Forecasting VaR and ES using a joint quantile regression and its implications in portfolio allocation

Luca Merlo
Department of Statistical Sciences, Sapienza University of Rome
and
Lea Petrella
MEMOTEF Department, Sapienza University of Rome
and
Valentina Raponi
IESE Business School, University of Navarra

July 12, 2021

Abstract

In this paper, we propose a multivariate quantile regression framework to forecast Value at Risk (VaR) and Expected Shortfall (ES) of multiple financial assets simultaneously, extending Taylor (2019). We generalize the Multivariate Asymmetric Laplace (MAL) joint quantile regression of Petrella and Raponi (2019) to a time-varying setting, which allows us to specify a dynamic process for the evolution of both the VaR and ES of each asset. The proposed methodology accounts for the dependence structure among asset returns. By exploiting the properties of the MAL distribution, we propose a new portfolio optimization method that minimizes portfolio risk and controls for well-known characteristics of financial data. We evaluate the advantages of the proposed approach on both simulated and real data, using weekly returns on three major stock market indices. We show that our method outperforms other existing models and provides more accurate risk measure forecasts than univariate methods.

Keywords: Quantile Regression, Multiple Quantiles, Multivariate Asymmetric Laplace Distribution, CAViaR, Value at Risk, Expected Shortfall

1 Introduction

The events of the ongoing credit crisis and past financial crises have emphasized the necessity for appropriate risk measures. The use of quantitative risk measures has become an essential management tool providing advice, analysis and support for financial and asset management decisions to market participants and regulators. The most widely used risk measure is Value at Risk (VaR). VaR measures the maximum loss that a financial operator can incur over a defined time horizon and for a given confidence level. Its clear meaning and computational ease made it very popular among practitioners, so much so that it has widely infiltrated the banking regulatory framework. However, VaR has a number of drawbacks (Artzner et al. 1997, 1999). First, VaR does not account for tail risk; i.e., it does not warn us about the size of the losses that occur with a probability lower than the predetermined confidence level. Second, VaR is not a "coherent" risk measure (Artzner et al. 1999) since it does not satisfy the sub-additivity property, and hence, it does not take into consideration the benefits of diversification. As a result, investors and risk managers are likely to construct positions with unintended weaknesses that result in greater losses under conditions beyond the VaR level (Yamai and Yoshiba 2005). Market participants could solve such problems by adopting the Expected Shortfall (ES) risk measure, which is defined as the conditional expectation of exceedances beyond VaR (see Acerbi and Tasche (2002) and Rockafellar and Uryasev (2000)). Unlike VaR, ES is a coherent risk measure and provides more information on the shape and the heaviness of the tails of the loss distribution. Therefore, ES has gained increasing attention from risk managers, banking regulators and investors as an alternative measure of risk, complementing the VaR measure.

However, despite its interesting properties, and in contrast with VaR, little work exists on modeling ES. This is in part because ES is not an "elicitable" measure, in the sense that there does not exist a loss function such that ES is the solution that minimizes the expected loss. Several works have been proposed in the literature to overcome the problem of elicitability (see, e.g., Engle and Manganelli (2004), Cai and Wang (2008), Taylor (2008), Zhu and Galbraith (2011), Du and Escanciano (2017), Patton et al. (2019) and Bu et al. (2019)). Recently, using the results of Fissler and Ziegel (2016), who show that ES is jointly elicitable with VaR, Taylor (2019) uses the Asymmetric Laplace (AL) distribution to jointly estimate dynamic models for both VaR and ES. In particular, Taylor (2019) shows that the negative of the log-likelihood associated with the AL distribution belongs to the class of loss functions presented in Fissler and Ziegel (2016), and hence it can be used to estimate and forecast the VaR and ES measures in one step. In his paper, the joint estimation of VaR and ES is obtained in a univariate quantile regression framework, exploiting the interesting result that ES can be expressed in terms of the scale parameter of the AL density.

The literature mentioned above, however, has mainly focused on univariate time series, which completely disregards the strong interrelation among assets in financial markets. To capture the

degree of tail interdependence between assets, several quantile-based methods have also been proposed to estimate VaR, but they do not specify a model for the ES component; see, for example, the relevant works of Baur (2013); Bernardi et al. (2015); White et al. (2015); Kraus and Czado (2017) and Bonaccolto et al. (2019).

