DEM Working Papers

N. 2019/11

An improved approach for estimating large losses in insurance analytics and operational risk using the g-and-h distribution

Marco Bee, Julien Hambuckers and Luca Trapin

UNIVERSITÀ DEGLI STUDI DI TRENTO

Dipartimento di Economia e Management

Università degli Studi di Trento

Department of Economics and Management, University of Trento, Italy.

Editors

Luciano ANDREOZZI luciano.andreozzi@unitn.it

Roberto GABRIELE roberto.gabriele@unitn.it

Technical officer

Marco TECILLA marco.tecilla@unitn.it

Guidelines for authors

Papers may be written in Italian or in English. Faculty members of the Department must submit to one of the editors in pdf format. Management papers should be submitted to R. Gabriele. Economics Papers should be submitted to L. Andreozzi. External members should indicate an internal faculty member that acts as a referee of the paper.

Typesetting rules:

1. papers must contain a first page with title, authors with emails and affiliations, abstract, keywords and codes. Page numbering starts from the first page;
2. a template is available upon request from the managing editors.

An improved approach for estimating large losses in insurance analytics and operational risk using the g-and-h distribution

Marco Bee
Department of Economics and Management, University of Trento - Italy

Julien Hambuckers
Department of Finance, HEC Liège, University of Liège, Belgium, and
and
Chair of Statistics, Georg-August-Universität Göttingen - Germany

Luca Trapin
Department of Economic Policy,
Università Cattolica del Sacro Cuore, Milan - Italy

June 11, 2019

Abstract. In this paper, we study the estimation of parameters for g-and-h distributions. These distributions find applications in modeling highly skewed and fat-tailed data, like extreme losses in the banking and insurance sector. We first introduce two estimation methods: a numerical maximum likelihood technique, and an indirect inference approach with a bootstrap weighting scheme. In a realistic simulation study, we show that indirect inference is computationally more efficient and provides better estimates in case of extreme features of the data. Empirical illustrations on insurance and operational losses illustrate these findings.

JEL codes. C15; C46; C51; G22.

Keywords. Intractable likelihood; indirect inference; skewed distribution; tail modeling; bootstrap.

1 Introduction

The g-and-h distribution (Tukey, 1977) is a model for data featuring non-zero skewness and/or excess kurtosis. It is defined by means of the following

The codes used for all the computations carried out in this paper are available at `http://marcobee.weebly.com/software.html`.

non-linear transformation of a standard normal random variable:

$$X = a + b\frac{e^{gZ} - 1}{g} e^{\frac{hZ^2}{2}}, \quad Z \sim N(0,1), \tag{1}$$

where $a \in \mathbb{R}$ is a location parameter, $b \in \mathbb{R}^+$ is a scale parameter, $g \in \mathbb{R}$ and $h \geq 0$ are shape parameters. It is a very flexible model, since, as shown by Dutta and Perry (2006, Fig. 3), its skewness-kurtosis region is very large compared to other commonly used distributions. If $g = 0$ the distribution is symmetric, whereas if $h = 0$ it becomes a scaled lognormal; see Cruz et al. (2015, Section 9.4.1).

The g-and-h distribution is also a member of the *quantile distribution* family. Hence, it can equivalently be defined via its quantile function, which is given by (Cruz et al., 2015, p. 318)

$$Q(p; \boldsymbol{\theta}) = Q(z_p; \boldsymbol{\theta}) = a + b\frac{e^{gz_p} - 1}{g} e^{\frac{hz_p^2}{2}}, \tag{2}$$

where $p \in (0,1)$ and z_p is the p-quantile of the standard normal distribution. Without loss of generality, we set $a = 0$ and $b = 1$ (Degen et al., 2007; Cruz et al., 2015).

The g-and-h distribution is particularly important when the main purpose of the analysis is the estimation of extreme quantiles or, more generally, of quantities related to the tail of the distribution. In general, Extreme Value Theory (EVT) is routinely used for this purpose, because under fairly mild conditions (see, e.g., McNeil et al., 2015, Sect. 5.2) the excesses converge to the Generalized Pareto Distribution (GPD). However, the convergence of the g-and-h to the GPD is extremely slow (Degen et al., 2007). Hence, as stressed by Cruz et al. (2015, Remark 9.6), if the true data generating process is g-and-h, estimating the distribution of extreme losses via the GPD approximation may yield imprecise results, even for large sample sizes. This distinguishing feature, with respect to other models for skewed and/or heavy-tailed data, makes a substantial difference in applications.

Although the rationale behind the construction of (1) is quite intuitive, practical application of the g-and-h distribution is hindered by the lack of closed-form density. As a consequence, the literature has mostly considered maximum likelihood estimation (MLE) as problematic, and most research has focused on estimation approaches based on either known features of the distribution or on computer-intensive techniques.

Along the first line, the earliest approach exploits the quantile function, which is explicitly known: the quantile-based method (Hoaglin, 1985) estimates the parameters by matching theoretical and empirical quantiles. More recently, since the r-th moment ($r = 1, \ldots, 4$) of the distribution exists if $h \in [0, 1/r)$ (Cruz et al., 2015, p. 320), Peters et al. (2016) estimate the parameters via L-moments. In the second group of methods, the straightforward simulation procedure of the g-and-h distribution has been used to

develop approximate maximum likelihood estimation (Bee and Trapin, 2016) and indirect inference (Bee et al., 2019).

In recent years, the increased computing power has made feasible the estimation of models with intractable likelihood function. These procedures are based on the maximization of a likelihood constructed from an approximation $\hat{f}$ of the density. We treat $\hat{f}$ as the true density and proceed to numerical maximization of the approximated log-likelihood, as for classical MLE. For quantile distributions, it is possible to rely on the quantile function to compute $\hat{f}$. This idea has been exploited by Rayner and MacGillivray (2002) and Prangle (2017) to find numerical approximations of the generalized g-and-h density.

One of the goals of the present paper is to extend this procedure to the g-and-h distribution. We develop a method for approximating the density and performing MLE of the parameters, thus establishing the feasibility of a likelihood-based approach to estimating the g-and-h distribution.

A second aim of this paper is to assess the improvement of the II estimation procedure using a bootstrap-based estimate of the optimal weighting matrix. Bee et al. (2019) find that Indirect Inference (II) outperforms both Hoaglin (1985) quantile-based method and Dupuis and Field (2004) robust approach. They use the identity matrix as weighting matrix, since the properties of the estimator are asymptotically independent from its specification. However, for finite sample sizes, the matrix does play a role, which may be particularly relevant in insurance and risk management applications, where the datasets are sometimes small.

Our simulation experiments suggest that II is better in terms of computational cost. On the other hand, from the point of view of statistical efficiency, an overall winner does not emerge. As for parameter estimation, MLEs mostly exhibit a smaller root-mean-squared-error (RMSE), but in some cases are more biased, especially when the true distribution is highly skewed and leptokurtic. When one considers Value-at-Risk (VaR) estimation, MLE is again better in the majority of setups, but with a significant exception when both g and h are large. This is an important result since most applications are characterized by such features.

The empirical analysis of actuarial and operational risk data confirms this fact: when the empirical distribution of the losses exhibits extreme skewness and kurtosis, II is more precise than numerical MLE.

The paper is organized as follows. In Section 2 we review the g-and-h distribution and develop the MLE and II estimation approaches; in Section 3 we describe the simulation experiments and comment the outcomes; in Section 4 we apply the methods to two real data-sets; in Section 5 we discuss the results and conclude.

