To appear in *Quantitative Finance*, Vol. 00, No. 00, Month 20XX, 1–21

Estimating large losses in insurance analytics and operational risk using the g-and-h distribution

M. BEE*†, J. HAMBUCKERS‡ and L. TRAPIN §

†Department of Economics and Management, University of Trento, Italy
‡Department of Finance, HEC Liège, University of Liège, Belgium
§Department of Statistical Sciences "P. Fortunati", University of Bologna

(*Received 00 Month 20XX; in final form 00 Month 20XX*)

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 than the maximum likelihood method in case of extreme features of the data. Empirical illustrations on insurance and operational losses illustrate these findings.

Keywords: Actuarial Science; Tail Analysis; Advanced Econometrics; Computational Finance; Extreme Risk & Insurance

JEL Classification: C15; C51; G22

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 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).

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

*Corresponding author. Email: marco.bee@unitn.it; ORCID: 0000-0002-9579-3650.

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 is to assess the improvement of the indirect inference (II) estimation procedure using a bootstrap-based estimate of the optimal weighting matrix. Bee et al. (2019) 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.

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 distribution of the excesses converges to the Generalized Pareto Distribution (GPD). EVT is often implemented via the Peaks-over-Threshold (POT) method: given iid observations $x_1, \ldots, x_n$ from some distribution and a threshold u large enough, the parameters are estimated employing the excesses $x_i - u$, which are assumed to be approximately GPD-distributed. The Value-at-Risk (VaR) is finally given in closed form by the quantile of the fitted 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: if the asymptotic GPD approximation is poor, the estimated VaR based on the GPD will be quite different from the true VaR, and the only way of computing an accurate estimate of the VaR is to fit the g-and-h distribution.

From a modeling perspective, the g-and-h distribution has often been found to be appropriate for operational risk measurement (e.g., Moscadelli 2004, Dutta and Perry 2006, Degen et al. 2007, Cruz 2018, Sect. 9.4.1). On the other hand, it has been mostly ignored in actuarial applications. Since this may mostly be due to its limited analytical tractability (Peters et al. 2016), one of the goals of our paper is to draw some attention on the potential of the g-and-h to model non-life insurance losses.

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 often 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 only exists if $h \in [0, 1/r)$ (Cruz et al. 2015, p. 320), Peters et al. (2016) fit the distribution 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).

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. The latter result is consistent with the unbiasedness property of II estimators (Gouriéroux et al. 2000). However, when one considers VaR estimation, the unbiasedness of II estimators does not necessarily extend to unbiasedness of the corresponding VaR estimate. The MLE method is better at estimating VaR 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 the preceding remarks: 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

Existing estimation methods for the g-and-h distribution include Hoaglin (1985) quantile-based method, Dupuis and Field (2004) robust approach, Peters et al. (2016) L-moments-based method and Bayesian techniques (see, e.g., Peters and Sisson 2006).

The derivation of the quantile-based estimators exploits the closed-form formula (2). The estimator of the location parameter a is just the median of the observations: $\hat{a} = \text{median}\{x_1, \ldots, x_n\}$.

To find $\hat{g}$ notice that, for $0 < p < 0.5$, the following holds:

$$\frac{x_{1-p} - x_{0.5}}{x_{0.5} - x_p} = \frac{a + b\frac{e^{-gz_p}-1}{g}e^{hz_p^2/2} - a}{a - \left(b\frac{e^{-gz_p}-1}{g}e^{hz_p^2/2} + a\right)} = e^{-gz_p}.$$

Accordingly, the value of g that matches the quantiles x_p and x_{1-p} is given by

$$g_p = -\frac{1}{z_p}\log\left(\frac{x_{1-p} - x_{0.5}}{x_{0.5} - x_p}\right).$$

The value of g_p depends on p, so that $\hat{g}$ is usually taken to be equal to the median of the $\hat{g}_p$'s over a grid of quantiles (Cruz et al. 2015, p. 325).

The estimates of b and h are found by means of linear regression. To do this, the upper half spread $\text{UHS}_p \stackrel{\text{def}}{=} x_{1-p} - x_{0.5}$ is divided by $(e^{-gz_p} - 1)$:

$$\text{UHS}_p^* = \frac{g(x_{1-p} - x_{0.5})}{(e^{-gz_p} - 1)} = be^{hz_p^2/2},$$

so that $\log(\text{UHS}_p^*) = \log(b) + hz_p^2/2$. Accordingly, $\widehat{\log(b)}$ and $\hat{h}$ are respectively the estimates of the intercept and of the slope in the corresponding linear regression model, and $\hat{b} = \exp(\widehat{\log(b)})$.

Dupuis and Field (2004) proposed an estimation method whose main advantage is the robustness with respect to outliers. They estimate the parameters g and h using Huber (1981) "Proposal 2".

After fixing a set of quantiles, Dupuis and Field (2004) employ the function $\rho((x-y)/y)$, where

$$\rho(x) = \begin{cases} x^2/2, & \text{if } |x| < c_1 \\ c_1|x| - c_1^2/2, & \text{otherwise} \end{cases}$$

is Huber's ρ function and $c_l \in \mathbb{R}^+$. The estimates are the solution of the problem

$$\min_{g,h} \sum_{i=1}^{n} \rho((x_i - y_i)/y_i)$$

over the chosen range of quantiles.

Both methods are easy to implement, but depend on the quantile levels they are based upon. Moreover, Bee et al. (2019) find via Monte Carlo experiments that II outperforms both Hoaglin (1985) quantile-based method and Dupuis and Field (2004) robust approach.

To introduce moment-based approaches, notice that, provided $0 < h < 1/r$, the moments $\mathrm{E}(X^r)$, $r = 1, \ldots, 4$, can be computed in closed form (Cruz et al. 2015, p. 320):

$$\begin{aligned}
\mathrm{E}(X) &= \frac{e^{\frac{g^2}{2-2h}} - 1}{g\sqrt{1-h}}, \\
\mathrm{E}(X^2) &= \frac{1 - 2e^{\frac{g^2}{2-4h}} + e^{\frac{2g^2}{1-2h}}}{g^2\sqrt{1-2h}}, \\
\mathrm{E}(X^3) &= \frac{3e^{\frac{g^2}{2-6h}} + e^{\frac{9g^2}{2-6h}} - 3e^{\frac{2g^2}{1-3h}} - 1}{g^3\sqrt{1-3h}}, \\
\mathrm{E}(X^4) &= s(g,h)\frac{e^{\frac{8g^2}{1-4h}}}{g^4\sqrt{1-4h}},
\end{aligned}$$

where

$$s(g,h) = 1 + 6e^{\frac{6g^2}{4h-1}} + e^{\frac{8g^2}{4h-1}} - 4e^{\frac{7g^2}{8h-2}} - 4e^{\frac{15g^2}{8h-2}}.$$

The classical method of moments has been developed by Headrick et al. (2008). However, it cannot be used when $h \geq 0.25$, because in this case $\mathrm{E}(X^4)$ does not exist. This drawback is overcome by the L-moments-based estimation method proposed by Peters et al. (2016): indeed, L-moments always exist (if $h < 1$) and L-moments-based estimation is known to be more efficient than classical moment-based estimation (Vogel and Fennessey 1993, Hodis et al. 2012). L-moments estimators are found by iterative minimization of the objective function

$$(\tau_3 - \hat{\tau}_3)^2 + (\tau_4 - \hat{\tau}_4)^2, \tag{3}$$

where $\tau_3 = \ell_3/\ell_2$ and $\tau_4 = \ell_4/\ell_2$ are the L-skewness and L-kurtosis respectively, ℓ_r is the r-th L-moment, $\hat{\tau}_3$ and $\hat{\tau}_4$ are the corresponding sample counterparts. The computation of L-moments of the g-and-h distribution requires numerical integration (Peters et al. 2016, Proposition 5.4), and further numerical work is required for the solution of (3).

