

To appear in *Quantitative Finance*, Vol. 00, No. 00, Month 20XX, 1–19

Estimating Value-at-Risk for the g-and-h distribution: an indirect inference approach

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, and

Chair of Statistics, Faculty of Economics, University of Goettingen, Goettingen, Germany

[§]Department of Economic Policy, Università Cattolica del Sacro Cuore, Milano, Italy

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

The g-and-h distribution is a flexible model with desirable theoretical properties. Especially, it is able to handle well the complex behavior of loss data. However, parameter estimation is difficult, because the density cannot be written in closed form. In this paper we develop an indirect inference method using the skewed- t distribution as instrumental model. We show that the skewed- t is a well suited auxiliary model and study the numerical issues related to its implementation. A Monte Carlo analysis and an application to operational losses suggest that the indirect inference estimators of the parameters and of the VaR outperform the quantile-based estimators.

Keywords: Value-at-Risk; g-and-h distribution; loss model; indirect inference; simulation; intractable likelihood.

JEL Classification: C15; C46; C51; G22

1. Introduction

The g-and-h distribution, first introduced by Tukey (1977), is constructed via a simple transformation of the standard normal random variable, and is a flexible model for skewed and heavy-tailed data. In its general version, it is characterized by a parameter vector $(a, b, g, h)' \in \Theta = \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}$. The parameters g and h determine skewness and kurtosis, whereas a and b are respectively a location and a scale parameter; see Cruz *et al.* (2015, Section 9.4.1) for more details.

The following stochastic representation of $X \sim \text{gh}(a, b, g, h)$ identifies the map from the normal to the g-and-h distribution:

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

If $h = 0$, one obtains the g distribution, which is a scaled lognormal. The case $g = 0$ is the h distribution, which is symmetric but leptokurtic.

Over the years, the distribution has been successfully used in several fields where a proper account of skewness and/or leptokurtosis is important. However, it has proven to be especially relevant as a loss model in risk measurement and insurance modeling (Dutta and Babbel 2002, Hübner *et al.* 2005, Peters and Sisson 2006, Degen *et al.* 2007, Jobst 2007, Galloppo and Rogora 2011, Jiménez

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

and Arunachalam 2011, Cruz *et al.* 2015, Peters *et al.* 2016). A precise assessment of both the probability and the magnitude of extreme losses is crucial in these frameworks, so that the tail behavior of the model employed plays a key role. As pointed out by Dutta and Perry (2006), it may happen that the overall fit of a certain distribution is good, but extreme events are estimated inaccurately, resulting in unrealistic capital estimates. The g-and-h distribution, on the contrary, appears to provide a good fit overall and in the tails.

A way to circumvent this issue is to rely on Extreme Value Theory (EVT). Indeed, EVT provides a well-grounded theoretical setup to study the statistical properties of the tail of a distribution. However, whereas the excesses of a distribution converge to the Generalized Pareto distribution (GPD) under rather mild conditions, Degen *et al.* (2007) show that the convergence to the GPD is extremely slow for the g-and-h distribution. Therefore, quantile estimation using EVT may lead to imprecise outcomes. Accordingly, if the true data-generating-process is g-and-h, it is important to compute tail probabilities and risk measures using the actual g-and-h distribution instead of relying on the GPD. Indeed, except if the sample size is very large, the GPD will be heavily misspecified.

The g-and-h distribution has two main distinct features with respect to possible competing size distributions. The first one is the peculiar tail behavior, studied at length in Degen *et al.* (2007). The main theoretical result is summarized in Table 1 of their paper: among the distributions considered, the g-and-h has the slowest rate of convergence to the Generalised Pareto distribution (GPD). The simulation experiments carried out by Degen *et al.* (2007) confirm that this feature has a relevant impact in practice.

The second one is the theoretical skewness-kurtosis range, which for the g-and-h is larger than for similar models: see Fig. 3 in Dutta and Perry (2006), where the g-and-h is compared to the normal, lognormal, gamma, beta of the second kind, log-logistic, Weibull and GPD. Hence, the g-and-h captures the largest variety of skewness-kurtosis combinations.

Accurate estimates of tail-risk measures require not only the appropriate model, but also reliable estimates of the parameters. Unfortunately, fitting the g-and-h distribution to data via maximum likelihood is difficult, because there is no closed-form density, so that it is impossible to write the likelihood function.

To overcome this problem, other estimation techniques have been developed. One way of proceeding is based on the explicit availability of the quantile function: accordingly, the earliest, and still most widespread, estimation method matches empirical and theoretical quantiles (Hoaglin 1985). Similarly, the r -th moment ($r = 1, \dots, 4$) of the g-and-h distribution exists and can be computed in closed form if $h \in [0, 1/r)$; see Cruz *et al.* (2015, p. 320). Hence, an approach using L-moments has been developed by Peters *et al.* (2016). Finally, it is possible to exploit the fact that the simulation of the g-and-h distribution is straightforward: it is enough to simulate standard normal random numbers and apply the stochastic representation (1). This opens the door to computer-intensive estimation methods such as Approximate Maximum Likelihood Estimation (Bee and Trapin 2016).

In this paper we tackle the estimation problem using the principle of indirect inference. The method has been proposed by Gourieroux *et al.* (1993) and is based on two steps. In the first one, the parameters are estimated using a misspecified model (usually termed *instrumental model*) for which the estimators are easily computed in a standard way, but are typically not even consistent. The second step, based on simulations from the model of interest and on standard optimization routines, gives an adjusted estimator called *indirect inference* estimator. Under mild regularity conditions, this estimator is consistent and asymptotically normal.

Here we use as instrumental distribution the skewed- t (Fernández and Steel 1998). This distribution is very well suited in the present case since both the skewed- t and the g-and-h have four parameters and, more importantly, the parameters have the same interpretation in terms of the moments of the distribution. The skewed- t has already been employed in a similar setup by Garcia *et al.* (2011) for indirect inference estimation of the stable distribution.

On the theoretical side, we find the exact relationship between the g-and-h and skewed- t parameters, which justifies the use of the skewed- t as auxiliary model. On the numerical side, we study the numerical issues related to the implementation of the algorithm. Our simulation experiments

compare the indirect inference estimators to the classical quantile-based estimators. The outcomes suggest that the former produces better parameter estimates in terms of root-mean-squared-error. The simulated parameter values are then used to assess the precision of the Value-at-Risk (VaR) estimates obtained with the two methods. The superiority of the indirect inference approach is clear in this exercise. Finally, we fit the g-and-h to two datasets of operational losses recorded in 2014 in the Italian bank Unicredit. The empirical analysis confirms the importance of using the g-and-h and of estimating the parameters via indirect inference.

The rest of the paper is organized as follows. Section 2 reviews the method of indirect inference. Section 3 gives some background on the g-and-h distribution and develops indirect inference estimation based on the skewed- t distribution. Section 4 contains the results of the Monte Carlo simulation. Section 5 computes VaR using operational losses recorded in the Unicredit bank. Section 6 concludes.

2. Indirect Inference: some background

Suppose to be interested in estimating the parameters of the distribution F^θ , $\theta \in \Theta \subseteq \mathbb{R}^p$, of a random variable Y for which it is difficult or impossible to write the likelihood function. Let M be an auxiliary model with density $f_M(m; \Psi)$, where Ψ is a vector of parameters. Given a random sample $\mathbf{y}_n = (y_1, \dots, y_n)'$ iid $\sim F^\theta$, the log pseudo-likelihood function, constructed using the density of the auxiliary model and the sample \mathbf{y}_n from F^θ , is denoted by $\ell_n(\Psi|F^\theta) = \sum_{i=1}^n \log f_M(y_i; \Psi)$.

