Modelling uncertainty in VaR

The Evaluation of Estimators with Heavy-tailed Distributions

Luis Fernando Corrêa da Costa

University of Minho

lfcorreacosta@gmail.com

Abstract

Potential failures of risk measures are generally associated with the uncertainty of not knowing the true data generating process (DGP) of financial series, observed either in the model specification error or parameters estimation error (Boucher et al., 2014). Neglecting these factors can lead to erroneous interpretations and imprecise risk coverage. In order to deal with the uncertainty of model specification error, this paper seeks to evaluate the performance of a set of heavy-tailed GARCH estimators to predict the risk of financial returns through the Value-at-risk (VaR). In a simulated study and employed to equity indexes returns, two versions of the ranked-based estimator are more efficient to forecast the VaR than the traditional Gaussian, Student's-t and GED procedures. Moreover, our results suggest that a higher accuracy can be achieved by performing heavy-tailed estimators since these procedures contribute to reducing the number of violations of returns from VaR predictions.

Keywords: estimators; forecast; GARCH(1,1); heavy-tails; uncertainty; Value-at-risk; volatility.

1 Introduction

The Basel accords have recommended the Value at Risk (VaR) as the main building block that financial institutions must adopt to monitor their capital exposure and control their position in risky assets. Nevertheless, potential failures of VaR forecasts can be mainly attributed to the uncertainty of not knowing the actual data generating process (DGP) of financial returns. Generally, these failures derive from different sources, such as the model specification error, which refers to making inappropriate assumptions about the (DGP) for the random variable, and the parameter estimation error related to the uncertainty in the parameter values of the chosen model (Boucher et al., 2014). Since several approaches are based on the estimation of the conditional volatility through GARCH models to obtain future VaR densities (see Nieto &

Ruiz (2016)), the choice of the most appropriate estimator contributes for reducing the model uncertainty improving risk predictions.

The GARCH (1,1) model introduced by Bollerslev (1986) has still been the primary benchmark for many extensions of autoregressive conditional heteroscedastic specifications and other competing approaches (Hansen & Lunde, 2005). Initial studies are primarily based on the estimation of GARCH parameters through the maximisation of a Gaussian quasi-likelihood function (GQML). However, many authors have documented that GQML estimator is consistent and asymptotically normal only under strong assumptions (Bollerslev & Wooldridge, 1992) such as the existence of a uniformly bounded kurtosis (Lee & Hansen, 1994; Lumsdaine, 1996). Therefore, given that the normality of errors distributions of financial time series is hardly ensured, estimating parameters via GQML may generate misspecified models. In this regard, some research arose to overcome the inefficiency of GQML. Bollerslev et al. (1987) first extended the GARCH(p,q) specification to the case of Student's-t distribution for assets returns. Engle & Gonzalez-Rivera (1991) developed a semi-parametric estimator for GARCH models that encompasses others conditional probability densities of innovations outperforming the GQML estimator. Lee & Lee (2009) suggested a normal-mixed quasi-maximum likelihood estimator that shows consistency compared to the standard GQML for skewed and heavy-tailed distribution when the estimator belongs to the same errors density family of the process.

Another literature stream has been concerned with the development of non-Gaussian estimators (NGQML) to improve the consistency of conditional heteroscedastic models for heavy-tailed distribution processes. Francq et al. (2011) proposed a two-stage NQMLE that jointly performed with a scale parameter produces consistent estimates under specific errors distributions assumptions. Andrews (2012) suggested a rank-based estimator obtained via a scale transformation of standard GARCH specifications to estimate the parameters of the conditional heteroscedastic model. They show that the rank-based residual technique outperforms the Gaussian quasi-maximum likelihood and shows similar asymptotic efficiency as the maximum likelihood estimator. Fan et al. (2014) introduced a three-step quasi-likelihood method that identifies an unknown scale parameter that reduces the bias of GARCH estimates. Francq & Zakoian (2014) point out that the procedure developed by Fan et al. (2014) are robust to estimate a GARCH(1,1) model, especially for heavy-tail errors densities. Furthermore, NGQML is efficient in most of the cases due to its adaption property (Ling & Zhu, 2014), which can be adequate to estimate models for assets returns and improve risk measures (Fiorentini & Sentana, 2014). On the other hand, comparatively to standard estimators, NGQMLE is computationally expensive since it involves three optimisation steps (Andrews, 2014; Francq & Zakoian, 2014; Fiorentini & Sentana, 2014).

More recently, Preminger & Storti (2017) introduced an alternative log-transform-based least-squares approach to estimate a GARCH(1,1) model for heavy-tailed errors distribution based on the identification of a tuning constant to scale the conditional volatility. They demonstrate that the log-transform-based least-squares is consistent and asymptotically normal for excess kurtosis series, outperforming both GQML and other non-Gaussian estimators under leptokurtic distribution conditions.

The estimators proposed by Andrews (2012), Fan et al. (2014) and Preminger & Storti (2017) are developed to deal with errors without finite fourth moments. Under this condition, Fan et al. (2014) highlight that the rank-based estimator for the heteroscedastic parameter shows a similar asymptotic performance when compared to the NGQMLE procedure, with the former being slightly more accurate. Andrews (2012) observe that the rank-based estimator produces more efficient parameters than the NQMLE, and suggests the use of the residuals of the rank estimator to identify the noise distributions before fitting a GARCH model through NGQML. Concerning the estimation method, both NGQMLE (Fan et al., 2014) and the log-transform-based least-squares approach (Preminger & Storti, 2017) are multi-steps procedures and require the previous identification of an unknown parameter that approximates the quasi-likelihood function to the distribution of the noise process. These two approaches use standardised residuals obtained from an initial Gaussian quasi-likelihood estimation to obtain the scale parameter η and the tuning constant $\hat{c}_n$ respectively.

This article aims to incorporate the uncertainty regarding the model specification error into VaR. We consider estimators for GARCH(1,1) that accommodate excess kurtosis distributions, generally observed in financial processes, here represented by the models described by Andrews (2012), Fan et al. (2014) and Preminger & Storti (2017). Namely heavy-tailed estimators, we also include in this group of models an estimator that combines the ranked-based technique of Andrews (2012) with the non-Gaussian method of Fan et al. (2014). The performance of these approaches are compared with the traditional Gaussian, Student's-t and Generalised Errors Distribution estimators to predict the risk for financial returns. Besides this introduction and conclusions, the remaining of this article is organised as follows; the econometric methodology of the heavy-tailed estimators is presented in section 2; comparisons between estimators' performance to obtain GARCH parameters are carried out in section 3, and VaR analyses are undertaken in section 4. VaR forecasts are evaluated in two contexts, by employing backtesting measures over simulated financial series, as well as by comparing the predictive ability of these estimators on equity indexes returns by applying the Model Confidence Set procedure developed by Hansen et al. (2011).

2 Estimators with heavy-tailed distributions

The estimators described in this section are performed to obtain the conditional volatility of returns of GARCH models as introduced by Bollerslev (1986). We consider a GARCH(1,1) specification given by:

$$\begin{aligned} \varepsilon_t &= \sigma_t z_t \\ \sigma_t^2 &= \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \end{aligned} \tag{1}$$

Where ε_t is a stochastic white noise process, with $z_t \sim N(0,1)$ and conditional variance equals to σ_t^2. Parameters in equation 1 are such that $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$ and $(\alpha + \beta) < 1$, whose persistence of the conditional variance is given by the sum of α and β.

2.1 Ranked-based estimator (RANK)

We start by describing the rank-based estimator proposed by Andrews (2012). The primary assumption of the model is that the errors distribution does not necessarily require finite variance to be estimated. Therefore, given a general GARCH(p,q) set of parameters equal to $\theta := (\alpha_1/\alpha_0, ..\alpha_q/\alpha_0, ..\beta_1, ..\beta_q)'$, the residuals $\epsilon_t(\theta)$ of the stochastic process is expressed as:

$$\epsilon_t(\theta) = \ln\left(\varepsilon_t^2\right) - \ln\left(\tilde{\sigma}_t^2(\theta)\right) \tag{2}$$

for t = 1,2...n, where the conditional variance is:

$$\tilde{\sigma}_t^2(\theta) = \begin{cases} 1 & , for \quad t = \min(1, p-q+1), ...q \\ 1 + \sum_{i=1}^{p} \theta_i \varepsilon_{t-i}^2 + \sum_{i=1}^{q} \theta_{p+j} \tilde{\sigma}_{t-j}^2(\theta), for & t = p+1, ..., n, \end{cases} \tag{3}$$

Given a non-constant and non-decreasing weight function $\lambda \in \Re$ from (0, 1), the rank function D_n is defined as:

$$D_n(\theta) = \sum_{t=p+1}^{n} \lambda\left(\frac{R_t(\theta)}{n-p+1}\right)\left[\epsilon_t(\theta) - \overline{\epsilon_t(\theta)}\right] \tag{4}$$

Where $\left[\epsilon_t(\theta) - \overline{\epsilon_t(\theta)}\right]$ is the mean-corrected residuals, and $R_t(\theta)$ is the ascending-ordered vector of residuals $\{\epsilon_t(\theta)\}_{t=p+1}^{n}$, such that the rank function can be written as:

$$D_n(\theta) = [\lambda((t-p/n-p+1))]\left[\epsilon_t(\theta) - \overline{\epsilon_t(\theta)}\right] \tag{5}$$

