Forecasting Value-at-Risk under Different Distributional Assumptions

Manuela Braione [1,†] and Nicolas K. Scholtes [1,2,*,†]

[1] Center for Operations Research and Econometrics (CORE), Université catholique de Louvain,
     34 Voie du Romans Pays, B-1348 Louvain-la-Neuve, Belgium; manuela.braione@uclouvain.be
[2] Center for Research in Finance and Management (CeReFiM), Université de Namur, 62 Rue de Bruxelles,
     B-5000 Namur, Belgium
[*] Correspondence: nicolas.scholtes@unamur.be; Tel.: +32-10-474-321
[†] These authors contributed equally to this work.

Abstract: Financial asset returns are known to be conditionally heteroskedastic and generally non-normally distributed, fat-tailed and often skewed. These features must be taken into account to produce accurate forecasts of Value-at-Risk (VaR). We provide a comprehensive look at the problem by considering the impact that different distributional assumptions have on the accuracy of both univariate and multivariate GARCH models in out-of-sample VaR prediction. The set of analyzed distributions comprises the normal, Student, Multivariate Exponential Power and their corresponding skewed counterparts. The accuracy of the VaR forecasts is assessed by implementing standard statistical backtesting procedures used to rank the different specifications. The results show the importance of allowing for heavy-tails *and* skewness in the distributional assumption with the skew-Student outperforming the others across all tests and confidence levels.

Keywords: Value-at-Risk; forecast accuracy; distributions; backtesting

JEL classification: C01; C22; C52; C58

1. Introduction

Value-at-Risk (VaR) is a quantitative tool used to measure the maximum potential loss in value of a portfolio of assets over a defined period for a given probability. Specifically, VaR construction requires a quantile estimate of the far-left tail of the unconditional returns distribution. Though widely-used as a risk measure in the past, standard methods of VaR construction assuming iid-ness and normality have come under criticism due to their failure to incorporate three stylized facts of financial returns, namely (*i*) the presence of volatility clustering, indicated by high autocorrelation of absolute and squared returns, (*ii*) excess kurtosis (fat tails) and (*iii*) skewness in the density of the unconditional returns distribution.

The ability to account for volatility clustering is one of the key strengths of the ARCH modelling approach developed in Engle [1] and extended in Bollerslev [2]. By combining this approach with a non-normal conditional distribution assumption for the returns, several papers have shown that univariate GARCH models can produce reliable out-of-sample volatility forecasts. For example, Angelidis *et al.* [3] combine three GARCH specifications with the univariate skew-Student and skew-GED (Generalized Error) distributions to show that these are able to produce superior VaR forecasts compared to the normal. Specifically, they apply the exponential GARCH (EGARCH) model of Nelson [4] and the threshold ARCH (TARCH) model to five univariate returns series and find that while the choice of a skewed, heavy-tailed distribution significantly improves the

forecasting performance, the choice of the volatility model appears to be irrelevant. These findings are echoed in Mittnik and Paolella [5] who combine the Asymmetric Power ARCH (APARCH) model of Ding *et al.* [6] with an asymmetric generalised Student distribution. Within the univariate framework the most complete study of VaR prediction methods is provided by Kuester *et al.* [7] who compare fully parametric models with VaR constructed using historical simulation, extreme-value theory and quantile regression. Their results show that considerable improvement over normality is achieved when using innovation distributions that allow for skewness and fat tails.

Another salient feature of financial returns series is the fact that comovements between markets increase during periods of high volatility, as shown for example by Longin and Solnik [8] and Brooks *et al.* [9]. In light of this, a number of studies predict the VaR of a portfolio using multivariate models for the system of individual asset returns in order to achieve forecast improvements due to the use of more information. However, as mentioned in Bauwens *et al.* [10], in high dimensional frameworks these models can suffer from the "curse of dimensionality" problem thus being more computationally intensive.

In order to shed some light on this issue, several papers focus on direct comparison of the predictive performance of univariate and multivariate GARCH (MGARCH) models under various distributional assumptions with the aim of providing evidence in favour of one of the two approaches. Key studies in this literature include Giot and Laurent [11] and Santos *et al.* [12]. In the first, the authors compare the univariate APARCH model with a multivariate TVC-APARCH combined with the normal, Student and skew-Student distributions, showing that in both the univariate and multivariate settings the latter produced superior VaR forecasts. In the second, the authors study a number of univariate and multivariate volatility models with normal and Student distributions, finding that the multivariate models combined with the Student offer superior out-of-sample performances.

This paper builds on their approach by widening the set of distributions used to model the error term in both the univariate and multivariate frameworks while maintaining a generic specification for the conditional volatility. Specifically, we consider three symmetric distributions, *i.e.*, normal, Student and Multivariate Exponential Power (MEP), and their corresponding skewed counterparts obtained by applying the transformation of Bauwens and Laurent [13] [1]. By incorporating skewness in the corresponding symmetric densities by means of new parameters, we can explicitly analyse its marginal contribution as well as its joint effect with heavy-tails in the model forecasting performance.

As for the choice of the volatility models, within the MGARCH literature we employ the Rotated BEKK (RBEKK) model of Noureldin *et al.* [15] as it is easy to estimate using covariance targeting even for moderately large cross sections, while in the univariate setting GARCH(1,1) specifications are used to modeling the conditional portfolio variance. Our choice of a relatively simple volatility model allows us to limit our focus entirely on whether the chosen distribution contributes to adequately capture features of the return data. Within the univariate scenario we also consider the NCT-APARCH model proposed by Krause and Paolella [16], as the authors developed an extremely fast method for parameter estimation that is found to outperform highly competitive models and can be easily compared to ours [2].

Both univariate and multivariate models are estimated employing the aforementioned set of distributional assumptions and their accuracy in producing out-of-sample VaR forecasts is assessed by means of statistical backtesting procedures. The selected tests include the Unconditional Coverage (UC), Independence (IND) and Conditional Coverage (CC) tests of Christoffersen [17], the Duration-Based Test of Independence (DBI) of Dumitrescu *et al.* [18], the Time Until First Failure (TUFF) test of Kupiec [19] and the Dynamic Quantile (DQ) test of Engle and Manganelli [20].

---

[1] The interested reader is referred to Ley and Paindaveine [14] who provide a detailed overview of this approach.
[2] We thank the authors for kindly providing us their MATLAB codes.

The results of the tests are summarized using a grading scheme based on the number of acceptances of the null hypothesis which determines the distributional assumption providing the most accurate VaR forecasts.

Results from VaR backtesting show that in the multivariate setup the skew-Student clearly outperforms all other distributions. Moreover, its univariate version produces more accurate VaR forecasts than the NCT-GARCH and is able to compete with the NCT-APARCH which incorporates asymmetry into the conditional volatility specification. Overall, our results show that allowing for heavy-tails *and* skewness produces the most accurate VaR forecasts in both the univariate and multivariate setups. By comparison, specifications including only heavy tails underperform relative to their skewed counterparts with the difference being more pronounced at 5% VaR. Interestingly, the univariate skew-normal distribution produces Var forecasts comparable to the more heavily-parametrized skew-Student and skew-MEP. As regards comparing the performance of univariate and multivariate models in general, our results do not allow us to clearly advocate the use of one methodology over the other. However, given that the hierarchy of distributions according to VaR forecast accuracy is preserved under both frameworks, the univariate approach may be preferred due to its lower computational burden.

The paper is organized as follows: Section 2 reviews the GARCH modelling framework, the theoretical methodology used for constructing the skewed distributions and the Maximum Likelihood (ML) estimation of the models with the selected distributional assumptions. Section 3 introduces the empirical methodology, comprising the portfolio construction, VaR estimation and backtesting procedures. Section 4 provides estimation results and outcomes of the VaR tests and Section 5 concludes with some final remarks.

2. Theoretical Framework

This section illustrates the key points of our theoretical framework. Namely, we outline the alternative approaches to obtain portfolio VaR forecasts using univariate and multivariate models, we describe the procedure used to construct skewed distributions from the corresponding symmetric counterparts and finally we provide an overview of the set of employed distributions, comprising their likelihood derivation.

Two things are worth mentioning. First, we focus on the portfolio VaR for a long position, implying that the predictive power of the models is linked to their ability in modelling large negative returns. Second, we define the asset allocation scheme via an N-dimensional vector of equal weights $\boldsymbol{w}_t$, where N denotes the number of assets in the portfolio and $\boldsymbol{w}_t = (w_{1,t}, \ldots w_{N,t})$, with $w_i = 1/N$, $\sum_{i=1}^{N} w_i = 1$. As shown in DeMiguel *et al.* [21], Tu and Zhou [22], Brown *et al.* [23] and Fugazza *et al.* [24], this "naive" diversification rule is able to consistently outperform more sophisticated methods. Moreover, it has the advantage of not being affected by the specified target return as in the Markowitz framework, being only driven by the number of assets.

2.1. VaR Estimation

Let $\boldsymbol{y}_t = (y_{1,t}, ..., y_{N,t})'$ denote the N-dimensional discrete time vector of de-meaned daily returns at time t and $\boldsymbol{w}_t$ the vector of equal weights known at time $t - 1$. The portfolio return is obtained as $r_{p,t} = \boldsymbol{w}'_{t-1}\boldsymbol{y}_t$, and the portfolio VaR at time t is equal to

$$VaR_{t,\alpha} = \mu_{p,t} + \sigma_{p,t}q_\alpha, \tag{1}$$

where $\mu_{p,t}$ and $\sigma_{p,t}$ are respectively the portfolio mean and standard deviation and q_α is the *left* quantile of the assumed conditional distribution at α%. As usually done in practical applications, the considered VaR confidence levels are $\alpha = 5\%$ and 1%. Note that in the remainder of the paper we will set $\mu_{p,t} = 0$, thus considering a simplified analytical formula for the computation of the VaR which only accounts for the portfolio conditional variance. Alternative approaches, as done for example in

Bauwens *et al.* [25], fit an ARMA-type structure to the portfolio conditional mean or just assume it to be constant over time (see Santos *et al.* [12]). Ultimately, when present, the dynamic dependence in the conditional means of portfolio returns is known to be weak and quite difficult to predict, thus assuming a zero mean will have a negligible effects on the VaR forecasts.

The specification of the portfolio standard deviation $\sigma_{p,t}$ depends on whether we consider a univariate or multivariate approach, with the difference between the two occurring in the conditioning set used.

In the univariate case, the portfolio standard deviation is obtained as the standard deviation of the portfolio returns conditional on past portfolio returns, *i.e.*, as the square root of

$$\sigma_{p,t}^2 = E\left[r_{p,t}^2 | r_{p,1}, \ldots r_{p,t-1}\right], \tag{2}$$

where the conditional variance $\sigma_{p,t}^2$ is estimated using a model chosen from the univariate GARCH class. More precisely, we use a simple GARCH(1,1) specification with variance targeting, which is written as

$$\sigma_{p,t}^2 = \bar{\sigma} + a_1 r_{p,t-1}^2 + b_1 \sigma_{t-1}^2, \tag{3}$$

where $\bar{\sigma} = (1 - a_1 - b_1)\bar{\omega}$ and $\bar{\omega}$ equals the unconditional variance of returns. Covariance stationarity requires that $a_1 + b_1 < 1$, with $\{a_1, b_1\} \geq 0$ scalar parameters to be estimated.

In the multivariate setup, the conditioning set is made up of the entire vector of past returns:

$$\begin{aligned}
\sigma_{p,t}^2 &= E\left[r_{p,t}^2 | Y_1, \ldots Y_{t-1}\right] \\
&= w_{t-1}' H_{t|t-1} w_{t-1}
\end{aligned} \tag{4}$$

such that $H_{t|t-1} = E(H_t | \Im_{t-1})$ is the conditional covariance matrix of returns given the information set available at time $t-1$. In this case, a multivariate model for H_t needs to be specified. Given its computational ease in practical application, we implement the rotated BEKK (RBEKK) model of Noureldin *et al.* [15], as it can be estimated for relatively large cross sections even with reach dynamics. The basic idea underlying the model is to transform the original data by performing a rotation and then to fit to the rotated returns the popular BEKK specification of Engle and Kroner [26]. Consequently, the model can be easily estimated with covariance targeting, where the long-run covariance is given by the identity matrix. More precisely, let $\bar{H}$ define the unconditional covariance of y_t and let the latter be rewritten as $y_t = \bar{H}^{1/2}\epsilon_t$. Using the spectral decomposition of $\bar{H}$, *i.e.*, $\bar{H} = P\Lambda P'$, where P is a squared matrix of eigenvectors and the eigenvalue matrix Λ is diagonal with nonnegative elements, we can get the symmetric square root of $\bar{H}$ as $\bar{H}^{1/2} = P\Lambda^{1/2}P'$. The series of rotated returns is thus defined as

$$\epsilon_t = \bar{H}^{-1/2}y_t = P\Lambda^{-1/2}P'y_t,$$

with $Var(\epsilon_t) = I_N$, the N-dimensional identity matrix. Note that the conditional covariance of the rotated returns is $Var(\epsilon_t | \Im_{t-1}) = G_t$, given that $Var(y_t | \Im_{t-1}) = H_t = \bar{H}^{1/2}G_t\bar{H}^{1/2}$. Hence, any dynamic covariance model for H_t can also be applied to model the dynamics of G_t. Since the rotated returns are orthogonal, a suitable parameterization is represented by the scalar BEKK, which ensures the positive definiteness of the matrix G_t being a function of only two scalar parameters, under the assumption of covariance stationarity. Its dynamic equation is expressed as

$$G_t = (1 - a_2 - b_2)I_N + a_2\epsilon_{t-1}\epsilon_{t-1}' + b_2 G_{t-1}, \tag{5}$$

where, as in the univariate setup, $\{a_2, b_2\} \geq 0$ and $a_2 + b_2 < 1$.