2 The g-and-h distribution and its estimation

The g-and-h random variable $X \sim gh(a, b, g, h)$ is defined in (1), and we let $\boldsymbol{\theta} = (a, b, g, h)'$ be the vector of its parameters. In this section, we detail two estimation techniques. The first one (numerical MLE) exploits an approximation of the density based on the quantile function (2). The second one relies on an auxiliary model, in the idea of Gourieroux et al. (1993).

2.1 Maximum likelihood estimation

The quantile function (2) can be used for computing a numerical approximation of the density.

Similarly to the approaches proposed for the generalized g-and-h by Rayner and MacGillivray (2002) and Prangle (2017), we employ the following basic result from probability theory: if the random variable V has density f_V and $h(v)$ is a differentiable 1-1 transformation, the density of $W = h(V)$ is equal to

$$f_W(w) = \frac{f_V(v)}{h'(v)}, \quad \text{where } v = h^{-1}(w). \tag{3}$$

Combining (1) and (3), and setting $h(\cdot)$ equal to the quantile function (2), the approximated density is given by

$$\hat{f}(x) = \frac{\phi(p)}{Q'(p; \boldsymbol{\theta})}, \quad p = F(x; \boldsymbol{\theta}), \tag{4}$$

where $\phi(\cdot)$ is the standard normal density, $Q'(p; \boldsymbol{\theta})$ is the derivative of (2) and $F(x; \boldsymbol{\theta})$ is the g-and-h distribution function.

$Q'(p; \boldsymbol{\theta})$ is known in closed form (Cruz et al., 2015, Eq. 9.33):

$$Q'(p; \boldsymbol{\theta}) = e^{gp + \frac{hp^2}{2}} + \frac{h}{g} p e^{\frac{hp^2}{2}} (e^{gp} - 1).$$

On the other hand, $F(x; \boldsymbol{\theta})$ has to be computed via numerical inversion of (2) using some root-finding technique.

Given a random sample $x_1, \ldots, x_n$, a pseudo-code describing this procedure is given by Algorithm 1 below.

Algorithm 1.

1. *For each observation x_i ($i = 1, \ldots, n$):*
   (a) *Evaluate $p_i = F(x_i; \boldsymbol{\theta})$ by numerical inversion of (2);*
   (b) *Compute $\hat{f}(x_i) = \phi(p_i)/Q'(p_i; \boldsymbol{\theta})$.*

2. *Treat $\hat{f}$ as the true density and proceed to numerical maximization with respect to $\boldsymbol{\theta}$ of the approximated log-likelihood function $\hat{\ell}(\boldsymbol{\theta}; x_1, \ldots, x_n) = \sum_{i=1}^{n} \hat{f}(x_i; \boldsymbol{\theta})$.*

Step $1(a)$ is the key computational issue, as numerical root-finding is typically rather slow. It follows that the total computational burden, which is almost entirely related to step $1(a)$, increases linearly with the number of observations.

2.2 Indirect inference

Indirect inference is a simulation-based estimation method introduced by Gourieroux et al. (1993). Analogously to other computer-intensive techniques, it only requires the ability to sample the distribution of interest. This makes the method particularly valuable when the density is not explicitly available. With respect to other simulation-based methods, its desirable asymptotic properties are another advantage (Gourieroux et al., 1993, Section 3; Calzolari et al., 2004).

Although II can be generalized to problems with dependent observations, here we describe it in the iid case, since this is the relevant setup in the present paper.

Consider a random sample $\boldsymbol{y}_n = (y_1, \ldots, y_n)' \overset{\text{iid}}{\sim} F^{\boldsymbol{\theta}}$ from a random variable Y with cumulative distribution function $F^{\boldsymbol{\theta}}$, $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^p$. Let M be an auxiliary random variable with density $f_M(m; \boldsymbol{\psi})$, where $\boldsymbol{\psi}$ is a vector of parameters. The log pseudo-likelihood function $\ell_n(\boldsymbol{\psi}|F^{\boldsymbol{\theta}}) = \sum_{i=1}^{n} \log f_M(y_i; \boldsymbol{\psi})$ is constructed using the density of M and the sample $\boldsymbol{y}_n$ from $F^{\boldsymbol{\theta}}$.

The auxiliary parameter vector $\boldsymbol{\psi}(\boldsymbol{\theta}) \in \boldsymbol{\Psi} \subseteq \mathbb{R}^q$, $q \geq p$, is implicitly defined by the relationship $\boldsymbol{\psi}(\boldsymbol{\theta}) = \arg\max_{\boldsymbol{\psi}} \ell_n(\boldsymbol{\psi}|F^{\boldsymbol{\theta}})$. The function $\boldsymbol{\theta} \to \boldsymbol{\psi}(\boldsymbol{\theta})$ is called binding function, and the pseudo-true values $\boldsymbol{\psi}(\boldsymbol{\theta})$ are expected to be highly informative about $\boldsymbol{\theta}$. In practice, the binding function must typically be estimated via simulation.

The description above implies that II is based on two steps. First, one computes the pseudo-maximum likelihood estimate $\hat{\boldsymbol{\psi}}(\boldsymbol{\theta})$ by solving the problem $\arg\max_{\boldsymbol{\psi}} \ell_n(\boldsymbol{\psi}|F^{\boldsymbol{\theta}})$ using the observed data $\boldsymbol{y}_n$.

Second, one simulates n_s observations $\tilde{y}_1, \ldots, \tilde{y}_{n_s}$ from the true model $F^{\boldsymbol{\theta}}$ and computes a new estimate $\hat{\boldsymbol{\psi}}^s(\boldsymbol{\theta}) = \arg\max_{\boldsymbol{\psi}} \ell_{n_s}(\boldsymbol{\psi}|F^{\boldsymbol{\theta}})$ using the simulated data. The II estimate $\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\psi}}(\boldsymbol{\theta})) \overset{\text{def}}{=} \hat{\boldsymbol{\theta}}$ is the solution of the problem $\min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} (\hat{\boldsymbol{\psi}}^s(\boldsymbol{\theta}) - \hat{\boldsymbol{\psi}}(\boldsymbol{\theta}))' \boldsymbol{\Xi}^{-1} (\hat{\boldsymbol{\psi}}^s(\boldsymbol{\theta}) - \hat{\boldsymbol{\psi}}(\boldsymbol{\theta}))$, where $\boldsymbol{\Xi}$ is a weighting matrix.

When the number of parameters of the true and auxiliary model is the same, i.e. when $p = q$, the limiting distribution of $\hat{\boldsymbol{\theta}}$ does not depend on $\boldsymbol{\Xi}$ (Gourieroux et al., 1993, Proposition 5), and this result can be invoked to justify the use of the identity matrix as weighting matrix. However, in finite

samples, $\hat{\boldsymbol{\theta}}$ depends on $\boldsymbol{\Xi}$, and Gourieroux et al. (1993) show that there is an optimal choice of this matrix.

2.3 Optimal indirect inference for the g-and-h distribution

II for the g-and-h distribution based on the skewed-t auxiliary model $M \sim St(\omega, \kappa, \lambda, \nu)$ with $\boldsymbol{\Xi} = \boldsymbol{I}$ has been proposed by Bee et al. (2019). Here we extend their approach by developing an estimate of $\boldsymbol{\Xi}$ and employing it in the II estimation program.