The pseudo-true value Ψ can be written as $\Psi(\theta) = \arg \max_{\Psi} \ell_n(\Psi|F^\theta)$. The map $\theta \rightarrow \Psi(\theta)$ is called binding function and the auxiliary parameters $\Psi(\theta)$ are likely to contain a large amount of information about θ .

The indirect inference approach is a two-step procedure. In a likelihood setup, the first step is the computation of the auxiliary (or *quasi-maximum likelihood*) estimate $\hat{\Psi}(\theta) \in \Theta^a \subseteq \mathbb{R}^q$, $q \geq p$, i.e. the solution of the problem $\arg \max_{\Psi} \ell_n(\Psi|F^\theta)$, using the observed data \mathbf{y}_n ¹.

The second step is based on the simulation of n_s observations y_1, \dots, y_{n_s} from the true model, which are then used to compute the estimate $\hat{\Psi}^s(\theta) = \arg \max_{\Psi} \ell_{n_s}(\Psi|F^\theta)$. The indirect inference estimate $\hat{\theta}(\hat{\Psi}(\theta)) \stackrel{\text{def}}{=} \hat{\theta}$ is finally obtained as solution of the problem $\min_{\theta} (\hat{\Psi}^s(\theta) - \hat{\Psi}(\theta))' \Xi^{-1} (\hat{\Psi}^s(\theta) - \hat{\Psi}(\theta))$, where Ξ is a weighting matrix.

In general, the asymptotic distribution of the indirect inference estimator $\hat{\theta}$ depends on Ξ . However, in a just-identified setup, i.e. when $p = q$ (which is the case of the present paper), the asymptotic covariance matrix of $\hat{\theta}$ does not depend on the weighting matrix (Gourieroux *et al.* 1993, Proposition 5). This remark simplifies the implementation of the method, implying that one can safely use the identity matrix as weighting matrix, i.e. put $\Xi = \mathbf{I}$. This also implies that, in the just-identified setting, the indirect inference estimator is identical to the score-matching estimator developed by Gallant and Tauchen (1996); see Gourieroux *et al.* (1993, Section 2.5) for more details.

A pseudo-code is as follows:

ALGORITHM 1

- (i) Let y_1, \dots, y_n iid $\sim F^\theta$ be the observed data. Compute $\hat{\Psi}(\theta) = \arg \max_{\Psi} \ell_n(\Psi|F^\theta)$, i.e. the maximizer of the log pseudo-likelihood function $\ell_n(\Psi; F^\theta) = \sum_{i=1}^n \log f_M(y_i; \Psi)$ based on the pdf of the auxiliary model M ;
- (ii) Compute the indirect inference estimator

$$\hat{\theta} = \arg \min_{\theta} (\hat{\Psi}^s(\theta) - \hat{\Psi}(\theta))' \Xi^{-1} (\hat{\Psi}^s(\theta) - \hat{\Psi}(\theta)), \quad (2)$$

¹In general, it is possible to define $\hat{\Psi}(\theta)$ as any functional, of dimension $q \geq p$, of the empirical distribution function $F^{(n)}$. See Jiang and Turnbull (2004) for a discussion of this possibility, which we will not pursue in this paper.

where $\hat{\Psi}^s(\boldsymbol{\theta}) = \arg \max_{\boldsymbol{\theta}} \ell_{n_s}(\Psi|F^{\boldsymbol{\theta}})$ is the log pseudo-likelihood based on n_s observations y_1, \dots, y_{n_s} sampled from $F^{\boldsymbol{\theta}}$. Large-sample considerations justify the use of $\Xi = \mathbf{I}$ if $\dim(\Psi) = \dim(\boldsymbol{\theta})$.

Asymptotic properties of indirect inference estimators are similar to MLE under some regularity conditions (Jiang and Turnbull 2004). The key requirement is that the auxiliary statistic $\hat{\Psi}(\boldsymbol{\theta})$ has an asymptotically normal distribution and is \sqrt{n} -consistent for $\Psi(\boldsymbol{\theta})$ for the correctly specified model. If this is the case, $\hat{\Psi}(\boldsymbol{\theta})$ remains asymptotically normal and \sqrt{n} -consistent for $\Psi(\boldsymbol{\theta})$ even under misspecification, i.e. if the observations y_1, \dots, y_n are generated by a model different from the auxiliary model. Formally, the three regularity conditions that must be satisfied are:

- (a) $\sqrt{n}\{\hat{\Psi}(\boldsymbol{\theta}) - \Psi(\boldsymbol{\theta})\} \xrightarrow{D} N(\mathbf{0}, \boldsymbol{\nu})$;
- (b) $\boldsymbol{\nu}$ is positive definite and symmetric and $n\boldsymbol{\nu} \xrightarrow{P} \boldsymbol{\nu}$, where $\boldsymbol{\nu}$ is an estimate of the asymptotic covariance matrix $\boldsymbol{\nu}$, which is given by Gourieroux *et al.* (1993, Prop. 5);
- (c) $\Psi(\cdot)$ is second-order continuously differentiable in a neighborhood of $\boldsymbol{\theta}$ and the derivative matrix Ψ' is of full rank at $\boldsymbol{\theta}$.

3. Indirect inference for the g-and-h distribution

3.1. Estimating the g-and-h distribution: some background

The benchmark estimators of the parameters are the quantile-based estimators (Q -estimators; Hoaglin 1985), which exploit the simple formula of the p -th quantile:

$$x_p = a + b \frac{e^{gz_p} - 1}{g} e^{hz_p^2/2},$$

where z_p is the p -th quantile of the standard normal distribution.

The Q -estimators are found as follows. The parameters g and h can be treated separately, so that they can be estimated sequentially. First, we compute the value of g that exactly matches the quantiles x_p and x_{1-p} . For $0 < p < 0.5$, the following holds:

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

Hence, g 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 g is usually taken to be equal to the median of the g_p 's. Given the estimate of g , we proceed to the estimation of h . Dividing the upper half spread $\text{UHS}_p = x_{1-p} - x_{0.5}$ by $(e^{-gz_p} - 1)$ and adjusting for the skewness, we obtain

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

A plot of $\log(\text{UHS}_p^*)$ against $z_p^2/2$ has intercept $\log(b)$ and slope h , hence the estimate of b is obtained by applying the exponential function to the intercept and the linear regression coefficient is an estimate of h .

Dupuis and Field (2004) developed another estimation method which is robust to the presence of

outliers. In particular, they propose an approach that fits the parameters g and h based on Huber (1981) ‘‘Proposal 2’’. Dupuis and Field (2004) suggest to fix a set of quantiles and use 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_1 is a positive constant. The estimated parameters are obtained by solving

$$\min_{g,h} \sum_i \rho((x_i - y_i)/y_i)$$

over a specified set of quantiles.

3.2. Indirect inference based on a skewed- t auxiliary model

We propose to specify indirect inference (II) for the g-and-h distribution using the skewed- t (Fernández and Steel 1998) auxiliary model, which is characterized by a parameter vector $\Psi = (\nu, \gamma, \lambda, \omega)'$, with $\nu \in (0, \infty)$, $\gamma \in (0, \infty)$, $\lambda \in (0, \infty)$ and $\omega \in \mathbb{R}$. The density is given by

$$f(x; \omega, \lambda, \gamma, \nu) = \frac{2}{\gamma + \frac{1}{\gamma}} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \lambda^{-1} \times \left[1 + \frac{(x - \omega)^2}{\nu \lambda^2} \left\{ \frac{1}{\gamma^2} \mathbb{1}_{[0, \infty)}(x - \omega) + \gamma^2 \mathbb{1}_{(-\infty, 0]}(x - \omega) \right\} \right]^{-(\nu+1)/2}, \quad (3)$$

where ν is the number of degrees of freedom and determines tail heaviness, γ is a skewness parameter and λ and ω are respectively a scale and a location parameter.