Finally, Bayesian estimation techniques have been used for the g-and-h distribution in operational risk applications. In particular, Peters and Sisson (2006) have developed a Markov Chain Monte Carlo - Approximate Bayesian Computation method; see Cruz et al. (2015, Sect. 9.4.4) for details.

In the following we detail two additional 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 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. 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

$$f_W(w) = \frac{f_V(v)}{h'(v)}, \quad \text{where } v = h^{-1}(w). \tag{4}$$

Combining (1) and (4), and setting $h(\cdot)$ equal to the quantile function (2), the approximated density is given by

$$\hat{f}(x) = \frac{\phi(z_p)}{Q'(z_p; \boldsymbol{\theta})}, \quad z_p = F(x; \boldsymbol{\theta}), \tag{5}$$

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

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

$$Q'(z_p; \boldsymbol{\theta}) = e^{gz_p + \frac{hz_p^2}{2}} + \frac{h}{g} z_p e^{\frac{hz_p^2}{2}} (e^{gz_p} - 1).$$

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 $z_{pi} = F(x_i; \boldsymbol{\theta})$ by numerical inversion of (2);*
   (b) *Compute $\hat{f}(x_i) = \phi(z_{pi}) / Q'(z_{pi}; \boldsymbol{\theta})$.*
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 the II approach 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 $x_1, \ldots, x_n \overset{\text{iid}}{\sim} F^{\boldsymbol{\theta}}$ from a random variable X with cumulative distribution function $F^{\boldsymbol{\theta}}$, $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^d$. 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 $\ell_n(\boldsymbol{\psi}|F^{\boldsymbol{\theta}}) = \sum_{i=1}^n \log f_M(x_i;\boldsymbol{\psi})$ is constructed using the density of M and the sample $x_1, \ldots, x_n$ from $F^{\boldsymbol{\theta}}$.

The auxiliary parameter vector $\boldsymbol{\psi}(\boldsymbol{\theta}) \in \boldsymbol{\Psi} \subseteq \mathbb{R}^q$, $q \geq d$, 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.

Given this setup, the II method is based on the two steps described in Algorithm 2.

ALGORITHM 2

(i) Compute the pseudo-maximum likelihood estimate $\hat{\boldsymbol{\Psi}}(\boldsymbol{\theta}) = \arg\max_{\boldsymbol{\Psi}} \ell_n(\boldsymbol{\Psi}|F^{\boldsymbol{\theta}})$ *by maximizing the log pseudo-likelihood function using the observed data* $x_1, \ldots, x_n$*;*

(ii) Simulate n_s *observations* $\tilde{x}_1, \ldots, \tilde{x}_{n_s}$ *from the true model* $F^{\boldsymbol{\theta}}$ *and compute a new pseudo-MLE* $\hat{\boldsymbol{\psi}}^s(\boldsymbol{\theta}) = \arg\max_{\boldsymbol{\psi}} \ell_{n_s}(\boldsymbol{\psi}|F^{\boldsymbol{\theta}})$ *using the simulated data. Compute the indirect inference estimator*

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} (\hat{\boldsymbol{\Psi}}^s(\boldsymbol{\theta}) - \hat{\boldsymbol{\Psi}}(\boldsymbol{\theta}))' \boldsymbol{\Xi}^{-1} (\hat{\boldsymbol{\Psi}}^s(\boldsymbol{\theta}) - \hat{\boldsymbol{\Psi}}(\boldsymbol{\theta})),$$

where $\boldsymbol{\Xi}$ *is a weighting matrix.*

When the number of parameters of the true and auxiliary model is the same, i.e. when $d = q$, 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*

The skewed-t distribution $St(x;\omega,\kappa,\lambda,\nu)$ of Fernández and Steel (1998) is a generalization of the classical Student-t distribution. $\omega \in \mathbb{R}$ is a location parameter, $\kappa \in \mathbb{R}^+$ is a scale parameter, $\lambda \in \mathbb{R}^+$ and $\nu \in \mathbb{R}^+$ are shape parameters accommodating skewness and kurtosis, respectively. An estimation method for the g-and-h distribution based on II, and using as an auxiliary model the skewed-t distribution with $\boldsymbol{\Xi} = \boldsymbol{I}$ has been proposed in Bee et al. (2019). Here we extend their approach by developing an estimate of $\boldsymbol{\Xi}$ and employing it in the II estimation program.

An estimation method for the g-and-h distribution based on II, and using as an auxiliary model the skewed-t distribution with ... = ... has been proposed in

In finite samples, the choice of the instrumental distribution M plays a very important role: as pointed out by Garcia et al. (2011), it is advisable to use an auxiliary model such that the parameters of the instrumental and true distribution have the same interpretation. λ and ν in the skewed-t distribution govern skewness and kurtosis, exactly as g and h in the g-and-h. Hence, the skewed-t model is a suitable candidate. Moreover, Garcia et al. (2011) have shown that it performs well in II estimation of the stable distribution.

In an exactly identified setup (i.e., when $d = q$), 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 $x_1, \ldots, x_n$ from $X \sim gh(\boldsymbol{\theta})$, the bootstrap estimate of $\boldsymbol{\Xi}$ is obtained as follows:

ALGORITHM 3 *(Non-parametric bootstrap estimate of the weighting matrix)*

(i) Sample with replacement n observations $x_1^*, \ldots, x_n^*$ from $x_1, \ldots, x_n$;
(ii) Use $x_1^*, \ldots, x_n^*$ 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.

The maximization of the pseudo-log-likelihood function at Step (ii) is performed by means of standard optimization techniques. We use the quantile estimates of the g-and-h parameters (computed using the observed data) as starting values for the skewed-t parameters. Some experiments with different initializations have always given the same parameter estimates. 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).

Specifically, 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. Since $\mathrm{E}(X^r)$ only exists if $h \in [0, 1/r)$, the kurtosis is undefined in Setup (6).

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 $n_s = 5000$. The numerical value of n_s has been chosen according to the outcomes of a preliminary simulation study where II estimation has been repeated with $n_s \in \{1000, 3000, 5000, 7500, 1000\}$: whereas moving from $n_s = 1000$ to $n_s = 3000$ and from $n_s = 3000$ to $n_s = 5000$ yields better estimators in terms of RMSE, values of n_s larger than 5000 did not produce any improvement. Hence, we employ $n_s = 5000$ throughout[1]

[1]The detailed results are available upon request.

3.1. *Parameter estimation*

Figures 1 and 2 show the bias and the RMSE 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 $n = 1000$. The II approach uses the bootstrap estimate of the optimal weighting matrix.