2.2. Constructing Skew Densities

As far as financial applications are concerned, modelling and inference based on the normal distribution have often been proven to be of limited usefulness, as it is possible to gain statistical efficiency by allowing for more involved distributions featuring heavy tails and skewness. As a way to capture higher moments, the literature offers several alternatives. For example, the multivariate noncentral t distribution has fat tails and is skewed; however, the skewness is linked directly to the location parameter, making it somewhat inflexible. The lognormal distribution has also been used to model asset returns, but its skewness is a function of its mean and variance, not a separate parameter. Others, such as the generalized hyperbolic (GH) or *mixtures* of distributions have also been employed in financial applications, despite being more computationally demanding (see Barndorff-Nielsen [27] for an introduction and Paolella and Polak [28] for a recent application).

In this respect, Fernández and Steel [29] in the univariate case and Bauwens and Laurent [13] in the multivariate one developed a practical procedure for constructing skewed densities from their symmetric unimodal counterparts. These densities can be defined by introducing skewness in the corresponding symmetric densities by means of new parameters, such that the symmetric density results as a particular case. We build on their findings in order to enlarge the set of distributions to be used for VaR forecasting. In the following, we briefly recall the main steps of the procedure, being the discussion restricted to the more general multivariate framework for sake of space.

We begin by defining the notion of symmetry of a standardized density used hereafter. In the univariate case, symmetry corresponds to $g(x) = g(-x)$, where $g(x)$ is a unimodal probability density function with zero mean. In the multivariate case, we rely on the general notion of M-symmetry stated in Definition 1 of Bauwens and Laurent [13], which encompasses the class of spherically symmetric densities. These can be obtained as a special case of the general family of multivariate *elliptical* distributions, denoted as

$$g(x; \mu, \Sigma, \eta) \propto h((x-\mu)'\Sigma^{-1}(x-\mu), \eta), \tag{6}$$

where x is a random vector with an integrable, positive function $h(\cdot): \mathbb{R}^+ \to \mathbb{R}^+$, and η captures the shape parameter of the distribution, when present. The spherically symmetric set of distributions, comprising the standard normal, Student and MEP, are obtained by setting μ and Σ equal to zero and I_N, respectively.

The idea of introducing skewness into an M-symmetric standardized distribution revolves around scaling it differently for negative (positive) values by multiplying (dividing) by a positive constant. The value of this scaling parameter (hereafter referred to as ξ) determines whether the resulting distribution is skewed to the left ($0 < \xi < 1$) or to the right ($\xi > 1$).

Definition 1. Given a random vector $z = (z_1, \ldots, z_N)'$ with multivariate symmetric standardized distribution $g(z; \eta)$ following Equation (6), the standardized skewed density $f(z|\xi; \eta)$ with vector of asymmetry parameters $\xi = (\xi_1, \ldots \xi_N)'$, is obtained as:

$$f(z|\xi, \eta) = 2^N \left(\prod_{i=1}^{N} \frac{\xi_i}{1+\xi_i^2} \right) g(z^\star; \eta) \tag{7}$$

with

$$z^\star = (z_1^\star, \ldots, z_N^\star)' \tag{8}$$

$$z_i^\star = z_i \xi_i^{I_i} \tag{9}$$

and

$$I_i = \begin{cases} -1 & \text{if } z_i \geq 0 \\ 1 & \text{if } z_i < 0 \end{cases}. \tag{10}$$

The marginal r^{th}-order moment of the obtained skewed distribution can be computed directly from the standardized r^{th} moment of the symmetric density $g(\cdot)$. This is accomplished by applying the following transformation function:

$$E(z_i^{\star r}|\boldsymbol{\xi}) = M_{i,r} \frac{\xi_i^{r+1} + \frac{(-1)^r}{\xi_i^{r+1}}}{\xi_i + \frac{1}{\xi_i}}, \tag{11}$$

where the rth-order moment of the marginal $g_i(\cdot)$, truncated to the positive real values, is given by

$$M_{i,r} = \int_0^\infty 2u^r g_i(u) du. \tag{12}$$

Since only the first two moments are required in the transformation process, their analytical expression for $r = 1, 2$ in Equation (11) is reported below:

$$m_i = E(z_i^\star|\xi_i) = M_{i,1}\left(\xi_i - \frac{1}{\xi_i}\right) \tag{13}$$

$$s_i^2 = \text{Var}(z_i^\star|\xi_i) = \left(M_{i,2} - M_{i,1}^2\right)\left(\xi_i^2 + \frac{1}{\xi_i^2}\right) + 2M_{i,1}^2 - M_{i,2}. \tag{14}$$

As shown in Appendix A, s_i^2 can also be expressed directly as a function of m_i^2. Note that the resulting skewed distribution, $f(z|\boldsymbol{\xi},\eta)$ from Definition 1, is not centered at 0 and the variance is a function of $\boldsymbol{\xi}$ (and, where is the case, of the shape parameter η). Given that the elements of $z^\star$ are uncorrelated (since those of x are uncorrelated by assumption), standardization of $z^\star$ is achieved by the following transformation:

$$z = (z^\star - m)./s, \tag{15}$$

where $m = (m_1, ..., m_N)$ and $s = (s_1, ..., s_N)$ are the vectors of unconditional means and standard deviations of $z^\star$ computed in Equations (13) and (14) respectively and "./" denotes element-by-element division. Consequently, the standardized form of Definition 1 requires replacing Equation (9) with the following one:

$$z_i^\star = (s_i z_i + m_i)\xi_i^{I_i}, \tag{16}$$

where

$$I_i = \begin{cases} -1 & \text{if } z_i \geq -\frac{m_i}{s_i} \\ 1 & \text{if } z_i < -\frac{m_i}{s_i} \end{cases} \tag{17}$$

2.3. Distributions

As already mentioned, three symmetric and three asymmetric distributions are considered in the univariate and multivariate framework. Again, for sake of brevity, we only report the multivariate log-likelihood functions as the univariate can be obtained as special cases [3]. The algebraic derivations of the formulas of the moments can be found in Appendix A.2. In all cases, estimation of the parameters is performed in one step by Maximum Likelihood (ML). Namely, the log-likelihood function for T observations is expressed as

$$\ell_T(\boldsymbol{\psi}) = \sum_{t=1}^{T} \log f(\cdot|\boldsymbol{\psi}, \mathcal{F}_{t-1}), \tag{18}$$

---

[3]   In Appendix D we provide a brief overwiew of the noncentral t distribution and its likelihood derivation.

where $\boldsymbol{\psi}$ is the finite-dimensional vector of model parameters and $f(\cdot|\boldsymbol{\psi}, \mathcal{F}_{t-1})$ denotes the assumed conditional density function of the portfolio returns, in the univariate case, or of the asset return vector in the multivariate one.

Multivariate normal distribution This is the most commonly employed distribution in the literature as it is uniquely identified by its conditional first and second moments, which renders ML estimation much simpler from a computational point of view. In addition, given that the score of the normal log-likelihood function has the martingale difference property when the first two conditional moments are correctly specified, the Quasi Maximum Likelihood (QML) estimates are still consistent and asymptotically normal even if the true DGP is not normally-distributed Bollerslev and Wooldridge [30]. The log-likelihood function for T observations is expressed as follows

$$\ell_T(\boldsymbol{\psi}) = -\frac{1}{2}\sum_{t=1}^{T}\left[N\log(2\pi) + \log|H_t| + \boldsymbol{y}_t'H_t^{-1}\boldsymbol{y}_t\right]. \tag{19}$$

As the multivariate normal does not incorporate extra parameters, here we have $\boldsymbol{\psi} = (a_2, b_2)'$ [4].

Multivariate Student distribution The Student distribution is a symmetric and bell-shaped distribution, with heavier tails than the normal. Under the multivariate Student assumption, the log-likelihood function is obtained as

$$\begin{aligned}\ell_T(\boldsymbol{\psi}) &=& -\frac{1}{2}\sum_{t=1}^{T}\left[\log|H_t| + (N+\nu)\log\left(1 + \frac{\boldsymbol{y}_t'H_t^{-1}\boldsymbol{y}_t}{\nu-2}\right)\right] \\ &+& T\left[\log\Gamma\left(\frac{\nu+N}{2}\right) - \log\Gamma\left(\frac{\nu}{2}\right) - \frac{N}{2}\log(\pi) - \frac{N}{2}\log(\nu-2)\right],\end{aligned} \tag{20}$$

where $\Gamma(\nu) = \int_0^{\infty} e^{-z}z^{\nu-1}dz$ denotes the Gamma function and $\nu > 2$ is the degree of freedom parameter representing the thickness of the distribution tails. As ν increases, the distribution converges to the multivariate normal. The vector of parameter to be estimated is $\boldsymbol{\psi} = (a_2, b_2, \nu, \xi)'$.

Multivariate Exponential Power (MEP) distribution This distribution belongs to the Kotz family of distributions (a particular class of symmetric and elliptical distributions discussed extensively in Fang *et al.* [31]) and is known to have several equivalent definitions in the literature. It can also include both the normal and the Laplace as special cases, as a function of the value of the non-normality parameter β dictating the tail-behaviour of the distribution. Given its simple implementation, in this paper we consider the pdf given in Solaro [32], which gives rise to the following log-likelihood function:

$$\begin{aligned}\ell_T(\boldsymbol{\psi}) &=& -\frac{1}{2}\sum_{t=1}^{T}\left[\log|H_t| + \left(\boldsymbol{y}_t'H_t^{-1}\boldsymbol{y}_t\right)^{\frac{\beta}{2}}\right] \\ &+& T\left[\log(N) + \log\Gamma\left(\frac{N}{2}\right) - \frac{N}{2}\log(\pi) - \log\Gamma\left(1 + \frac{N}{\beta}\right) - \left(1 + \frac{N}{\beta}\right)\log(2)\right], \tag{21}\end{aligned}$$

where $\boldsymbol{\psi} = (a_2, b_2, \beta)'$. Note the restriction $\beta > 0$. When $\beta = 2$, the distribution reduces to the multivariate normal, while for $\beta = 1$ it corresponds to the multivariate Laplace. Whenever $\beta < 2$ (> 2), the distribution exhibits thicker (thinner) tails than the normal.

---

[4] In the univariate case it corresponds to $\boldsymbol{\psi} = (a_1, b_1)'$.

Multivariate skew-normal distribution The multivariate skew-normal is the first non-symmetric distribution we consider herein; it accounts for the skewness of the return distribution without taking into account its kurtosis (as it does not involve a tail parameter). Applying Definition 1 we derive the skew-normal density function, with corresponding log-likelihood function equal to

$$
\begin{aligned}
\ell_T(\boldsymbol{\psi}) \;=\; & -\frac{1}{2} \sum_{t=1}^{T} \left[\log |H_t| + \sum_{i=1}^{N} \left(s_i \sum_{j=1}^{N} p_{ij,t} y_{j,t} + m_i \right)^2 \xi_i^{2I_i} \right] \\
& + \; T \left[\sum_{i=1}^{N} (\log \xi_i + \log s_i) - \log(1 + \xi_i^2) \right] + \frac{TN}{2} \left[\log(2) - \log(\pi) \right],
\end{aligned}
\tag{22}
$$

where $p_{ij,t}$ corresponds to the jth element of the ith row of the inverse cholesky factor of the matrix H_t (the full derivation is provided in Appendix A.1), ξ_i represents the asymmetry parameter of each marginal and I_i is the indicator function defined in Equation (17). The vector of parameters of interest is $\boldsymbol{\psi} = (a_2, b_2, \boldsymbol{\xi})'$.

Multivariate skew-Student distribution Applying the same procedure as for the skew-normal, the log-likelihood function of the skew-Student distribution is given by the following expression:

$$
\begin{aligned}
\ell_T(\boldsymbol{\psi}) \;=\; & -\frac{1}{2} \sum_{t=1}^{T} \left[\log |H_t| + (\nu + N) \log \left(1 + \frac{\sum_{i=1}^{N} \left(s_i \sum_{j=1}^{N} p_{ij,t} y_{j,t} + m_i \right)^2 \xi_i^{2I_i}}{\nu - 2} \right) \right] \\
& + \; T \left[N \log(2) - \frac{N}{2} \log(\pi) + \log \Gamma \left(\frac{\nu + N}{2} \right) - \log \Gamma \left(\frac{\nu}{2} \right) - \frac{N}{2} \log(\nu - 2) \right] \\
& + \; T \left[\sum_{i=1}^{N} (\log \xi_i + \log s_i) - \log(1 + \xi_i^2) \right],
\end{aligned}
\tag{23}
$$

where the parameter ν dictates the thickness of the tails and ξ_i is again the asymmetry parameter of each marginal. Notice that the univariate means and standard deviations are functions of ξ_i and ν and need not be estimated. Consequently, the skew-Student parametrization requires $N + 1$ parameters to be estimated in addition to those stemming from the RBEKK specification, *i.e.*, $\boldsymbol{\psi} = (a_2, b_2, \nu, \boldsymbol{\xi})'$.