Given a random sample y_1, \dots, y_n from $Y \sim \text{gh}(a, b, g, h)$, the pseudo log-likelihood corresponding to (3) can be written as:

$$\ell_n = \log \left(\frac{2h(\nu)}{\sqrt{\pi\nu}} \frac{1}{\lambda \left(\frac{\gamma}{1+\gamma} \right)} \right) - \frac{\nu+1}{2} \sum_{i=1}^n \log \left(1 + \frac{1}{\nu} \left(\frac{y_i - \omega}{\lambda} \right)^2 g_\omega(y_i, \gamma) \right), \quad (4)$$

with

$$h(\nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})}, \quad \text{and} \quad g_\omega(y, \gamma) = \begin{cases} \frac{1}{\gamma^2} & \text{if } y \geq \omega \\ \gamma^2 & \text{if } y < \omega \end{cases}.$$

The skewed- t distribution is a well-suited model in this setup for at least three reasons: it has the same number of parameters of the g-and-h, the binding function is one-to-one and there is a nice and easily interpretable correspondence between the skewed- t and the g-and-h parameters, as the following proposition shows.

PROPOSITION 1 For all $(a, b, g, h)' \in \Theta$:

$$\begin{aligned}\omega(a, b, g, h) &= a + \omega(0, b, g, h); \\ \lambda(a, b, g, h) &= b\lambda(a/b, 1, g, h); \\ \gamma(a, b, g, h) &= \gamma(0, 1, g, h) = \frac{1}{\gamma(0, 1, -g, h)}; \\ \nu(a, b, g, h) &= \nu(0, 1, g, h) = \nu(a, b, -g, h).\end{aligned}$$

Proof: see Appendix A.

To find the II estimator one has to solve the problem (2). In general, this requires the use of deterministic numerical routines, so that the performance of II estimation is expected to be better when (2) is low-dimensional. Moreover, the location and scale parameters are less important for tail probability estimation. Accordingly, we prefer to use the Q -estimators of a and b , say \hat{a}^q and \hat{b}^q , and employ indirect inference for estimating the parameters of the distribution of the standardized observations $\tilde{y}_i \stackrel{\text{def}}{=} (y_i - \hat{a}^q)/\hat{b}^q$, $i = 1, \dots, n$.

With standardized observations, the estimator $\hat{\Psi}(\boldsymbol{\theta})$ is the maximizer of the pseudo log-likelihood (4) with $\omega = 0$ and $\lambda = 1$, which can be found numerically.

The pseudo-code of the estimation procedure is Algorithm 1 with a standardized skewed- t auxiliary model and observations following the standardized g-and-h distribution. For the sake of completeness, we detail it here.

ALGORITHM 2

- (i) Let $\tilde{y}_1, \dots, \tilde{y}_n$ iid $\sim \text{gh}^{\boldsymbol{\theta}}$, with $\boldsymbol{\theta} = (g, h)'$, be the standardized observed data. Compute $\hat{\Psi}(\boldsymbol{\theta}) = (\hat{\gamma}, \hat{\nu}) = \arg \max_{\Psi} \ell_n(0, 1, \gamma, \nu; \tilde{y}_1, \dots, \tilde{y}_n)$, i.e. the maximizer of the log pseudo-likelihood function ℓ_n (4) with $\omega = 0$ and $\lambda = 1$ based on the pdf of the standard skewed- t auxiliary model;
- (ii) Compute the indirect inference estimator

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= (\hat{g}, \hat{h}) = \arg \min_{\boldsymbol{\theta}} (\hat{\Psi}^s(\boldsymbol{\theta}) - \hat{\Psi}(\boldsymbol{\theta}))' \Xi^{-1} (\hat{\Psi}^s(\boldsymbol{\theta}) - \hat{\Psi}(\boldsymbol{\theta})) = \\ &= \arg \min_{g, h} ((\hat{\gamma}^s, \hat{\nu}^s) - (\hat{\gamma}, \hat{\nu}))' \mathbf{I}_2 ((\hat{\gamma}^s, \hat{\nu}^s) - (\hat{\gamma}, \hat{\nu}))\end{aligned}$$

where $\hat{\Psi}^s(\boldsymbol{\theta}) = (\hat{\gamma}^s, \hat{\nu}^s) = \arg \max_{\Psi} \ell_{n_s}$, where ℓ_{n_s} is the log pseudo-likelihood (4) based on n_s observations y_1, \dots, y_{n_s} sampled from $\text{gh}(0, 1, g, h)$. Since $\dim(\Psi) = \dim(\boldsymbol{\theta})$, large-sample considerations justify the use of $\Xi = \mathbf{I}$.

It is easy to check that the conditions (a)-(c) of Section 2 are satisfied because the skewed- t log-likelihood (4) is regular; see the proof of asymptotic normality of the unconstrained quasi-maximum likelihood estimator in the Appendix of Garcia *et al.* (2011).

As mentioned in Section 2, Gouriou *et al.* (1993) show how the asymptotic covariance matrix of $\hat{\Psi}$ can be computed. In the present setup, the formula requires the evaluation of the cross-derivatives of ℓ_n with respect to Ψ and $\boldsymbol{\theta}$. Whereas the differentiability of ℓ_n follows from standard Lebesgue dominated convergence arguments, the cross-derivatives cannot be computed in closed form. It is possible to approximate them via simulation, but in this way the computation of the covariance matrix suffers from both the simulation error and the error caused by the use of a limiting formula in a finite-sample setup. Hence, we prefer to use the non-parametric bootstrap to evaluate the standard deviation of the estimators (Garcia *et al.* 2011, p. 330).

3.3. Computational issues

The maximization of (4) can be carried out with standard optimization routines such as the Nelder-Mead algorithm as implemented in the R `optim` function. On the other hand, solving (2) is more difficult.

We have experimented with various algorithms and the most efficient ones have proved to be the `uobyqa` (unconstrained optimization by quadratic approximation) and `newuoa` algorithms implemented in the `minqa` R package. The latter algorithm is supposed to be computationally more efficient when the dimension of the problem is large. In the present case, probably because the dimension is small, the computing times of the two algorithms are almost identical. Unreported simulation experiments suggest that `uobyqa` is slightly more precise, so that we always use `uobyqa` in the following.

4. Simulation experiments

In this section, the finite sample performance of the II estimation procedure is investigated by means of simulations.

The design of the simulation experiment is motivated by two remarks. First, we focus on the typical parameter values estimated when fitting the g-and-h to loss data. Second, the parameter values are such that the skewness and kurtosis of the corresponding distributions cover large ranges. The latter requirement is quite relevant since skewness and kurtosis may vary a lot in different financial datasets and their accurate estimation is crucial. With operational risk data, Dutta and Perry (2006) find $g \in [1.79, 2.3]$ and $k \in [0.1, 0.35]$.

As pointed out by Dutta and Perry (2006), a and b are of minor importance for the computation of risk measures, so that one can just use the standard observations $(y_i - \hat{a}^g)/\hat{b}^g$, $i = 1, \dots, n$ for estimating g and h via indirect inference. We do not use parameter values that give distribution shapes or skewness/kurtosis combinations unlikely to be found when fitting the g-and-h loss datasets.

We first sample observations from the g-and-h distribution in each of the following seven parameter configurations:

- (1) $a = 0$, $b = 1$, $g = 0.2$, $h = 0.1$ (*Skewness*=1.07; *Kurtosis*=8.48);
- (2) $a = 0$, $b = 1$, $g = 0.5$, $h = 0.1$ (*Skewness*=3.35; *Kurtosis*=38.51);
- (3) $a = 0$, $b = 1$, $g = 0.8$, $h = 0.1$ (*Skewness*=8.43; *Kurtosis*=278.18);
- (4) $a = 0$, $b = 1$, $g = 0.2$, $h = 0.05$ (*Skewness*=0.78; *Kurtosis*=5.09);
- (5) $a = 0$, $b = 1$, $g = 0.2$, $h = 0.15$ (*Skewness*=1.59; *Kurtosis*=19.05);
- (6) $a = 0$, $b = 1$, $g = 0.2$, $h = 0.2$ (*Skewness*=2.71; *Kurtosis*=76.14);
- (7) $a = 0$, $b = 1$, $g = 2$, $h = 0.2$ (*Skewness*=376.06; *Kurtosis*=242 000.6).