In an exactly identified setup, the optimal weighting matrix is the asymptotic covariance matrix of $\hat{\boldsymbol{\psi}}$ (Jiang and Turnbull, 2004), and is given by a sandwich formula that requires the first two derivatives of the auxiliary log-likelihood; see Gourieroux et al. (1993, p. S112) for details.

An appealing alternative, which avoids the computation of the first two derivatives of $\ell_n(\boldsymbol{\psi}|F^{\boldsymbol{\theta}})$ at the expense of a modest computational effort, is the use of the non-parametric bootstrap. Given a sample $y_1, \ldots, y_n$ from $Y \sim St(\boldsymbol{\theta})$, the bootstrap estimate of $\boldsymbol{\Xi}$ is obtained as follows:

Algorithm 2. *(Non-parametric bootstrap estimate of the weighting matrix)*

(i) Sample with replacement n observations $y_1^*, \ldots, y_n^*$ from $y_1, \ldots, y_n$;

(ii) Use $y_1^*, \ldots, y_n^*$ to maximize the pseudo log-likelihood function and compute the pseudo-maximum likelihood estimators $\hat{\boldsymbol{\psi}}(\boldsymbol{\theta}) = (\hat{\omega}, \hat{\kappa}, \hat{\lambda}, \hat{\nu})'$ of the skewed-t distribution;

(iv) Repeat steps (i)-(ii) B times;

(iv) The bootstrap estimate of $\boldsymbol{\Xi}$ ($\hat{\boldsymbol{\Xi}}^*$, say) is the empirical covariance matrix of the skewed-t parameter estimates obtained in the B replications.

Jiang and Turnbull (2004) show that the II and ML estimators have similar asymptotic properties under some regularity conditions. However, the finite-sample behavior of the estimators is likely to be different. Moreover, in the present paper we are maximizing an approximation of the likelihood function, and this may to have an impact on the precision of the MLEs. Hence, we explore via simulation the relative efficiency of the two approaches.

3 Simulation experiments

In this section we perform a Monte Carlo analysis of the II and numerical MLE approaches outlined above. We use some of the setups employed in Bee et al. (2019) as well as one additional parameter configuration with values of the parameters similar to those found in the first empirical application (see Section 4 below).

Skewness (sk.) and kurtosis (kurt.) are computed in closed form using the formulas for the first four moments of the g-and-h given by Cruz et al. (2015, p. 320). Since $\mathrm{E}(X^r)$ only exists if $h \in [0, 1/r)$, the kurtosis is undefined in Setup (6). All in all, we sample the $gh(0, 1, g, h)$ distribution in the following six configurations of the parameters.

(1) $g = 0.5,\ h = 0.1$ (sk. $= 3.41$; kurt. $= 44.24$);

(2) $g = 0.8,\ h = 0.1$ (sk. $= 9.27$; kurt. $= 606.61$);

(3) $g = 0.2,\ h = 0.05$ (sk. $= 0.79$; kurt. $= 5.10$);

(4) $g = 0.2,\ h = 0.2$ (sk. $= 2.81$; kurt. $= 155.98$);

(5) $g = 2,\ h = 0.2$ (sk. $= 7.76 \times 10^{10}$; kurt.$= 1.08 \times 10^{58}$).

(6) $g = 2.5,\ h = 0.3$ (sk.$= 9.76 \times 10^{101}$; kurt.$= +\infty$).

Note that setups (5) and (6), which according to Dutta and Perry (2006) are likely to be relevant for operational risk modeling, are extreme in terms of skewness and kurtosis.

All experiments are carried out with sample size $n \in \{100, 1000\}$; the number of replications is $B = 200$ and the number of observations simulated from the auxiliary model is $n_s = 5000$.

3.1 Parameter estimation

Figures 1 and 2 show the bias and the RMSE respectively of the II and ML estimators of g and h in the six setups. In both figures, panels (a) and (b) refer to the $n = 100$ case, (c) and (d) are based on $n = 1000$.

Before commenting the results, it is worth pointing out that MLE has failed (i.e., aborted without convergence) 6 times in Setup 6 with $n = 100$. Even though, in the same instances, II has always converged, we have discarded the samples and replaced them with new ones.

In terms of bias (see Fig. 1) II is better than MLE, more notably in setups 5 and 6 and for $n = 100$. This is not surprising, since it is well known that II is a bias-correction method, whereas MLEs are consistent but, in general, biased.

On the other hand, in terms of RMSE (see Fig. 2), MLE is better for $n = 100$ in all setups except 4. For $n = 1000$ the two approaches are approximately equivalent in the first 4 parameter configurations, and MLE is preferable in the last two setups.

Table 1 shows the average computational cost of the two procedures for each of the six parameter configurations. The computing times are similar when $n = 100$, whereas II is much faster than ML in all setups when $n = 1000$.

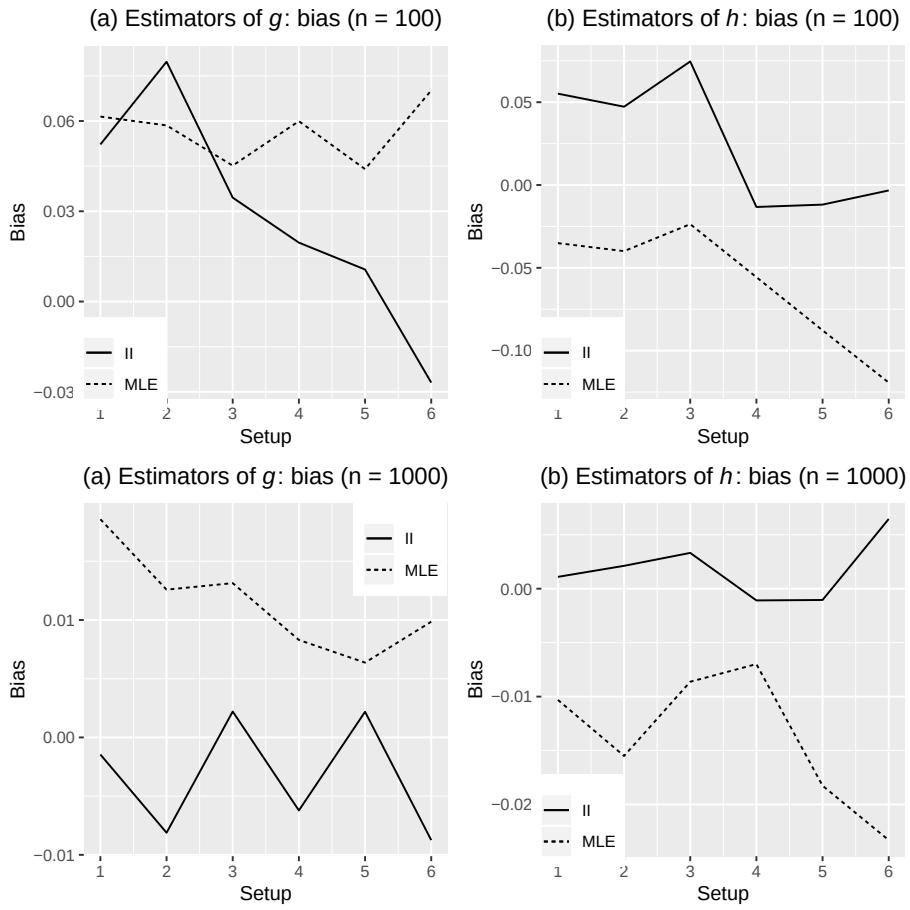

Figure 1: Bias of the II and MLE estimators of g (panels (a) and (c)) and h (panel (b) and (d)); panels (a) and (b) are based on $n = 100$, (c) and (d) on $n = 1000$.