The vector of parameters of the GARCH model(θ) is obtained by using the ranked residuals in $\{\epsilon_t(\theta)\}_{t=p+1}^{n}$ to minimise $D_n(\theta)$. The optimisation requires the choice of the appropriate weight function λ. By analysing its statistical properties, Andrews (2012) defines λ as shown in equation 6, with the distribution of the standardised residuals $(z = \epsilon_t/\tilde{\sigma}_t)$ symmetric around zero.

$$\lambda_f(x) \propto -f'\left(F^{-1}(x)\right)/f\left(F^{-1}(x)\right) \tag{6}$$

Giving f and F the density and distribution functions for $log(z^2)$ respectively, Andrews (2012) compares the asymptotic relative efficiency of the rank estimations by scaling the error distribution process through Laplace QML, Logistic, Normal and Student-t distributions. The Student's-t with 7 degrees of freedom shows higher performance among all distributions types and errors with different heaviness levels. Andrews (2012) finds that the RANK estimator should be performed with a similar weight function when the error distribution is unknown. Following Andrews (2012), we conduct the estimations in this study considering the weight function in equation 7:

$$\lambda_{t7}(x) = -\left[7\left\{F_{t7}^{-1}\left((x+1)/2\right)\right\}^2 - 5\right] / \left[\left\{F_{t7}^{-1}\left((x+1)/2\right)\right\}^2 + 5\right] \tag{7}$$

λ_{t7} follows a Student-t distribution with 7 degrees of freedom, where F_{t7} is the distribution function of standardised residuals under this specification.

2.2 Non-Gaussian estimator (NGQML)

The three-step quasi-maximum likelihood estimator with non-Gaussian likelihood functions (Fan et al., 2014) assumes that a GARCH process is specified as follows:

$$\begin{aligned} \epsilon_t =& \xi\sigma_t z_t \\ \sigma_t^2 =& \omega_0 + \sum_{i=1}^{q}\alpha_i\epsilon_{t-i}^2 + \sum_{i=1}^{q}\beta_j\sigma_{t-j}^2 \end{aligned} \tag{8}$$

Given a vector of parameters for $\theta = \{\xi, \gamma\}$ in equation 8, ξ is the scale coefficient and γ contains the GARCH parameters ω, α and β . The model is based on the existence of an unknown parameter η that minimises the difference between the actual innovation density $g(\cdot)$, and the quasi-likelihood function f belonging to a family of functions $\frac{1}{\eta}f\left(\frac{\cdot}{\eta}\right)$, in such a way that η is estimated from equation 9:

$$\eta_f = argmax_{\eta>0}\left\{-log\left(\eta\right) + log(f)\left(\frac{z}{\eta}\right)\right\} \tag{9}$$

The first step consists of using the standardised residuals obtained from a Gaussian quasi-likelihood estimator (GQML) as substitutes for the true innovation sample, which is estimated through the following likelihood function:

$$\widetilde{\theta}_T = argmax_{\theta}\frac{1}{T}\sum_{t=1}^{T}\left(-\log\left(\xi\sigma\right) - \frac{\epsilon_t^2}{2\xi^2\sigma_t^2}\right) \tag{10}$$

The standardised residuals from GQMLE (equation 10) is given by $\widetilde{z} = \epsilon_t / \left(\tilde{\xi}\sigma_t \left(\widetilde{\gamma}_t \right) \right)$ and it is used to estimate the scale parameter η in the second step:

$$\hat{\eta}_f = argmax_{\eta>0} E \left\{ -\log\left(\eta\right) + \log\left(f\right) \left(\frac{\widetilde{z}}{\eta} \right) \right\} \tag{11}$$

The maximisation of the non-Gaussian quasi-maximum likelihood is performed by plugging $\hat{\eta}_f$ from equation 11 into equation 10. The NGQML estimator $\hat{\theta}_T$ is obtained in the last step:

$$\hat{\theta}_T = argmax_{\theta} \frac{1}{T} \sum_{t=1}^{T} \left(-\log\left(\hat{\eta}_f \xi \sigma\right) + \log \frac{\epsilon_t}{\hat{\eta}_f \xi \sigma_t} \right) \tag{12}$$

The scale parameter η is generic characteristic of non-Gaussian likelihoods of true innovations (Fan et al., 2014). η encompasses different families of likelihood functions and varies according to the distribution thickness, assuming values lower than 1 for heavy-tailed errors distributions and higher than 1, otherwise. Therefore, η works as a mechanism that adjusts the innovations errors to the quasi-likelihood function.

2.3 Log-transform-based least-squares estimator(LLSE)

Preminger & Storti (2017) introduced a log-transform-based least-squares approach to estimate a GARCH(1,1) model, whose scale of the estimated volatility is dependent on an unknown tuning constant. From a standard GARCH process, the tuning constant is set as $c_0 = E\left[\ln \epsilon_t^2\right]$, such that the GARCH parametrisation is given by:

$$\begin{aligned} \epsilon_t &= \sigma_t z_t \\ \ln \epsilon_t^2 - c_0 &= \ln \sigma_t^2 + \zeta_t \end{aligned} \tag{13}$$

Where $(\zeta_t = \ln \eta_t^2 - c_0)$ and $(\ln \epsilon_t^2 - c_0)$ follow a zero mean i.i.d. random variable. The objective function in 14 is estimated via non-linear least-squares.

$$\tilde{Q}_n\left(\theta\right) = \frac{1}{2n} \sum_{t=1}^{n} \left(\ln \epsilon_t^2 - c_0 - \ln \tilde{\sigma}_t^2\left(\theta\right) \right)^2 \tag{14}$$

With $\theta = (\omega, \alpha, \beta)$, the vector of GARCH(1,1) parameters:

The standard GARCH(1,1) model described 1 is estimated through the GQML estimator in the first step where the residuals are computed as follows:

$$ln\left(z^2\right) = ln\left(\frac{\epsilon^2}{\sqrt{\sigma^2}} \right) \tag{15}$$

The tuning constant $\hat{c}_n$ is calculated using the transformed standardised residuals $\hat{c}_n = n^{-1}\sum_{t=1}^{n} ln\left(z_t^2\right)$. By plugin $\hat{c}_n$ in equation 14, the least-squared estimator $\hat{\theta}_n$ is obtained from the minimisation of :

$$\begin{aligned}\tilde{Q}_n\left(\theta,\hat{c}_n\right) &= \frac{1}{2n}\sum_{t=1}^{n}\left(\ln\epsilon_t^2 - \hat{c}_n - \ln\tilde{\sigma}_t^2\left(\theta\right)\right)^2 \\ \hat{\theta}_n &= argmin\tilde{Q}_n\left(\theta,\hat{c}_n\right)\end{aligned} \tag{16}$$

It is worth noting that objective function in 16 is indirectly identified by the location parameter $\hat{c}_n$, which contains information of the density of z^2 estimated via GQML. In this case, as the scale parameter η of NGQM, $\hat{c}_n$ is the term that allows for the adjustment of the estimator to the innovations errors. In both cases, the identification of η and $\hat{c}_n$ provides for adapting the estimator to the distribution form, improving the efficiency of the model.

2.4 Non-Gaussian estimator with rank selection for errors innovations (RANK$_2$)

We also perform a method that combines the rank estimator suggested by Andrews (2012) and the NGQML proposed by Fan et al. (2014). Estimations are carried out by substituting standardised residuals obtained via GQML for standardised residuals retrieved form a rank estimation. This technique is used to identify the noise distributions of the process, which can help to improve the model accuracy (see a discussion in (Andrews, 2014)).

The first step consists of optimising the rank function $D_n\left(\theta\right)$ defined in equations 4 and 5, whose weight function is given by a Student's-t distribution with 7 degrees of freedom as described in equation 7. The resulting standardised residuals of the rank estimation $\hat{z}_t^R$ are given by:

$$\hat{z}_t^R = \frac{\epsilon_t^2}{\sqrt{\hat{\sigma}_t^2\left(\hat{\kappa}^R\right)}} \tag{17}$$

Where $\sqrt{\hat{\sigma}_t^2\left(\hat{\kappa}^R\right)}$ is the conditional standard deviation estimated over the mean-corrected residuals $\left[\epsilon_t\left(\theta\right) - \overline{\epsilon_t\left(\theta\right)}\right]$, weighted by the rank residuals$\left\{R_t\left(\theta\right)\right\}_{t=p+1}$, here represented by $\hat{\kappa}^R$.