Multivariate skew-MEP distribution The log-likelihood function to be maximized is given by

$$
\begin{aligned}
\ell_T(\boldsymbol{\psi}) \;=\; & -\frac{1}{2} \sum_{t=1}^{T} \left[\log |H_t| + \left(\sum_{i=1}^{N} \left(s_i \sum_{j=1}^{N} p_{ij,t} y_{j,t} + m_i \right)^2 \xi_i^{2I_i} \right)^{\frac{\beta}{2}} \right] \\
& + \; T \left[\sum_{i=1}^{N} (\log \xi_i + \log s_i) - \log(1 + \xi_i^2) \right] \\
& + \; T \left[\log(N) + \log \Gamma \left(\frac{N}{2} \right) - \frac{N}{2} \log(\pi) \right. \\
& \left. - \; \log \Gamma \left(1 + \frac{N}{\beta} \right) - \left(1 + \frac{N}{\beta} \right) \log(2) \right],
\end{aligned}
\tag{24}
$$

where β is a parameter determining the tail-thickness of the density function, as in the symmetric case. Considering the skewness parameters implies the parameter vector $\boldsymbol{\psi} = (a_2, b_2, \beta, \boldsymbol{\xi})'$.

3. Empirical Application

3.1. Data and Forecasting Scheme

Our dataset (used in the paper of Noureldin *et al.* [33]) [5] comprises daily open-to-close returns of 10 stocks from the Dow Jones Industrial Average: Bank of America (BAC), JP Morgan (JPM), International Business Machines (IBM), Microsoft (MSFT), Exxon Mobil (XOM), Alcoa (AA), American Express (AXP), Du Pont (DD), General Electric (GE) and Coca Cola (KO). Each univariate vector of returns is calculated as $y_t = 100 \times (\log p_t - \log p_{t-1})$ and covers a period of 2200 days, from February 2001 to November 2009. Univariate descriptive statistics over the period of interest are provided in Table 1.

Table 1. Univariate descriptive statistics.

Stock	Mean	Std.dev.	Skewness	Kurtosis	KS Test	JB Test
Estimation sample: 1 February 2001 to 23 January 2007 (1500 observations)						
BAC	0.09	1.09	−0.18	7.45	0.00	0.00
JPM	0.00	1.68	0.90	31.02	0.00	0.00
IBM	−0.04	1.24	0.01	5.96	0.01	0.00
MSFT	−0.01	1.37	0.37	6.01	0.00	0.00
XOM	−0.01	1.13	0.05	8.27	0.82	0.00
AA	0.01	1.59	0.14	4.74	0.00	0.00
AXP	−0.02	1.44	0.33	7.73	0.00	0.00
DD	0.02	1.21	0.37	6.76	0.21	0.00
GE	−0.01	1.34	0.13	7.90	0.02	0.00
KO	0.01	0.99	0.16	5.53	0.00	0.00
Forecasting sample: 24 January 2007 to 30 October 2009 (700 observations)						
BAC	−0.18	3.95	0.37	9.36	0.00	0.00
JPM	0.01	3.06	0.36	8.53	0.00	0.00
IBM	0.08	1.45	−0.02	6.31	0.00	0.00
MSFT	0.02	1.60	0.08	5.90	0.00	0.00
XOM	0.03	1.61	−0.39	11.31	0.00	0.00
AA	−0.04	2.93	−0.83	7.50	0.00	0.00
AXP	0.04	3.06	0.22	6.96	0.00	0.00
DD	−0.04	1.89	−0.12	5.70	0.00	0.00
GE	0.02	2.17	0.21	8.96	0.00	0.00
KO	−0.03	1.22	0.07	7.68	0.06	0.00
Full sample: 1 February 2001 to 30 October 2009 (2200 observations)						
BAC	0.01	2.40	0.33	21.72	0.00	0.00
JPM	0.00	2.21	0.57	16.90	0.00	0.00
IBM	0.00	1.31	0.02	6.24	0.02	0.00
MSFT	0.00	1.45	0.25	6.08	0.00	0.00
XOM	0.00	1.30	−0.20	11.56	0.04	0.00
AA	0.00	2.11	−0.69	9.95	0.00	0.00
AXP	0.00	2.09	0.32	11.23	0.00	0.00
DD	0.00	1.46	0.03	7.25	0.00	0.00
GE	0.00	1.65	0.22	10.85	0.00	0.00
KO	0.00	1.07	0.11	6.89	0.00	0.00

Descriptive statistics of the stock return time series used in the empirical application. The three panels report the statistics for the in-sample period, the out-of-sample period and the full sample period, respectively. "KS test" and "JB test" denotes the Kolmogorov-Smirnov test and Jarque Bera test, with corresponding *p*-values in column.

Across the three panels, the values of skewness and kurtosis show that the assets are far from being unconditionally normally distributed, thus supporting the conjecture that more flexible distributional assumptions can be conducive to enhanced model performance.

To this extent, one-step ahead forecasts of the conditional portfolio variance (in the univariate case) and of the conditional covariance matrix of returns (in the multivariate one) are recursively obtained as:

$$\hat{\sigma}^2_{p,t+1} = E(\sigma^2_{p,t+1}|\Im_t), \tag{25}$$

$$\hat{H}_{t+1} = E(H_{t+1}|\Im_t), \tag{26}$$

where $\Im_t$ is the information set at time t and $\sigma^2_{p,t}, H_t$ are defined as in Equations (2) and (5), respectively. Using a rolling-fixed-window scheme, the parameters are estimated over a window length of 1500 observations and used to predict the conditional variance process for the following 20 days. Each time the window is shifted forward by 20 observations and the parameters are re-estimated over the new period in order to compute the next set of forecasts. We iterate this process until the end of the dataset for a total of 35 parameter estimates and 700 one-step ahead forecasts. Table A1 in Appendix B reports the complete list of windows and forecast horizons along with their corresponding calendar dates.

For each model, the portfolio VaR forecast at $\alpha\%$ confidence level is then obtained as

$$VaR_{t+1,\alpha} = \hat{\sigma}_{p,t+1} q_\alpha. \tag{27}$$

For the symmetric distributions in our analysis (normal, Student and MEP), one can easily compute the long VaR of the portfolio by applying Equation (27) and the inverse of each CDF at $\alpha\%$. However, for the non-symmetric distributions this is not straightforward. In order to bypass this complication, for each non-symmetric distribution we apply a simple Monte-Carlo simulation approach, as widely used in VaR computations. Namely, we draw $10,000$ random vectors (numbers) from each symmetric multivariate (univariate) standardized distribution z_t and then we use the estimated skewness parameters to construct the corresponding skewed distribution $z_t^\star$. By assuming $r_j = \hat{H}^{1/2}_{t|t-1} z_j^\star$ ($r_j = \hat{\sigma}_{t|t-1} z_j^\star$) as the true DGP, we obtain a set of 10,000 simulated returns over the period of interest. Finally, the simulated return distribution is used to derive the 5% and 1% quantiles for the one-step-ahead VaR.

3.2. Testing the Accuracy of VaR Forecasts

The models accuracy in predicting VaR is assessed using multiple statistical backtesting methods. A common starting point for this procedure is the so-called hit function, or indicator function, which is equal to

$$I_t(\alpha) = \begin{cases} 1 \text{ if } r_t \le VaR(\alpha) \\ 0 \text{ if } r_t > VaR(\alpha) \end{cases} \tag{28}$$

i.e., it takes the value one if the ex-post portfolio loss exceeds the VaR predicted at time $t-1$ and the value zero otherwise. According to Christoffersen [34], in order to be accurate, the hit sequence has to satisfy the two properties of correct failure rate and independence of exceptions. The former implies that the probability of realizing a VaR violation should be equal to $\alpha \times 100\%$, while the latter further requires the violations to be independent of each other. These properties can be combined together into one single statement assessing that the hit function has to be an i.i.d. Bernoulli random variable with probability p, *i.e.*, $I_t(p) \overset{i.i.d}{\sim} B(p)$.

This represents the key foundation to many of the backtesting procedures developed in recent years and particularly to the accuracy tests being used in this paper. We focus on tests included in the following three categories:

- Evaluation of the Frequency of Violations
- Evaluation of the Independence of Violations
- Evaluation of the Duration between Violations.

Their properties are briefly described below.

Frequency of Violations The first way of testing the VaR accuracy is to test the number or the frequency of margin exceedances. A test designed to this aim is the Kupiec test (Kupiec [35]), also known as the Unconditional Coverage (UC) test. Its null hypothesis is simply that the percentage of violated VaR forecasts or failure rate p is consistent with the given confidence level α, *i.e.*, $H_0 : p = \alpha$.

Denoting by F the length of the forecasting period and with v the number of violations occurred throughout this period, the log-likelihood ratio test statistic is defined as

$$UC = -2 \left(\ln \left(\frac{p^v (1-p)^{F-v}}{\hat{p}^v (1-\hat{p})^{F-v}} \right) \right), \tag{29}$$

where $\hat{p} = v/F$ is the maximum likelihood estimator under the alternative hypothesis. This ratio test statistic is asymptotically $\chi^2(1)$ distributed and the null hypothesis is rejected if the critical value at the $\alpha\%$ confidence level is exceeded.

A similar useful test is the TUFF (Time Until First Failure) test (Kupiec [35]). Under the null, the probability of an exception is equal to the inverse probability of the VaR confidence level, namely $H_0 : p = \hat{p} = 1/v$. Its basic assumptions are similar to those of the Kupiec test and the t-statistic under the null is obtained as

$$TUFF = -2 \left(\ln \left(\frac{p(1-p)^{v-1}}{\frac{1}{v} \left(1 - \frac{1}{v} \right)^{(v-1)}} \right) \right). \tag{30}$$

The TUFF statistic is also asymptotically $\chi^2(1)$ distributed.

Independence of Violations A limitation of the Kupiec test is that it is only concerned with the coverage of the VaR estimates without accounting for any clustering of the violations. This aspect is crucial for VaR practitioners, as large losses occurring in rapid succession are more likely to lead to disastrous events than individual exceptions.

The Independence test (IND) of Christoffersen [34] uses the same likelihood ratio framework as the previous tests but is designed to explicitly detect clustering in the VaR violations. Under the null hypothesis of independence, the IND test assumes that the probability of an exceedance on a given day t is not influenced by what happened the day before. Formally, $H_0 : p_{10} = p_{11}$, where p_{ij} denotes the probability of an i event on day $t-1$ being followed by a j event on day t.

The relevant IND test statistic can be derived as

$$IND = -2 \left(\ln \left(\frac{\hat{p}^v (1-\hat{p})^{F-v}}{\hat{p}_{11}^{v_{11}} (1-\hat{p}_{11})^{v_{01}} \hat{p}_{10}^{v_{10}} (1-\hat{p}_{10})^{v_{00}}} \right) \right), \tag{31}$$

where v_{ij} is the number of violations with value i at time $t-1$ followed by j at time t. Under the null, the IND statistic is also asymptotically distributed as a $\chi^2(1)$ random variable.

Although the aforementioned test has received support in the literature, Christoffersen [34] noted that it was not complete on its own. For this reason, he proposed a joint test, the Conditional Coverage (CC) test, which combines the properties of both UC and IND tests.

Formally, the CC ratio statistic can be proven to be the sum of the UC and the IND statistics:

$$
\begin{aligned}
CC & = -2(ln(L_0^{UC)} - ln(L_1^{IND})) \\
& = -2(ln(L_0^{UC}) - ln(L_1^{UC}) + ln(L_1^{UC}) - ln(L_1^{IND})) \\
& = -2(ln(L_0^{UC}) - ln(L_1^{UC}) + ln(L_0^{IND}) - ln(L_1^{IND})) \\
& = -2\underbrace{(ln(L_0^{UC}) - ln(L_1^{UC}))}_{UC} -2\underbrace{(ln(L_0^{IND}) - ln(L_1^{IND}))}_{IND},
\end{aligned}
\tag{32}
$$

where we added and subtracted the quantity $\ln(L_1)^{UC}$ and substituted $\ln(L_0)^{IND}$ for $\ln(L_1)^{UC}$. CC is also χ^2 distributed, but with two degrees of freedom since there are two separate statistics in the test. According to Campbell [36], in some cases it is possible that a VaR model passes the joint test while still failing either the independence test or the unconditional coverage test. Thus it is advisable to run them separately even when the joint test yields a positive result.

A second test belonging to this class is the Regression-based test of Engle and Manganelli [37], also known as Dynamic Quantile (DQ) test. Instead of directly considering the hit sequence, the test is based on its associated quantile process $H_t(\alpha) = I_t(\alpha) - \alpha$ which assumes the following values:

$$
H_t(\alpha) = \begin{cases} 1 - \alpha \text{ if } I_t = 1 \\ -\alpha \quad \text{ if } I_t = 0 \end{cases}.
\tag{33}
$$

The idea of this approach is to regress current violations on past violations in order to test for different restrictions on the parameters of the model. That is, we estimate the linear regression model $H_t(\alpha) = \delta + \sum_{k=1}^{K} \beta_k H_{t-k}(\alpha) + \epsilon_t$ and then we test the joint hypothesis $H_0(DQ_{cc}) : \delta = \beta_1 = ... = \beta_K = 0$. This assumption coincides with the null of Christoffersen's CC test. It is also possible to split the test and separately test the independence hypothesis and the unconditional coverage hypothesis, respectively as $H_0(DQ_{ind}) : \beta_1 = ... = \beta_K = 0$ and $H_0(DQ_{uc}) : \delta = 0$. $(DQ_{cc}), (DQ_{ind})$ and (DQ_{uc}) are asymptotically χ^2 distributed with respectively $\{K+1\}$, K and one degrees of freedom.