4.1. Parameter estimation

The II approach developed in Section 3 is used to estimate g and h , whereas the estimators of a and b are the Q -estimators. We also employ the robust method proposed by Dupuis and Field (2004) (see Section 3.1), denoted by R in the following. The experiments are repeated for sample sizes $n \in \{100, 1000\}$, and are based on $B = 100$ replications. Tables 1 to 7 display the bias and the root-mean-squared-error (RMSE) in each setup for both sample sizes.

In terms of RMSE, the II estimators are clearly preferable to the Q estimators. For both g and h , the improvement of II with respect to Q -estimation is generally larger when $n = 1000$. In addition, both when $n = 100$ and when $n = 1000$, the gain of the II estimator is larger for parameter h than for parameter g . Considering that h is related to kurtosis, this implies that tail heaviness is estimated more precisely with II, and should entail a more accurate evaluation of VaR.

As for the R estimators, the tables show that they are always severely biased. Turning to the comparison with the II estimators, we first notice when $n = 1000$, \hat{h}^{II} is always preferable to \hat{h}^R

Table 1.: Bias and RMSE for the Q , R and II estimators in setup (1)

Sample size	Parameter	n_s	Bias			RMSE		
			Q	R	II	Q	R	II
$n = 100$	g	100			0.038			0.256
		1000	-0.001	-0.064	0.021	0.262	0.109	0.188
		5000			0.021			0.210
	h	100			-0.014			0.123
		1000	-0.028	0.123	-0.034	0.184	0.088	0.106
		5000			-0.043			0.111
$n = 1000$	g	100			0.004			0.175
		1000	-0.004	0.062	< 0.001	0.085	0.069	0.082
		5000			-0.008			0.062
	h	100			0.021			0.065
		1000	-0.005	-0.049	-0.002	0.059	0.051	0.024
		5000			0.002			0.019

Table 2.: Bias and RMSE for the Q , R and II estimators in setup (2).

Sample size	Parameter	n_s	Bias			RMSE		
			Q	R	II	Q	R	II
$n = 100$	g	100			-0.024			0.317
		1000	-0.034	-0.173	-0.010	0.294	0.206	0.232
		5000			-0.006			0.230
	h	100			< 0.001			0.145
		1000	-0.014	-0.047	-0.030	0.197	0.073	0.104
		5000			-0.037			0.093
$n = 1000$	g	100			0.007			0.219
		1000	0.010	-0.157	0.003	0.088	0.161	0.089
		5000			-0.008			0.075
	h	100			0.023			0.089
		1000	-0.013	-0.022	0.001	0.070	0.028	0.034
		5000			< 0.001			0.025

Table 3.: Bias and RMSE for the Q , R and II estimators in setup (3).

Sample size	Parameter	n_s	Bias			RMSE		
			Q	R	II	Q	R	II
$n = 100$	g	100			-0.007			0.326
		1000	-0.023	-0.286	-0.005	0.251	0.311	0.250
		5000			0.029			0.240
	h	100			-0.006			0.164
		1000	-0.054	< 0.001	-0.011	0.206	0.081	0.114
		5000			-0.046			0.122
$n = 1000$	g	100			-0.006			0.258
		1000	-0.264	-0.023	-0.001	0.095	0.268	0.112
		5000			-0.005			0.087
	h	100			0.025			0.112
		1000	-0.005	0.038	< 0.01	0.067	0.048	0.040
		5000			-0.005			0.033

both in terms of bias and RMSE. In particular, the bias of \hat{h}^R is, in general, much larger. Similar remarks hold for \hat{g} : \hat{h}^{II} is preferable to \hat{g}^R in most cases in terms of RMSE, and in all cases in terms of bias. When $n = 100$ the outcomes are analogous, although there are a few cases where R is better than II in terms of RMSE (but not of bias). The clear advantage of the II estimator of h with respect to both the Q and R estimators are likely to be quite important for VaR computation, where a precise estimation of the tail heaviness is crucial. Finally, the outcomes for the case $n = 50$, reported in tables A.1 to A.7 of the supplementary material, suggest that II is preferable in terms

Table 4.: Bias and RMSE for the Q , R and II estimators in setup (4).

Sample size	Parameter	n_s	Bias			RMSE		
			Q	R	II	Q	R	II
$n = 100$	g	100			0.006			0.230
		1000	< 0.001	-0.076	0.002	0.248	0.108	0.206
		5000			0.010			0.196
	h	100			-0.029			0.120
		1000	-0.028	-0.049	-0.056	0.177	0.064	0.128
		5000			-0.035			0.093
$n = 1000$	g	100			< 0.001			0.130
		1000	-0.019	-0.071	-0.021	0.094	0.076	0.082
		5000			-0.012			0.080
	h	100			0.015			0.054
		1000	0.002	-0.024	0.003	0.056	0.028	0.028
		5000			-0.004			0.019

Table 5.: Bias and RMSE for the Q , R and II estimators in setup (5).

Sample size	Parameter	n_s	Bias			RMSE		
			Q	R	II	Q	R	II
$n = 100$	g	100			0.003			0.266
		1000	-0.011	-0.065	-0.014	0.265	0.110	0.214
		5000			0.007			0.213
	h	100			-0.016			0.123
		1000	-0.044	-0.109	-0.024	0.192	0.120	0.099
		5000			-0.028			0.079
$n = 1000$	g	100			0.009			0.222
		1000	0.002	-0.050	0.004	0.095	0.059	0.092
		5000			-0.002			0.065
	h	100			0.028			0.078
		1000	0.004	-0.072	0.001	0.060	0.075	0.028
		5000			-0.001			0.025

Table 6.: Bias and RMSE for the Q , R and II estimators in setup (6).

Sample size	Parameter	n_s	Bias			RMSE		
			Q	R	II	Q	R	II
$n = 100$	g	100			0.016			0.314
		1000	-0.003	-0.064	< -0.001	0.265	0.124	0.256
		5000			-0.005			0.244
	h	100			-0.017			0.130
		1000	-0.050	-0.138	-0.035	0.192	0.150	0.106
		5000			-0.035			0.094
$n = 1000$	g	100			0.005			0.248
		1000	0.012	-0.050	0.004	0.098	0.062	0.110
		5000			-0.003			0.082
	h	100			0.026			0.087
		1000	-0.103	-0.003	-0.002	0.064	0.105	0.029
		5000			< 0.001			0.026

of bias, whereas R is mostly better in terms of RMSE.

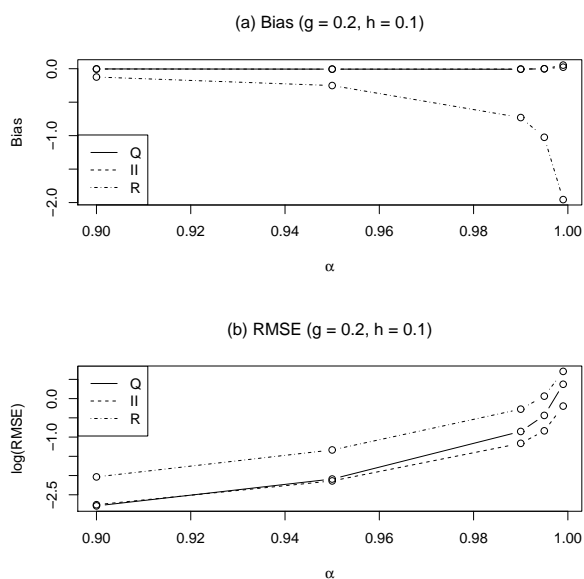
4.2. VaR estimation

To assess the quality of the two procedures in terms of VaR estimates, we continue the simulation experiments carried out above. At the end of the i -th ($i = 1, \dots, 200$) replication of the simulation,

Table 7.: Bias and RMSE for the Q , R and II estimators in setup (7).