The new scale parameter η^R is estimated through the optimisation of the objective function in equation 18, with $\hat{z}_t^R$ the standardised residuals of the rank estimation .

$$\hat{\eta}_f^R = argmax_{\eta>0}E\left\{-\log\left(\eta\right) + \log\left(f\right)\left(\frac{\hat{z}_t^R}{\eta}\right)\right\} \tag{18}$$

The estimator is obtained by plugging $\hat{\eta}_{Rf}$ from 18 into the likelihood function. Therefore, the NGQML with rank selection for errors innovations $\hat{\theta}_T^R$ is estimated as follows:

$$\hat{\theta}_T^R = argmax_\theta \frac{1}{T} \sum_{t=1}^{T} \left(-\log\left(\hat{\eta}_f^R \xi \sigma\right) + \log \frac{\epsilon_t}{\hat{\eta}_f^R \xi \sigma_t} \right) \tag{19}$$

Despite their methods, all estimators described above were developed to deal with errors distributions far from the gaussianity. Their respective authors analysed the estimators' properties to obtain GARCH parameters. Except Preminger & Storti (2017), who investigated the log-transform-based least-squares estimator to predict risk, the comparison between these methods has not yet been made within the VaR framework, and therefore is undertaken in the next sections.

3 Comparison of the estimators' efficiency

In this section, we carry out an extensive simulation study to compare the efficiency of a set of estimators to obtain ω, α and β parameters of GARCH(1,1) processes, evaluated through the root mean squared errors (RMSE). The Student's-t estimator (STD) is the benchmark for the other models since it follows the innovation distribution of the simulated processes. Conjointly with STD, we perform estimates through the Gaussian quasi-maximum likelihood (GQML) and the Generalised Error Distributions (GED) estimators, which we refer as the standard estimators in this study. These approaches are compared to the heavy-tailed errors distributions estimators previously described, that is, the three-steps estimator (NGQML) proposed by Fan et al. (2014), the rank-based estimator (RANK) developed by Andrews (2012), the Log-transform-based least-squares estimator (LLSE) from Preminger & Storti (2017), and the modified version of the NGQML (Fan et al., 2014) that uses the rank technique (Andrews, 2012) prior to the volatility scale parameter estimation (RANK$_2$).

We simulate high persistent processes with true parameters equal to $\omega = 0.01, \alpha = 0.09$ and, $\beta = 0.9$, with innovations following a Student-t distribution with 20, 9, 6, 4 and 3 degrees of freedom. For each degree of freedom -type we reproduce 4000 paths, divided equally in samples sizes with 250, 500, 1000 and 2500 observations, where the 1000 first are discarded as burn-in sample. Tables 1, 2 and 3 reports the RMSE of $\hat{\alpha}$ and $\hat{\beta}$ and $\hat{\omega}$ of all estimators respectively. The first columns show the degrees of freedom of the innovation distributions in descending order, from the thin-tailed (t_{20}) to the heaviest-tailed distribution (t_3). Apart from the STD, which is the benchmark, the most efficient estimator of each case (rows) are in boldface.

For RMSE of $\hat{\omega}$ in table 1, GQML is the most accurate estimator when the sample size contains 250 observations and the innovations are thin-tailed with 20 degrees of freedom, whereas GED is superior for n=500 and t_{20}. Generally, GED shows better performance to estimate ω in small samples (n=250 and n=500) and also in the case of samples with 1000 observations for t_9

and t_4. On the other hand, NGQML performs nearly equal to STD in larger samples, being the best estimator mainly for n=2500, and for innovations with degrees of freedom equal to 20, 6 and 3 of samples with n=1000.

Table 1: Comparison of RMSE of $\hat{\omega}$ with Student's-t innovations

	N	STD	GQML	GED	NGQLE	RANK	$RANK_2$	LLSE
t_{20}	n=250	0.6443	**0.7203**	0.7324	1.0203	3.7244	1.0512	0.9071
	n=500	0.2725	0.2882	**0.2745**	0.3068	0.7092	0.3529	0.3752
	n=1000	0.0982	0.1003	0.0987	**0.0976**	0.1993	0.0996	0.1748
	n=2500	0.0427	0.0438	0.0432	**0.0430**	0.0702	0.0439	0.0619
t_9	n=250	0.6035	0.7134	**0.5641**	0.9944	1.8549	0.9203	0.9833
	n=500	0.2422	0.2342	**0.2189**	0.4184	0.7940	0.4298	0.3304
	n=1000	0.1000	0.1296	**0.1013**	0.1346	0.2027	0.1362	0.1799
	n=2500	0.0397	0.0429	0.0402	**0.0398**	0.0612	0.0406	0.0539
t_6	n=250	0.5955	0.5695	**0.5654**	0.9775	1.7366	0.9358	0.8727
	n=500	0.2342	0.2451	**0.2243**	0.2975	0.5391	0.3052	0.2994
	n=1000	0.0775	0.0995	0.0782	**0.0762**	0.1346	0.0786	0.1058
	n=2500	0.0379	0.0453	0.0385	**0.0378**	0.0550	0.0386	0.0479
t_4	n=250	0.3450	0.5850	**0.4099**	0.7684	1.1016	0.7525	0.6360
	n=500	0.1484	0.3038	**0.1584**	0.2330	0.3799	0.2351	0.2742
	n=1000	0.0647	0.1541	**0.0669**	0.1159	0.1158	0.1193	0.1236
	n=2500	0.0337	0.0616	0.0346	**0.0335**	0.0473	0.0362	0.0462
t_3	n=250	0.3914	1.2419	**0.2710**	0.5735	0.7873	0.6443	1.2411
	n=500	0.1281	0.2701	**0.1294**	0.2150	0.3317	0.1990	0.2565
	n=1000	0.0649	0.1426	0.0808	**0.0553**	0.1454	0.1133	0.0965
	n=2500	0.0308	0.0749	0.0414	**0.0289**	0.0543	0.0535	0.0444

Note: The innovation distribution follow a Student's-t with degrees of freedom equal to 20, 9, 6, 4, and 3. N is the number of observations of the corresponding simulated process.

As seen in table 2, GED has the best performance to obtain α in several cases. GED shows superiority for innovations with t_{20} and t_9, and it is the most accurate estimator for n $\leq$ 1000, except in the case of t_6 with n=250, when $RANK_2$ performs better. Similarly for $\hat{\omega}$, NGQML presents higher accuracy in larger sample sizes. NGQML is the most efficient estimator to estimate α for t_6, t_4, t_3 and n=2500, as well as t_6 and n=1000.

Regarding RMSE of $\hat{\beta}$ in table 3, GED also outperforms other estimators for series with 250 and 500 observations, and t_4 for series with 1000 observations. NGQML and $RANK_2$ are the most efficient estimators when $t = 20$, $t = 9$ with n=2500. Different from $\hat{\omega}$ and $\hat{\alpha}$, on which NGQML performs better in larger sample sizes, $RANK_2$ is the most accurate estimator to estimate β for n=2500 generally. RANK is also superior in the case of 1000 observations for

20, 6 and 2 degrees of freedom.

Table 2: Comparison of RMSE of $\hat{\alpha}$ with Student's-t innovations

	N	STD	GQML	GED	NGQLE	RANK	$RANK_2$	LLSE
t_{20}	n=250	0.5653	0.5614	**0.5546**	0.5772	0.7132	0.5756	0.7853
	n=500	0.3130	0.3150	**0.3132**	0.3291	0.4318	0.3407	0.4415
	n=1000	0.2005	0.2043	**0.2018**	0.2044	0.2685	0.2093	0.2960
	n=2500	0.1242	0.1268	**0.1252**	0.1260	0.1461	0.1273	0.1777
t_9	n=250	0.5324	0.5702	**0.5264**	0.5622	0.7231	0.5589	0.7831
	n=500	0.3234	0.3666	**0.3314**	0.3406	0.4402	0.3430	0.4697
	n=1000	0.2132	0.2341	**0.2175**	0.2182	0.2750	0.2258	0.2898
	n=2500	0.1323	0.1416	**0.1336**	0.1338	0.1538	0.1364	0.1799
t_6	n=250	0.5590	0.6283	0.5764	0.5890	0.7233	**0.5761**	0.8103
	n=500	0.3374	0.3997	**0.3400**	0.3424	0.4465	0.3578	0.4744
	n=1000	0.2153	0.2587	**0.2189**	**0.2189**	0.2724	0.2279	0.2740
	n=2500	0.1351	0.1654	0.1391	**0.1362**	0.1597	0.1458	0.1687
t_4	n=250	0.5941	0.8405	**0.6416**	0.6734	0.7736	0.7197	0.8265
	n=500	0.3418	0.5810	**0.3607**	0.3636	0.5014	0.4048	0.5398
	n=1000	0.2369	0.4177	**0.2486**	0.2552	0.3011	0.2889	0.3760
	n=2500	0.1434	0.2389	0.1572	**0.1495**	0.1785	0.1715	0.1790
t_3	n=250	0.6619	1.2739	**0.7109**	0.7956	2.9678	1.0333	0.9609
	n=500	0.3928	0.8743	**0.4308**	0.4455	0.5589	0.5415	0.6558
	n=1000	0.2762	0.5356	**0.2845**	0.3011	0.9441	0.6655	0.3753
	n=2500	0.1655	0.4021	0.2278	**0.1920**	0.3575	0.3772	0.2371

Note: The innovation distribution follow a Student's-t with degrees of freedom equal to 20, 9, 6, 4, and 3. N is the number of observations of the corresponding simulated process.