Duration between Violations One of the drawbacks of Christoffersen's CC test is that it is not capable of capturing dependence in all forms, since it only considers the dependence of observations between two successive days. To address this, Christofferson and Pelletier [38] introduced the Duration-Based test of independence (DBI), which is an improved test for both independence and coverage. Its basic intuition is that if exceptions are completely independent of each other, then the upcoming VaR violations should be independent of the time that has elapsed since the occurrence of the last exceedance (Campbell [36]). The duration (in days) between two exceptions is defined via the no-hit-duration $D_i = t_i - t_{i-1}$, where t_i is the day of ith violation.

A correctly specified model should have an expected conditional duration of $1/p$ days and the no-hit duration should have no memory. The authors construct the ratio statistic considering different distributions for the null and the alternative hypotheses, namely the exponential, since it is the only memory-free (continuous) random distribution, and the Weibull, which allows for duration dependence. The likelihood ratio statistic is derived as

$$
DBI = -2 \left(\ln \left(\frac{L_0}{L_1} \right) \right) = -2 \left(\ln \left(\frac{p \exp\{-pD\}}{a^b b D^{b-1} \exp\{-(aD)^b\}} \right) \right),
\tag{34}
$$

which has a χ^2 distribution with one degree of freedom.

Under the null hypothesis of independent violations, $b = 1$ and a is estimated via numerical maximization of $ln(L_1)$. Whenever $b < 1$, the Weibull function has a decreasing path which

corresponds to an excessive number of very long durations (very calm period) while $b > 1$ corresponds to an excessive number of very short durations, namely very volatile periods.

4. Results

4.1. Parameter Estimates

Before turning to the out of sample analysis, it is worth first looking at some parameter estimates obtained by fitting the models on the data. Table 2 reports full sample estimation results over the period January 2001 to February 2009 for a total of 2200 daily observations.

Table 2. Full sample parameter estimates.

	Normal	Student	MEP	Skew-Normal	Skew-Student	Skew-MEP	NCT-APARCH	NCT-GARCH
				Univariate models				
$\bar{\sigma}$	0.005	0.004	0.002	0.005	0.004	0.010	0.010	0.010
a_1	0.07 (0.016)	0.06 (0.013)	0.07 (0.016)	0.07 (0.029)	0.06 (0.013)	0.07 (0.015)	0.05	0.05
b_1	0.92 (0.017)	0.93 (0.013)	0.92 (0.017)	0.92 (0.031)	0.93 (0.014)	0.93 (0.017)	0.90	0.90
ν		8.44 (1.358)			8.42 (1.366)		7.20	5.20
γ							-0.360	-0.120
β			1.875 (0.041)			1.873 (0.038)		
$\bar{\xi}$				0.938 (0.026)	0.935 (0.023)	0.946 (0.02)		
LogLik	-2971	-2946	-2965	-2968	-2943	-2963	-2944	-2963
AIC	2.703	2.681	2.698	2.701	2.679	2.697	2.680	2.696
				Multivariate models				
a_2	0.021 (0.001)	0.015 (0.001)	0.022 (0.001)	0.032 (0.002)	0.015 (0.001)	0.026 (0.001)		
b_2	0.976 (0.001)	0.983 (0.001)	0.976 (0.001)	0.966 (0.002)	0.983 (0.001)	0.972 (0.002)		
ν		8.42 (0.370)			8.30 (0.37)			
$\bar{\xi}$				1.016 (0.035)	1.018 (0.022)	1.018 (0.034)		
β			1.919 (0.013)			2.036 (0.031)		
LogLik	$-32,154$	$-31,352$	$-32,193$	$-32,121$	$-31,330$	$-32,067$		
AIC	29.232	28.504	29.278	29.211	28.449	29.154		

Note: The table reports test statistics and robust standard errors obtained from full sample parameter estimation, for $T = 2200$. Note that $\bar{\xi}$ denotes the value of the skewness parameters averaged across univariate series with Mean Asymptotic Square Errors (MASE) reported in brackets. For both the NCT-GARCH and NCT-APARCH models $\psi = (\mu, \nu, \gamma)'$, see Appendix D. The AIC is rescaled by T.

A common feature emerging from both univariate and multivariate panels is that the use of skewed distribution assumptions seems to be justified, as all asymmetric coefficients are significant at standard levels. Moreover, the direct comparison of models fit via the Akaike Information Criteria (AIC) highlights that the models incorporating skewed distributions consistently outperform their symmetric counterparts, with the skew-Student achieving the best fit. In both cases the Student has the lowest AIC among the symmetric densities, while the normal possesses the highest.

In the univariate setting, the NCT-APARCH model substantially improves over the NCT-GARCH due to the introduction of skewness into the volatility model, but it performs slightly worse than its closest competitor, the skew-Student GARCH model. This performance is even more impressive considering that the AIC penalizes the NCT-based models for only three parameters, given that remaining ones are fixed *prior to* maximization of the likelihood function according to the fast procedure of Krause and Paolella [16] (see Appendix D).

The evolution of the parameters across re-estimations $\tau = 1, ..., 35$ allows us to discern their sensitivity to the financial crisis by comparing their dynamics against the key dates outlined in Table A1 in Appendix B. From Figure 1, which looks at the tail parameter estimates for the MEP and Student distributions along with their skewed counterparts, a number of commonalities are immediately apparent. Panel (a) reveals a small increase followed by a sharp drop in β between $\tau = 18$ and 22 for both the univariate and multivariate MEP, indicating a thickening of the tails.

Since the time frame corresponds to the windows including the events leading up to the collapse of Lehman Brothers on 15 September 2008, we interpret our results as β adjusting to incorporate the extreme negative events associated to this period.
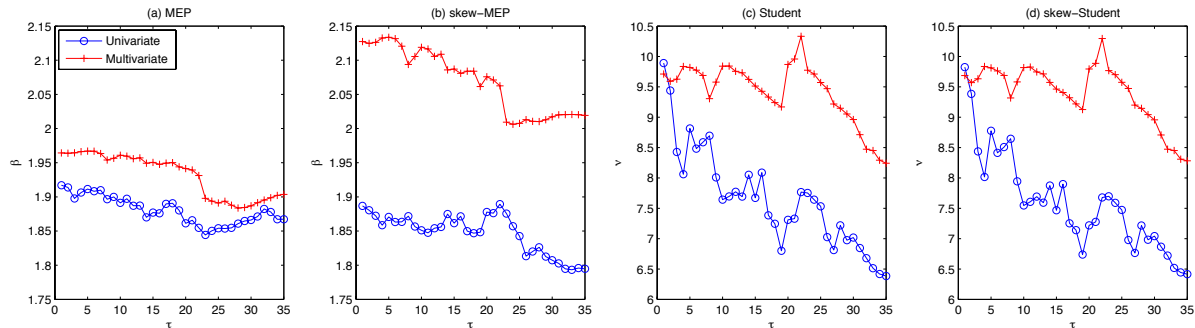

Figure 1. Tail parameter evolution across re-estimations.

Panel (b) exhibits the same dynamics as before albeit at a reduced magnitude. This occurs due to the introduction of skewness which is able to capture some of the negative returns associated to the crisis, thereby reducing the need for such a large increase in tail thickness. Evidence of such *compensating effects* between skewness and heavy-tails is given in Figure 2 which provides the evolution of the skew-parameter ξ across re-estimations (given as an average over the assets in the multivariate case for ease of exposition). Specifically, the dynamics of ξ for the univariate skew-MEP in Panel (a) appear to track the tail dynamics: as β decreases on average between τ equal 1 and 18, ξ exhibits a gradual increase, indicating a reduction of negative skewness. Moreover, the aforementioned drop is accompanied by ξ attaining its maximum value.
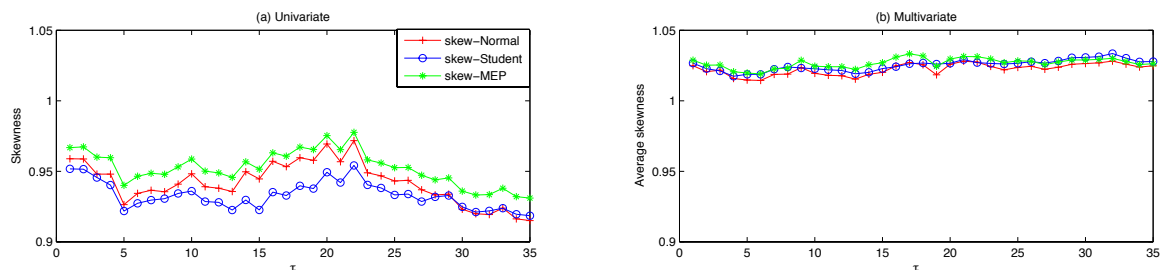

Figure 2. Skew parameters evolution across re-estimations.

Turning to the Student and skew-Student distributions, the dynamics are somewhat different along a number of dimensions. First, there is now a marked difference between the univariate and multivariate cases. In the former, the dynamics are similar to the MEP and skew-MEP results wherein the tails thicken at around $\tau = 18$ as evidenced by the sharp drop in ν. This is then followed by a short-lived increase after which the tails thicken gradually over the remaining estimation period. The reasoning for this, whereby the tails thicken to accommodate the negative returns associated to the crisis, is the same as the MEP-case.

Unlike the MEP and skew-MEP cases, where the multivariate specification exhibits thinner tails than the univariate but both exhibit congruent dynamics, the Student and skew-Student exhibit fewer similarities between specifications. In contrast, moving from the Student to skew-Student setting reveals almost identical tail parameter dynamics. Combined with the positive skew exhibited in

Panel (b) of Figure 2, this indicates that in this case tail thickness does not adjust to incorporate the addition of skewness. The effect of these features on VaR forecast accuracy will be seen in Section 4.2.

4.2. VaR Backtesting Results

In the multivariate setting, the out-of-sample covariance matrix predictions are used to construct equally-weighted portfolios for the computation of the one-step-ahead VaR. Table 3 compares portfolios standard deviation obtained using the univariate and multivariate approaches over both the in- and out-of-sample periods.

Table 3. Portfolios descriptive statistics.

	Univariate Models							
	Normal	**Student**	**MEP**	**Skew-Normal**	**Skew-Student**	**Skew-MEP**	**NCT-APARCH**	**NCT-GARCH**
	In-sample: 1 February 2001 to 23 January 2007 (1500 observations)							
$\bar{\sigma}_p$	0.8716	0.8722	0.8662	0.8709	0.8718	0.8780	0.8222	0.8720
$\min\{\sigma_p\}$	0.4159	0.4155	0.3899	0.4134	0.4141	0.4457	0.3727	0.3852
$\max\{\sigma_p\}$	2.7805	2.6827	2.8234	2.8147	2.6965	2.7095	2.5242	2.5826
	Forecasting sample: 24 January 2007 to 30 October 2009 (700 observations)							
$\bar{\sigma}_p$	1.4659	1.4601	1.4800	1.4656	1.4550	1.4573	1.4569	1.4644
$\min\{\sigma_p\}$	0.4398	0.4448	0.4139	0.4334	0.4427	0.4718	0.3745	0.3899
$\max\{\sigma_p\}$	3.9589	3.8967	3.9743	3.9656	3.8033	3.9281	4.2310	3.1567
	Multivariate Models							
	Normal	**Student**	**MEP**	**Skew-Normal**	**Skew-Student**	**Skew-MEP**		
	In-sample: 1 February 2001 to 23 January 2007 (1500 observations)							
$\bar{\sigma}_p$	0.9021	0.9115	0.8983	0.9018	0.9115	0.8924		
$\min\{\sigma_p\}$	0.5226	0.5474	0.5077	0.5212	0.5478	0.4840		
$\max\{\sigma_p\}$	1.9067	1.7923	1.9391	1.9097	1.7935	1.9972		
	Forecasting sample: January 24, 2007 to October 30, 2009 (700 observations)							
$\bar{\sigma}_p$	1.4721	1.4646	1.4875	1.4740	1.4549	1.4691		
$\min\{\sigma_p\}$	0.5226	0.5474	0.5105	0.5233	0.5479	0.4908		
$\max\{\sigma_p\}$	3.1733	3.0657	3.2108	3.1728	3.0674	3.2683		

Note: The table reports average, minimum and maximum value of portfolio standard deviation over the in- and out-of-sample periods.

As already noted, the financial crisis features heavily in the summary statistics. Since this period is included in the forecasting sample (starting from observation 1921 according to Table A1), we notice a sharp increase in the portfolio standard deviation of all the models (see also the figures reported in Appendix C). On a general basis, heavy-tailed or skewed distributions do not lead to remarkable gains over the in-sample period, as all the models report very similar values on average. This pattern is partly reversed in the forecasting period: while the normal and skew-normal remain very close, the skew-Student and skew-MEP exhibit a lower portfolio standard deviation than their symmetric counterparts. Remarkably good is the performance of the models featuring the NCT distribution; the NCT-APARCH achieves the lowest standard deviation among the univariate alternatives and compares favorably also to many multivariate models, being only outperformed by the skew-Student RBEKK. It appears that the uncertainty due to the larger number of parameters to be estimated is not fully compensated by the gain coming from a better representation of the volatility dynamics in multivariate models.