3.2 A comparison of standard and optimal indirect inference

Figures 3 and 4 compare the performance of the II methods based on $\boldsymbol{\Xi} = \hat{\boldsymbol{\Xi}}^*$ (from now on "optimal") and $\boldsymbol{\Xi} = \boldsymbol{I}$ (from now on "standard"). The plots show the RMSEs of the estimators of g and h obtained in the two cases with $B = 200$ and $n_s = 5000$, when $n = 100$ (Figure 3) and $n = 1000$ (Figure 4).

When $n = 100$, optimal II is better than standard in the last two setups, especially for h. The outcomes are more similar when $n = 1000$, which is justified by the asymptotic equivalence of the two approaches. The only relevant difference is observed in Setup 3, where the RMSE of the optimal estimator of g is approximately 30% smaller.

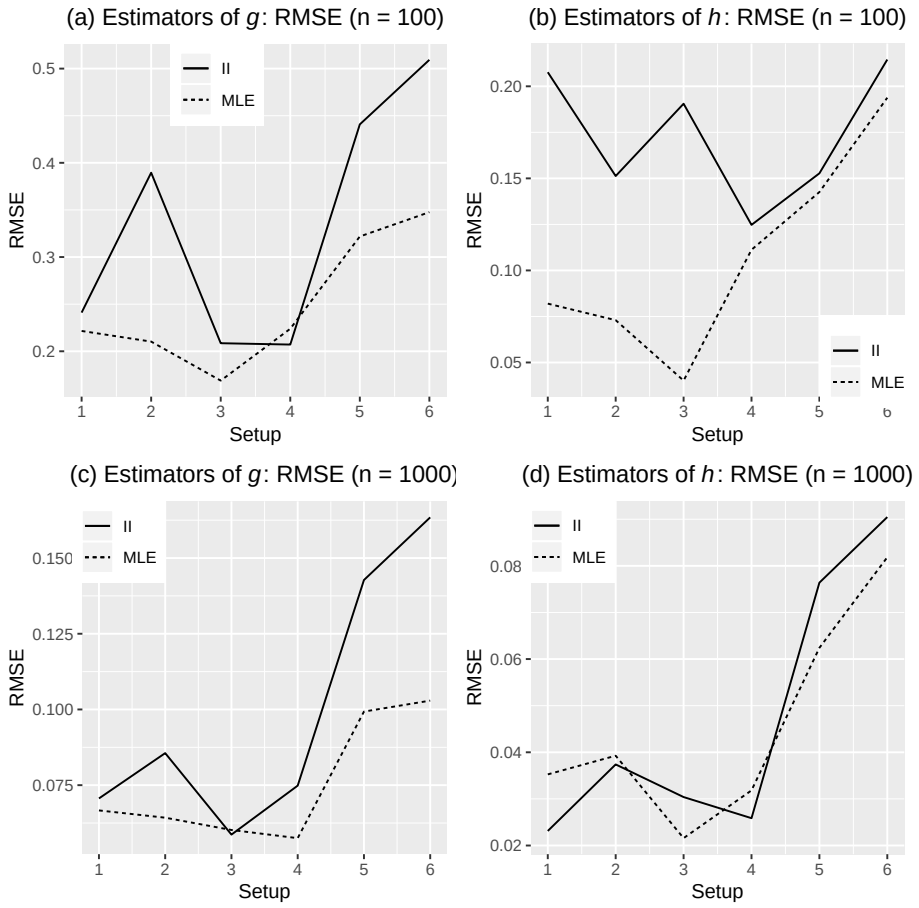

Figure 2: RMSE of the II and MLE estimators of g (panels (a) and (c)) and h (panel (b) and (d)); panels (a) and (b) are based on $n = 100$, (c) and (d) on $n = 1000$.

3.3 VaR estimation

Another interesting quantity to look at is the VaR (or quantile) estimate, for a level far in the tail. Indeed, most applications of the g-and-h distribution (and in particular those considered in Section 4) eventually aim at computing such quantities.

Once the parameters of the g-and-h distribution have been estimated, the VaR can be computed in closed form by plugging the estimates into (2). Figures 5 to 10 show the relative bias $RB_\alpha \stackrel{\text{def}}{=} \text{bias}(\widehat{\text{VaR}}_\alpha)/\text{VaR}_\alpha$ and the RMSE of the II and MLE estimators of the VaR in each of the six setups. The RMSE graphs (panels (b)) are on logarithmic scale.

In terms of bias, according to panel (a) of figures 5 to 10, the two estimators are overall equivalent when $n = 1000$, with minor differences in specific setups. On the other hand, when $n = 100$ the II estimator of VaR is

Table 1: Computing times (in seconds) of II and MLE in the six setups.

Sample size	Method	Setup					
		1	2	3	4	5	6
$n = 100$	II	15.31	12.14	19.59	16.44	15.25	9.28
	MLE	11.98	13.68	10.87	14.67	10.02	14.01
$n = 1000$	II	14.15	13.32	18.24	15.17	10.38	9.90
	MLE	155.41	145.16	152.86	151.60	167.84	182.97

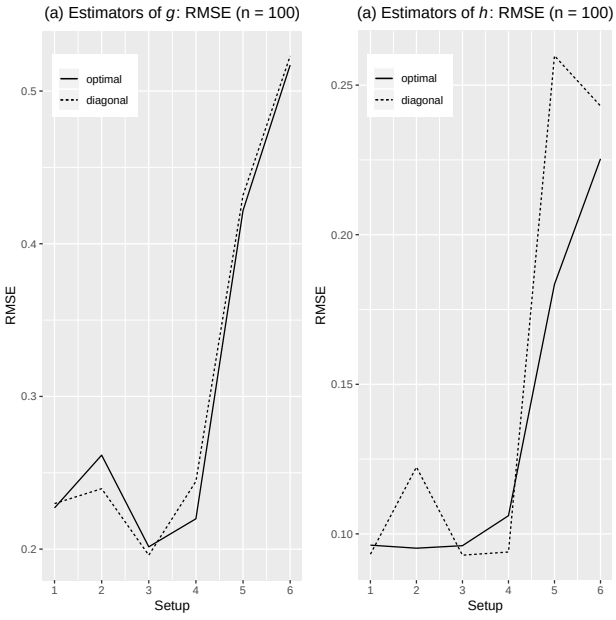

Figure 3: RMSE of the optimal and standard II estimators of g (panel (a)) and h (panel (b)) for $n = 100$.

biased for α close to 0 in the first 5 setups, whereas the results are reversed in Setup 6 (Fig. 10), where $\widehat{\text{VaR}}^{MLE}$ has a large positive bias.

Figures 5 to 10 (panel (b)) suggest that, when $n = 100$, $\widehat{\text{VaR}}^{MLE}$ has a smaller RMSE in the first 5 settings. On the other hand, in Setup 6, the RMSE of $\widehat{\text{VaR}}^{II}$ is smaller than $\widehat{\text{VaR}}^{MLE}$. When $n = 1000$, the outcomes obtained with the two methods are not very different from each other: II is better in Setup 1 (at least for the smallest values of α) and 5, MLE is preferable in setups 2, 3, 4, whereas in Setup 6 there is no clear winner.