Table 3: Comparison of RMSE of $\hat{\beta}$ with Student's-t innovations

	N	STD	GQML	GED	NGQLE	RANK	RANK_2	LLSE
t_{20}	n=250	1.0230	1.1116	**1.0978**	1.9423	2.9181	1.9043	1.5579
	n=500	0.5874	0.6153	**0.5917**	0.6564	1.3410	0.7140	0.8117
	n=1000	0.2487	0.2561	0.2519	0.2473	0.4492	**0.2472**	0.4370
	n=2500	0.1387	0.1432	0.1407	**0.1391**	0.1950	**0.1391**	0.2098
t_9	n=250	1.0076	1.0910	**1.0173**	2.0986	2.9345	1.8267	1.5579
	n=500	0.5275	0.5376	**0.4964**	0.7967	1.4626	0.7870	0.8253
	n=1000	0.2654	0.3334	**0.2731**	0.3285	0.4779	0.3284	0.4660
	n=2500	0.1428	0.1537	0.1449	**0.1439**	0.1904	**0.1439**	0.2037
t_6	n=250	0.9910	1.0180	**1.0051**	2.1006	2.7944	1.9785	1.4562
	n=500	0.6309	0.6541	**0.6082**	0.8176	1.3556	0.8283	0.7910
	n=1000	0.2537	0.3194	0.2631	0.2522	0.3903	**0.2517**	0.3295
	n=2500	0.1456	0.1787	0.1518	0.1455	0.1854	0.**1454**	0.1893
t_4	n=250	0.9523	1.5105	**1.1908**	2.4563	2.8619	2.1227	1.6616
	n=500	0.5172	0.8194	**0.5661**	0.7749	1.1970	0.7638	0.7382
	n=1000	0.2628	0.5089	**0.2851**	0.4065	0.4049	0.4056	0.4280
	n=2500	0.1444	0.2567	0.1592	0.1473	0.1790	**0.1468**	0.1887
t_3	n=250	1.1191	1.7969	**1.1789**	2.9637	3.0943	2.6515	1.7037
	n=500	0.5976	1.2028	**0.7298**	1.0519	1.5431	0.9584	1.0125
	n=1000	0.2966	0.6877	0.3215	0.3098	0.4290	**0.3000**	0.4583
	n=2500	0.1550	0.4237	0.1701	0.1621	0.1781	**0.1603**	0.2331

Note: The innovation distribution follow a Student's-t with degrees of freedom equal to 20, 9, 6, 4, and 3. N is the number of observations of the corresponding simulated process.

4 Value-at-risk (VaR)

We have observed that heavy-tailed estimators present asymptotic efficiency to estimate GARCH parameters, especially for processes with leptokurtic distribution. Once financial returns generally display this characteristic, we aim to verify whether the ability of these estimators to accommodate excess kurtosis can help to predict the risk at the left tail of returns' distributions. Therefore, in the remaining of this article, we analyse the performance of the standard and the heavy-tailed estimators to forecast the Value-at-risk (VaR), which is the most popular measure used by financial institutions to monitor their minimum capital risk requirements. Empirically, VaR is defined as a statistical method that summarises the expectation of a loss from an asset or a portfolio, associated with a confidence interval over a specific time horizon (Jorion, 2006).

Moreover, VaR is the negative quantile of returns distribution, as described in equation 20:

$$VaR_{t+1}(\alpha) = -Q_{\alpha}(r_{t+1}|I_t) = -inf\{x \in \Re : P_r(r_{t+1} \leq x|I_t) \geq \alpha\} \tag{20}$$

Where $Q_{\alpha}(\cdot)$ is the cumulative distribution function, α is the significance level, $r_{t+1} = ln(P_{t+1}/P_t)$ are the log returns at each period t, and I_t is the α-quantile of the cumulative distribution function calculated according to the density of returns.

In this study, some backtesting procedures are employed to evaluate and compare the performance of estimators to forecast the VaR. We calculate the ratio of actual to the expected VaR violations (A/E), which measures VaR failures to cover the risk of financial returns for a certain period. It is expected that extreme returns exceeds the VaR coverage at some α level so that VaR violations is determined through a "hit sequence" (Hit_{t+1}) (Christoffersen & Gonçalves, 2005), such that:

$$Hit_{t+1} = \begin{cases} 1, & if \quad r_{t+1} < -VaR^{\alpha}_{t+1} \\ 0, & if \quad r_{t+1} \geq -VaR^{\alpha}_{t+1} \end{cases} \tag{21}$$

The rule of the hit sequence Hit_{t+1} returns 1 when the VaR prediction is smaller than the asset return, and 0 otherwise. A/E is the ratio of actual VaR violations to expected VaR violations:

$$A/E = \frac{\sum_{j=1}^{p} d_{t+j}}{p} \tag{22}$$

with $d_{t+1} = 1(r_{t+1} < VaR^{\alpha}_{t+1})$ in equation 21, $p = \alpha\%P$ the proportion of expected violations, where α is the coverage level and P the number of observations. The highest VaR accuracy is achieved when A/E=1.

We also test the null hypothesis of correct coverage of VaR forecasts through the tests developed by Kupiec (1995), Christoffersen (1998), and Engle & Manganelli (2004). The unconditional coverage test (LRuc) proposed by Kupiec (1995) compares the observed number of violations to the expected number of violations determined by the significance level α. LRuc is based on the likelihood ratio test derived from a Bernoulli hit sequence, whose statistics is:

$$LR_{uc} = -2ln\left[L(p)/L(\hat{\pi})\right] \tag{23}$$

where $\hat{\pi}$ represents a fraction of observed violations in the hit sequence of 0's and 1's, assuming 1 if $r_{t+1} < -VaR^{\alpha}_{t+1}$, and 0, otherwise.

The conditional coverage test (LRcc) introduced by Christoffersen (1998) tests if VaR violations are independently distributed over time:

$$LR_{cc} = -2ln\left[L(p)/L(\hat{\pi})\right] - 2ln\left[L(\hat{\pi})/L(\hat{\pi}_1)\right] \tag{24}$$

where the first element of the equation is the LRuc test as shown in 23, and the second element is the ratio of LRuc from two different periods.

The dynamic quantile test (DQ) developed by Engle & Manganelli (2004) has under the null hypothesis the correct unconditional and conditional number of violations of VaR forecast, where the statistical significance of parameters of the following regression is tested:

$$Hit_t^{\alpha} = \delta_0 + \sum_{l=1}^{L} \delta_1 Hit_{t-l}^{\alpha} + \delta_{L+1} VaR_{t-1}^{\alpha} + \varepsilon_t \tag{25}$$

where Hit_t is the hit sequence, and the the joint significance of the parameters are tested through a Wald statistic following a Chi-squared distribution with $L + 2$ degrees of freedom.

In the next sections, we compare VaR forecasts produced by the standard and heavy-tailed estimators in two different contexts; first, by analysing the ability of the estimators to forecast the VaR in a simulated study, and second, by testing the methods in tracking the VaR for six equity indexes returns. Backtesting measures described above are used to evaluate VaR forecasts.

4.1 Simulated VaR

As previously mentioned, heavy-tailed estimators are more efficient to estimate GARCH parameters when performed over large sample sizes and for excess kurtosis innovation series. In this section, their ability to forecast the VaR are compared in a simulated study. We generate high persistence processes with GARCH (1,1) parameters equal to $\omega = 0.01, \alpha = 0.09, \beta = 0.9$, and Student's-t innovation errors with 3 degrees of freedom, which resembles many financial series. We simulate 100 realisations where the initial 1000 observations are discarded as a burn-in sample, and the last 1000 observations are left for VaR predictions. Forecasts are performed considering estimation windows length equal to 250, 500, 1000 and 2500. The procedure is carried out as follows:

1. Performing one-step ahead rolling window scheme of a GARCH (1,1) model to obtain $\hat{\sigma}_{t+1}$ through each of the standard and heavy-tailed estimators:

$$\hat{\sigma}_{(t+1)} = \hat{\omega} + \hat{\alpha}\varepsilon_t + \hat{\beta}\sigma_t \tag{26}$$

2. Calculating VaR forecasts at 5% and 1% coverage levels (α):

$$V\hat{a}R_{(t+1)}^{\alpha} = \hat{\sigma}_{(t+1)} Q_{\alpha} \tag{27}$$

Where Q_{α} is the α quantile for a given error distribution, and $\hat{\sigma}_{t+1}$ is the predicted conditional standard deviation. For each estimator, the outcome is a 1000x100 matrix, corresponding to 100 series of VaR predictions series with length equals to 1000.

A/E columns in tables are the mean ratio of actual to expected vaR violations per 100 simulations. Regarding LRuc, LRcc and DQ tests, the figures are interpreted as the percentage of null hypothesis rejections from the underlying tests out of 100 generated process. In this regard, more accurate models are expected to present A/E closer to 1 and the smallest percentage of test rejections. The upper panels refer to backtesting outcomes for VaR at 5% coverage level, and the bottom panels bring the results for VaR at 1% coverage level. Shadowed rows indicate the best models of each panel. STD in the second row follows the same innovation error of the simulated process so that the estimator is used as a benchmark for the other models.