Overall, according to this table, the univariate and multivariate approaches deliver quite similar portfolio summary statistics. If the focus was on the predicted portfolio variance alone, then the ideal choice would be to use a univariate volatility model coupled with either the skew-Student or the NCT distribution, as they are easy to estimate and computationally faster than the multivariate specifications. Ultimately, we are interested in the models accuracy in forecasting the one-step-ahead portfolio VaR, so we move to analyse the outcomes of the statistical backtesting procedures.

Tables 4 and 5 report the results from the TUFF, UC, IND, CC and DBI tests in the univariate and multivariate cases, respectively.

All statistical tests are computed for the 5% and 1% VaR confidence level. For each portfolio we report test statistics along with their corresponding *p*-values in brackets. Since the applied tests measure the models accuracy in forecasting VaR along several dimensions (as detailed in Section 3.2), the overall results are summarized using a performance measure which considers the percentage of acceptances of the null hypothesis across the different tests at the standard 5% significance level. To this end, rejections of the null are highlighted in bold with the total grade across distributions reported in Table 6.

Table 4. VaR backtesting results—Univariate.

	Norm	Skew-Norm	Student	Skew-Student	MEP	Skew-MEP	NCT-APARCH	NCT-GARCH
5% VaR								
violation/frequency	52 (0.074)	42 (0.060)	53 (0.076)	29 (0.041)	52 (0.074)	40 (0.057)	53 (0.076)	25 (0.036)
TUFF	5.991 **(0.014)**	5.991 **(0.014)**	5.991 **(0.014)**	5.991 **(0.014)**	5.991 **(0.014)**	5.991 **(0.014)**	0.022 (0.883)	5.991 **(0.014)**
UC	7.611 **(0.006)**	1.389 (0.239)	8.476 **(0.004)**	1.147 (0.284)	7.611 **(0.006)**	0.720 (0.396)	3.326 (0.068)	8.476 **(0.004)**
IND	8.829 **(0.003)**	2.646 (0.104)	8.815 **(0.003)**	1.155 (0.283)	7.870 **(0.005)**	1.676 (0.195)	5.151 **(0.023)**	9.861 **(0.002)**
CC	16.440 **(0.000)**	4.035 (0.133)	17.290 **(0.000)**	2.302 (0.316)	15.481 **(0.000)**	2.397 (0.302)	8.477 **(0.014)**	18.337 **(0.000)**
DBI	0.023 (0.880)	0.079 (0.778)	0.087 (0.768)	0.830 (0.362)	0.046 (0.831)	0.100 (0.751)	0.797 (0.372)	0.303 (0.582)
1% VaR								
violation/frequency	15 (0.021)	12 (0.017)	12 (0.017)	5 (0.007)	14 (0.020)	11 (0.016)	11 (0.016)	8 (0.011)
TUFF	1.426 (0.232)	1.426 (0.232)	1.426 (0.232)	1.426 (0.232)	1.426 (0.232)	1.426 (0.232)	1.426 (0.232)	1.426 (0.232)
UC	6.957 **(0.008)**	2.972 (0.085)	2.972 (0.085)	0.641 (0.423)	5.479 **(0.019)**	1.967 (0.161)	0.138 (0.710)	1.967 (0.161)
IND	7.638 **(0.006)**	3.406 (0.065)	3.406 (0.065)	0.707 (0.400)	6.072 **(0.014)**	2.330 (0.127)	0.326 (0.568)	2.330 (0.127)
CC	14.595 **(0.001)**	6.378 **(0.041)**	6.378 **(0.041)**	1.348 (0.510)	11.551 **(0.003)**	4.297 (0.117)	0.464 (0.793)	4.297 (0.117)
DBI	0.518 (0.472)	0.126 (0.723)	0.002 (0.969)	0.000 (0.985)	0.547 (0.459)	0.281 (0.596)	0.501 (0.479)	0.281 (0.596)

Table 5. VaR backtesting results—Multivariate.

	Norm	Skew-Norm	Student	Skew-Student	MEP	Skew-MEP
5% VaR						
violation/frequency	53 (0.076)	51 (0.073)	56 (0.080)	38 (0.054)	52 (0.074)	44 (0.063)
TUFF	5.991 **(0.014)**	5.991 **(0.014)**	5.991 **(0.014)**	0.022 (0.883)	5.991 **(0.014)**	5.991 **(0.014)**
UC	8.476 **(0.004)**	6.789 **(0.009)**	11.311 **(0.001)**	0.264 (0.608)	7.611 **(0.006)**	2.260 (0.133)
IND	9.861 **(0.002)**	7.849 **(0.005)**	13.261 **(0.000)**	0.273 (0.601)	8.829 **(0.003)**	2.516 (0.113)
CC	18.337 **(0.000)**	14.637 **(0.001)**	24.572 **(0.000)**	0.536 (0.765)	16.440 **(0.000)**	4.776 (0.092)
DBI	0.858 (0.354)	0.818 (0.366)	1.052 (0.305)	1.501 (0.221)	0.797 (0.372)	1.593 (0.207)
1% VaR						
violation/frequency	19 (0.027)	18 (0.026)	17 (0.024)	8 (0.011)	17 (0.024)	15 (0.021)
TUFF	1.426 (0.232)	1.426 (0.232)	1.426 (0.232)	1.426 (0.232)	1.426 (0.232)	1.426 (0.232)
UC	14.153 **(0.000)**	12.176 **(0.000)**	10.313 **(0.001)**	0.138 (0.710)	10.313 **(0.001)**	4.051 **(0.044)**
IND	14.568 **(0.000)**	13.106 **(0.000)**	11.190 **(0.001)**	0.326 (0.568)	10.978 **(0.001)**	1.652 (0.199)
CC	28.721 **(0.000)**	25.282 **(0.000)**	21.503 **(0.000)**	0.464 (0.793)	21.292 **(0.000)**	5.702 (0.057)
DBI	3.109 (0.078)	0.902 (0.342)	3.798 (0.055)	0.282 (0.595)	1.573 (0.210)	0.774 (0.379)

Table 6. Grade comparison—VaR backtesting.

	Norm	Skew-Norm	Student	Skew-Student	MEP	Skew-MEP	NCT-APARCH	NCT-GARCH
				5% VaR				
Univariate grade	20%	80%	20%	80%	20%	80%	60%	20%
Multivariate grade	20%	20%	20%	100%	20%	80%		
				1% VaR				
Univariate grade	40%	80%	80%	100%	80%	100%	100%	100%
Multivariate grade	40%	40%	40%	100%	40%	80%		

The first distinguishing feature from the VaR backtesting results is the clear predominance of the skew-Student distribution. This holds for both the univariate and multivariate frameworks with the former producing VaR forecasts that outperform the NCT-APARCH at 5% VaR. Similarly, the skew-MEP distribution produces highly accurate VaR forecasts across the board. These findings exemplify a second feature of the results namely, the impact on VaR forecast accuracy of introducing skewness into a heavy-tailed distribution. Clearly, the performance of both the skew-Student and skew-MEP distributions improves compared to their symmetric counterparts, but this effect is less pronounced at 1% VaR. Indeed, the performance of the Student and MEP distributions improves when moving from the 5 % to the 1% VaR scenario, suggesting that despite the improvement arising from the introduction of skewness, heavy-tails remain useful in capturing larger swings in returns, *i.e.*, those events located further out in the tails. As a final remark on the effect of skewness, we observe the large difference in performance between the NCT-APARCH and NCT-GARCH at 5% VaR. Though both feature a skewed distributional specification, recall that the NCT-APARCH also features asymmetry in the conditional volatility model which explains its superior performance. An analogous result is obtained in Giot and Laurent [11].

Analysis of the differences between the univariate and multivariate models reveals two key points. First, while the skew-Student and skew-MEP retain their dominance under both frameworks, the performance of their symmetric counterparts at 1% VaR is worse under the multivariate specification. Second, within the univariate framework, the skew-normal is capable of producing VaR forecasts comparable to the high performance skewed and heavy-tailed distributions at both VaR confidence levels. This does not hold in the multivariate setup where the skew-normal offers no improvements in VaR accuracy over the normal. Besides, we observe that in both frameworks the empirical failure rate (*i.e.*, the frequency of violations) of the skewed distributions is closer to the nominal value than their symmetric counterparts, which are oversized. The NCT-GARCH represents the only exception, being considerably more conservative than the NCT-APARCH at both 5% and 1% VaR confidence levels.

With respect to the previous backtesting methods, the DQ test takes into account a more general temporal dependence between the series of violations and is considered the most reliable in assessing VaR accuracy [6]. The DQ test results are reported in Tables 7 and 8. In order to compare the different distributional assumptions, Table 9 summarises the percentage of null hypothesis acceptances over the two lag-lengths for each confidence level.

The DQ test results tell a similar story to the VaR backtesting procedures. As before, the skew-Student outperforms its competitors at both 5% and 1% VaR under both univariate and multivariate setups. The univariate skew-normal continues to produce VaR forecasts with an accuracy comparable to the top-performers. Again, this does not extend to the multivariate case.

---

[6]   We thank an anonymous referee who pointed out the possibility of better comparing the outcomes of the DQ and Christoffersen's CC and IND tests by allowing the latter to be computed in an extended framework than the standard one described in Section 3.2. As results did not lead to significant improvements, this issue is briefly covered in Appendix E.

Finally, forecast accuracy at 5% VaR again reveals a more pronounced improvement when moving from symmetric to skewed distributions than in the 1% case. Overall, introducing skewness into heavy-tailed distributions continues to offer the highest VaR forecast accuracy.

Table 7. Dynamic Quantile test results—Univariate models.

	Norm	Skew-Norm	Student	Skew-Student	MEP	Skew-MEP	NCT-APARCH	NCT-GARCH
				5% VaR				
				K = 1				
DQ_{UC}	9.319 (**0.002**)	1.594 (0.207)	10.092 (**0.001**)	1.079 (0.299)	8.965 (**0.003**)	0.809 (0.368)	3.188 (0.074)	10.485 (**0.001**)
DQ_{IND}	1.433 (0.231)	1.168 (0.280)	0.389 (0.533)	0.020 (0.886)	0.281 (0.596)	0.867 (0.352)	0.698 (0.403)	1.651 (0.199)
DQ_{CC}	10.189 (**0.006**)	2.665 (0.264)	10.201 (**0.006**)	1.087 (0.581)	9.036 (**0.011**)	1.636 (0.441)	3.680 (0.159)	11.463 (**0.003**)
				K = 3				
DQ_{UC}	7.028 (**0.008**)	1.196 (0.274)	6.821 (**0.009**)	1.240 (0.265)	6.193 (**0.013**)	0.525 (0.469)	3.024 (0.082)	8.068 (**0.005**)
DQ_{IND}	0.633 (0.426)	0.251 (0.616)	4.536 (**0.033**)	0.712 (0.399)	3.738 (0.053)	0.042 (0.838)	0.102 (0.749)	0.382 (0.537)
DQ_{CC}	12.280 (**0.015**)	2.398 (0.663)	18.210 (**0.001**)	2.633 (0.621)	15.467 (**0.004**)	1.624 (0.805)	5.220 (0.265)	12.955 (**0.011**)
				1% VaR				
				K = 1				
DQ_{UC}	9.622 (**0.002**)	3.744 (0.053)	3.744 (0.055)	0.580 (0.446)	7.357 (**0.007**)	2.393 (0.122)	0.151 (0.698)	2.393 (0.122)
DQ_{IND}	0.713 (0.398)	0.363 (0.547)	0.363 (0.557)	0.026 (0.872)	0.579 (0.447)	0.280 (0.597)	0.107 (0.744)	0.280 (0.597)
DQ_{CC}	9.985 (**0.007**)	3.991 (0.136)	3.991 (0.136)	0.598 (0.741)	7.680 (**0.021**)	2.603 (0.272)	0.254 (0.881)	2.603 (0.272)
				K = 3				
DQ_{UC}	9.078 (**0.003**)	3.376 (0.066)	3.376 (0.066)	0.373 (0.541)	6.848 (**0.009**)	2.118 (0.146)	0.126 (0.723)	2.118 (0.146)
DQ_{IND}	0.013 (0.910)	0.528 (0.468)	0.528 (0.468)	0.617 (0.432)	0.100 (0.751)	0.876 (0.349)	2.605 (0.106)	0.876 (0.349)
DQ_{CC}	13.749 (**0.008**)	9.654 (**0.047**)	9.654 (**0.047**)	19.505 (**0.001**)	11.961 (**0.018**)	9.171 (0.057)	10.840 (**0.028**)	9.171 (0.057)

Table 8. Dynamic Quantile test results—Multivariate.