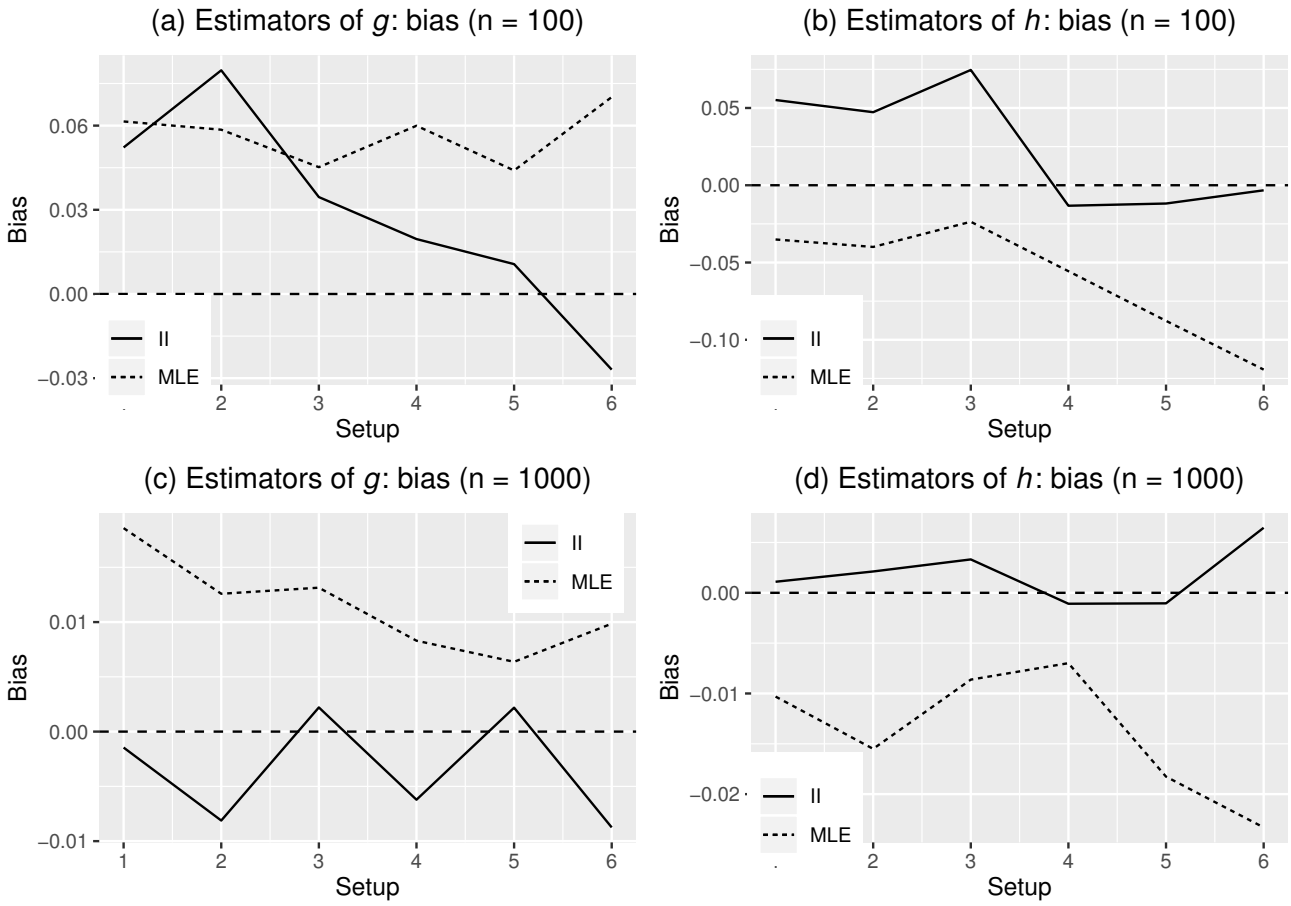


Figure 1.: Bias of the II and MLE estimators of g (panels (a) and (c)) and h (panel (b) and (d)); panels (a) and (b) are based on $n = 100$, (c) and (d) on $n = 1000$.

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, the II algorithm has always converged, we have discarded the samples and replaced them with new ones. The outcomes of unreported simulation experiments suggest that MLE failure becomes more and more frequent when the sample size decreases: for example, for $n = 20$, MLE does not converge in about 20% of the cases, whereas we are always able to compute an II estimate.

In terms of bias (see Fig. 1), II is better than MLE, more notably in setups 5 and 6 and for $n = 100$. This seems to suggest that in settings characterized by large (or even infinite) skewness and kurtosis, MLE is more biased for small sample size. This result is not surprising, since it is well known that II is a bias-correction method (Gouriéroux et al. 2000), whereas MLEs are consistent but 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

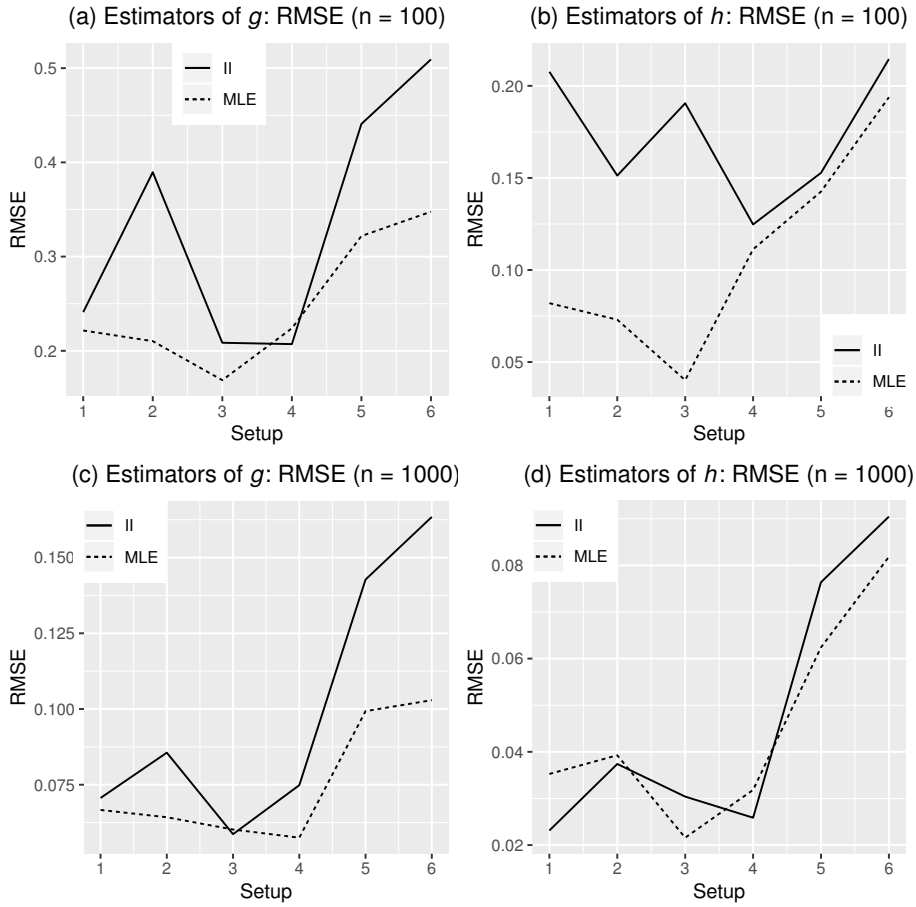


Figure 2.: RMSE of the II and MLE estimators of g (panels (a) and (c)) and h (panel (b) and (d)); panels (a) and (b) are based on $n = 100$, (c) and (d) on $n = 1000$.

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

in all setups when $n = 1000$.

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 $n = 100$, the optimal II method 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.