Table 4 reports the results for VaR predictions with the estimation windows length with 250 observations. According to the four backtesting measures, GED estimator present superior performance to forecast the VaR at 5% and 1% coverage levels. GED produces the closest A/E ratios to the unity in both cases, underestimating VaR at 5% (0.936) and overestimating VaR at 1% (1.248) on average. The estimator also has the lowest proportion of test rejections in these cases, with percentages for LRuc, LRcc and DQ tests equal to 7%, 4% and 6% for VaR at 5% coverage level, and 2%, 1% and 12% for VaR at 1% coverage level.

Table 4: VaR Backtesting of simulated processes. Estimation window length with 250 observations. A/E is the mean ratio of VaR violations. LRuc, LRcc and DQ are the proportion of rejections of the null hypothesis of correct VaR converge out of 100 simulations.

VaR	estimator	A/E	LRuc	LRcc	DQ
5%	GQML	0.862	0.120	0.140	0.090
	STD	1.280	0.400	0.340	0.450
	GED	0.936	0.070	0.040	0.060
	NGQML	1.184	0.380	0.430	0.620
	RANK	0.791	0.410	0.340	0.140
	RANK_2	0.792	0.400	0.320	0.130
	LLSE	0.889	0.130	0.080	0.060
1%	GQML	1.942	0.460	0.380	0.560
	STD	1.421	0.090	0.040	0.150
	GED	1.248	0.020	0.010	0.120
	NGQML	2.663	0.670	0.650	0.780
	RANK	1.345	0.040	0.040	0.160
	RANK_2	1.363	0.040	0.040	0.190
	LLSE	1.537	0.130	0.070	0.260

NOTE: A/E is the mean ratio of actual to expected VaR violations. LRuc is the unconditional coverage test (Kupiec, 1995), LRcc is the conditional coverage test(Christoffersen, 1998) and DQ is the dynamic quantile test (Engle & Manganelli, 2004) (the figures are the proportion of rejections out of 100 simulations). Results of 100 simulated processes generated with Student's-t distribution innovations errors with 3 degrees of freedom.

In table 5 are depicted backtesting results for VaR forecasts when the estimation length is set to 500 observations. STD and GED estimators present similar performance in this case. Considering VaR at 5% coverage level, the A/E ratio of both estimators are equivalent, with the former overestimating the risk and the latter underestimating it. GED shows higher accuracy with the least proportion of rejections of the null of correct VaR coverage at 5%. Furthermore, GED is outperformed by STD by a small margin regarding the A/E ratio for VaR at 1% coverage level .

Table 5: VaR Backtesting of simulated processes. Estimation window length with 500 observations. A/E is the mean ratio of VaR violations. LRuc, LRcc and DQ are the proportion of rejections of the null hypothesis out of 100 simulations.

VaR	estimator	A/E	LRuc	LRcc	DQ
5%	GQML	0.805	0.370	0.300	0.100
	STD	1.117	0.220	0.190	0.230
	GED	0.892	0.180	0.170	0.070
	NGQML	0.846	0.280	0.260	0.090
	RANK	0.739	0.530	0.440	0.150
	$RANK_2$	0.731	0.580	0.530	0.200
	LLSE	0.840	0.290	0.210	0.060
1%	GQML	1.801	0.320	0.280	0.420
	STD	1.163	0.030	0.000	0.100
	GED	1.174	0.010	0.000	0.110
	NGQML	1.460	0.150	0.120	0.240
	RANK	1.230	0.020	0.010	0.100
	$RANK_2$	1.222	0.030	0.020	0.120
	LLSE	1.401	0.040	0.040	0.190

NOTE: A/E is the mean ratio of actual to expected VaR violations. LRuc is the unconditional coverage test (Kupiec, 1995), LRcc is the conditional coverage test(Christoffersen, 1998) and DQ is the dynamic quantile test (Engle & Manganelli, 2004) (the figures are the proportion of rejections out of 100 simulations). Results of 100 simulated processes generated with Student's-t distribution innovations errors with 3 degrees of freedom.

Heavy-tailed estimators tend to perform better in larger estimation windows as observed VaR backtesting results in table 6 and 7. Nevertheless, GED shows superior accuracy to forecast VaR at 5% generally. Analysing VaR at 1% coverage level and estimation window length with 1000 observations in table 6, $RANK_2$ outperforms the others estimators in terms of A/E ratio and also has the lowest proportion of rejections for the three coverage tests. RANK is the second most efficient estimator being nearly as accurate as $RANK_2$ to cover the VaR at 1% level. NGQML and LLSE estimators present a relatively small proportion of test rejections to GED despite higher A/E ratios.

Table 6: VaR Backtesting of simulated processes. Estimation window length with 1000 observations. A/E is the mean ratio of VaR violations. LRuc, LRcc and DQ are the proportion of rejections of the null hypothesis out of 100 simulations.

VaR	estimator	A/E	LRuc	LRcc	DQ
5%	GQML	0.767	0.540	0.470	0.100
	STD	1.051	0.160	0.140	0.150
	GED	0.917	0.340	0.280	0.110
	NGQML	0.797	0.480	0.430	0.080
	RANK	0.693	0.660	0.600	0.170
	$RANK_2$	0.690	0.680	0.630	0.210
	LLSE	0.795	0.460	0.440	0.110
1%	GQML	1.711	0.250	0.200	0.320
	STD	1.104	0.020	0.020	0.050
	GED	1.250	0.080	0.070	0.120
	NGQML	1.342	0.100	0.060	0.140
	RANK	1.159	0.050	0.020	0.130
	$RANK_2$	1.149	0.040	0.010	0.110
	LLSE	1.335	0.090	0.060	0.170

NOTE: A/E is the mean ratio of actual to expected VaR violations. LRuc is the unconditional coverage test (Kupiec, 1995), LRcc is the conditional coverage test(Christoffersen, 1998) and DQ is the dynamic quantile test (Engle & Manganelli, 2004) (the figures are the proportion of rejections out of 100 simulations). Results of 100 simulated processes generated with Student's-t distribution innovations errors with 3 degrees of freedom.

Table 7: VaR Backtesting of simulated processes. Estimation window length with 2500 observations. A/E is the mean ratio of VaR violations. LRuc, LRcc and DQ are the proportion of rejections of the null hypothesis out of 100 simulations.

VaR	estimator	A/E	LRuc	LRcc	DQ
5%	GQML	0.731	0.650	0.550	0.180
	STD	1.024	0.210	0.170	0.110
	GED	0.958	0.350	0.300	0.160
	NGQML	0.778	0.500	0.450	0.140
	RANK	0.701	0.640	0.560	0.170
	$RANK_2$	0.698	0.640	0.560	0.170
	LLSE	0.777	0.510	0.430	0.130
1%	GQML	1.644	0.180	0.130	0.240
	STD	1.067	0.030	0.010	0.080
	GED	1.331	0.120	0.130	0.170
	NGQML	1.327	0.070	0.060	0.140
	RANK	1.156	0.020	0.020	0.090
	$RANK_2$	1.153	0.020	0.010	0.080
	LLSE	1.315	0.060	0.050	0.140

NOTE: A/E is the mean ratio of actual to expected VaR violations. LRuc is the unconditional coverage test (Kupiec, 1995), LRcc is the conditional coverage test(Christoffersen, 1998) and DQ is the dynamic quantile test (Engle & Manganelli, 2004) (the figures are the proportion of rejections out of 100 simulations). Results of 100 simulated processes generated with Student's-t distribution innovations errors with 3 degrees of freedom.

4.2 Empirical VaR

In this section, we compare the performance of the estimators to predict the risk of the following equity indexes: Nasdaq (USA), Ftse (UK) , Dax (Germany), Nikkei (Japan), Hang Seng (Hong Kong) and Ibovespa (Brazil). Daily returns are calculated as $r_t = log(p_t/p_{t-1})$, where p_t is the adjusted closing price[1]. The data was collected from December 01, 2005, to July 02, 2020. We employ a rolling window scheme of 1000 trading days to forecast 1-day-ahead VaR covering the period ranging from January 04, 2010, to July 02, 2020. Table 8 reports descriptive statistics of the equity indexes returns for the 1000 first days of each series, whereas table 9 contains descriptive statistics for the remaining data set for which VaR predictions are undertaken. Except for the Jarque-Bera test (JB) for normality, all figures in tables are in percentage.

Table 8: Descriptive statistics of market indexes returns (in %). 1000 first trading days until December 30, 2009

	nobs	Mean	Stdev	Min	Med	Max	Skew.	Kurt.	JB
Nasdaq	1000	-0.003	1.731	-9.588	0.084	11.159	-0.122	5.989	1506.05
Ftse	1000	-0.004	1.510	-8.710	0.034	8.811	-0.148	6.086	1556.08
Dax	1000	0.011	1.651	-7.433	0.108	10.797	0.206	7.204	2181.53
Nikkei	1000	-0.039	1.900	-12.111	0.036	13.235	-0.379	7.293	2252.60
HSeng	1000	0.038	2.125	-13.582	0.110	13.407	0.085	6.244	1635.09
Ibov	1000	0.073	2.234	-12.096	0.163	13.678	-0.010	5.103	1091.84

NOTE: nobs are the trading days. Stdev, Skew and Kurt are the standard deviation, skewness and excess kurtosis of distribution returns. JB is the value of the Jarque-Bera statistic for normality.