	Norm	Skew-norm	Student	Skew-Student	MEP	Skew-MEP
				5% VaR		
				K = 1		
DQ_{UC}	10.485 (**0.001**)	8.226 (**0.004**)	14.416 (**0.000**)	0.280 (0.597)	9.319 (**0.002**)	2.538 (0.111)
DQ_{IND}	1.651 (0.199)	1.232 (0.267)	2.406 (0.121)	0.000 (0.993)	1.433 (0.231)	0.261 (0.610)
DQ_{CC}	11.463 (**0.003**)	8.990 (**0.011**)	15.751 (**0.000**)	0.280 (0.869)	10.189 (**0.006**)	2.727 (0.256)
				K = 3		
DQ_{UC}	7.733 (**0.005**)	6.096 (**0.014**)	10.767 (**0.001**)	0.277 (0.599)	7.028 (**0.008**)	1.607 (0.205)
DQ_{IND}	1.006 (0.316)	0.914 (0.339)	0.690 (0.406)	0.789 (0.374)	0.633 (0.426)	3.148 (0.076)
DQ_{CC}	14.578 (**0.006**)	11.839 (**0.019**)	18.650 (**0.001**)	2.662 (0.616)	12.280 (**0.015**)	9.201 (0.056)
				1% VaR		
				K = 1		
DQ_{UC}	19.549 (**0.000**)	18.210 (**0.000**)	15.080 (**0.000**)	0.151 (0.698)	13.358 (**0.000**)	9.622 (**0.002**)
DQ_{IND}	1.278 (0.258)	1.167 (0.280)	1.041 (0.308)	0.107 (0.744)	2.095 (0.148)	0.713 (0.398)
DQ_{CC}	22.121 (**0.000**)	18.684 (**0.000**)	15.521 (**0.000**)	0.254 (0.881)	16.575 (**0.000**)	9.985 (**0.007**)
				K = 3		
DQ_{UC}	13.935 (**0.000**)	14.056 (**0.000**)	9.885 (**0.002**)	0.126 (0.723)	10.157 (**0.001**)	6.809 (**0.009**)
DQ_{IND}	20.104 (**0.000**)	6.375 (0.094)	15.843 (**0.000**)	2.605 (0.106)	14.404 (**0.000**)	4.056 (0.255)
DQ_{CC}	44.966 (**0.000**)	35.949 (**0.000**)	47.702 (**0.000**)	8.842 (0.065)	33.192 (**0.000**)	9.271 (0.054)

Table 9. Grade comparison—DQ tests.

	Norm	Skew-Norm	Student	Skew-Student	MEP	Skew-MEP	NCT-APARCH	NCT-GARCH
				5% VaR				
Univariate grade	33%	100%	17%	100%	33%	100%	100%	33%
Multivariate grade	33%	33%	33%	100%	33%	100%		
				1% VaR				
Univariate grade	33%	83%	83%	83%	33%	100%	83%	100%
Multivariate grade	17%	33%	17%	100%	17%	50%		

5. Conclusions

Given its importance in risk management, practitioners must be capable of forecasting the Value-at-Risk of their asset portfolios to a high degree of accuracy. This requires taking into account a number of properties of financial returns namely, non-normality, heavy-tails, skewness and the possibility of comovements between assets. In this article, we focus primarily on the effect of varying the distributional assumption used to forecast VaR. Moreover, we addressed the still open question of whether univariate or multivariate models are most appropriate for the problem of portfolio VaR forecasting.

The distributions treated in the paper comprised three symmetric and three skewed distributional assumptions (*i.e.*, normal, Student, MEP and their skewed counterparts) which were coupled with the RBEKK model in the multivariate framework and GARCH in the univariate. In addition, we compared our specification with a novel method for fast estimation of the (univariate) NCT-GARCH and NCT-APARCH models. We then proceeded to the models' accuracy in predicting equally-weighted portfolio VaR.

Employing a series of standard backtesting methods to compare the distribution-based model performance, the results reveal that the skew-Student specification produces the most accurate one-step ahead VaR forecast across all multivariate specifications and is able to compete with the high-performance NCT-APARCH in the univariate setup. This finding is echoed in the univariate MEP results wherein the skewed version outperforms the symmetric across all tests and confidence levels. By contrast, the multivariate skew-MEP exhibits thinner tails than the normal and thus, performs poorly at the 1% confidence level. More generally, The test results reveal a clear hierarchy of distributional assumptions *within* the univariate and multivariate setups, with that hierarchy being preserved when moving from one to the other. However, attempting to compare the performance of univariate to multivariate distributions in general does not reveal any key differences with the exception of certain cases in the skew-normal and skew-MEP specifications. Consequently, the additional computational burden of estimating multivariate models does not seem to be justified.

There are several possible avenues of research extending from this work. Given the focus on distributions, we limited our attention to relatively simple parametric models namely, the RBEKK and GARCH models. As the NCT-GARCH/APARCH results showed, capturing asymmetries in returns volatility improves forecast accuracy. Considering their multivariate versions would provide a useful contribution and allow for a detailed study of the optimal model-distribution-dimension combination. Another possibility would be to consider higher forecast horizons for the VaR in order to check if the inclusion of skewness and asymmetric forms of dependence can lead to significant improvements in the long run.

Author Contributions: Manuela Braione and Nicolas K. Scholtes contributed equally to the writing of the MATLAB code as well as the preparation of the current manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Appendixes

A. Derivations

A.1. Transformation

The transformation $z_t = H_t^{-1/2} y_t$ is incorporated into the symmetric, standardised pdfs as follows:

$$
\begin{aligned}
\kappa^{\star\prime}\kappa^{\star} &= (\kappa_1^{\star}, ..., \kappa_N^{\star})'(\kappa_1^{\star}, ..., \kappa_N^{\star}) \\
&= \left(\ldots (s_i z_i + m_i) \xi_i^{I_i} \ldots \right)' \left(\ldots (s_i z_i + m_i) \xi_i^{I_i} \ldots \right) \\
&= \left(\ldots (s_i \sum_{j=1}^{N} p_{ij} y_j + m_i) \xi_i^{I_i} \ldots \right)' \left(\ldots (s_i \sum_{j=1}^{N} p_{ij} y_j + m_i) \xi_i^{I_i} \ldots \right) \\
&= \sum_{i=1}^{N} \left(s_i \sum_{j=1}^{N} p_{ij} y_j + m_i \right) \xi_i^{2 I_i},
\end{aligned}
$$

where p_{ij} corresponds to the jth element of the ith row of $H_t^{-1/2}$. Note that the t subscript is dropped for simplicity. The matrix square root operation is carried out by applying the Cholesky decomposition of H_t such that $BB' = H_t$. As a result, each z_i is obtained by multiplying the row vector of $H_t^{-1/2}$ corresponding to asset i with the demeaned return vector (giving us the inner summation above) which is then multiplied by the univariate standard deviation and added to the univariate mean. The presence of skewness is factored in by the term $\xi_i^{I_i}$, where the factor I_i is defined as in Equation (17).

A.2. Distributions Moments

We report the first two moments of the univariate symmetric normal, Student and MEP distributions used to compute the log-likelihood function as given in Section 2.3.

Note that the following relation holds:

$$
m_i = M_{i,1} \left(\xi_i - \frac{1}{\xi_i} \right) \Rightarrow M_{i,1}^2 = m_i^2 \left(\frac{\xi_i^2}{(\xi_i^2 - 1)^2} \right).
$$

Hence, substituting this result into Equation (14), the 2nd order moment of the skewed distributions can be obtained as a function of the first:

$$
\begin{aligned}
s_i^2 &= M_{i,1}^2 \left(-\xi_i^2 - \frac{1}{\xi_i^2} + 2 \right) + M_{i,2} \left(\xi_i^2 + \frac{1}{\xi_i^2} - 1 \right) \\
&= \frac{\xi_i^2}{(\xi_i^2 - 1)^2} \left(\frac{-\xi_i^4 + 2\xi_i^2 - 1}{\xi_i^2} \right) m_i^2 + M_{i,2} \left(\frac{\xi_i^4 - \xi_i^2 + 1}{\xi_i^2} \right) \\
&= \frac{\xi_i^2}{(\xi_i^2 - 1)^2} \left(\frac{(-\xi_i^2 + 1)(\xi_i^2 - 1)}{\xi_i^2} \right) m_i^2 + M_{i,2} \frac{\xi_i^2 \left(\xi_i^2 - 1 + \frac{1}{\xi_i^2} \right)}{\xi_i^2} \\
&= m_i^2 \left(\frac{1 - \xi_i^2}{\xi_i^2 - 1} \right) + M_{i,2} \left(\xi_i^2 + \frac{1}{\xi_i^2} - 1 \right).
\end{aligned}
$$

Skew-Normal

$$M_{i,1} = \int_0^\infty \frac{2}{\sqrt{2\pi}} u \exp\left\{-\frac{1}{2}u^2\right\} du = \sqrt{\frac{2}{\pi}} \int_0^\infty u \exp\left\{-\frac{1}{2}u^2\right\} du = \sqrt{\frac{2}{\pi}}$$

$$M_{i,2} = \int_0^\infty \frac{2}{\sqrt{2\pi}} u^2 \exp\left\{-\frac{1}{2}u^2\right\} du = \sqrt{\frac{2}{\pi}} \int_0^\infty u^2 \exp\left\{-\frac{1}{2}u^2\right\} du = 1$$

Skew-Student

$$M_{i,1} = \frac{2\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi(\nu-2)}} \int_0^\infty u \left(1+\frac{u^2}{\nu-2}\right)^{-\frac{1+\nu}{2}} du$$

$$= \frac{2\sqrt{\nu-2}\left(\frac{\nu-1}{2}\right)\Gamma\left(\frac{\nu-1}{2}\right)}{\sqrt{\pi}(\nu-1)\Gamma\left(\frac{\nu}{2}\right)}$$

$$= \frac{2\sqrt{\nu-2}\,\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi}(\nu-1)\Gamma\left(\frac{\nu}{2}\right)}$$

$$= \frac{\Gamma\left(\frac{\nu-1}{2}\right)\sqrt{\nu-2}}{\sqrt{\pi}\,\Gamma\left(\frac{\nu}{2}\right)}$$

$$M_{i,2} = \frac{2\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi(\nu-2)}} \int_0^\infty u^2 \left(1+\frac{u^2}{\nu-2}\right)^{-\frac{1+\nu}{2}} du$$

$$= \frac{(\nu-2)\Gamma\left(\frac{\nu}{2}-1\right)}{2\Gamma\left(\frac{\nu}{2}\right)}$$

$$= \frac{(\nu-2)\Gamma\left(\frac{\nu}{2}-1\right)}{2\left(\frac{\nu-2}{2}\right)\Gamma\left(\frac{\nu}{2}-1\right)} = 1$$

Skew-MEP

$$M_{i,1} = \frac{2^{-\frac{1}{\beta}}}{\Gamma\left(1+\frac{1}{\beta}\right)} \int_0^\infty u \exp\left\{-\frac{1}{2}u^\beta\right\} du$$

$$= \frac{2^{-1+\frac{1}{\beta}}\Gamma\left(\frac{2+\beta}{\beta}\right)}{\Gamma\left(1+\frac{1}{\beta}\right)}$$

$$M_{i,2} = \frac{2^{-\frac{1}{\beta}}}{\Gamma\left(1+\frac{1}{\beta}\right)} \int_0^\infty u^2 \exp\left\{-\frac{1}{2}u^\beta\right\} du$$

$$= \frac{4^{\frac{1}{\beta}}\Gamma\left(\frac{3}{\beta}\right)}{\beta\Gamma\left(1+\frac{1}{\beta}\right)}$$

B. Tables

Table A1. Windows length and corresponding calendar time.

|	Rolling Fixed-Window		Forecast Horizon	
It.	Observations	Days	Observations	Days
1	1–1500	1 Febraury 2001–23 January 2007	1501–1520	24 January 2007–21 Febraury 2007
2	21–1520	2 March 2001–21 Febraury 2007	1521–1540	22 Febraury 2007–21 March 2007
3	41–1540	30 March 2001–21 March 2007	1541–1560	22 March 2007–19 April 2007
4	61–1560	30 April 2001–19 April 2007	1561–1580	20 April 2007–17 May 2007
5	81–1580	29 May2001–17 May 2007	1581–1600	18 May 2007–15 June 2007
6	101–1600	26 June 2001– 15 June 2007	1601–1620	18 June 2007–16 July 2007
7	121–1620	25 July 2001–16 July 2007	1621–1640	17 July 2007–13 August 2007
8	141–1640	22 August 2001–13 August 2007	1641–1660	14 August 2007–11 September 2007
9	161–1660	26 September 2001–11 September 2007	1661–1680	12 September 2007–9 October 2007
10	181–1680	24 October 2001–9 October 2007	1681–1700	10 October 2007–6 November 2007
11	201–1700	21 November 2001–6 Novembe 2007	1701–1720	7 November 2007–5 December 2007
12	221–1720	20 December 2001–5 December 2007	1721–1740	6 December 2007–4 January 2008
13	241–1740	22 January 2002–4 January 2008	1741–1760	7 January 2008–4 Febraury 2008
14	261–1760	20 Febraury 2002–4 Febraury 2008	1761–1780	5 Febraury 2008– 4 March 2008
15	281–1780	20 March 2002–4 March 2008	1781–1800	5 March 2008–2 April 2008
16	301–1800	18 April 2002–2 April 2008	1801–1820	3 April 2008–30 April 2008
17	321–1820	16 May 2002–30 May 2008	1821–1840	1 May 2008–29 May 2008
18	341–1840	14 June 2002–29 May 2008	1841–1860	30 May 2008–26 June 2008
19	361–1860	15 July 2002–26 June 2008	1861–1880	27 June 2008–25 July 2008
20	381–1880	12 August 2002–25 July 2008	1881–1900	28 July 2008– 22 August 2008
21	401–1900	10 September2002–22 August 2008	1901–1920	25 August 2008– 22 September 2008
22	421–1920	8 October 2002–22 September 2008	1921–1940	23 September 2008–20 October 2008
23	441–1940	5 November 2002–20 October 2008	1941–1960	21 October 2008–17 November 2008
24	461-1960	4 December 2002–17 November 2008	1961–1980	18 November 2008–16 December 2008
25	481–1980	3 January 2003–16 December 2008	1981–2000	17 December 2008–15 January 2009
26	501–2000	3 Febraury 2003–15 January 2009	2001–2020	16 January 2009–13 Febraury 2009
27	521–2020	4 March 2003–13 Febraury 2009	2021–2040	17 Febraury 2009–16 March 2009
28	541–2040	1 April 2003–16 March 2009	2041–2060	17 March 2009–14 April 2009
29	561–2060	30 April 2003–14 April 2009	2061–2080	15 April 2009–12 May 2009
30	581–2080	29 May 2003–12 May 2009	2081–2100	13 May 2009–10 June 2009
31	601–2100	26 June 2003–10 June 2009	2101–2120	11 June 2009–9 July 2009
32	621–2120	25 July 2003–9 July 2009	2121–2140	10 July 2009–6 August 2009
33	641–2140	22 August 2003–6 August 2009	2141–2160	7 August 2009–3 September 2009
34	661–2160	22 September 2003–3 September 2009	2161–2180	4 September 2009–2 October 2009
35	681-2180	20 October 2003–2 October 2009	2181–2200	5 October 2009–30 October 2009