Sample size	Parameter	n_s	Bias			RMSE		
			Q	R	II	Q	R	II
$n = 100$	g	100			-0.017			0.591
		1000	-0.053	-0.724	0.025	0.419	0.752	0.449
		5000			-0.015			0.432
	h	100			0.002			0.369
		1000	-0.092	0.198	-0.056	0.378	0.251	0.272
		5000			-0.055			0.260
$n = 1000$	g	100			-0.014			0.493
		1000	-0.007	-0.691	0.003	0.132	0.696	0.193
		5000			-0.007			0.136
	h	100			0.047			0.264
		1000	-0.014	0.275	0.006	0.127	0.283	0.091
		5000			< 0.001			0.071

we use the vector of parameters estimated with each method for computing the VaR_α , obtained in closed form via the quantile function. Figures 1 to 7 show the bias and the RMSE of the VaR_α estimates in the seven setups considered in the experiments, for $\alpha \in \{0.9, 0.95, 0.99, 0.995, 0.999\}$. In all plots, the II estimates are based on $n_s = 5000$ and the sample size is $n = 1000$. The y -axis of the RMSE plots is on a logarithmic scale to improve readability. Results for different sample sizes are reported in figures A.1 to A.14 of the supplementary material.

**Figure 1.:** Bias (panel (a)) and RMSE (panel (b)) of the Q -VaR, II-VaR and R -VaR estimates in setup (1).

The graphs show that the RMSE of the VaR_α based on the indirect inference estimates (II-VaR $_\alpha$) is always smaller than the RMSE of the VaR_α based on the Q estimates (Q -VaR $_\alpha$) when $\alpha > 0.95$, and essentially identical when $\alpha \leq 0.95$.

As for the bias, the outcomes are more mixed, but the simulated distributions of the VaR_α estimators suggest that the distribution of the II-VaR $_\alpha$ is much less skewed than the Q -VaR $_\alpha$.

Finally, the R approach is always the worst one, mainly because of a very large negative bias. Hence, it looks like the estimation bias of this method noted in Section 4.1 has a strong impact on VaR computation.

The plots A.1 to A.14 in the supplementary material, corresponding to the experiments with $n \in \{50, 100\}$ show that II-VaR is more precise for $n = 100$ and sometimes also for $n = 50$. This

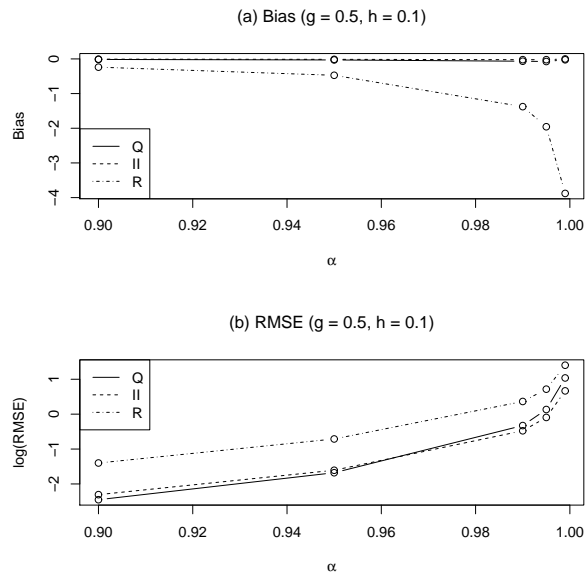


Figure 2.: Bias (panel (a)) and RMSE (panel (b)) of the Q -VaR, II-VaR and R -VaR estimates in setup (2).

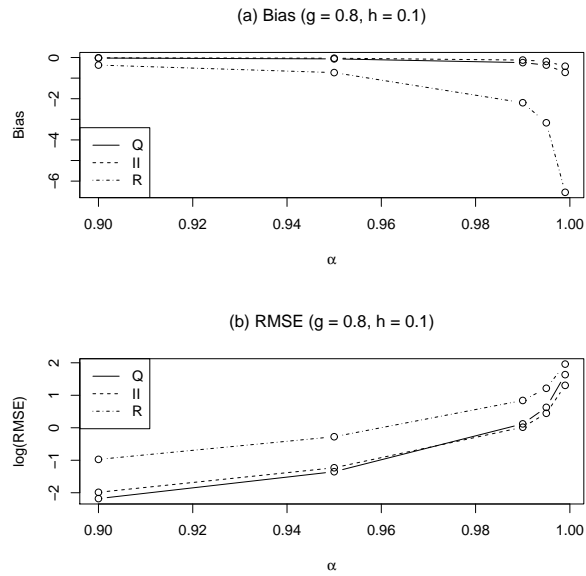


Figure 3.: Bias (panel (a)) and RMSE (panel (b)) of the Q -VaR, II-VaR and R -VaR estimates in setup (3).

outcome suggests that more precise parameters estimates do not automatically imply more precise VaR estimates.

Figure 8 shows the distributions of the three VaR_α in setup (4) for $\alpha = 0.999$ and $n = 1000$ (note that, for comparison purposes, the scale of the x -axis is the same in the three histograms). As can be seen from Figure 4, in this case the bias of $Q\text{-VaR}_\alpha$ is smaller than the bias of $II\text{-VaR}_\alpha$, but the $Q\text{-VaR}_\alpha$ distribution is certainly more dispersed and farther away from the normal. The $D\text{-VaR}$ distribution is by far less dispersed, but has the very undesirable feature that all the values are smaller than the true VaR.

The graphs for the remaining settings and sample sizes (figures A.15 to A.34 in the supplementary material), convey the same message mostly for $n = 100$ and $n = 1000$, but to a lesser extent also for $n = 50$.

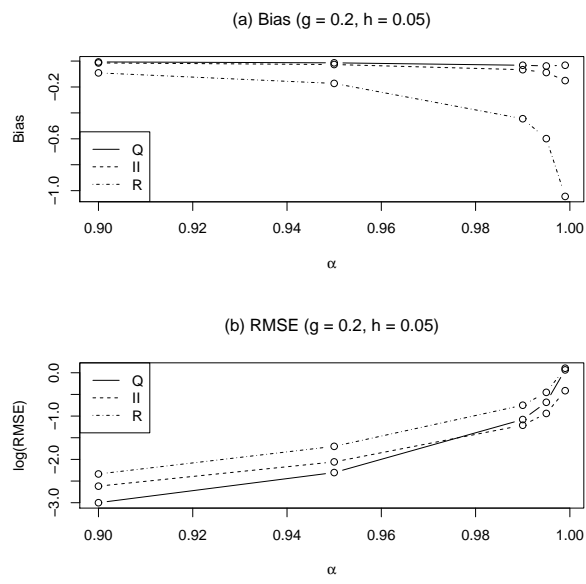


Figure 4.: Bias (panel (a)) and RMSE (panel (b)) of the Q -VaR, II-VaR and R -VaR estimates in setup (4).