Table 9: Descriptive statistics of market indexes returns (in %) between January 04, 2010 and July 02, 2020

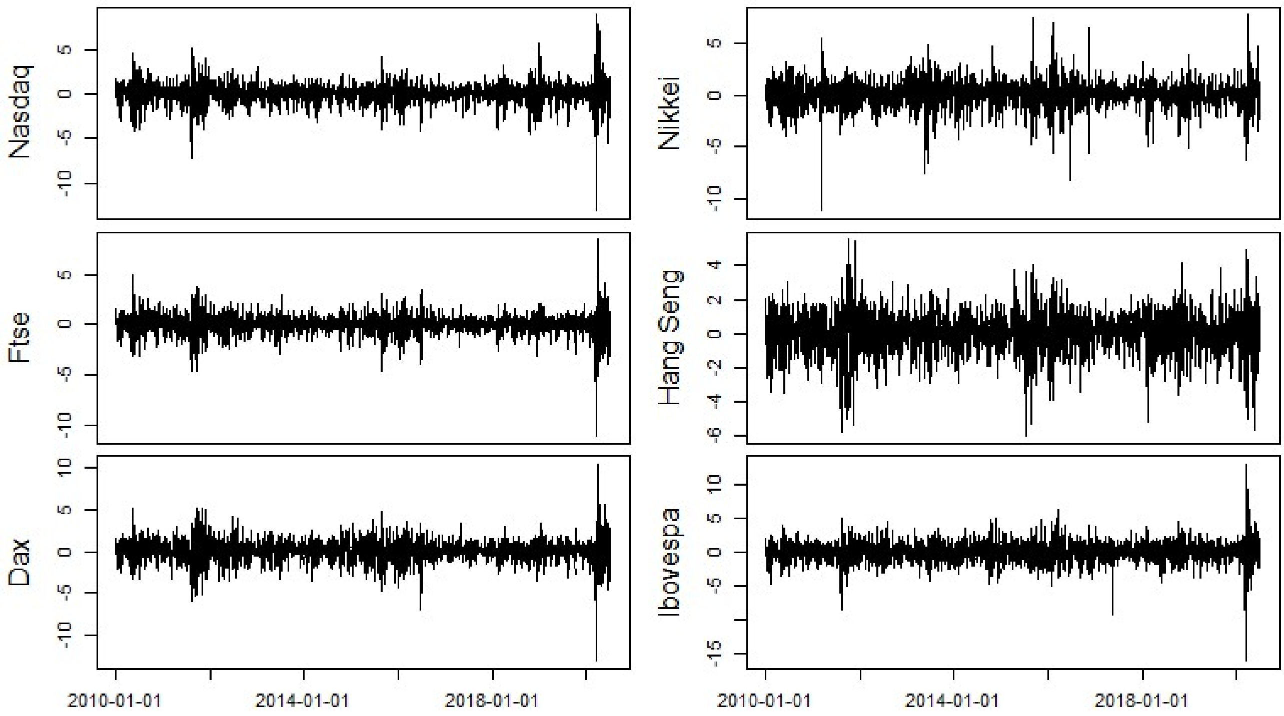
	nobs	Mean	Stdev	Min	Med	Max	Skew.	Kurt.	JB
Nasdaq	2643	0.057	1.222	-13.149	0.104	8.935	-0.790	11.591	15099.26
Ftse	2653	0.008	1.009	-11.084	0.048	8.490	-0.776	10.862	13333.86
Dax	2667	0.028	1.294	-13.055	0.082	10.414	-0.551	8.026	7307.48
Nikkei	2570	0.029	1.355	-11.153	0.066	7.731	-0.453	5.414	3234.31
HSeng	2586	0.005	1.177	-6.018	0.051	5.519	-0.349	2.452	702.80
Ibov	2597	0.013	1.603	-15.994	0.029	13.010	-0.858	12.310	16748.46

NOTE: nobs are the trading days. Stdev, Skew and Kurt are the standard deviation, skewness and excess kurtosis of distribution returns. JB is the value of the Jarque-Bera statistic for normality.

[1]the data was collected from the Thomson Reuters Datastream

As seen in both tables, the null hypothesis of normality is rejected by the Jarque-Bera test (JB) for all equity indexes-cases. Likewise, the excess kurtosis (kurt) reinforces the fact of non-normality of financial returns. Extreme returns in (table8) are regards to the Subprime crisis. In the forecasting sample (table 9), Nasdaq and Dax are the only series characterised by positive skewness. It is important to highlight that this period coincides with the spread of the COVID-19 over the world, whose impacts on financial markets are illustrated in figure 1. Furthermore, both periods, in-sample and forecasting sample, are highly volatile with an annualised standard deviation of Nasdaq returns, for example, equals to 27.51% (1.731%x$\sqrt{252}$) for the Subprime crisis period and 19.39% (1.122%x$\sqrt{252}$) for the recent COVID-19 crisis.

Figure 1: Equity indexes returns



Given the leptokurtic characteristics of returns, it is expected that estimators that deal with heavy-tailed innovations stand out from the standard Gaussian estimator. In order to analyse the estimators' performance, the procedure here implemented follows the steps used in the simulated study. We fit GARCH(1,1) models through the proposed estimators to obtain predicted VaR densities at 1% and 5% coverage levels. We set the estimation window length to 1000 trading days. The accuracy of the VaR forecasts are tested through the LRuc test (Kupiec, 1995), LRcc test (Christoffersen, 1998) and DQ test (Engle & Manganelli, 2004) as well as the ratio of actual to the expected VaR violations are calculated (A/E). Tables 10 and 11 report VaR at 5% and 1% coverage levels, respectively. Shadowed rows in tables highlight estimators whose A/E ratio is the closest to 1. Columns 4 to 6 are the p-values relative to coverage tests.

It is observed in table 10 that the standard estimators outperform the heavy-tailed estimators for most of equity indexes cases when considered VaR at 5% level. GQML has the lowest A/E ratio for Ftse, Dax and Nikkei, and it is as efficient as GED to predict the risk for Hang Seng returns. P-values of LRuc and LRcc tests indicate that GQML produces the correct VaR coverage in these cases, even though the null hypothesis related to DQ test is rejected for Dax and Nikkei indexes. Student's-t (STD) is the most accurate estimator in the case of Ibovespa index, with A/E ratio almost equals to the unity. Nevertheless, NGQML is outperformed by STD only by a small margin in this case. GED and RANK estimators have the best performance for Nasdaq, the only market where a heavy-tailed estimator presents similar performance to GED in tracking VaR at 5% coverage level.

Heavy-tailed estimators stand out when the risk level of VaR is set to 1%. Examining table 11, RANK is the estimator with the best performance when analysed A/E ratio for all equity indexes, except in the case of Ibovespa returns. Despite its superiority, the null hypothesis of correct VaR coverage is rejected for Nasdaq by LRuc, LRcc and DQ tests and Ftse by DQ test. RANK_2 stands as the second-best estimator for these markets, whereas STD shows the same performance as RANK for Nikkei. In the case of Ibovespa returns, GED demonstrates high accuracy to forecast VaR, with A/E equals to 1. Generally, for the equity indexes-cases, the estimators tend to underestimate the number of expected VaR violations, with $A/E > 1$. Conversely, the risk is overestimated by the majority of estimators for Ibovespa returns, with heavy-tailed estimators exhibiting similar performance for VaR at 1% coverage level.

Table 10: Backtesting measures for VaR at 5% coverage level

	Models	A/E	LRuc	LRcc	DQ
Nasdaq	GQML	1.0972	0.2585	0.3944	0.2503
	STD	1.1578	0.0692	0.1516	0.0111
	GED	1.0897	0.2969	0.4462	0.1354
	NGQML	1.1653	0.0571	0.1258	0.0057
	RANK	1.0897	0.2969	0.4462	0.1847
	$RANK_2$	1.1502	0.0833	0.1811	0.0050
	LLSE	1.1124	0.1926	0.3000	0.0327
Ftse	GQML	1.0856	0.3183	0.3171	0.1581
	STD	1.1308	0.1297	0.0973	0.0620
	GED	1.0931	0.2781	0.1987	0.1137
	NGQML	1.1760	0.0427	0.0058	0.0033
	RANK	1.1383	0.1094	0.1402	0.1061
	$RANK_2$	1.1383	0.1094	0.0540	0.0225
	LLSE	1.1609	0.0634	0.0064	0.0041
Dax	GQML	1.1349	0.1177	0.0562	0.0079
	STD	1.2401	0.0061	0.0046	0.0002
	GED	1.1424	0.0990	0.0885	0.0371
	NGQML	1.2627	0.0028	0.0017	0.0000
	RANK	1.1725	0.0467	0.0608	0.0372
	$RANK_2$	1.2101	0.0159	0.0134	0.0004
	LLSE	1.2552	0.0036	0.0006	0.0000
Nikkei	GQML	1.0428	0.6210	0.4642	0.0047
	STD	1.0973	0.2650	0.3817	0.0002
	GED	1.0584	0.5011	0.4610	0.0015
	NGQML	1.1128	0.1970	0.3310	0.0001
	RANK	1.0895	0.3046	0.4041	0.0004
	$RANK_2$	1.1051	0.2292	0.3571	0.0004
	LLSE	1.1595	0.0701	0.0997	0.0000
Hang Seng	GQML	1.0596	0.4912	0.3023	0.6999
	STD	1.1137	0.1924	0.2067	0.5356
	GED	1.0596	0.4912	0.3023	0.6966
	NGQML	1.1292	0.1395	0.1474	0.4868
	RANK	1.0905	0.2978	0.1787	0.6099
	$RANK_2$	1.0905	0.2978	0.1787	0.4298
	LLSE	1.1678	0.0562	0.0917	0.1640
Ibovespa	GQML	0.9472	0.5339	0.5504	0.3205
	STD	1.0012	0.9892	0.6207	0.0663
	GED	0.9241	0.3693	0.2433	0.0386
	NGQML	1.0166	0.8469	0.6611	0.0637
	RANK	0.9781	0.7968	0.7440	0.5013
	$RANK_2$	0.9858	0.8674	0.5596	0.0652
	LLSE	1.0320	0.7100	0.4725	0.0143