C. Figures
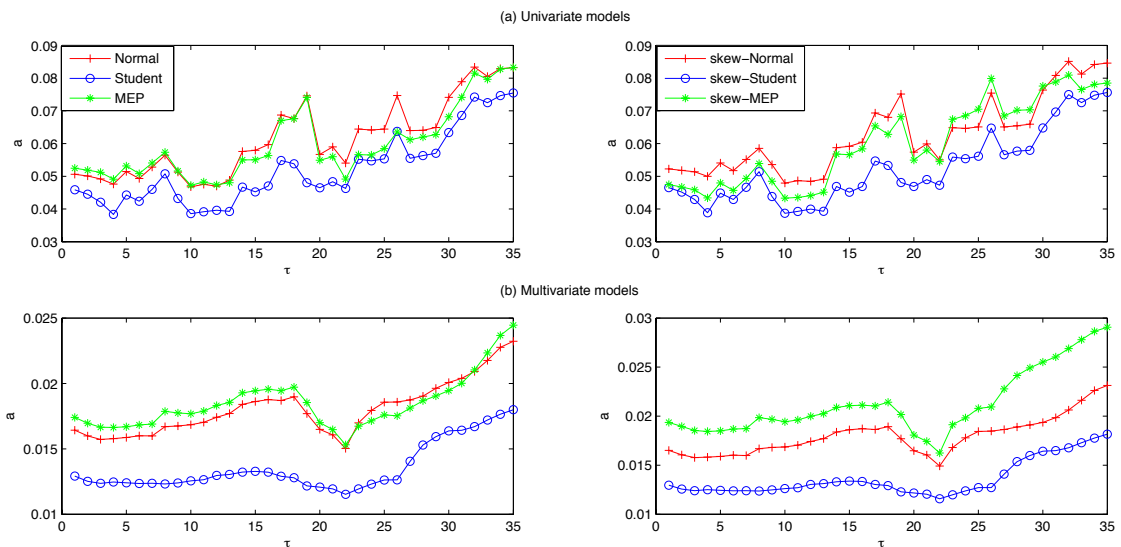
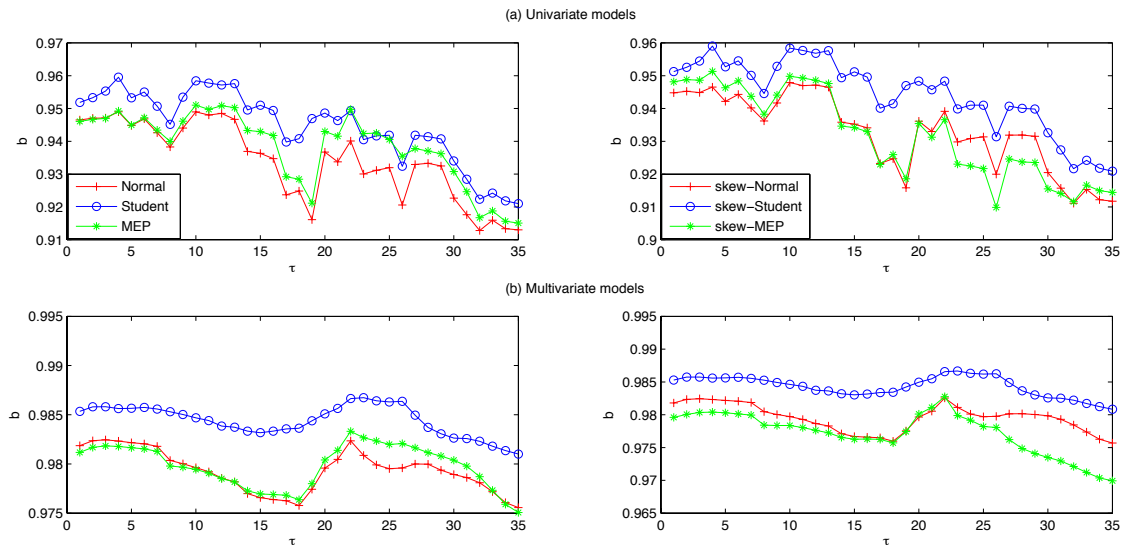**Figure A1.** Evolution across re-estimations of the GARCH/RBEKK parameter, *a*.

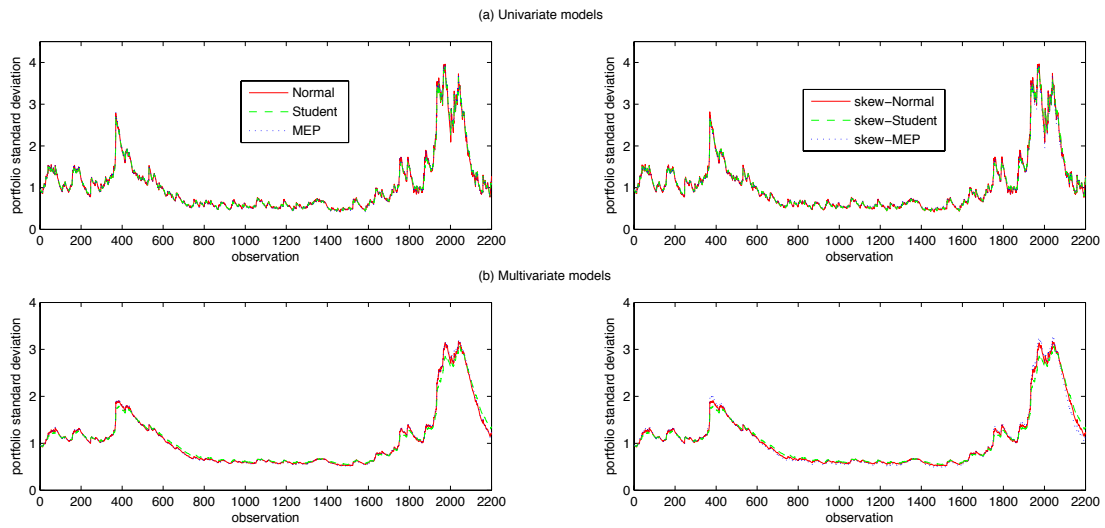**Figure A2.** Evolution across re-estimations of the GARCH/RBEKK parameter, *b*.

Figure A3. Portfolio standard deviations.
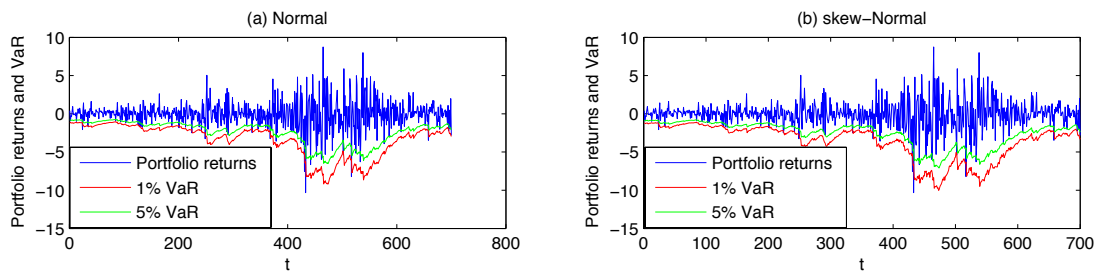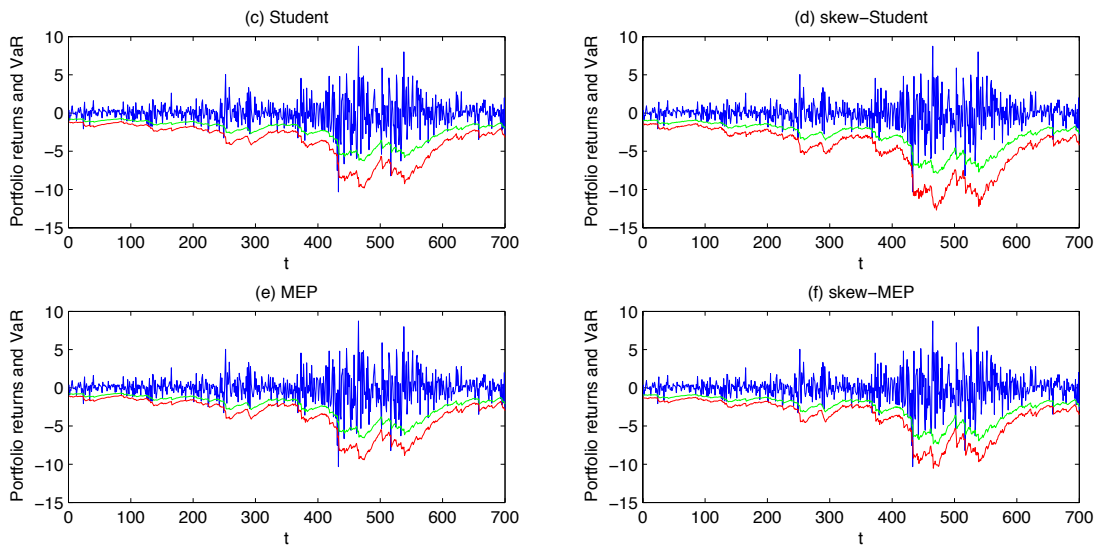

Figure A4. *Cont.*
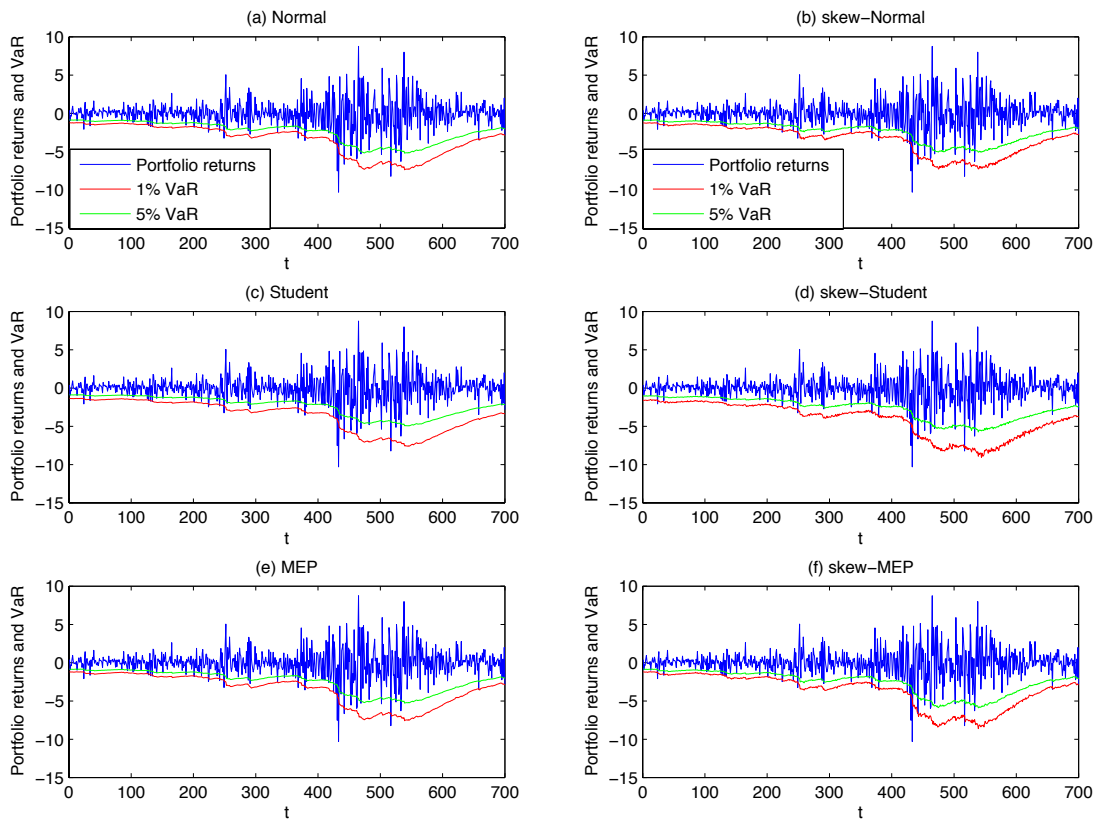
Figure A4. VaR: Univariate models.