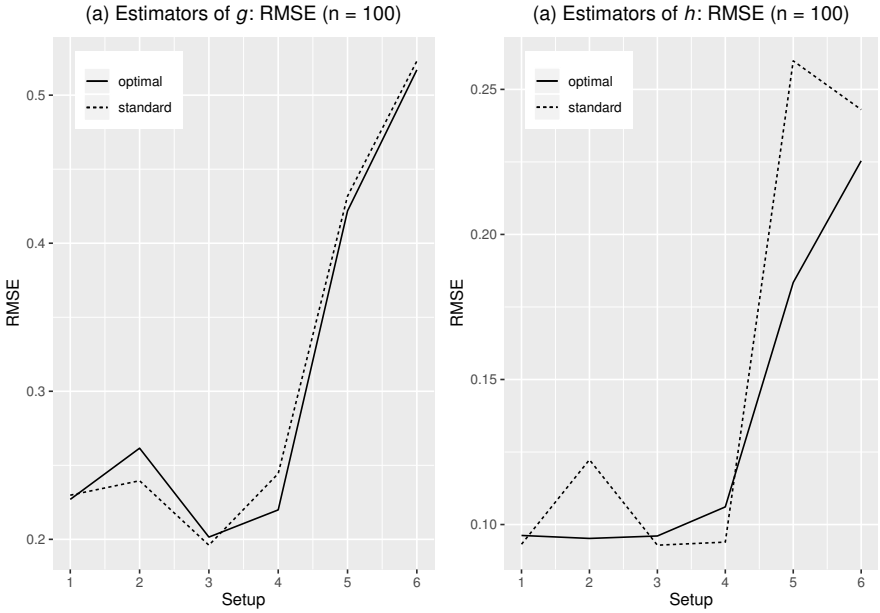


Figure 3.: RMSE of the optimal and standard II estimators of g (panel (a)) and h (panel (b)) for $n = 100$.

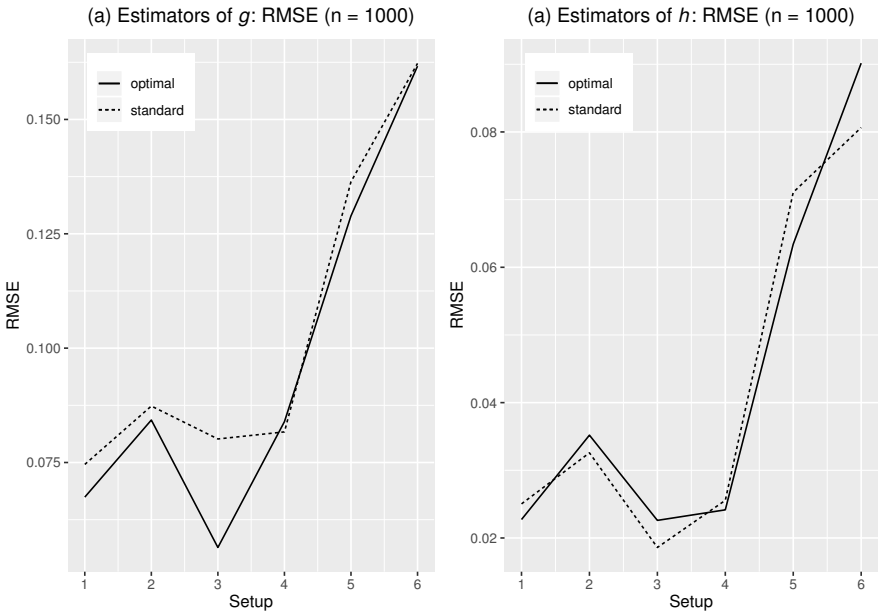


Figure 4.: RMSE of the optimal and standard II estimators of g (panel (a)) and h (panel (b)) for $n = 1000$.

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 are estimated, the VaR is 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

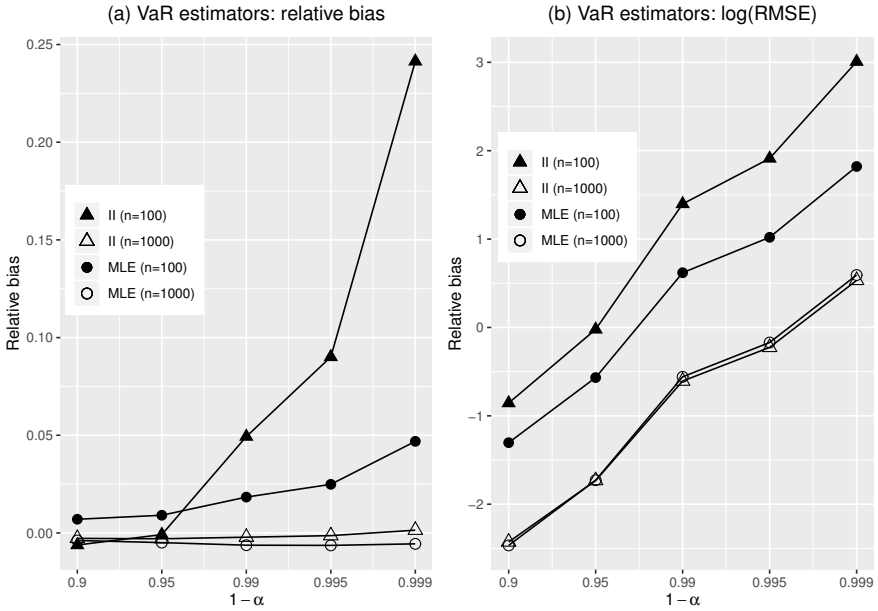


Figure 5.: Setup 1. Relative bias (panel (a)) and log-RMSE (panel (b)) of the II and MLE estimators of the VaR for different values of α.

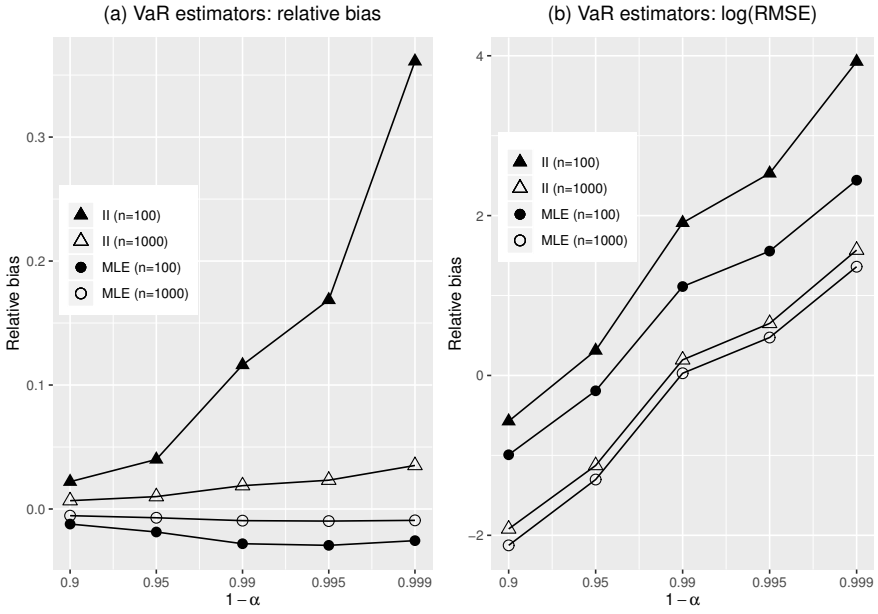


Figure 6.: Setup 2. Relative bias (panel (a)) and log-RMSE (panel (b)) of the II and MLE estimators of the VaR for different values of α.

when $n = 1000$, with minor differences in specific setups. On the other hand, when $n = 100$ the II estimator of VaR is 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. This is likely to be related to the bias of $\hat{g}^{MLE}$ and especially of $\hat{h}^{MLE}$ (see Fig. 1).

At first sight, the non-negligible bias of the II estimators in the first 5 setups with $n = 100$ is surprising, since $\hat{g}^{II}$ and $\hat{h}^{II}$ are less biased than the corresponding MLEs (see Fig. 1). However, since the g-and-h VaR is computed via (2), which is a non-linear function of g and h, unbiasedness of the estimators of g and h does not imply unbiasedness of the estimator of the VaR.