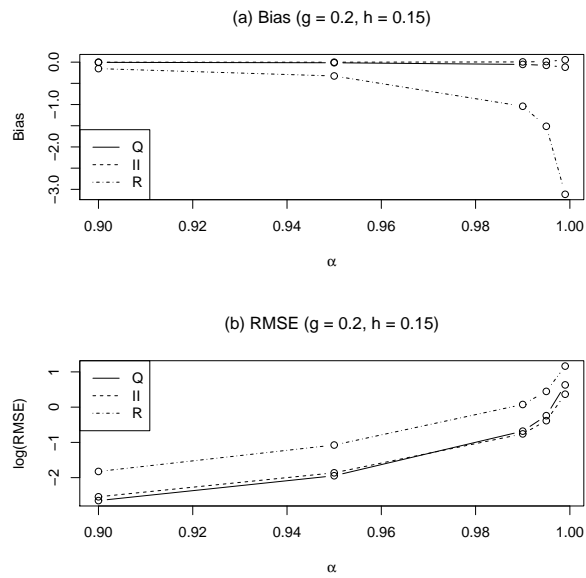


Figure 5.: Bias (panel (a)) and RMSE (panel (b)) of the Q -VaR, II-VaR and R -VaR estimates in setup (5).

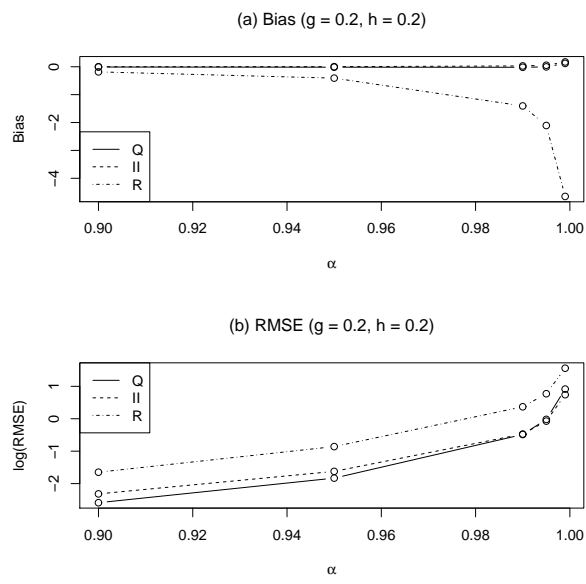


Figure 6.: Bias (panel (a)) and RMSE (panel (b)) of the Q -VaR, II-VaR and R -VaR estimates in setup (6).

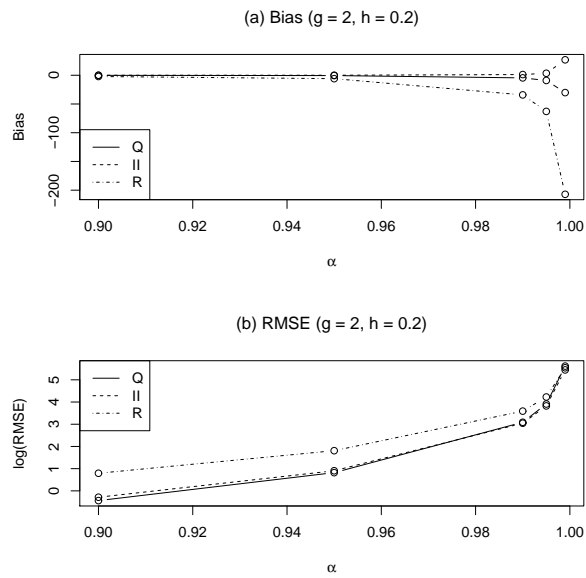


Figure 7.: Bias (panel (a)) and RMSE (panel (b)) of the Q -VaR, II-VaR and R -VaR estimates in setup (7).

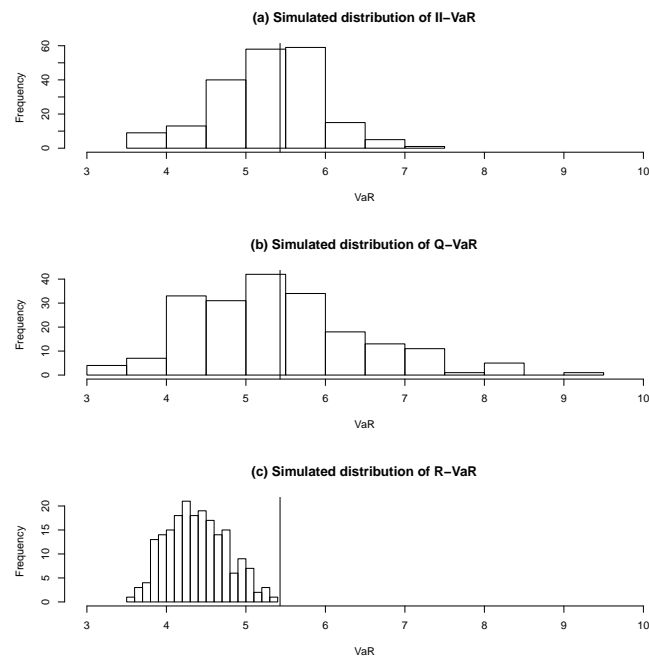


Figure 8.: Simulated distributions of II-VaR_α (panel (a)), of Q-VaR_α (panel (b)) and of D-VaR_α (panel (c)) in setup (4) for $\alpha = 0.999$ and $n = 1000$. The vertical line is the true VaR.

5. Empirical Analysis

In this section we illustrate the proposed approach by studying a dataset of operational losses recorded at the Italian bank UniCredit between January and June 2014. Operational losses are defined by the Basel Committee for Banking Supervision as “direct or indirect losses resulting from inadequate or failed internal processes, people and systems or from external events” (BCBS 2004). According to the Basel II Accord, in the advanced measurement approach banks are asked to compute VaR_α of their operational loss distribution, with α close to one (e.g. 0.999)¹. Consequently, it is particularly important for them to use a distribution that models accurately the probabilities far in the tail.

Here, we focus on losses belonging to two event types, reflecting the physical processes behind the loss generation: *execution, delivery, and process management* (EDPM, 417 losses) and *clients, products, and business practices* (CPBP, 583 losses). The data, which are scaled by an unknown factor for anonymity reasons, have been provided by the UniCredit operational risk department and are displayed in Figure 9. Losses are adjusted for inflation by means of the monthly Italian consumer price index (base 100: December 2013), and the scaled collection threshold is 2000€. Around one-fourth of the losses are related to financial instruments and derivative products. More information related to these data can be found in Hambuckers *et al.* (2018).

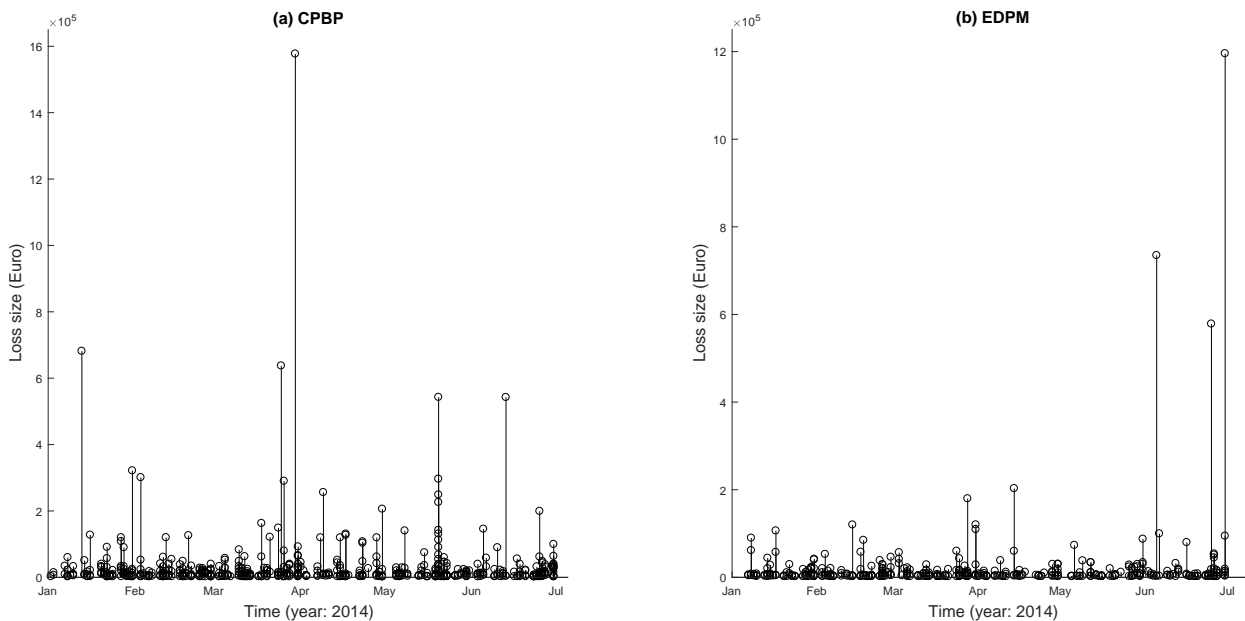


Figure 9.: UniCredit operational losses for the first semester of 2014, for the CPBP (panel (a)) and EDPM (panel (b)) event type.