NOTE: A/E is the ratio of actual to expected VaR violations. p-values of the unconditional coverage test(LRuc) (Kupiec, 1995), conditional coverage test (LRcc) Christoffersen (1998) and dynamic quantile test (DQ) (Engle & Manganelli, 2004).

Table 11: Backtesting measures for VaR at 1% coverage level

	Models	A/E	LRuc	LRcc	DQ
Nasdaq	GQML	2.4593	0.0000	0.0000	0.0000
	STD	1.8540	0.0001	0.0003	0.0000
	GED	1.8161	0.0002	0.0004	0.0001
	NGQML	1.8918	0.0000	0.0001	0.0000
	RANK	1.5513	0.0084	0.0117	0.0028
	$RANK_2$	1.6648	0.0017	0.0034	0.0000
	LLSE	1.8161	0.0002	0.0004	0.0000
Ftse	GQML	2.0354	0.0000	0.0000	0.0000
	STD	1.4700	0.0230	0.0659	0.0000
	GED	1.4700	0.0230	0.0659	0.0000
	NGQML	1.4323	0.0356	0.0938	0.0000
	RANK	1.3570	0.0797	0.0546	0.0000
	$RANK_2$	1.3946	0.0539	0.1297	0.0000
	LLSE	1.4323	0.0356	0.0329	0.0000
Dax	GQML	1.8038	0.0002	0.0005	0.0000
	STD	1.2401	0.2301	0.3216	0.0996
	GED	1.2026	0.3088	0.4035	0.0974
	NGQML	1.2777	0.1674	0.2483	0.0883
	RANK	1.1650	0.4046	0.4902	0.1023
	$RANK_2$	1.2026	0.3088	0.4035	0.0890
	LLSE	1.3153	0.1190	0.1860	0.1085
Nikkei	GQML	2.2179	0.0000	0.0000	0.0000
	STD	1.3230	0.1169	0.0648	0.1062
	GED	1.3619	0.0805	0.0529	0.0934
	NGQML	1.4786	0.0228	0.0236	0.0360
	RANK	1.3230	0.1169	0.0648	0.0883
	$RANK_2$	1.3619	0.0805	0.0529	0.0772
	LLSE	1.5175	0.0143	0.0169	0.0257
Hang Seng	GQML	2.2042	0.0000	0.0000	0.0000
	STD	1.7788	0.0003	0.0016	0.0000
	GED	1.7015	0.0011	0.0048	0.0000
	NGQML	1.5081	0.0157	0.0478	0.0000
	RANK	1.3534	0.0865	0.1831	0.0500
	$RANK_2$	1.4308	0.0386	0.0993	0.0001
	LLSE	1.6241	0.0034	0.0129	0.0000
Ibovespa	GQML	1.3092	0.1306	0.2031	0.0089
	STD	0.9626	0.8474	0.7698	0.5238
	GED	1.0012	0.9953	0.7687	0.5795
	NGQML	0.9241	0.6939	0.7398	0.4719
	RANK	0.9241	0.6939	0.7398	0.4569
	$RANK_2$	0.9241	0.6939	0.7398	0.4721
	LLSE	0.9241	0.6939	0.7398	0.4765

NOTE: A/E is the ratio of actual to expected VaR violations. p-values of the unconditional coverage test(LRuc) (Kupiec, 1995), conditional coverage test (LRcc) Christoffersen (1998) and dynamic quantile test (DQ) (Engle & Manganelli, 2004).

4.3 Model Confidence Set

In the previous section, we evaluated the performance of the seven estimators to forecast the risk of a set of equity indexes returns. We employed VaR backtesting procedures to compare the efficiency of the estimators to track the VaR for 5% and 1% coverage levels. In this section, VaR forecasts produced by the standard and heavy-tailed estimators are compared through the Model Confidence Set (MCS) procedure developed by Hansen et al. (2011). The period of evaluation ranges from January 4, 2010, to July 2, 2020, when VaR predictions are performed with an estimation window length of 1000 trading days. Estimators are selected by MCS algorithm according to their ability to forecast the VaR using the asymmetric loss function suggested by González-Rivera et al. (2004):

$$L\left(r_t, VaR_t^{\alpha}\right) = \left(\alpha - d_t^{\alpha}\right)\left(r_t - VaR_t^{\alpha}\right) \tag{28}$$

where r_t are the returns, α is the coverage level and $d_t^{\alpha} \equiv 1(r_t < VaR_t^{\alpha})$, whose (1-$\alpha$) is a weight that penalises more heavily observations when $r_t < VaR_t^{\alpha}$.

From an initial group of models (M_0) and a confidence level α, the MCS procedure aims to deliver a superior set of estimators $\hat{M}^{*}_{SSM}$ that produces the most accurate VaR predictions. Based on the loss function 28, the equal predictive ability (EPA) hypothesis of the models is tested by comparing the average loss between pairs of models and the average loss between a single model and all elements in (M_0) through the $T_{R,M}$ and $T_{Max,M}$ statistics, respectively. P-values of these corresponding statistics are assessed using 5000 bootstrap replications. We refer to the original paper of Hansen et al. (2011) for the complete description of the methodology, and Bernardi & Catania (2016) for MCS implementation within the VaR context. The MCS procedure in this study was applied through the R package developed by Catania & Bernardi (2017).

Table 12 and 13 report the final set of estimators determined by the MCS procedure for VaR forecasts at 5% and 1% coverage level. The estimators are order in columns according to p-values of $T_{R,M}$ (columns 2 and 3) and $T_{Max,M}$ (columns 4 and 5) tests, and the loss function of González-Rivera et al. (2004) ((columns 7 and 8). We set the confidence level of MCS to 0.70 to select the most accurate estimators to predict the VaR.

According to MCS results in table 12, RANK, GED, GQML and RANK_2 exhibit equivalent predictive ability to forecast the VaR at 5% coverage level in case of Nasdaq and Nikkei equity indexes. Besides RANK, GED, GQML and RANK_2, LLSE is also considered an efficient estimation to cover the risk for FTSE, whereas GED and GQML do so for Dax index returns. In the case of Hang Seng, the single estimator selected by MCS is RANK, and for Ibovespa, RANK and GQML are within the superior set of models. For MCS of VaR at 5% coverage level, it is observed that RANK has the highest p-values and the lowest mean loss function, except in the case of DAX where the estimator does not be part of the final set of models.

Furthermore, RANK is considered an accurate estimator to forecast the VaR at 1% coverage level for all markets, as seen in table 13. RANK is the single estimator selected by the MCS procedure in case of Nasdaq and Nikkei. Alongside with RANK, STD is included in the superior set for FTSE, whereas $RANK_2$ do so for Hang Seng index. The most heterogeneous sets of estimators determined by the MCS procedure are reported to Dax and Ibovespa index returns. RANK, LLSE, GED and STD belongs to the superior set of MCS for the former, whereas the predictive ability to forecast VaR at 1% cannot be rejected toward any estimator in the case of Ibovespa.

Table 12: Model Confidence set for VaR forecasts at 5% coverage level

	Estimator	p-value$_{R,M}$	Estimator	p-value$_{Max,M}$	Estimator	Loss x10
Nasdaq	RANK	1.000	RANK	1.000	RANK	1.3511
	GED	1.000	GED	1.000	GED	1.3521
	GQML	0.945	GQML	0.892	GQML	1.3533
	$RANK_2$	0.771	$RANK_2$	0.627	$RANK_2$	1.3538
Ftse	RANK	1.000	RANK	1.000	RANK	1.0931
	GED	1.000	GED	1.000	GED	1.0956
	$RANK_2$	1.000	LLSE	0.798	$RANK_2$	1.0959
	GQML	0.964	$RANK_2$	0.657	GQML	1.0966
	LLSE	0.865	GQML	0.573	LLSE	1.0982
Dax	GED	1.000	GED	1.000	GED	1.4190
	GQML	0.935	GQML	0.935	GQML	1.4191
Nikkei	RANK	1.000	RANK	1.000	RANK	1.3511
	GED	1.000	GED	1.000	GED	1.3521
	GQML	0.942	GQML	0.893	GQML	1.3533
	$RANK_2$	0.762	$RANK_2$	0.617	$RANK_2$	1.3538
H. Seng	RANK	1.000	RANK	1.000	RANK	1.3530
Ibovespa	RANK	1.000	RANK	1.000	RANK	1.6586
	GQML	0.903	GQML	0.903	GQML	1.6592

NOTE: MCS is performed considering 0.7 confidence level. For each equity index, estimators in columns 2,4 and 6 are ordered according to p-values in columns 3 and 5, and the loss function in column 7. p-value$_{R,M}$ and p-value$_{Max,M}$ refer to the p-values of $T_{R,M}$ and $T_{Max,M}$ statistics, respectively. Loss is the asymmetric loss function suggested by González-Rivera et al. (2004). VaR forecasts are performed with a estimation window length of 1000 observations. Evaluation period between January 03, 2010, to July 02, 2020.