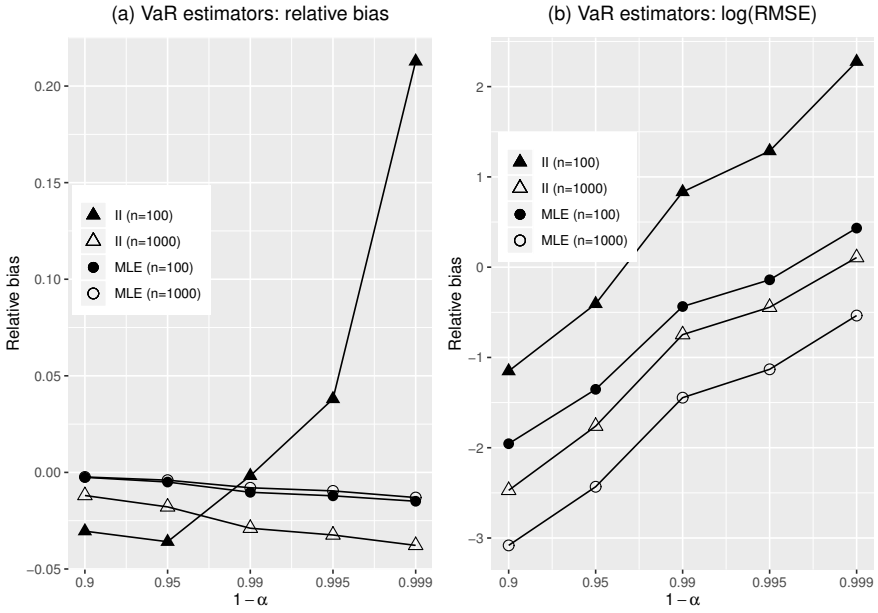


Figure 7.: Setup 3. Relative bias (panel (a)) and log-RMSE (panel (b)) of the II and MLE estimators of the VaR for different values of α.

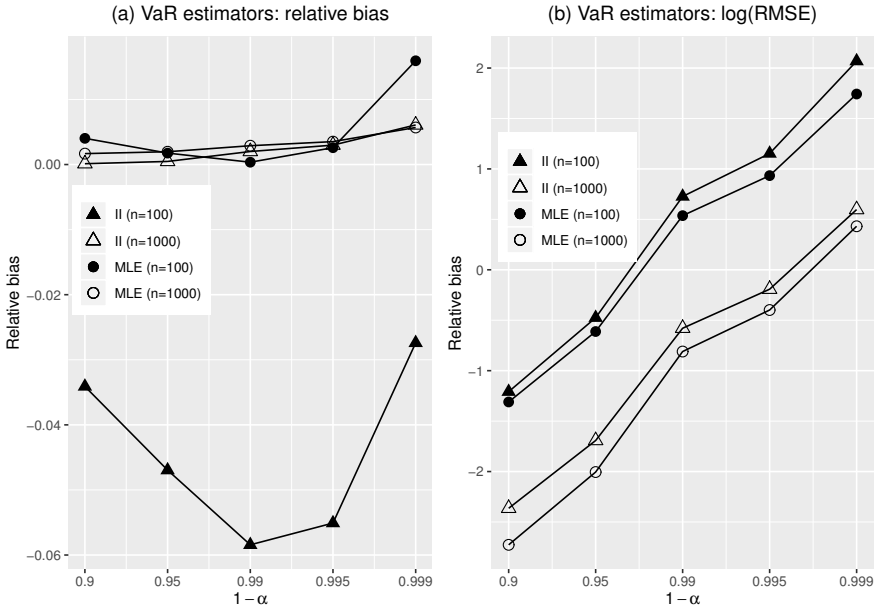


Figure 8.: Setup 4. Relative bias (panel (a)) and log-RMSE (panel (b)) of the II and MLE estimators of the VaR for different values of α.

An intuitive explanation is as follows. Since, from Fig. 2, $\text{RMSE}(\hat{g}^{II}) > \text{RMSE}(\hat{g}^{MLE})$ and $\text{RMSE}(\hat{h}^{II}) > \text{RMSE}(\hat{h}^{MLE})$, it must also be true that $\text{var}(\hat{g}^{II}) > \text{var}(\hat{g}^{MLE})$ and $\text{var}(\hat{h}^{II}) > \text{var}(\hat{h}^{MLE})$. Figure 11 shows the function:

$$DQ_\alpha(g,h) = Q(\alpha;0,1,g,h) - Q(\alpha;0,1,g^*,h^*),$$

which measures the difference between the quantile corresponding to parameters $(g,h) \in \mathrm{R} \times \mathrm{R}^+$ and the quantile obtained with $g^* = h^* = 0.2$ for $\alpha = 0.995$. It is clear that the difference increases

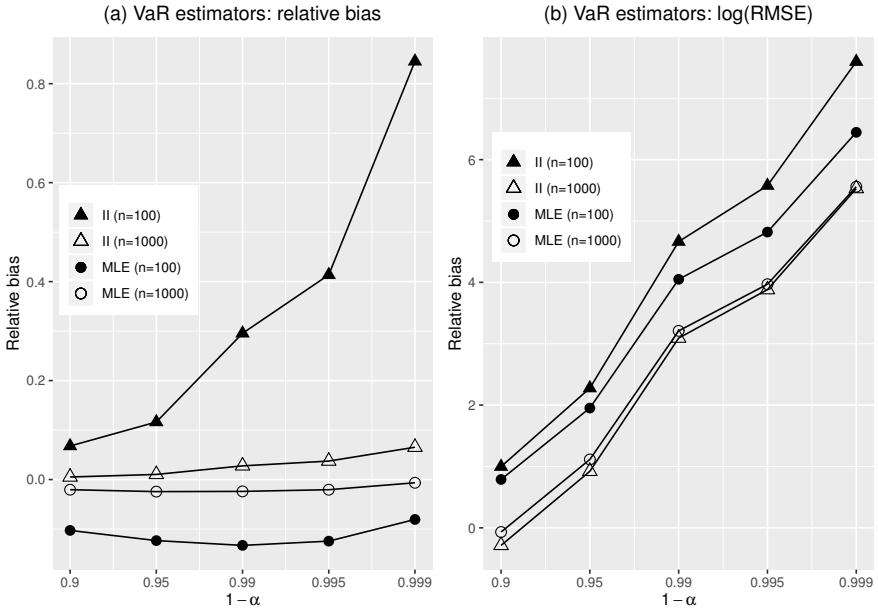


Figure 9.: Setup 5. Relative bias (panel (a)) and log-RMSE (panel (b)) of the II and MLE estimators of the VaR for different values of α.

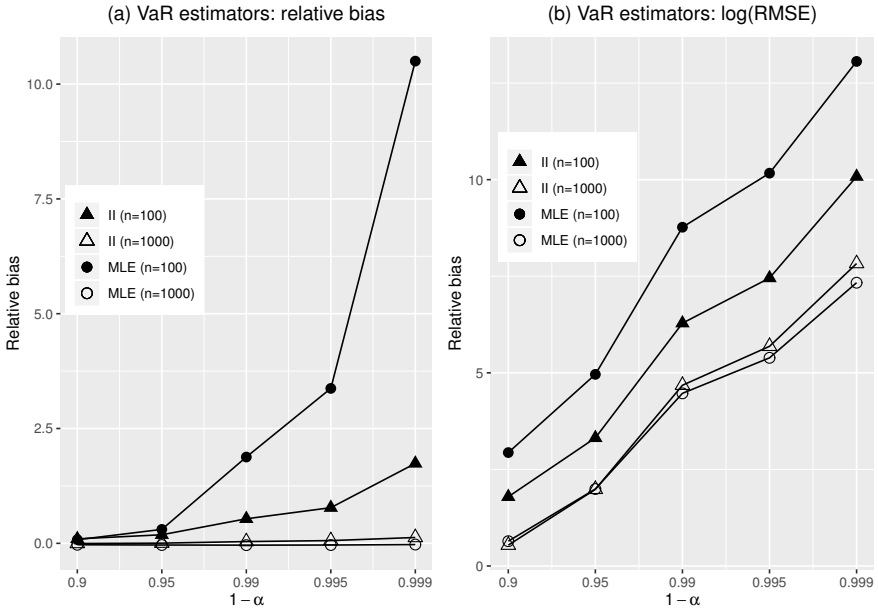


Figure 10.: Setup 6. Relative bias (panel (a)) and of the log-RMSE (panel (b)) of the II and MLE estimators of the VaR for different values of α.

sharply when both g and h become larger than the true parameters g^* and h^*. The larger variance of the II estimators implies that this happens more frequently for II than for MLE, so that the II (but not the MLE) VaR has a positive average bias. Notice also that this bias disappears when $n = 1000$, since for larger sample size the variance of the II estimators is smaller.