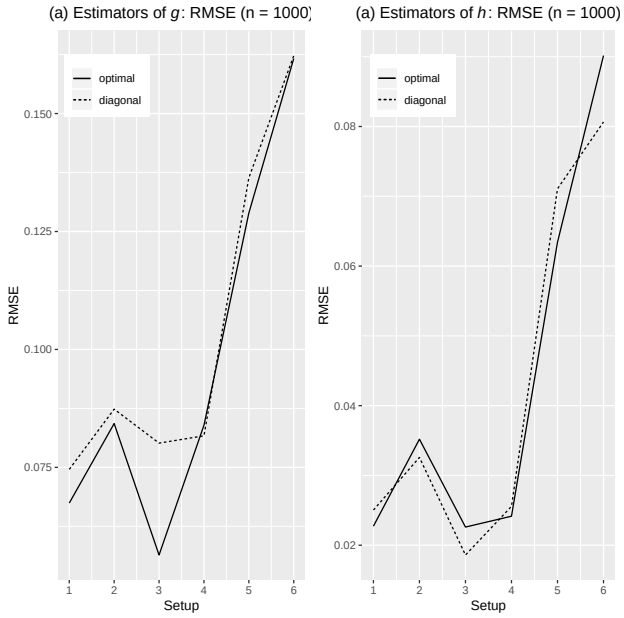

Figure 4: RMSE of the optimal and standard II estimators of g (panel (a)) and h (panel (b)) for $n = 1000$.

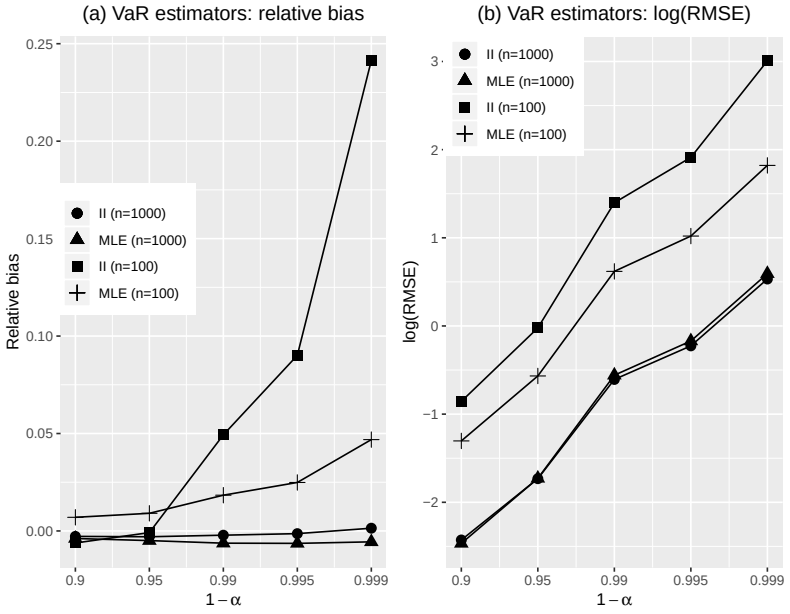

Figure 5: Setup 1. Relative bias (panel (a)) and log-RMSE (panel (b)) of the II and MLE estimators of the VaR for different values of α.

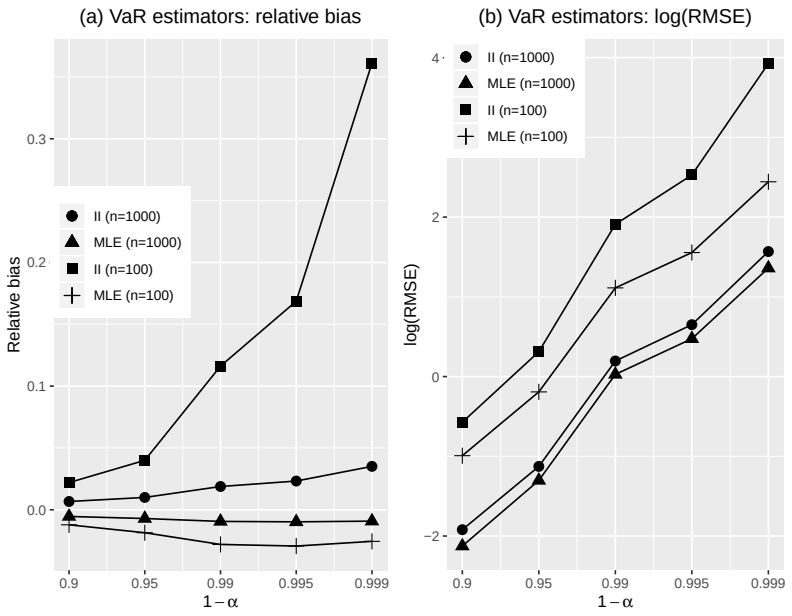

Figure 6: Setup 2. Relative bias (panel (a)) and log-RMSE (panel (b)) of the II and MLE estimators of the VaR for different values of α.

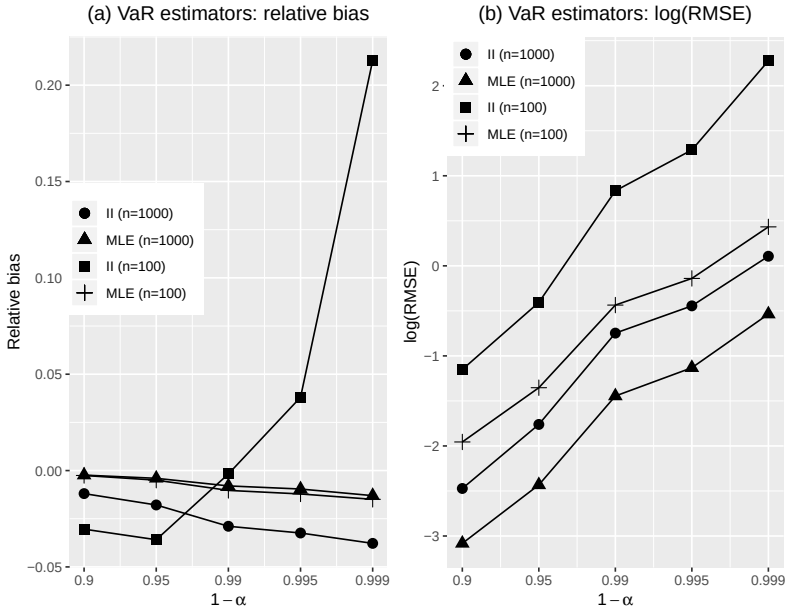

Figure 7: Setup 3. Relative bias (panel (a)) and log-RMSE (panel (b)) of the II and MLE estimators of the VaR for different values of α.

