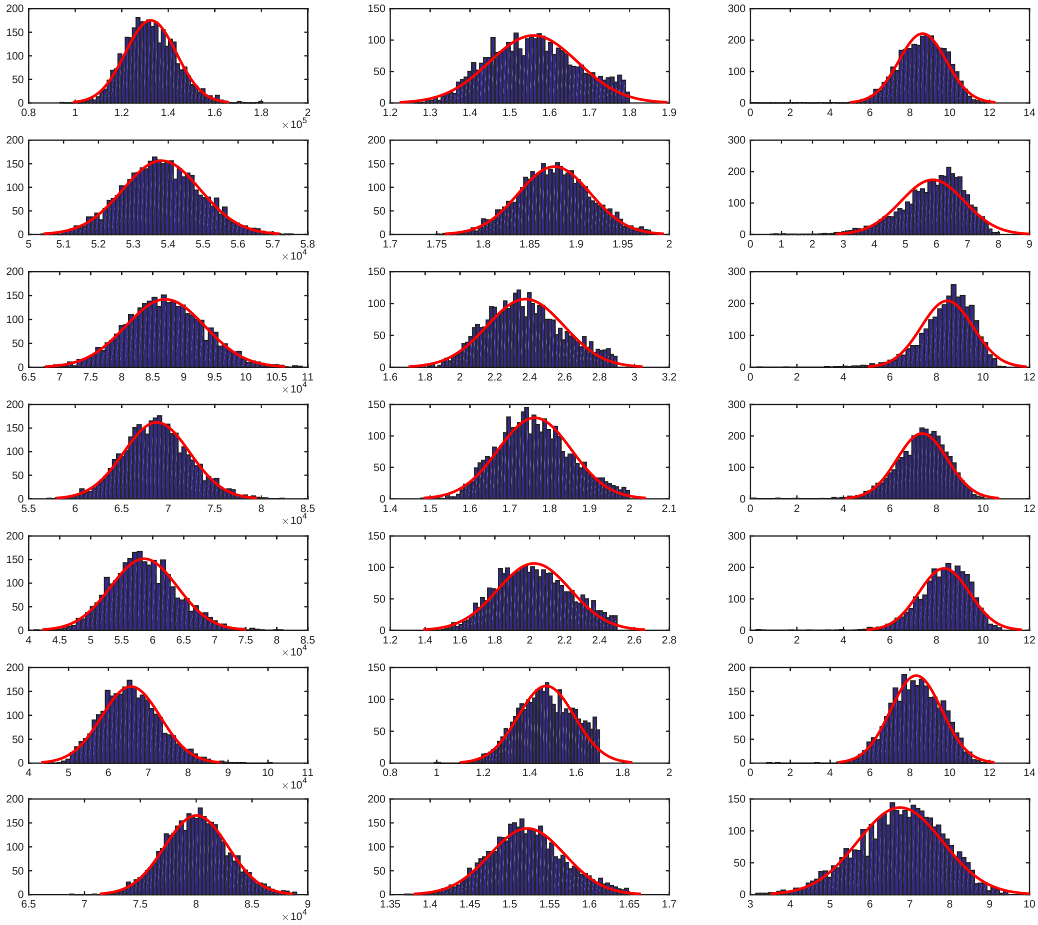

Figure A.10: *Prior distribution of the generalized Champernowne parameters with positive c*

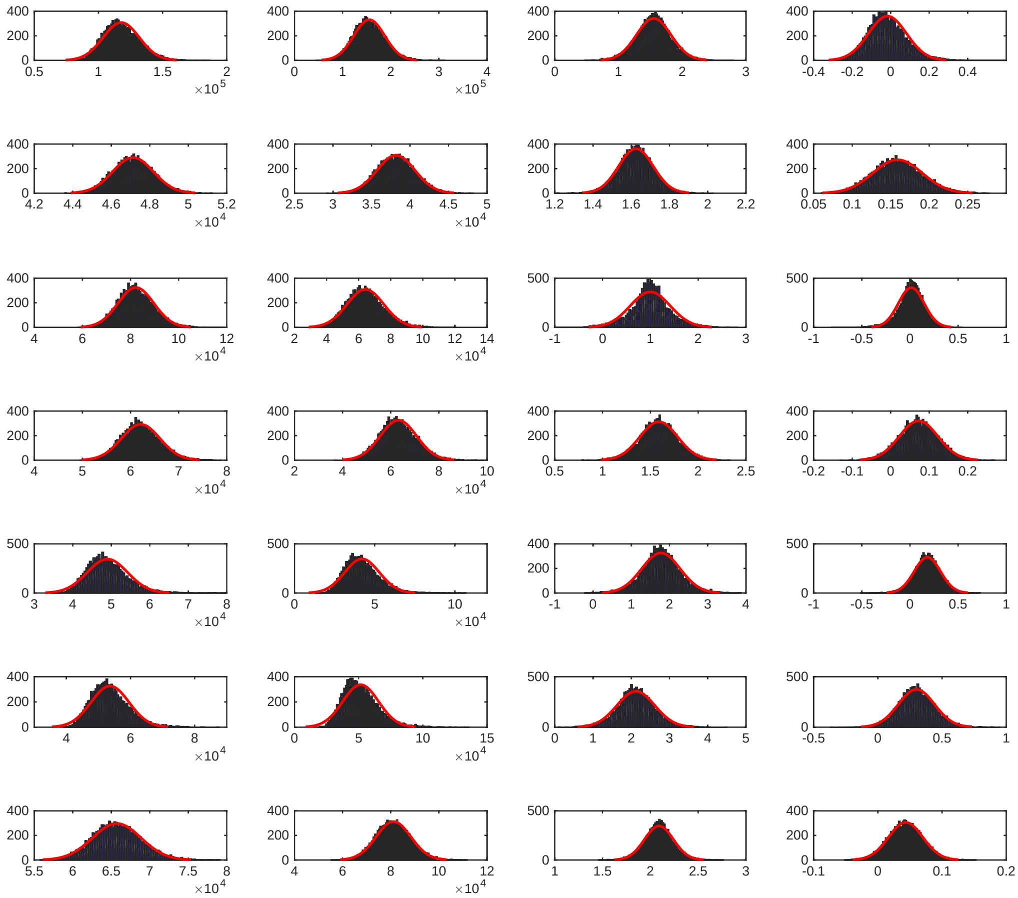

Figure A.11: *Prior distribution of the g-and-h parameters*