Figure 12 illustrates this phenomenon in Setup 1 with $n = 100$ and $\alpha = 0.999$: panels (a), (b) and (c) of the plot display the histograms of the simulated distributions of $\hat{g}^{II}$, $\hat{h}^{II}$ and $\widehat{VaR}^{II}$ respectively. The relative biases of each estimator are also reported in the three panels. The distributions of $\hat{g}^{II}$ and $\hat{h}^{II}$ are approximately unbiased, whereas the distribution of $\widehat{VaR}^{II}$ has a

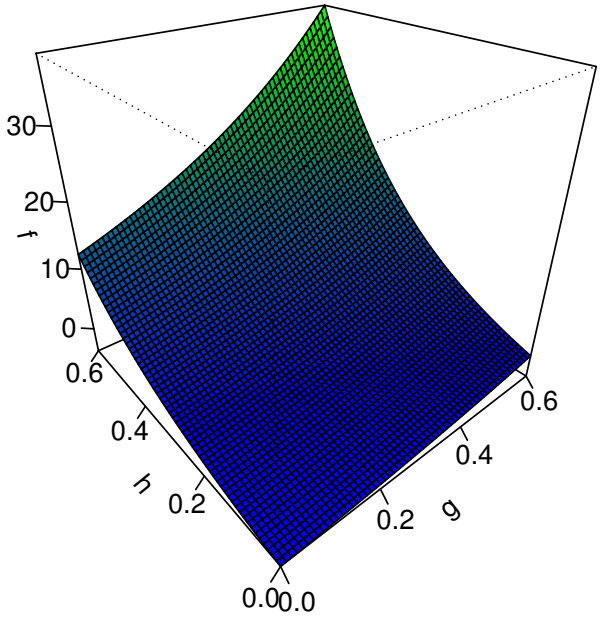


Figure 11.: Quantile difference as a function of g and h for $\alpha = 0.995$.

positive bias. As an aid to the interpretation, it is interesting to note that the two largest values of $\widehat{VaR}^{II}$ in panel (c) correspond to the two largest values of $\hat{g}^{II}$ in (a).

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: the II approach 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.

4. Empirical applications

In this section we outline the results of applications to actuarial data (Section 4.1) and to operational risk measurement (Section 4.2). In both cases we employ the II and MLE approaches and compare the outcomes. The II implementation is based on the bootstrap estimate of the weighting matrix.

4.1. *AON Re Belgium fire losses*

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

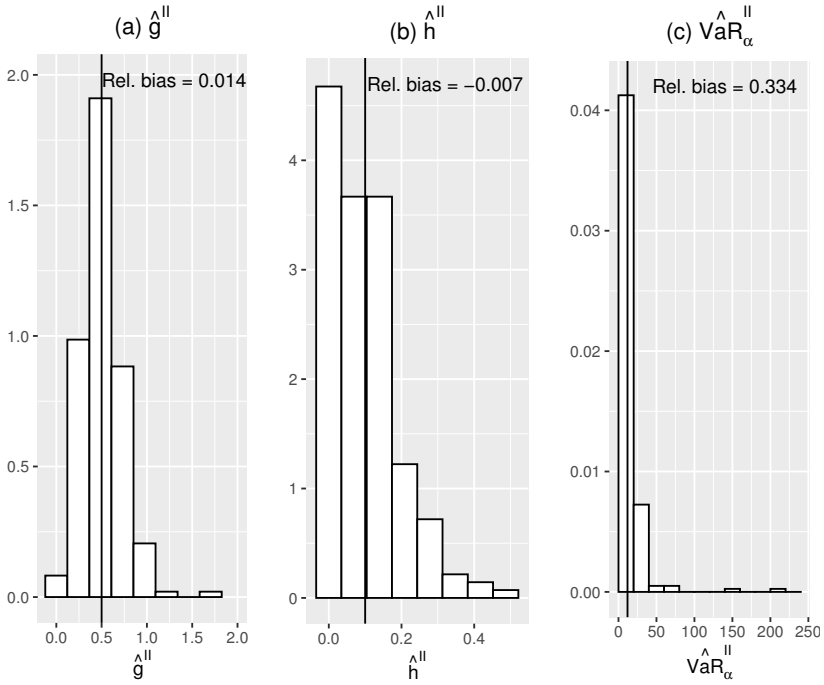


Figure 12.: Histograms of the simulated distributions of $\hat{g}^{II}$ (panel (a)), $\hat{h}^{II}$ (panel (b)) and $\widehat{VaR}_{\alpha}^{II}$ in Setup 1 with $n = 100$ and $\alpha = 0.999$. The vertical lines represent the true values of g (panel (a)), h (panel (b)) and VaR_{α} (panel (c)).

(variable `ClaimCost` in `beaonre`). For the subsequent analysis we standardize the observations[1]. The data are shown in Figure 13, where we also report the empirical skewness and kurtosis.

The quantile-based estimates of the location and scale parameters are $\hat{a} = 8.798$ and $\hat{b} = 15.656$. Point estimates of g and h obtained by means of the optimal II and numerical MLE approaches are shown in Table 2. Standard errors are computed via non-parametric bootstrap with 200 replications. Table 2 also reports the p-value of the Anderson-Darling (AD) test for the null hypothesis that the observed data $x_1, \ldots, x_n$ arise from the g-and-h distribution with parameters $(\hat{g}^{II}, \hat{h}^{II})$ and $(\hat{g}^{MLE}, \hat{h}^{MLE})$, respectively. To double-check the results, we show the p-values obtained by means of both the asymptotic approximation and the simulation of the null distribution; see the documentation of the `kSamples R` package (Scholz and Zhu 2019) for details.

Table 2.: Fire losses: Parameter estimates, bootstrap standard errors and p-value of the Anderson-Darling (AD) test.

	g	h	Asym. AD p-value	Sim. AD p-value
II	2.500 (0.130)	0.268 (0.097)	0.079	0.081
MLE	2.237 (0.103)	0.070 (0.028)	0.061	0.058

Looking at Table 2, we see that the II and MLE estimates of the parameters are quite different. In particular, $\hat{h}^{II}$ is much larger than $\hat{h}^{MLE}$. We discuss this issue in detail at the end of the section. Standard errors of II estimators are larger, similarly to the next application (see Table 4 below). This may be related to the additional sampling variability associated to any simulation-based procedure. In both cases, the p-values of the AD test do not yield a clear-cut result about the g-and-h distributional assumption.

[1]Standardized observations are defined as $(y_i - \hat{a})/\hat{b}$, $i = 1, \ldots, n$, where $\hat{a}$ and $\hat{b}$ are the Hoaglin (1985) quantile estimators (see Section 2).