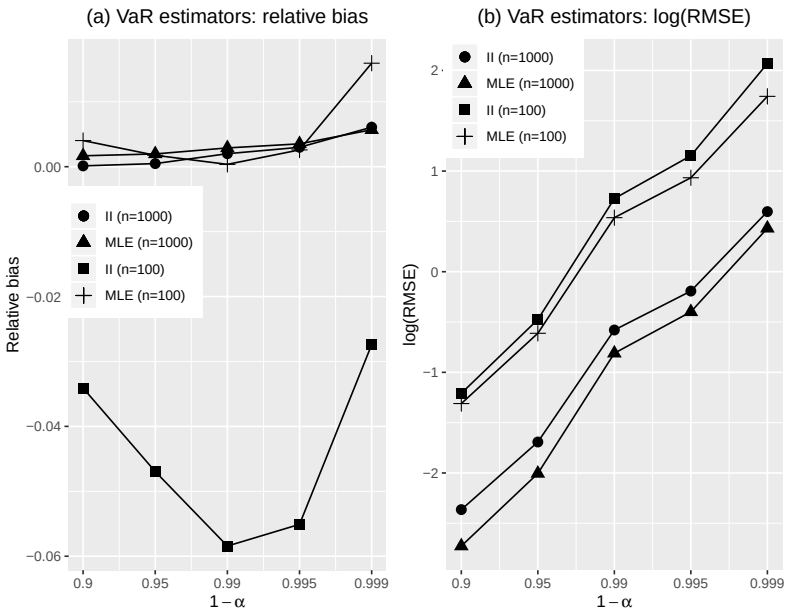

Figure 8: Setup 4. Relative bias (panel (a)) and log-RMSE (panel (b)) of the II and MLE estimators of the VaR for different values of α.

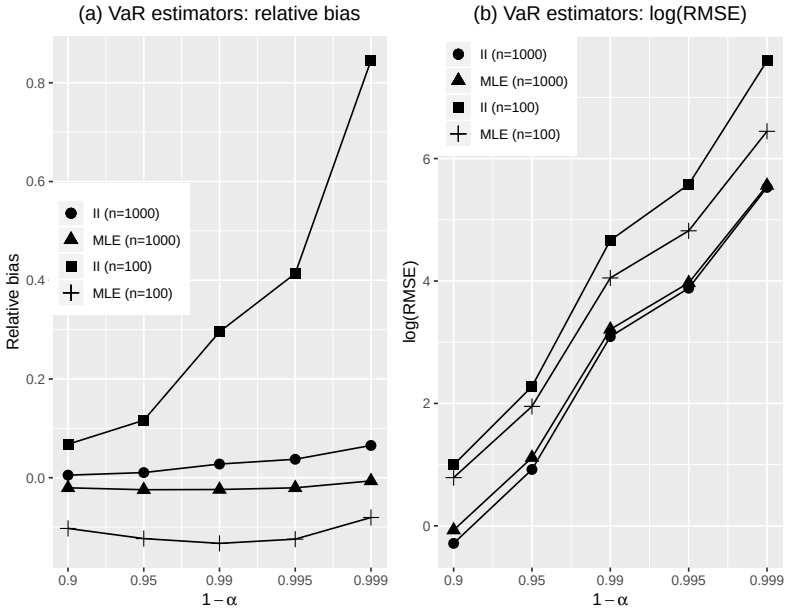

Figure 9: Setup 5. Relative bias (panel (a)) and log-RMSE (panel (b)) of the II and MLE estimators of the VaR for different values of α.

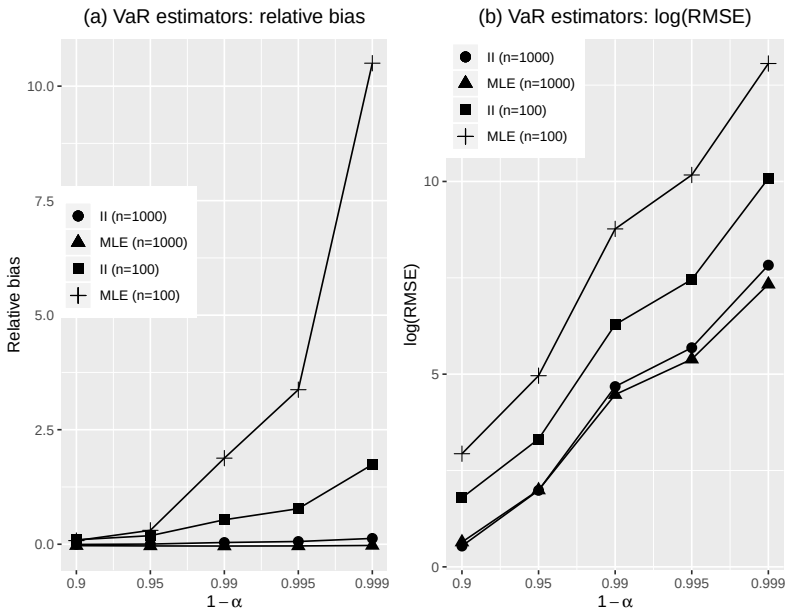

Figure 10: Setup 6. Relative bias (panel (a)) and of the log-RMSE (panel (b)) of the II and MLE estimators of the VaR for different values of α.

4 Empirical applications

4.1 AON Re Belgium fire losses

In this section we outline the results of an application to actuarial data. The `beaonre` dataset from the `CASdatasets R` package contains 1823 fire losses collected by the reinsurance broker AON Re Belgium and first used by Beirlant et al. (1999). With the aim of ascertaining the difference between the two proposed approaches when the sample size is moderately large, we use a random sample of 700 observations of the loss amount in thousand of Danish Krone (variable `ClaimCost` in `beaonre`). For the subsequent analysis we standardize the observations[1]. A stem-and-leaf plot of the data is shown in Figure 11 along with the empirical skewness and kurtosis.

Point estimates are obtained by means of both the II and numerical MLE approaches presented in Section 2. Standard errors are computed via non-parametric bootstrap with 200 replications.

Table 2 shows that the II and MLE estimates of the parameters are quite different. In particular, h is found to be very different between the two methods. We discuss this issue in detail at the end of the section.

[1]Standardized observations are given by $(y_i - \hat{a})/\hat{b}$, $i = 1, \ldots, n$, where $\hat{a}$ and $\hat{b}$ are the Hoaglin (1985) quantile estimators of a and b; see, e.g., Bee et al. (2019) for details about the computation of the estimators.

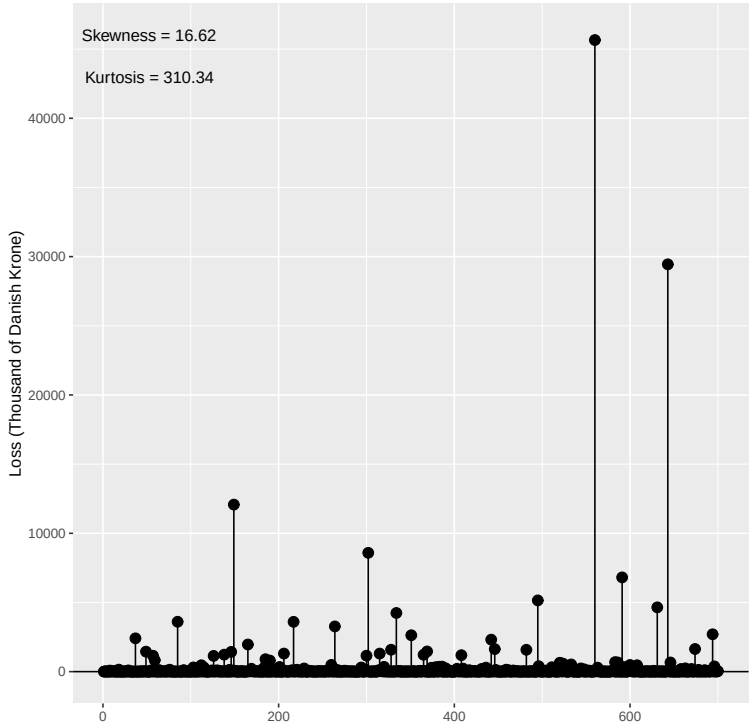

Figure 11: AON Re Belgium fire losses.