In this paper, we extend the univariate approach of Taylor (2019) to a multivariate framework, with the objective of obtaining joint estimates of both VaR and ES for multiple financial assets simultaneously, accounting for their dependence structure. To this end, we generalize the Multivariate Asymmetric Laplace (MAL) quantile regression approach of Petrella and Raponi (2019) to a time-varying setting by allowing the parameters of the MAL distribution to vary over time. For each asset, we model the evolution of VaR and ES as functions of the location and scale parameters of the distribution. In particular, for the VaR component, we adopt a Conditional Autoregressive Value at Risk (CAViaR) specification (Engle and Manganelli (2004)).

The advantages of our methodology are manifold. First, our approach is a joint modeling framework where both the model parameters and the pair (VaR, ES) of multiple returns are estimated simultaneously, generalizing the univariate results of Bassett et al. (2004) and Taylor (2019). Second, our theory captures empirical characteristics of financial data such as peakedness, skewness, and heavy tails (see e.g., Kraus and Litzenberger (1976); Friend and Westerfield (1980) and Barone-Adesi (1985)) without relying on the limitation of normally distributed returns.

The inferential problem is solved by developing a suitable Expectation-Maximization (EM) algorithm, which exploits the mixture representation of the MAL distribution (see Petrella and Raponi (2019)) properly generalized to the case of time-varying parameters. The finite sample properties of the proposed estimation method are also evaluated using a simulation exercise, where we show the validity and the robustness of our procedure under different data generating processes.

A further contribution of the paper concerns the evaluation of VaR and ES in the context of portfolio optimization (see, e.g., Yiu (2004) and Alexander and Baptista (2008)). In recent years, the MAL density has attracted wide attention in the literature because of its flexibility in modeling financial data (Mittnik and Rachev (1991); Kotz et al. (2012) and Paolella (2015)) and for its interesting properties, which can be exploited to derive optimal portfolio allocations (see Zhao et al. (2015) and Shi et al. (2018)). In the classic Mean-Variance (MV) methodology of Markowitz (1952), portfolio risk is measured using the standard deviation of the portfolio. However, the MV approach is reasonably applicable only in cases where either the returns follow a Gaussian distribution or the investor utility function is quadratic. Given the empirical evidence showing that market participants have a preference for positive skewness and they are more concerned about the downside risk (see Arditti (1971) and Konno and Suzuki (1995), among others), the MAL distribution could represent a more effective tool for selecting optimal portfolio allocations in the case of risk-averse agents. Therefore, in this paper, we exploit the MAL properties to

incorporate skewness directly into the portfolio optimization method and to identify the optimal allocation weights. We then compute the corresponding portfolio VaR and ES as a function of the multivariate structure of the data. We prove that this result follows directly from the fact that any linear combination of the MAL components is still AL distributed, with location, skew and scale parameters that are functions of the MAL parameters and the portfolio weights. Therefore, once we obtain the Maximum Likelihood (ML) estimates of the MAL parameters from the proposed dynamic quantile regression model, we fix a desired level of risk for any target portfolio and search for the optimal allocation weights according to the adopted strategy.

Specifically, we consider the Skewness Mean-Variance (SMV) strategy of Zhao et al. (2015), where the optimal allocation is obtained by minimizing the portfolio variance, while controlling for the skewness of the asset returns. However, Zhao et al. (2015) employed the method of moments to estimate the portfolio variance; in contrast, we estimate the MAL parameters in a ML framework by using an EM algorithm.

Empirically, we analyze the weekly returns of the FTSE 100, NIKKEI 225 and Standard & Poor's 500 (S&P 500) market indices from April 1985 to February 2021. In a first out-of-sample exercise, we jointly estimate the VaR and ES of the three stock market indices using the proposed dynamic joint quantile regression model, hence taking into account the correlation among the three indices. To evaluate VaR and ES forecasts and to show the main advantages of the proposed method, we use different backtesting procedures, where we compare the out-of-sample VaR and ES predictions with the ones obtained by applying the univariate method of Taylor (2019). In particular, to perform a joint evaluation of VaR and ES, we follow Fissler et al. (2015), Nolde et al. (2017), Patton et al. (2019) and Taylor (2019) and extend their approach by introducing a new scoring function based on the MAL distribution. We find that our multivariate method always provides more accurate VaR and ES predictions than other well-known approaches, such as the Quantile AutoRegression of Koenker and Xiao (2006) and the dynamic quantile regression of Taylor (2019). Moreover, in line with Taylor (2019), our results show that the Asymmetric Slope CAViaR specification of Engle and Manganelli (2004) yields the best VaR and ES forecasts for all three indices at different quantile levels, confirming the existence of relevant asymmetries in the impact of positive and negative returns.