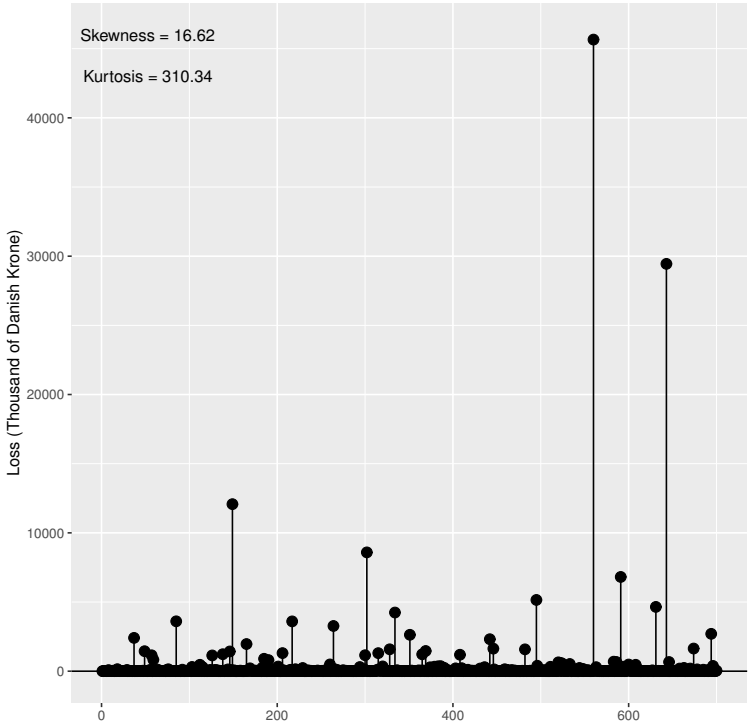


Figure 13.: AON Re Belgium fire losses.

Figure 14 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 II estimation method fits better the majority of data, but underestimates the largest observation. On the other hand, MLE is apparently more affected by the largest observation, and consequently fits less well the rest of the data.

In summary, the outcomes of the AD-test and the evidence of the QQ-plots in Figure 14 suggest that the two estimated g-and-h distributions yield a similar fit to the whole distribution, but the II-based distribution works better in the tail.

Table 3 reports the VaR measures estimated by means of the g-and-h distribution. For comparison purposes, we employ two additional techniques: the POT method, which is the state-of-the-art approach to the measurement of VaR in operational risk applications at large confidence levels (Moscadelli 2004, Dutta and Perry 2006), and the naïve approach based on the lognormal approximation of the loss distribution.

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

	$\alpha = 0.9$	$\alpha = 0.95$	$\alpha = 0.99$	$\alpha = 0.995$
II-VaR	12.656	35.232	292.427	658.164
	(1.586)	(6.691)	(95.780)	(256.696)
MLE-VaR	7.649	18.685	95.308	173.541
	(1.469)	(6.965)	(110.288)	(352.609)
POT-VaR	12.764	45.985	302.077	633.827
	(3.071)	(9.988)	(92.917)	(296.564)
Logn-VaR	9.672	20.495	81.022	133.404
	(1.410)	(3.381)	(16.776)	(29.843)
Emp.	12.765	42.022	270.568	491.888

The outcomes convey three messages. First, the II VaR is much closer to the empirical quantile

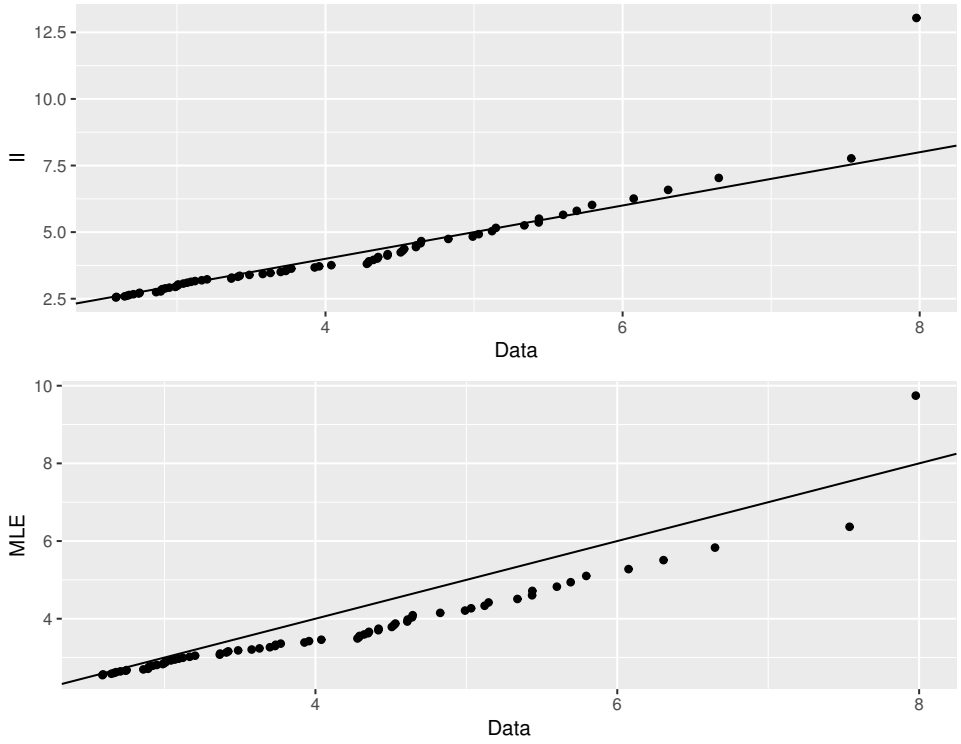


Figure 14.: QQ-plot of the observed losses vs. observations simulated from the II-estimated g-and-h (upper panel) and of the observed losses vs. observations simulated from the MLE-estimated g-and-h (lower panel).

than the MLE VaR, as expected given the better fit to the tail in Figure 14. 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. Finally, the lognormal VaR is too small at the highest levels.

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, in line with the properties of the method (Jiang and Turnbull 2004, Sect. 2.5), is that II might be more robust than MLE with respect to possible misspecifications of the model: as suggested by the outcomes of the AD test in Table 2, 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. Figure 15 displays the data, scaled by an unknown factor for confidentiality reasons, along with the sample skewness and kurtosis.

The estimates of the location and scale parameters computed via the quantile-based method are $\hat{a} = 4989.7$ and $\hat{b} = 5503.3$. The II and MLE estimates of the parameters g and h obtained with standardized observations are shown in Table 4. The two approaches yield quite similar results, and the null hypothesis of identical distributions is not rejected. Figure 16 displays the QQ plot of true vs. observations simulated form the estimated models. Analogously to the previous section, on the upper panel we display the results obtained with simulated data from the II-based estimated

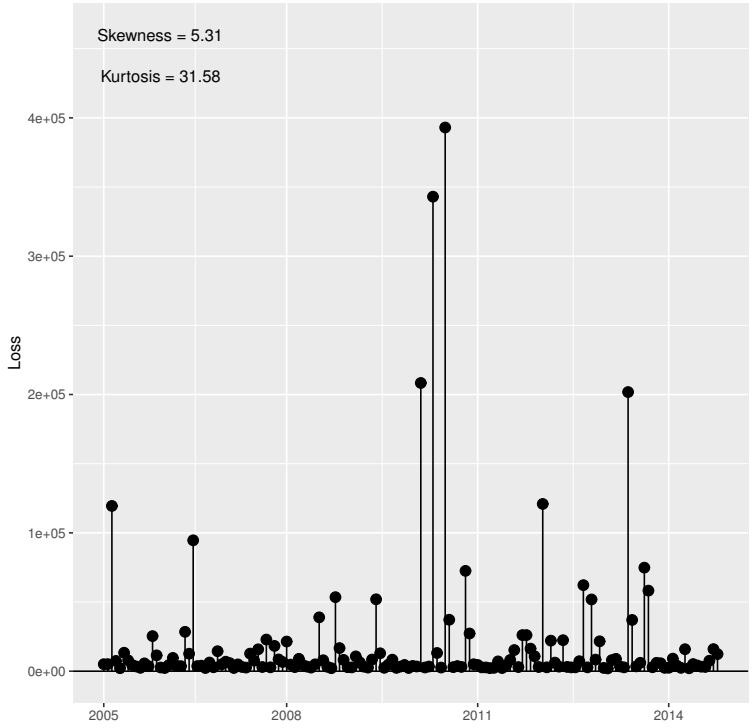


Figure 15.: Operational risk losses.