The analysis compares the three VaR_α computed by means of:

- the g-and-h distribution with parameters estimated by the Q method;
- the g-and-h distribution with parameters estimated by the II method;
- the Peaks-over-Threshold (POT) method based on the GPD approximation (Davison and Smith 1990).

¹In December 2017, the Basel III Accord prescribed the use of the standardized measurement approach, thus ruling out the calculation of operational risk regulatory capital using internal models. Under Basel III, operational risk capital will be determined via a predetermined formula. As pointed out by many academics and practitioners, the new approach implies that operational risk capital is no longer risk sensitive and that the foundations of operational risk management are seriously weakened; see, for example, Cruz (2018). We support this standpoint and hope that the industry will continue to use internal models.

The purpose is twofold: checking whether the II method gives any advantage over the Q method, and comparing it to the EVT-based approach often used for modeling the tail of loss distributions. We do not use the R approach because we know from the outcomes in Section 4 that its performance is not so good for sample sizes such as those of the datasets considered here.

We carry out the analysis for $\alpha \in \{0.9, 0.95, 0.99, 0.995, 0.999\}$. The POT method uses a threshold equal to the 90% quantile and the standard errors and confidence intervals of the VaR_α are computed via non-parametric bootstrap in all cases. The results are displayed in tables 8 and 9, where, for comparison purposes, we also show the empirical quantiles (given the rather small sample size, the 0.995 and 0.999 quantiles are omitted because they are likely to be poor estimates of the true quantiles).

Table 8.: The VaR_α for the 2014 EDPM Unicredit losses computed by means of the g-and-h estimated with the II and Q estimators and by means of the POT method. In all cases, we use the standardize losses $(y - \hat{a}^Q)/\hat{b}^Q$. *Emp.* is the empirical quantile of the losses. Standard errors are in parentheses and 95% confidence intervals are in brackets. The sample size is $n = 417$.

	$\alpha = 0.9$	$\alpha = 0.95$	$\alpha = 0.99$	$\alpha = 0.995$	$\alpha = 0.999$
II-VaR	5.348 (0.745) [4.08,6.96]	10.621 (2.256) [7.25,15.77]	34.363 (14.841) [17.54,74.38]	51.407 (28.517) [22.99,134.35]	113.457 (106.865) [37.56,428.17]
Q-VaR	5.923 (1.018) [3.97,7.68]	12.173 (3.132) [6.97,18.31]	41.990 (21.411) [15.21,91.60]	64.270 (41.919) [18.63,167.44]	148.436 (165.851) [26.27,639.70]
POT-VaR	5.034 (0.716) [3.79,7.12]	10.005 (1.828) [6.84,14.56]	40.205 (14.726) [20.63,86.96]	70.762 (39.912) [28.09,196.07]	256.412 (809.913) [47.53,1943.44]
Emp.	4.995	10.272	32.584	-	-

Table 9.: The VaR_α for the 2014 CPBP Unicredit losses computed by means of the g-and-h estimated with the II and Q estimators and by means of the POT method. In all cases, we use the standardize losses $(y - \hat{a}^Q)/\hat{b}^Q$. *Emp.* is the empirical quantile of the losses. Standard errors are in parentheses and 95% confidence intervals are in brackets. The sample size is $n = 583$.

	$\alpha = 0.9$	$\alpha = 0.95$	$\alpha = 0.99$	$\alpha = 0.995$	$\alpha = 0.999$
II-VaR	4.653 (0.449) [3.79,5.58]	9.069 (1.343) [6.72,12.08]	29.348 (8.752) [17.41,52.34]	44.482 (16.855) [23.68,90.10]	103.293 (64.137) [42.71,275.67]
Q-VaR	4.612 (0.456) [3.84,5.65]	8.976 (1.403) [6.89,12.53]	29.019 (9.878) [18.08,56.95]	43.996 (19.724) [24.60,101.09]	102.349 (82.804) [44.53,339.16]
POT-VaR	4.352 (0.578) [3.10,5.11]	9.016 (1.464) [6.68,12.23]	30.970 (6.903) [19.77,47.80]	49.020 (15.887) [27.44,93.85]	133.994 (166.562) [50.75,445.96]
Emp.	4.287	10.251	27.433	-	-

As can be seen from Table 8, for EDPM the point estimates of the II- and Q-VaR $_\alpha$ are similar, with II-VaR $_\alpha$ closer to the empirical quantiles than Q-VaR $_\alpha$. As of the standard error and the confidence interval, they are significantly smaller for II-VaR $_\alpha$. The POT-VaR $_\alpha$ is similar to the two g-and-h VaR $_\alpha$ for $\alpha < 0.99$ and becomes larger for $\alpha \geq 0.99$. However, the standard error and the confidence interval get rapidly larger when $\alpha \geq 0.99$, suggesting in particular that the estimator of VaR $_{0.999}$ is quite unreliable. One can see here the consequence of only using a fraction of the data.

Lastly, it appears that the II-VaR provides more reasonable estimates than the POT-VaR. This is appealing for practitioners, since the POT often leads to unrealistic estimates that cannot be used for regulatory purposes; see, e.g., Chavez-Demoulin and Embrechts (2010).

Turning now to the CPBP outcomes in Table 9, the estimates obtained with the II and Q methods are similar to each other, but again the II-VaR standard error and confidence interval are smaller.

Analogously to the preceding case, the POT-VaR gets larger for $\alpha \geq 0.99$, and both the standard error and the confidence interval suggest that the estimate is less reliable than the II and Q-VaR.

6. Conclusion

In this paper we propose a novel estimation method for the g-and-h distribution, using an indirect inference approach based on the skewed- t auxiliary model. We show that the skewed- t has desirable properties in this setup, and investigate the finite-sample performance of the method via simulation. According to the outcomes, our indirect inference approach outperforms the classical quantile-based estimation method, in terms of both parameter and VaR estimation accuracy.

When compared to the robust method of Dupuis and Field (2004), our approach is preferable, in terms of RMSE of parameter estimators, only when $n \in \{100, 1000\}$. However, when it comes to VaR estimation, II is better also for $n = 50$. This is probably due to the large bias of the R estimators and confirms that the relationship between parameter estimates and VaR estimates is highly non-linear, so that it is difficult to predict the impact of parameter estimation error on estimated VaR. It is therefore crucial to study not only the distribution of the parameters estimators, but also of the estimated VaR. Finally, the empirical application confirms that the VaR computed via II is more stable than both the VaR based on the Q-estimators and the VaR based on the POT method.

A related approach that may be worth investigating is the method of simulated quantiles proposed by Dominicy and Veredas (2013). Given the explicit availability of the quantile function of the g-and-h distribution, this is likely to be well suited in this setup. We leave to future work the analysis of this technique.