In a second empirical analysis, we aggregate the stock market indices to form a financial portfolio with a predetermined level of risk, and we estimate its optimal allocation weights by implementing our new optimization procedure. We then compute the out-of-sample portfolio's VaR and ES and evaluate the predictions using the univariate backtesting procedures of Taylor (2019), Nolde et al. (2017) and Patton et al. (2019). The empirical analysis reveals that our multivariate method produces the lowest average losses compared to other existing strategies based on the multivariate Normal and Student-t distributions, regardless of the scoring function used. In addition, we

find that the proposed methodology yields the highest Sharpe Ratio overall as well as the least concentrated portfolio allocations.

The rest of this paper is organized as follows. In Section 2, we introduce the dynamic multiple quantile regression and propose a joint model for VaR and ES. We then illustrate the EM-based ML approach for the simultaneous estimation of VaR and ES. Section 3 develops the portfolio allocation problem. Section 4 introduces a new scoring function for the joint evaluation of VaR and ES forecasts. In Section 5 we discuss the main empirical results, while Section 6 concludes the paper. All the proofs are provided in Appendix A, and the simulation study is presented in Appendix B.

2 Multivariate framework

In this paper, we generalize the univariate regression approach of Taylor (2019). Specifically, by extending the MAL density of Petrella and Raponi (2019) – allowing the location and scale parameters of the MAL distribution to vary over time – we estimate the pair of VaR and ES associated with each asset using a joint quantile regression framework. In this way, we are able to calculate the time-varying VaR and ES simultaneously for all marginal response variables, accounting for possible correlation among the considered assets. For the VaR components, we assume a CAViaR specification (see Engle and Manganelli (2004)). Parameter estimation is carried out using a suitable EM algorithm as in Petrella and Raponi (2019), properly extended to deal with the time-varying setting. In this way, the estimated parameters account for tail interdependence among multiple returns, and they convey this information to the VaR and ES estimates.

We start by introducing the time-varying joint quantile regression model in Section 2.1, where we consider a dynamic generalization of the MAL density proposed in Petrella and Raponi (2019). We then show in Section 2.2 how the resulting time-varying scale parameter of the MAL distribution can be used to model the ES vector and derive a parsimonious approach for the simultaneous estimation of VaR and ES in a multidimensional setting. The parameter estimation and the EM algorithm are described in Section 2.3.

2.1 Dynamic joint quantile regression

Let $\boldsymbol{Y}_t = [Y_{t1}, Y_{t2}, ..., Y_{tp}]'$ be a p-variate response variable and denote by $\mathcal{Q}_{Y_{tj}}(\tau_j|\mathcal{F}_{t-1})$ the τ_j-quantile function of the j-th component of $\boldsymbol{Y}_t$ conditional on the information set $\mathcal{F}_{t-1}$ available at time $t-1$, for $j = 1, ..., p$ and $t = 1, ..., T$. Then, for a given τ_j, we consider the following autoregressive dynamic:

$$\mathcal{Q}_{Y_{tj}}(\tau_j|\mathcal{F}_{t-1}) = \omega_j + \eta_j \mathcal{Q}_{Y_{t-1j}}(\tau_j|\mathcal{F}_{t-2}) + \ell(\boldsymbol{\beta}_j, Y_{t-1j}), \tag{1}$$

where $\omega_j = \omega_j(\tau_j)$, $\eta_j = \eta_j(\tau_j)$ and $\boldsymbol{\beta}_j = \boldsymbol{\beta}_j(\tau_j) = [\beta_{1j}, ..., \beta_{Kj}]'$ are model parameters that depend on the chosen level τ_j and where we suppress the index τ_j for simplicity of notation. The dynamic specification in (1) is well-known in the literature as the CAViaR model of Engle and Manganelli (2004), which aims to compute the τ-th level VaR by estimating the τ-th level quantile of the asset returns through a conditional autoregressive structure. The function $\ell(\cdot)$ represents the so-called News Impact Curve (NIC), originally introduced by Engle and Ng (1993). For each j-th component, the NIC function essentially feeds back the last available observation (Y_{t-1j}) into the present value of the conditional quantile through the $K \times 1$ parameter vector $\boldsymbol{\beta}_j$. Following the CAViaR literature, we will consider different specifications for $\ell(\cdot)$ to model the marginal quantiles, which will be described in the next section.