Figure A5. VaR: Multivariate models.

D. The Univariate NCT-APARCH Model

The main ingredients of the NCT-APARCH model discussed in the paper by Krause and Paolella [16] are, as the name may suggest, the use of a univariate noncentral t-distribution assumption for the error term and of an APARCH specification for the conditional variance equation.

The NCT density function is given as follows:

$$
\begin{aligned}
f_Z\left(t;\nu,\gamma\right) = \quad & e^{-\gamma^2/2} \quad \frac{\Gamma((\nu+1)/2)\nu^{\nu/2}}{\sqrt{\pi}\Gamma(\nu/2)} \left(\frac{1}{\nu+t^2}\right)^{\frac{\nu+1}{2}} \\
& \times \quad \left[\sum_{i=0}^{\infty} \frac{(t\gamma)^i}{i!}\left(\frac{2}{t^2+\nu}\right)^{i/2}\frac{\Gamma\{(\nu+i+1)/2\}}{\Gamma\{(\nu+1)/2\}}\right],
\end{aligned} \tag{35}
$$

where $\nu > 0$ denotes the degrees of freedom parameter and $\gamma \in \Re$ is the noncentrality parameter dictating the degree of asymmetry (with the Student distribution recovered when $\gamma = 0$).

The evolution of the conditional variance is modeled according to the APARCH model, which allows for both heavy-tails and asymmetry. It is defined as follows:

$$
\sigma_t^\delta = \bar{\sigma} + a_1\left(|\epsilon_{t-1}| - \gamma_1\epsilon_{t-1}\right)^\delta + b_1\sigma_{t-1}^\delta
$$

where $a_1 > 0$, $b_1 \geq 0$, $\delta > 0$ and $|\gamma_1| < 1$. In order to speed up the estimation procedure, the authors suggest to fix some parameters before the maximization of the likelihood function. In this paper, we calibrated the model parameters on our dataset in order to achieve the highest in-sample likelihood value (this applies also for the model incorporating the GARCH specification). Specifically, we set $\bar{\sigma} = 0.01$, $a_1 = 0.05$, $b_1 = 0.90$, $\delta = 2$ and $\gamma_1 = 0.4$. The vector of parameters to be estimated, then, reduces to $\boldsymbol{\psi} = (\mu, \nu, \gamma)'$, where $E(\boldsymbol{y}_t) = \mu$ denotes the location coefficient which is fastly estimated using the method of trimmed mean. We refer to the original paper by Krause and Paolella [16] for further details on the implemented estimation procedure.

E. Backtesting VaR: Augmented Independence and Conditional Coverage Tests

Tables A2 and A3 report test-statistics and corresponding p-values obtained from the Independence (IND) and Conditional Coverage (CC) tests of Christoffersen (see Section 4.2) computed with a number of lagged observations $K > 1$.

The tests are built in an extended Markov framework that allows for higher, or K-th order, dependence in the VaR observations, thus highlighting possible clustering in the series of violations. The distribution of the *generalized* IND and CC tests is asymptotically $\chi^2_{(1)}$ and $\chi^2_{(2)}$, respectively. For further details on the construction of the tests we remand the interested reader to theoretical framework developed in Pajhede [39].

In order to make these results directly comparable with those given in Tables 7 and 8, we set the number of lags $K = 3$ and run the tests for the 5% and 1% VaR confidence level.

Results for both univariate and multivariate specifications are reported below.

Table A2. Augmented IND and CC test results. Univariate models.

	Norm	Skew-Norm	Student	Skew-Student	MEP	Skew-MEP	NCT-APARCH	NCT-GARCH
				5% VaR				
IND ($K = 3$)	0.843 (0.358)	0.1494 (0.699)	2.294 (0.130)	1.0807 (0.299)	1.802 (0.179)	0.0091 (0.924)	0.182 (0.670)	0.576 (0.448)
CC ($K = 3$)	7.779 (**0.020**)	1.233 (0.540)	10.062 (**0.007**)	2.596 (0.273)	8.739 (**0.013**)	0.511 (0.775)	3.419 (0.181)	8.344 (**0.015**)
				1% VaR				
IND ($K = 3$)	0.003 (0.955)	0.2371 (0.626)	0.237 (0.626)	3.0345 (0.082)	0.038 (0.844)	0.408 (0.523)	1.288 (0.256)	0.408 (0.523)
CC ($K = 3$)	7.030 (**0.030**)	3.253 (0.197)	3.253 (0.197)	3.658 (0.161)	5.579 (0.061)	2.410 (0.300)	1.436 (0.488)	2.410 (0.300)

Table A3. Augmented IND and CC test results. Multivariate models.

	Norm	Skew-norm	Student	Skew-Student	MEP	Skew-MEP
			5% VaR			
IND ($K = 3$)	1.204	1.105	0.940	1.142	0.843	3.453
	(0.272)	(0.293)	(0.332)	(0.285)	(0.358)	(0.063)
CC ($K = 3$)	8.972	7.251	11.451	1.434	7.779	5.327
	(**0.011**)	(**0.027**)	(**0.003**)	(0.488)	(**0.020**)	(0.070)
			1% VaR			
IND ($K = 3$)	6.349	1.965	4.765	1.288	4.896	3.234
	(0.012)	(0.161)	(0.029)	(0.256)	(0.027)	(0.072)
CC ($K = 3$)	20.608	14.237	15.166	1.436	15.298	5.671
	(**0.000**)	(**0.001**)	(**0.001**)	(0.488)	(**0.000**)	(0.058)

Overall, adding more lags in the regression equation does not seem to crucially affect the outcome of the tests, as the new results are still in line with those obtained under $K = 1$ in Tables 4 and 5. The main difference registered for both univariate and multivariate approaches is in the outcome of the IND test, which now all models in the tables pass at the standard 5% level for both the 5% and 1% VaR. Apparently, violations are not clustered in time for more than one or two lags.

As for the CC test, results concerning the multivariate models are basically the same obtained by setting $K = 1$, while in the univariate case we observe an improvement in the performance of the models (excluding the normal) for the 1% VaR.

Ultimately, univariate distributions featuring skewness and heavy tails confirm their predominance over the remaining alternatives in terms of better predictive ability. In the multivariate case this is particularly true for the skew-Student and the skew-MEP, as the skew-normal is not leading to remarkable improvements over the corresponding normal assumption.

References

1. Engle, R. Autoregressive Conditional Heteroskedasticity with Estimates of United Kindgom Heteroskedasticity. *Econometrica* **1982**, *50*, 987–1007.
2. Bollerslev, T. Generalized Autoregressive Conditional Heteroskedasticity. *J. Econom.* **1986**, *31*, 307–327.
3. Angelidis, T.; Benos, A.; Degiannakis, S. The Use of GARCH Models in VaR Estimation. *Stat. Methodol.* **2004**, *1*, 105–128.
4. Nelson, D.B. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* **1991**, *59*, 347–370.
5. Mittnik, S.; Paolella, M. Conditional density and value-at-risk prediction of Asian currency exchange rates. *J. Forecast.* **2000**, *19*, 313–333.
6. Ding, Z.; Granger, C.W.; Engle, R.F. A long memory property of stock market returns and a new model. *J. Empir. Financ.* **1993**, *1*, 83–106.
7. Kuester, K.; Mittnik, S.; Paolella, M. Value-at-risk prediction: A comparison of alternative strategies. *J. Financ. Econom.* **2006**, *4*, 53–89.
8. Longin, F.; Solnik, B. Is the correlation in international equity returns constant: 1960–1990? *J. Int. Money Financ.* **1995**, *14*, 3–26.
9. Brooks, C.; Persand, G.; Value-at-risk and market crashes. *J. Risk* **2000**, *2*, 5–26.
10. Bauwens, L.; Laurent, S.; Rombouts, J. Multivariate GARCH models: A Survey. *J. Appl. Econom.* **2006**, *21*, 79–109.
11. Giot, P.; Laurent, S. Value-at-risk for long and short trading positions. *J. Appl. Econom.* **2003**, *18*, 641–663.
12. Santos, A.A.; Nogales, F.J.; Ruiz, E. Comparing univariate and multivariate models to forecast portfolio value-at-risk. *J. Financ. Econom.* **2013**, *11*, 400–441.
13. Bauwens, L.; Laurent, S. A New Class of Multivariate Skew Densities, with Application to GARCH Models. *J. Bus. Econ. Stat.* **2005**, *23*, 346–354.
14. Ley, C.; Paindaveine, D. Multivariate skewing mechanisms: A unified perspective based on the transformation approach. *Stat. Probab. Lett.* **2010**, *80*, 1685–1694.

15. Noureldin, D.; Shephard, N.; Sheppard, K. Multivariate rotated ARCH models. *J. Econom.* **2014**, *179*, 16–30.

16. Krause, J.; Paolella, M.S. A fast, accurate method for value-at-risk and expected shortfall. *Econometrics* **2014**, *2*, 98–122.

17. Christoffersen, P.F. Evaluating interval forecasts. *Int. Econ. Rev.* **1998**, *39*, 841–862.

18. Dumitrescu, E.; Hurlin, C.; Pham, V. Backtesting Value-at-Risk: From Dynamic Quantile to Dynamic Binary Tests. *Finance* **1995**, *33*, 79–111.

19. Kupiec, P.H. Techniques for verifying the accuracy of risk measurement models. *J. Deriv.* **1995**, *3*, 73–84.

20. Engle, R.F.; Manganelli, S. CAViaR: Conditional autoregressive value at risk by regression quantiles. *J. Bus. Econ. Stat.* **2004**, *22*, 367–381.

21. DeMiguel, V.; Garlappi, L.; Uppal, R. Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Rev. Financ. Stud.* **2009**, *22*, 1915–1953.

22. Tu, J.; Zhou, G. Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies. *J. Financ. Econ.* **2011**, *99*, 204–215.

23. Brown, S.J.; Hwang, I.; In, F. Why Optimal Diversification Cannot Outperform Naive Diversification: Evidence from Tail Risk Exposure, **2013**, doi:10.2139/ssrn.2242694 .

24. Fugazza, C.; Guidolin, M.; Nicodano, G. Equally Weighted vs. Long-Run Optimal Portfolios. *Eur. Financ. Manag.* **2014**, *21*, 742–789.

25. Bauwens, L.; Omrane, W.; Rengifo, E. *Intra-Daily FX Optimal Portfolio Allocation*; CORE Discussion Paper 2006/10; CORE: Louvain-la-Neuve, Belgium, 2006.

26. Engle, R.F.; Kroner, K.F. Multivariate simultaneous generalized ARCH. *Econometr. Theory* **1995**, *11*, 122–150.

27. Barndorff-Nielsen, O. Exponentially decreasing distributions for the logarithm of particle size. *Proc. R. Soc. Lond. A Math. Phys. Eng. Sci.* **1977**, *353*, 401–419.

28. Paolella, M.S.; Polak, P. COMFORT: A Common Market Factor Non-Gaussian Returns Model. *J. Econom.* **2015**, *187*, 593–605.

29. Fernández, C.; Steel, M.F. On Bayesian modeling of fat tails and skewness. *J. Am. Stat. Assoc.* **1998**, *93*, 359–371.

30. Bollerslev, T.; Wooldridge, J. Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time-Varying Covariances. *Econom. Rev.* **1992**, *1*, 143–172.

31. Fang, K.; Kotz, S.; Ng, K. *Symmetric Multivariate and Related Distributions*; Chapman and Hall: London, UK, 1990.

32. Solaro, N. Random variate generation from multivariate exponential power distribution. *Stat. Appl. II* **2004**, *2*, 25–44.

33. Noureldin, D.; Shephard, N.; Sheppard, K. Multivariate high-frequency-based volatility (HEAVY) models. *J. Appl. Econom.* **2012**, *27*, 907–933.

34. Christoffersen, P. Evaluating Interval Forecasts. *Int. Eco. Rev.* **1998**, *39*, 841–862.

35. Kupiec, P. Techniques for Verifying the Accuracy of Risk Management Models. *J. Deriv.* **1995**, *23*, 73–84.

36. Campbell, S. A Review of Backtesting and Backtesting Procedures. *J. Risk* **2006**, *9*, 1–17.

37. Engle, R.; Manganelli, S. CaViaR: Conditional Autoregressive Value at Risk by Regression Quantiles. *J. Bus. Econ. Stat.* **2004**, *22*, 367–381.

38. Christofferson, P.; Pelletier, D. Backtesting Value-at-Risk: A Duration-Based Approach. *J. Financ. Econom.* **2004**, *2*, 84–108.

39. Pajhede, T. *Backtesting Value-at-Risk: The generalized Markov Test*; U. Copenhagen Economics Discussion Paper 15-18; University of Copenhagen: Copenhagen, Denmark, 2015.