g-and-h distribution, whereas on the lower panel we use data simulated from the MLE-based g-and-h distribution. 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, the POT method and the naïve lognormal approach, as well as the empirical quantile[1]. In this case the VaRs obtained from the g-and-h estimated via II ($\widehat{VaR}^{II}$) and MLE ($\widehat{VaR}^{MLE}$) are similar, but again $\widehat{VaR}^{II}$ is closer to the empirical quantile. On the other hand, both $\widehat{VaR}^{II}$ and $\widehat{VaR}^{MLE}$ are in line with the empirical quantiles reported in the last line of the table. In terms of standard deviation, their performance is mostly better than POT. Analogously to the previous application, the lognormal underestimates the tail of the distribution.

Table 4.: Operational risk: Parameter estimates, bootstrap standard errors and p-value of the Anderson-Darling (AD) test.

	g	h	Asym. AD p-value	Sim. AD p-value
II	1.969	0.029	0.807	0.800
	(0.264)	(0.199)		
MLE	2.050	0.028	0.954	0.956
	(0.140)	(0.012)		

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

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

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.

	$\alpha = 0.9$	$\alpha = 0.95$	$\alpha = 0.99$	$\alpha = 0.995$
II-VaR	6.187	12.933	52.997	88.506
	(0.872)	(3.990)	(32.760)	(70.726)
MLE-VaR	6.371	14.259	61.500	104.708
	(0.760)	(3.396)	(25.474)	(54.535)
POT-VaR	6.351	14.384	45.926	68.128
	(0.908)	(3.818)	(40.645)	(83.807)
Logn-VaR	5.142	9.936	32.459	49.684
	(1.212)	(2.536)	(10.02)	(16.497)
Emp.	5.848	12.461	48.940	-

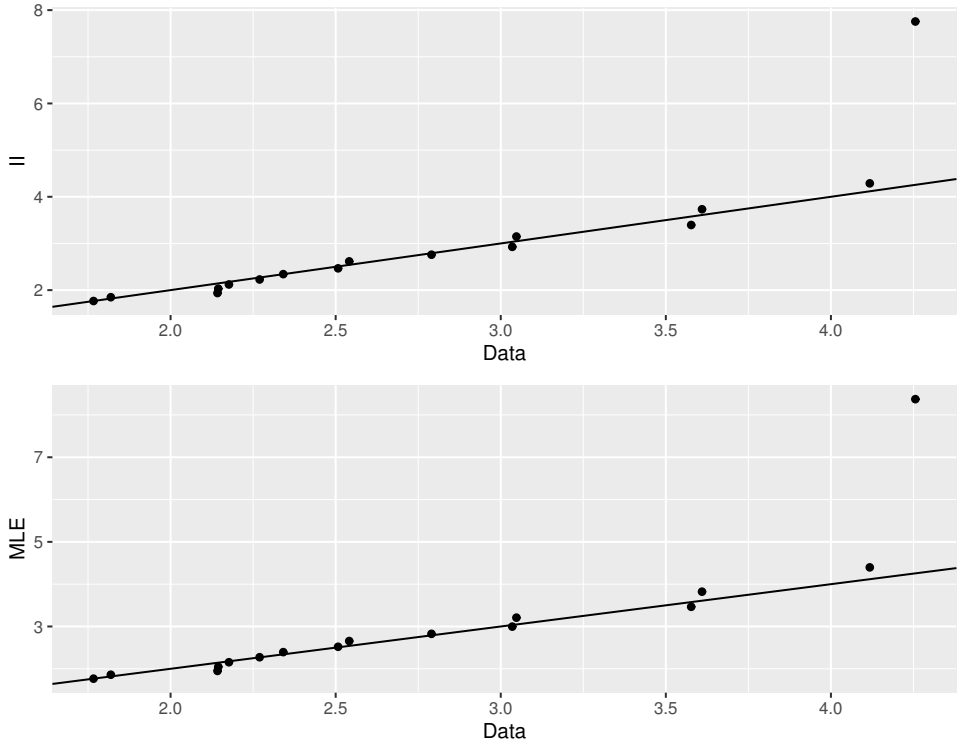


Figure 16.: QQ-plot of the observed losses vs. observations simulated from the II-estimated g-and-h (upper panel) and vs. observations simulated from the MLE-estimated g-and-h (lower panel).

likelihood method is more efficient 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. Moreover, the g-and-h distribution seems particularly well suited to model highly asymmetric and heavy-tailed data. For moderately skewed data, the scaled lognormal obtained when $h = 0$ may be a more parsimonious 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.

It may be worth investigating the use of instrumental models other than the skewed-t distribution. In so doing, one should probably restrict the analysis to models with two shape parameters, related to skewness and kurtosis respectively. A thorough analysis of this issue is open to 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.

Acknowledgements

We thank two anonymous reviewers for their valuable comments. We also thank Fabio Piacenza (UniCredit SpA) for providing us with the operational risk data. The codes used for all the computations carried out in this paper are available at `http://marcobee.weebly.com/software.html`.

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*, 19:1255–1266.

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.

Cruz, M. G. (2018). Editor's letter. *Journal of Operational Risk*, 13(1).

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.

Fernández, C. and Steel, M. (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, 93(441):359–371.

Garcia, R., Renault, E., and Veredas, D. (2011). Estimation of stable distributions by indirect inference. *Journal of Econometrics*, 161(2):325 – 337.

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.

Gouriéroux, C., Renault, E., and Touzi, N. (2000). Calibration by simulation for small sample bias correction. In Mariano, R., Schuermann, T., and Weeks, M. J., editors, *Simulation-Based Inference in Econometrics: Methods and Applications*, pages 328–358. Cambridge University Press, Cambridge.

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.

Headrick, T., Kowalchuk, R., and Sheng, Y. (2008). Parametric probability densities and distribution functions for Tukey g-and-h transformations and their use for fitting data. *Applied Mathematical Sciences*, 2:449–462.

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

Hodis, F., Headrick, T., and Sheng, Y. (2012). Power method distributions through conventional moments and L-moments. *Applied Mathematical Sciences*, 6(44):2159–2193.

Huber, P. J. (1981). *Robust Statistics*. New York, Wiley.

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.

Moscadelli, M. (2004). The modelling of operational risk: experiences with the analysis of the data collected by the basel committee. Technical report, Bank of Italy, Working Paper No 517.

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.

Peters, G. W. and Sisson, S. (2006). Bayesian inference, Monte Carlo sampling and operational risk. *Journal of Operational Risk*, 1(3):27–50.

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.

Scholz, F. and Zhu, A. (2019). *kSamples: K-Sample Rank Tests and their Combinations*. R package version 1.2-9.

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

Vogel, R. and Fennessey, N. (1993). L-moment diagrams should replace product moment diagrams. *Water Resources Research*, 29:1745–1752.