Using matrix notation, the representation in (1) can be embedded in the following multivariate linear regression model:

$$\boldsymbol{Y}_t = \boldsymbol{\mu}_t + \boldsymbol{\epsilon}_t, \qquad t = 1, 2, ..., T \tag{2}$$

where $\boldsymbol{\epsilon}_t$ denotes a $p \times 1$ vector of error terms, with each marginal quantile (at fixed levels $\tau_1, .., \tau_p$) equal to zero, to ensure that $\boldsymbol{\mu}_t = \mathcal{Q}_{\boldsymbol{Y}_t}(\boldsymbol{\tau}|\mathcal{F}_{t-1})$.

To estimate the regression model in (2), we consider a dynamic generalization of the MAL distribution introduced in Petrella and Raponi (2019) and Kotz et al. (2012); i.e., we consider the time-varying distribution $\text{MAL}_p\left(\boldsymbol{\mu}_t, \boldsymbol{D}_t\tilde{\boldsymbol{\xi}}, \boldsymbol{D}_t\tilde{\boldsymbol{\Sigma}}\boldsymbol{D}_t\right)$, with density function:

$$f_{\boldsymbol{Y}_t}(\boldsymbol{y}_t|\boldsymbol{\mu}_t, \boldsymbol{D}_t\tilde{\boldsymbol{\xi}}, \boldsymbol{D}_t\tilde{\boldsymbol{\Sigma}}\boldsymbol{D}_t, \mathcal{F}_{t-1}) = \frac{2\exp\left\{(\boldsymbol{y}_t - \boldsymbol{\mu}_t)'\boldsymbol{D}_t^{-1}\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\boldsymbol{\xi}}\right\}}{(2\pi)^{p/2}|\boldsymbol{D}_t\tilde{\boldsymbol{\Sigma}}\boldsymbol{D}_t|^{1/2}} \left(\frac{\tilde{m}_t}{2+\tilde{d}}\right)^{\nu/2} K_\nu\left(\sqrt{(2+\tilde{d})\tilde{m}_t}\right). \tag{3}$$

In (3), $\boldsymbol{\mu}_t$ represents the location parameter vector, $\boldsymbol{D}_t\tilde{\boldsymbol{\xi}} \in \mathcal{R}^p$ is the scale (or skew) parameter, with $\boldsymbol{D}_t = \text{diag}[\delta_{t1}, \delta_{t2}, ..., \delta_{tp}]$, $\delta_{tj} > 0$ and $\tilde{\boldsymbol{\xi}} = [\tilde{\xi}_1, \tilde{\xi}_2, ..., \tilde{\xi}_p]'$, having generic element $\tilde{\xi}_j = \frac{1-2\tau_j}{\tau_j(1-\tau_j)}$. $\tilde{\boldsymbol{\Sigma}}$ is a $p \times p$ positive definite matrix such that $\tilde{\boldsymbol{\Sigma}} = \tilde{\boldsymbol{\Lambda}}\boldsymbol{\Psi}\tilde{\boldsymbol{\Lambda}}$, with $\boldsymbol{\Psi}$ having the structure of a correlation matrix[1] and $\tilde{\boldsymbol{\Lambda}} = \text{diag}[\tilde{\sigma}_1, \tilde{\sigma}_1, ..., \tilde{\sigma}_p]$, with $\tilde{\sigma}_j^2 = \frac{2}{\tau_j(1-\tau_j)}$, $j = 1, ..., p$. Moreover, $\tilde{m}_t = (\boldsymbol{y}_t - \boldsymbol{\mu}_t)'(\boldsymbol{D}_t\tilde{\boldsymbol{\Sigma}}\boldsymbol{D}_t)^{-1}(\boldsymbol{y}_t - \boldsymbol{\mu}_t)$, $\tilde{d} = \tilde{\boldsymbol{\xi}}'\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\boldsymbol{\xi}}$, and $K_\nu(\cdot)$ denotes the modified Bessel function of the third kind with index parameter $\nu = (2-p)/2$.