Figure 12 shows the QQ-plot of the real data vs. observations simulated from the g-and-h distribution with parameters estimated via II (panel (a)) and via MLE (panel (b)); both plots are restricted to quantiles larger than 0.9. The fit is better with the II estimation method.

Table 3 reports the VaR measures estimated by means of the g-and-h distribution as well as via the state-of-the-art Peaks-over-Threshold (POT) method (see, e.g., McNeil et al., 2015, Sect. 5.3.2).

The outcomes convey two messages. First, the II VaR is much closer to the empirical quantile, as expected given the better fit in Figure 12. Second, the VaR computed by means of the g-and-h distribution estimated via II is approximately as precise as the VaR obtained by means of the POT method.

Table 2: Fire losses: Parameter estimates and bootstrap standard errors.

	g	h
II	2.500	0.268
	(0.130)	(0.097)
MLE	2.237	0.070
	(0.103)	(0.028)

Table 3: Fire losses: VaR estimates and bootstrap standard errors obtained by means of the g-and-h distribution estimated via II and MLE and by means of the POT method. For comparison purposes, the empirical quantile is reported in the last line.

	$\alpha = 0.95$	$\alpha = 0.99$	$\alpha = 0.995$
II-VaR	35.232	292.427	658.164
	(6.691)	(95.780)	(256.696)
MLE-VaR	18.685	95.308	173.541
	(6.965)	(110.288)	(352.609)
POT-VaR	45.985	302.077	633.827
	(9.988)	(92.917)	(296.564)
Emp.	42.022	270.568	491.888

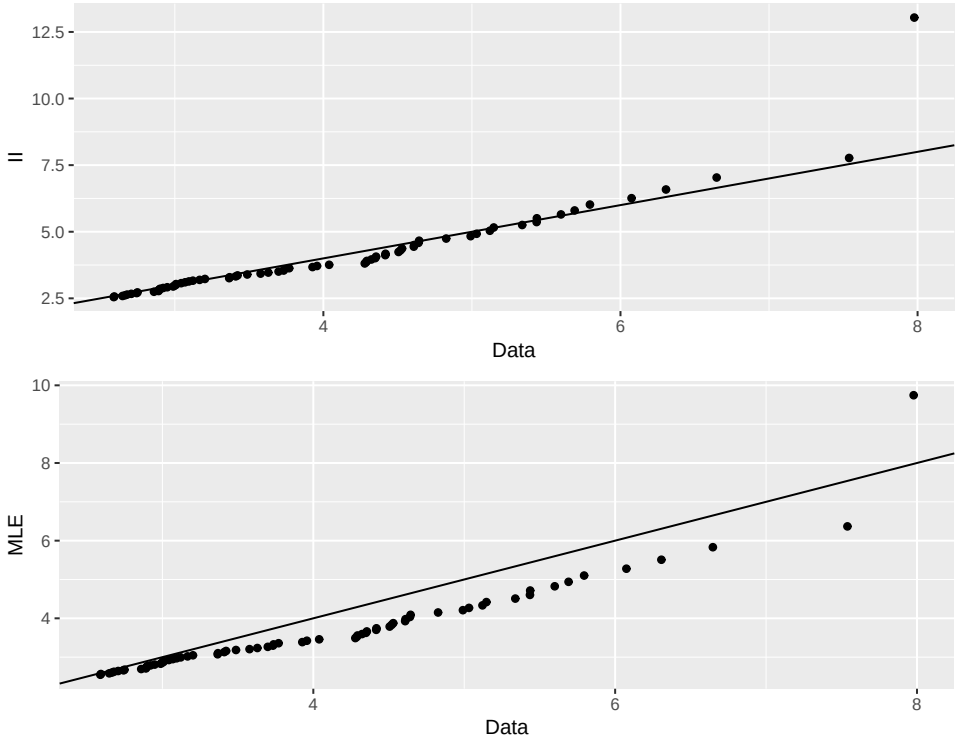

Figure 12: QQ-plot of the observed losses vs. observations simulated from the II-estimated g-and-h (upper panel) and of the observed losses vs. observations simulated from the MLE-estimated g-and-h (lower panel).

The large difference between the II and MLE approach is rather surprising at first sight, since, in the simulation experiments of Section 3, the outcomes are never so far away from each other. A tentative explanation is that II might be more robust than MLE with respect to possible misspecifications of the model: the distribution of the data considered in the present application is likely to be not exactly g-and-h, and II may outperform MLE

in a setup where the true data-generating process is not g-and-h. Further investigation of this conjecture is beyond the scope of this paper, but is one of the issues in our future research agenda.

4.2 Operational risk

In this section we analyze operational risk losses recorded at the Italian bank Unicredit; a detailed description of the data can be found in Hambuckers et al. (2018). Here we use the 152 losses observed in business line BDSF (Business Disruption and System Failures) between 2005 and 2014. The data, scaled by an unknown factor for confidentiality reasons, are displayed in Figure 13.

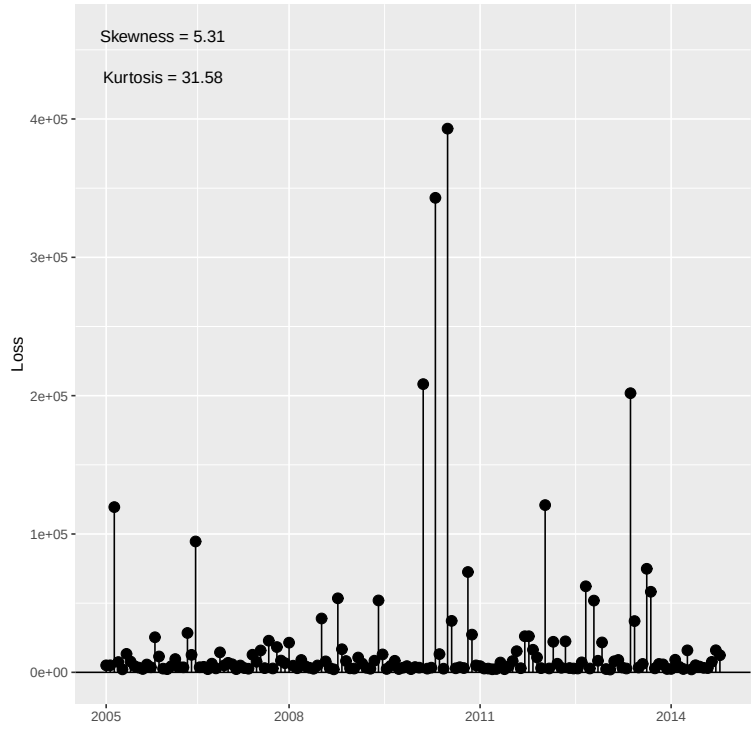

Figure 13: Operational risk losses.

The II and MLE parameter estimates obtained with standardized observations are shown in Table 4. The two approaches yield quite similar results. Figure 14 displays the quantile-quantile plot of true vs. simulated observations. Analogously to the preceding section, the upper panel simulates data from the II-based estimated g-and-h, the lower panel samples the MLE-based g-and-h, and the plots are restricted to quantiles larger than 0.9. The graphs confirm that the two estimated g-and-h distributions are

almost identical.