Table 13: Model Confidence set for VaR forecasts at 1% coverage level

	Estimator	p-vlue	Estimator	p-vlue	Estimator	Loss x10
Nasdaq	RANK	1.000	RANK	1.000	RANK	0.3880
Ftse	RANK	1.000	RANK	1.000	RANK	0.3218
	STD	0.807	STD	0.807	STD	0.3222
Dax	RANK	1.000	RANK	1.000	RANK	0.4048
	LLSE	0.980	LLSE	0.974	STD	0.4070
	GED	0.949	STD	0.839	GED	0.4070
	STD	0.927	GED	0.755	LLSE	0.4073
Nikkei	RANK	1.000	RANK	1.000	RANK	0.3880
H. Seng	RANK	1.000	RANK	1.000	RANK	0.3820
	$RANK_2$	0.855	$RANK_2$	0.855	$RANK_2$	0.3829
Ibovespa	RANK	1.000	RANK	1.000	RANK	0.5110
	GED	1.000	GED	1.000	GED	0.5137
	STD	1.000	LLSE	0.985	STD	0.5154
	LLSE	0.995	STD	0.983	NGQML	0.5162
	GQML	0.944	GQML	0.939	$RANK_2$	0.5167
	NGQML	0.921	NGQML	0.907	LLSE	0.5167
	$RANK_2$	0.846	$RANK_2$	0.808	GQML	0.5181

NOTE: NOTE: MCS is performed considering 0.7 confidence level. For each equity index, estimators in columns 2,4 and 6 are ordered according to p-values in columns 3 and 5, and the loss function in column 7. p-value$_{R,M}$ and p-value$_{Max,M}$ refer to the p-values of $T_{R,M}$ and $T_{Max,M}$ statistics, respectively. Loss is the asymmetric loss function suggested by González-Rivera et al. (2004). VaR forecasts are performed with a estimation window length of 1000 observations. Evaluation period between January 03, 2010, to July 02, 2020.

5 Concluding remarks

In this paper, we have evaluated the performance of a set of GARCH estimators with heavy-tailed errors distributions to forecast VaR for financial returns processes. These estimators were represented by the approaches of Andrews (2012) (RANK), Fan et al. (2014) (NGQML) and Preminger & Storti (2017) (LLSE) and a modified version combining RANK and NGQML methods ($RANK_2$). Furthermore, the efficiency of these procedures in tracking the VaR was compared with standard Gaussian (GQML), Student's-t(STD) and Generalised Errors Distribution (GED) estimators.

Heavy-tailed estimators are asymptotic efficient, improving GARCH parameters estimates in larger samples. The results of the simulated study show that for $\hat{\omega}$ and $\hat{\alpha}$ parameters, NGQML dominates the other estimators for series with 2500 observations, and with 1000 observations in some cases. As Fan et al. (2014) highlight, NGQMLE reduces the standard deviation of estimated parameters due to the scale parameter η that adjusts the innovations errors to the quasi-likelihood function. On the other hand, $RANK_2$ estimator demonstrates superior performance to obtain $\hat{\beta}$, which indicates that use of the rank technique before the NGQML when the noise distribution is unknown can help to increase the model accuracy, as highlighted by Andrews (2014).

Regarding VaR predictions performed over several simulated high persistent process, RANK_2 is the most efficient estimator when considered VaR at 1% coverage level, followed by RANK, which is outperformed by the formed by a small margin. These results are valid only for larger estimation windows, with heavy-tailed estimators performing poorly in small sample sizes.

Similar results were obtained when VaR forecasts are employed for equity indexes returns. Standard estimators perform relative well to predict the risk at 5% coverage level, including the GQML in some cases. Different form simulated process, in which RANK_2 exhibited higher accuracy at 1% VaR coverage level, RANK outperforms the others estimators to forecast the risk of financial returns, except for the Ibovespa case. Nevertheless, the superior predictive ability of RANK to cover the VaR is corroborated by the Model Confidence Set procedure of Hansen et al. (2011) with the estimator being selected as the first model generally.

To summarise, this article has shown that the use of heavy-tailed estimators to deal with the model specification error uncertainty helps to achieve higher accuracy of VaR coverages comparatively to standard methods. This result has empirical implication for risk management by reducing violations caused by extreme returns from VaR forecasts. As future research, we suggest taking the uncertainty of parameter estimation error into account to improve VaR predictions.

References

Andrews, B. (2012). Rank-based estimation for garch processes. *Econometric Theory*, *28*(5), 1037–1064.

Andrews, B. (2014). Comment. *Journal of Business & Economic Statistics*, *32*(2), 191-193.

Bernardi, M., & Catania, L. (2016). Comparison of value-at-risk models using the mcs approach. *Computational Statistics*, *31*(2), 579–608.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, *31*(3), 307–327.

Bollerslev, T., et al. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *Review of economics and statistics*, *69*(3), 542–547.

Bollerslev, T., & Wooldridge, J. M. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric reviews*, *11*(2), 143–172.

Boucher, C. M., Daníelsson, J., Kouontchou, P. S., & Maillet, B. B. (2014). Risk models-at-risk. *Journal of Banking & Finance*, *44*, 72–92.

Catania, L., & Bernardi, M. (2017). Mcs: Model confidence set procedure [Computer software manual]. (R package version 0.1.3)

Christoffersen. (1998). Evaluating interval forecasts. *International economic review*, 841–862.

Christoffersen, & Gonçalves, S. (2005). Estimation risk in financial risk management. *The Journal of Risk*, *7*(3), 1.

Engle, & Gonzalez-Rivera, G. (1991). Semiparametric arch models. *Journal of Business & Economic Statistics*, *9*(4), 345–359.

Engle, & Manganelli, S. (2004). Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, *22*(4), 367–381.

Fan, J., Qi, L., & Xiu, D. (2014). Quasi-maximum likelihood estimation of garch models with heavy-tailed likelihoods. *Journal of Business & Economic Statistics*, *32*(2), 178–191.

Fiorentini, G., & Sentana, E. (2014). Comment. *Journal of Business & Economic Statistics*, *32*(2), 193-198.

Francq, C., Lepage, G., & Zakoian, J.-M. (2011). Two-stage non gaussian qml estimation of garch models and testing the efficiency of the gaussian qmle. *Journal of Econometrics*, *165*(2), 246–257.

Francq, C., & Zakoian, J. (2014). Comment on" quasi-maximum likelihood estimation of garch models with heavy tailed likelihoods" by j. fan, l. qi et d. xiu. *Journal of Business & Economic Statistics*, *32*, 198–201.

González-Rivera, G., Lee, T.-H., & Mishra, S. (2004). Forecasting volatility: A reality check based on option pricing, utility function, value-at-risk, and predictive likelihood. *International Journal of forecasting*, *20*(4), 629–645.

Hansen, & Lunde. (2005). A forecast comparison of volatility models: does anything beat a garch (1, 1)? *Journal of applied econometrics*, *20*(7), 873–889.

Hansen, Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica*, *79*(2), 453–497.

Jorion, P. (2006). *Value at risk: The new benchmark for managing financial risk* (3rd ed.). McGraw-Hill.

Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models. *The J. of Derivatives*, *3*(2).

Lee, & Hansen, B. E. (1994). Asymptotic theory for the garch (1, 1) quasi-maximum likelihood estimator. *Econometric theory*, *10*(1), 29–52.

Lee, & Lee, S. (2009). Normal mixture quasi-maximum likelihood estimator for garch models. *Scandinavian Journal of Statistics*, *36*(1), 157–170.

Ling, S., & Zhu, K. (2014). Comment. *Journal of Business & Economic Statistics*, *32*(2), 202-203.

Lumsdaine, R. L. (1996). Consistency and asymptotic normality of the quasi-maximum likelihood estimator in igarch (1, 1) and covariance stationary garch (1, 1) models. *Econometrica: Journal of the Econometric Society*, 575–596.

Nieto, M. R., & Ruiz, E. (2016). Frontiers in var forecasting and backtesting. *International Journal of Forecasting*, *32*(2), 475–501.

Preminger, A., & Storti, G. (2017). Least-squares estimation of garch (1, 1) models with heavy-tailed errors. *The Econometrics Journal*, *20*(2), 221–258.