Notice that, as stressed in Petrella and Raponi (2019), the specification in (3) should not be viewed as a parametric assumption in model (2) but rather as a convenient tool to jointly estimate the marginal dynamic quantiles of a multivariate response variable in a quantile regression

[1] In greater detail, $\boldsymbol{\Psi}$ represents the correlation matrix of the (latent) Gaussian process that defines the mixture representation of the MAL distribution (see Equation (9) in Petrella and Raponi (2019)). Moreover, by simple calculations, it is possible to show that the covariance matrix of $\boldsymbol{Y}$ depends on $\boldsymbol{\Psi}$ through the following relationship: $\text{cov}(\boldsymbol{Y}) = \boldsymbol{D}(\tilde{\boldsymbol{\xi}}\tilde{\boldsymbol{\xi}}' + \tilde{\boldsymbol{\Lambda}}\boldsymbol{\Psi}\tilde{\boldsymbol{\Lambda}})\boldsymbol{D}$. In other words, $\boldsymbol{\Psi}$ represents a shifted and scaled version of the sample correlation matrix of $\boldsymbol{Y}$ through the vector $\tilde{\boldsymbol{\xi}}$ and the matrix $\boldsymbol{D}$, respectively.

framework. Moreover, as clarified in their paper, the constraints $\tilde{\xi}_j = \frac{1-2\tau_j}{\tau_j(1-\tau_j)}$ and $\tilde{\sigma}_j^2 = \frac{2}{\tau_j(1-\tau_j)}$ must be imposed to guarantee model identifiability (see Petrella and Raponi (2019), Proposition 2) and to ensure that the dynamic quantile specification in (1) holds, i.e., that $\mathbb{P}(Y_{tj} < \mu_{tj}) = \tau_j$ holds for each $j = 1, 2, ..., p$.

In addition, when these constraints are satisfied, each marginal component of the MAL distribution in (3) follows a univariate AL distribution, that is, $Y_{tj} \sim \text{AL}(\mu_{tj}, \tau_j, \delta_{tj})$, where δ_{tj} represents the time-varying scale parameter of Y_{tj}. This allows us to exploit the result of Taylor (2019), who showed the link between the scale parameter of the AL distribution and the ES risk measure in a univariate framework. By extending these results, we provide new insights on how to estimate the conditional VaR and ES jointly in a multidimensional setting, which accounts for correlations between marginals. This is explained in detail in the next section.

2.2 Modeling VaR and ES jointly

Following Engle and Manganelli (2004), the CAViaR specification in (1) allows us to derive the VaR of an asset at level τ_j by estimating the corresponding quantile at the τ_j-th level, through a conditional autoregressive structure. In what follows, we consider several CAViaR formulations, depending on the choice of the NIC function $\ell(\cdot)$. We then extend the idea of Taylor (2019) to a multivariate setting in order to model and estimate the ES component dynamically.

The CAViaR specifications that we consider are the following:

$$\mathcal{Q}_{Y_{tj}}(\tau_j|\mathcal{F}_{t-1}) = \omega_j + \eta_j \mathcal{Q}_{Y_{t-1j}}(\tau_j|\mathcal{F}_{t-2}) + \beta_{1j}|Y_{t-1j}|, \qquad \textit{Symmetric Absolute Value (SAV)} \tag{4}$$

$$\mathcal{Q}_{Y_{tj}}(\tau_j|\mathcal{F}_{t-1}) = \omega_j + \eta_j \mathcal{Q}_{Y_{t-1j}}(\tau_j|\mathcal{F}_{t-2}) + \beta_{1j}Y^+_{t-1j} + \beta_{2j}Y^-_{t-1j}, \qquad \textit{Asymmetric Slope (AS)} \tag{5}$$

$$\mathcal{Q}_{Y_{tj}}(\tau_j|\mathcal{F}_{t-1}) = \left(\omega_j + \eta_j \mathcal{Q}^2_{Y_{t-1j}}(\tau_j|\mathcal{F}_{t-2}) + \beta_{2j}Y^2_{t-1j}\right)^{1/2} \qquad \textit{Indirect GARCH(1,1) (IG)} \tag{6}$$

where $\boldsymbol{\omega} = [\omega_1, ..., \omega_p]', \boldsymbol{\eta} = [\eta_1, ..., \eta_p]'$ and $\boldsymbol{\beta} = [\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_p]'$, with $\boldsymbol{\beta}_j = [\beta_{1j}, \beta_{2j}]'$, are unknown parameters to be estimated, and where $y^+ = \max(y, 0)$ and $y^- = -\min(y, 0)$, denote the positive and negative parts of y, respectively.