Table 5 reports the VaR measures computed via the g-and-h distribution and the POT method as well as the empirical quantile[2]. In this case the two g-and-h VaRs are similar, but again $\widehat{VaR}^{II}$ is closer to the empirical quantile. On the other hand, the two g-and-h VaRs have a performance comparable to the POT VaR and in line with the empirical quantiles reported in the last line of the table.

Table 4: Operational risk: Parameter estimates and bootstrap standard errors.

	g	h
II	1.969	0.029
	(0.264)	(0.199)
MLE	2.050	0.028
	(0.140)	(0.0121)

Table 5: Operational risk: VaR estimates and bootstrap standard errors obtained by means of the g-and-h distribution estimated via II and MLE and by means of the POT method. For comparison purposes, the empirical quantile is reported in the last line.

	$\alpha = 0.95$	$\alpha = 0.99$	$\alpha = 0.995$
II-VaR	12.933	52.997	88.506
	(3.990)	(32.760)	(70.726)
MLE-VaR	14.259	61.500	104.708
	(3.396)	(25.474)	(54.535)
POT-VaR	14.384	45.926	68.128
	(3.818)	(40.645)	(83.807)
Emp.	12.461	48.940	-

5 Conclusion

In this paper we have developed two approaches to the estimation of the parameters of the g-and-h distribution. The results of the simulation experiments suggest that the numerical maximum likelihood method is more efficient mostly in terms of RMSE, but suffers from large bias when the distribution is highly skewed. Indirect inference performs better in the empirical applications and has a lighter computational burden, in particular when the sample size gets large.

The g-and-h distribution seems to be especially well suited for highly asymmetric and heavy-tailed empirical distributions. For moderately skewed data, the scaled lognormal obtained when $h = 0$ may be a more parsimonious

[2]We do not report the 99.5% empirical quantile because it makes little sense with a sample size as small as $n = 152$.

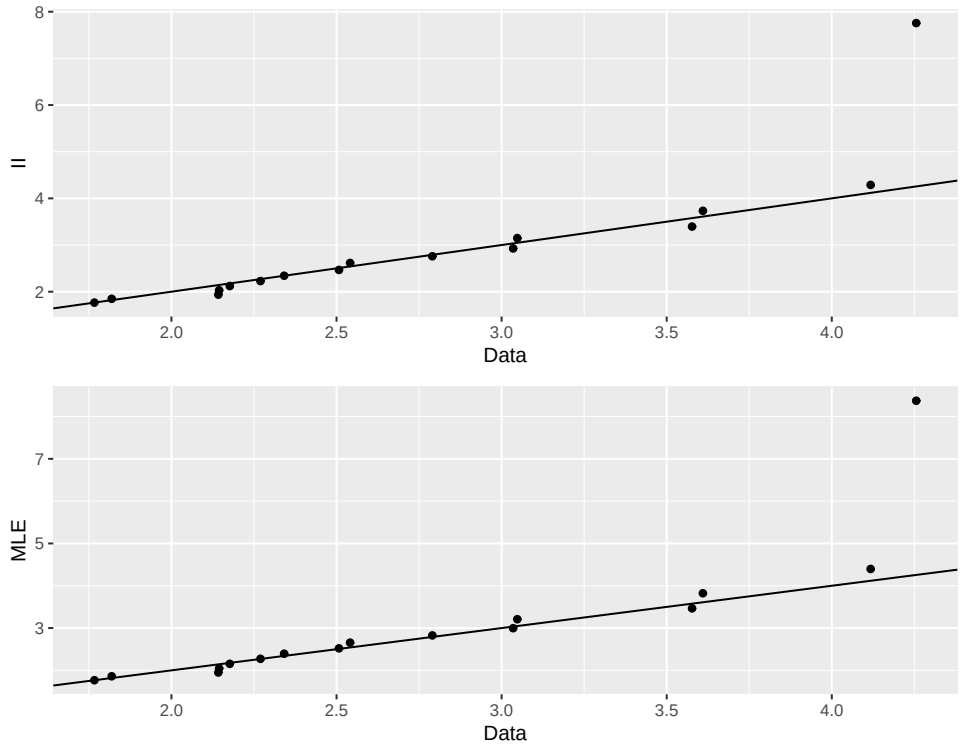

Figure 14: QQ-plot of the observed losses vs. observations simulated from the II-estimated g-and-h (upper panel) and vs. observations simulated from the MLE-estimated g-and-h (lower panel).

option. Hence, it would be important to devise a test of the hypothesis $H_0 : h = 0$, which would provide the investigator with a data-driven model selection tool. This issue requires further research.

On the empirical front, the practicability of the proposed estimation techniques opens the door for a wider use of the g-and-h distribution. Beyond insurance and operational loss data, hedge funds returns (Ding and Shawky, 2007) or health data (Rigby and Stasinopoulos, 2005) could be considered in future applications.

Acknowledgments. We thank Fabio Piacenza (UniCredit SpA) for providing us with the operational risk data.

References

Bee, M., Hambuckers, J., and Trapin, L. (2019). Estimating Value-at-Risk for the g-and-h distribution: an indirect inference approach. *Quantitative Finance*, forthcoming.

Bee, M. and Trapin, L. (2016). A simple approach to the estimation of Tukey's gh distribution. *Journal of Statistical Computation and Simulation*, 86(16):3287–3302.

Beirlant, J., Dierckx, G., Goegebeur, Y., and Matthys, G. (1999). Tail index estimation and an exponential regression model. *Extremes*, 2(2):177–200.

Calzolari, G., Fiorentini, G., and Sentana, E. (2004). Constrained indirect estimation. *Review of Economic Studies*, 71(4):945–973.

Cruz, M., Peters, G., and Shevchenko, P. (2015). *Fundamental Aspects of Operational Risk and Insurance Analytics: A Handbook of Operational Risk*. Wiley.

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

Ding, B. and Shawky, H. (2007). The performance of hedge fund strategies and the asymmetry of return distributions. *European Financial Management*, 13(2):309–331.

Dupuis, D. and Field, C. (2004). Large wind speeds: modeling and outlier detection. *Journal of Agricultural, Biological, and Environmental Statistics*, 9(1):105–121.

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

Gourieroux, C., Monfort, A., and Renault, E. (1993). Indirect inference. *Journal of Applied Econometrics, Supplement: Special Issue on Econometric Inference Using Simulation Techniques*, 8:S85–S118.

Hambuckers, J., Groll, A., and Kneib, T. (2018). Understanding the economic determinants of the severity of operational losses: A regularized Generalized Pareto regression approach. *Journal of Applied Econometrics*, 33(6):898–935.

Hoaglin, D. C. (1985). Summarizing shape numerically: The g-and-h distributions. *Exploring data tables, trends, and shapes*, pages 461–513.

Jiang, W. and Turnbull, B. (2004). The indirect method: Inference based on intermediate statistics - a synthesis and examples. *Statistical Science*, 19(2):239–263.

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

Peters, G. W., Chen, W. Y., and Gerlach, R. H. (2016). Estimating quantile families of loss distributions for non-life insurance modelling via L-moments. *Risks*, 4(2):14.

Prangle, D. (2017). gk: An R Package for the g-and-k and generalised g-and-h distributions. *ArXiv e-prints*, 1706.06889v1.

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.

Rigby, R. and Stasinopoulos, D. (2005). Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, 54(3):507–554.

Tukey, J. W. (1977). Modern techniques in data analysis. In *NSF-Sponsored Regional Research Conference at Southern Massachusetts University, North Dartmouth.*