Appendix A: Proof of Proposition 1

By definition, the pseudo log-likelihood function is given by $\ell(\nu, \gamma, \lambda, \omega | a, b, g, h) = E(\ell_1(Y; \nu, \gamma, \lambda, \omega))$, where $Y \sim \text{gh}(a, b, g, h)$ and the pseudo-true values are the maximizers of $\ell(\nu, \gamma, \lambda, \omega | a, b, g, h)$. We note the following.

Let $Y \sim \text{gh}(a, b, g, h)$. For $k_1 \in \mathbb{R}$, $Y + k_1 \sim \text{gh}(a + k_1, b, g, h)$, and $\ell_1(y + k_1; \nu, \gamma, \lambda, \omega + k_1) = \ell_1(y; \nu, \gamma, \lambda, \omega)$. Similarly, for $k_2 \in \mathbb{R}^+$, $k_2 Y \sim \text{gh}(k_2 a, k_2 b, g, h)$, and $\ell_1(k_2 y; \nu, \gamma, k_2 \lambda, k_2 \omega) = \ell_1(y; \nu, \gamma, \lambda, \omega)$. As a result, $\ell(\nu, \gamma, \lambda, \omega + k_1 | a + k_1, b, g, h) = \ell(\nu, \gamma, \lambda, \omega | a, b, g, h)$ and $\ell(\nu, \gamma, k_2 \lambda, k_2 \omega | a k_2, b k_2, g, h) = \ell(\nu, \gamma, \lambda, \omega | a, b, g, h)$. Using these results with $k_1 = -a$ and $k_2 = 1/b$ we obtain

$$\ell(\nu, \gamma, \lambda, \omega | a, b, g, h) = \ell(\nu, \gamma, \lambda, \omega - a | 0, b, g, h);$$

$$\ell(\nu, \gamma, \lambda, \omega | a, b, g, h) = \ell\left(\nu, \gamma, \frac{\lambda}{b}, \frac{\omega}{b} \left| \frac{a}{b}, 1, g, h\right.\right).$$

Finally, combining the last two results,

$$\ell(\nu, \gamma, \lambda, \omega | a, b, g, h) = \ell\left(\nu, \gamma, \frac{\lambda}{b}, \frac{\omega - a}{b} \left| 0, 1, g, h\right.\right). \quad (\text{A1})$$

From (A1) it is easy to deduce that

$$\begin{aligned}\omega(a, b, g, h) &= a + \omega(0, b, g, h); \\ \lambda(a, b, g, h) &= b\lambda\left(\frac{a}{b}, 1, g, h\right); \\ \gamma(a, b, g, h) &= \gamma(0, 1, g, h); \\ \nu(a, b, g, h) &= \nu(0, 1, g, h).\end{aligned}$$

From (1) it can also be shown that $Y \sim \text{gh}(a, b, g, h) \implies (-Y) \sim \text{gh}(a, b, -g, h)$ and $\ell_1(-y; \nu, 1/\gamma, \lambda, -\omega) = \ell_1(y; \nu, \gamma, \lambda, \omega)$, so that

$$\begin{aligned}\gamma(0, 1, -g, h) &= \frac{1}{\gamma(0, 1, g, h)}; \\ \nu(0, 1, -g, h) &= \nu(0, 1, g, h).\end{aligned}$$

Acknowledgements

We thank Fabio Piacenza (UniCredit SpA) for providing us with the data and two anonymous reviewers for valuable comments that considerably improved an earlier version of this paper. J. Hambuckers acknowledges the financial support of the Research Training Group 1644, Scaling Problems in Statistics, funded by the German Research Foundation (DFG).

References

- BCBS, Basel II: international convergence of capital measurement and capital standards. A revised framework. Technical report, Bank of International Settlements, Basel, Switzerland, 2004.
- Bee, M. and Trapin, L., A simple approach to the estimation of Tukey's gh distribution. *Journal of Statistical Computation and Simulation*, 2016, **86**, 3287–3302.
- Chavez-Demoulin, V. and Embrechts, P., Revisiting the edge, ten years on. *Communications in Statistics - Theory and Methods*, 2010, **39**, 1674–1688.
- Cruz, M.G., Editor's letter. *Journal of Operational Risk*, 2018, **13**.
- Cruz, M., Peters, G. and Shevchenko, P., *Fundamental Aspects of Operational Risk and Insurance Analytics: A Handbook of Operational Risk*, 2015, Wiley.
- Davison, A.C. and Smith, R.L., Models for exceedances over high thresholds. *Journal of the Royal Statistical Society: Series B*, 1990, **52**, 393–442.
- Degen, M., Embrechts, P. and Lambrigger, D.D., The quantitative modeling of operational risk: between g-and-h and EVT. *Astin Bulletin*, 2007, **37**, 265–291.
- Dominicy, Y. and Veredas, D., The method of simulated quantiles. *Journal of Econometrics*, 2013, **172**, 235–247.
- Dupuis, D. and Field, C., Large wind speeds: modeling and outlier detection. *Journal of agricultural, biological, and environmental statistics*, 2004, **9**, 105–121.
- Dutta, K.K. and Babbel, D.F., On measuring skewness and kurtosis in short rate distributions: The case of the US dollar London inter bank offer rates. *Wharton School Center for Financial Institutions, University of Pennsylvania*, 2002, **02-25**.
- Dutta, K.K. and Perry, J., 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, 2006.
- Fernández, C. and Steel, M., On Bayesian Modeling of Fat Tails and Skewness. *Journal of the American Statistical Association*, 1998, **93**, 359–371.

- Gallant, A. and Tauchen, G., Which moments to match?. *Econometric Theory*, 1996, **12**, 657–681.
- Galloppo, G. and Rogora, A., What has worked in operational risk?. *Global Journal of Business Research*, 2011, **5**, 1–17.
- Garcia, R., Renault, E. and Veredas, D., Estimation of stable distributions by indirect inference. *Journal of Econometrics*, 2011, **161**, 325 – 337.
- Gourieroux, C., Monfort, A. and Renault, E., Indirect inference. *Journal of Applied Econometrics, Supplement: Special Issue on Econometric Inference Using Simulation Techniques*, 1993, **8**, S85–S118.
- Hambuckers, J., Groll, A. and Kneib, T., Understanding the economic determinants of the severity of operational losses: a regularized Generalized Pareto regression approach. *Journal of Applied Econometrics (forthcoming)*, 2018.
- Hoaglin, D.C., Summarizing Shape Numerically: The g-and-h Distributions. *Exploring data tables, trends, and shapes*, 1985, pp. 461–513.
- Huber, P.J., *Robust Statistics*, 1981, New York, Wiley.
- Hübner, G., Peters, J.P. and Plunus, S., Measuring operational risk in financial institutions: Contribution of credit risk modeling,. Technical report, SSRN Working Paper Series, 2005.
- Jiang, W. and Turnbull, B., The Indirect Method: Inference Based on Intermediate Statistics - A Synthesis and Examples. *Statistical Science*, 2004, **19**, 239–263.
- Jiménez, J.A. and Arunachalam, V., Using Tukey’s g and h family of distributions to calculate Value-at-Risk and conditional Value-at-Risk. *Journal of Risk*, 2011, **13**, 95–116.
- Jobst, A.A., Consistent Quantitative Operational Risk Measurement and Regulation: Challenges of Model Specification, Data Collection, and Loss Reporting. Technical report, IMF Working Paper, 2007.
- Peters, G.W., Chen, W.Y. and Gerlach, R.H., Estimating Quantile Families of Loss Distributions for Non-Life Insurance Modelling via L-Moments. *Risks*, 2016, **4**, 14.
- Peters, G.W. and Sisson, S., Bayesian inference, Monte Carlo sampling and operational risk. *Journal of Operational Risk*, 2006, **1**, 27–50.
- Tukey, J.W., Modern techniques in data analysis. In *Proceedings of the NSF-Sponsored Regional Research Conference*, 1977.