For the ES component, we exploit the interesting link provided in Bassett et al. (2004), which relates univariate quantile regression to conditional ES through the following relation:

$$ES_{tj} = \mathbb{E}[Y_{tj}] - \frac{\mathbb{E}[(Y_{tj} - \mathcal{Q}_{Y_{tj}})(\tau_j - \mathbf{1}_{(Y_{tj} < \mathcal{Q}_{Y_{tj}})})]}{\tau_j} \tag{7}$$

where $\mathbf{1}(\cdot)$ is the indicator function. Following Taylor (2019), the expression in (7) can be rearranged so that the conditional ES can be expressed in terms of the conditional AL scale parameter δ_{tj}.

Specifically, recalling that each marginal of the MAL distribution has a univariate AL density with a conditional scale parameter equal to $\delta_{tj} = \mathbb{E}[(Y_{tj} - \mathcal{Q}_{Y_{tj}})(\tau_j - \mathbf{1}_{(Y_{tj} < \mathcal{Q}_{Y_{tj}})})]$, (7) reduces to:

$$ES_{tj} = \mathbb{E}[Y_{tj}] - \frac{\delta_{tj}}{\tau_j}, \qquad t = 1, 2, ..., T, \quad j = 1, 2, ..., p \tag{8}$$

implying that

$$\delta_{tj} = \tau_j \left(\mathbb{E}[Y_{tj}] - ES_{tj}\right), \qquad t = 1, 2, ..., T, \quad j = 1, 2, ..., p \tag{9}$$

To ensure that each estimated ES does not cross the corresponding estimated quantile, we model the ES in (9) as the product of the quantile and a constant factor (see, e.g., Gourieroux et al. (2012) and Taylor (2019)) as follows:

$$ES_{tj} = (1 + e^{\gamma_{0j}})\mathcal{Q}_{Y_{tj}}(\tau_j), \qquad t = 1, 2, ..., T, \quad j = 1, 2, ..., p, \tag{10}$$

where γ_{0j} is an unconstrained parameter to be estimated such that $1+e^{\gamma_{0j}}$ is greater than 1 and $\boldsymbol{Y}_t$ is assumed to have zero mean. We collect the unknown parameters in the vector $\boldsymbol{\gamma}_0 = [\gamma_{01}, ..., \gamma_{0p}]'$. As explained in Taylor (2019), this formulation correctly describes the relationship between ES and VaR for different data generating processes, such as a GARCH process with a Student-t distribution. Therefore, the representation in (10) provides a simple and parsimonious approach to estimating VaR and ES simultaneously in a dynamic framework.

In (10), however, only the quantile is dynamic, while the factor $1 + e^{\gamma_{0j}}$ remains constant over time. Therefore, to generalize this approach, we also consider the alternative formulation for the ES presented in Taylor (2019), where the difference between the ES and the VaR is modeled using an AutoRegressive (AR) specification as follows:

$$ES_{tj} = \mathcal{Q}_{Y_{tj}}(\tau_j) - x_{tj}, \qquad t = 1, 2, ..., T, \quad j = 1, 2, ..., p, \tag{11}$$

$$x_{tj} = (\gamma_{1j} + \gamma_{2j}(\mathcal{Q}_{Y_{t-1j}}(\tau_j) - Y_{t-1j}) + \gamma_{3j} x_{t-1j})\mathbf{1}_{(Y_{tj} \leq \mathcal{Q}_{Y_{tj}})} + x_{t-1j}\mathbf{1}_{(Y_{tj} > \mathcal{Q}_{Y_{tj}})}, \tag{12}$$

where we define the nonnegative parameter $\boldsymbol{\gamma} = [\boldsymbol{\gamma}_1, ..., \boldsymbol{\gamma}_p]'$, with $\boldsymbol{\gamma}_j = [\gamma_{1j}, \gamma_{2j}, \gamma_{3j}]'$, to ensure that the VaR and ES estimates do not cross.

In the next section, we show how to estimate the model parameters using a ML approach based on a dynamic modification of the EM algorithm proposed by Petrella and Raponi (2019).

2.3 Parameter estimation using the EM algorithm

Before describing the main steps of the EM algorithm, we introduce the notation $\boldsymbol{\mathcal{Q}}_t = \mathcal{Q}_{\boldsymbol{Y}_t}(\boldsymbol{\tau}|\mathcal{F}_{t-1})$, $\boldsymbol{D}_t(\boldsymbol{\gamma})$ and $\tilde{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})$ to clarify that the vector $\mathcal{Q}_{\boldsymbol{Y}_t}(\boldsymbol{\tau}|\mathcal{F}_{t-1})$ and the matrices $\boldsymbol{D}_t$ and $\tilde{\boldsymbol{\Sigma}}$ depend on the unknown parameters $\boldsymbol{\omega}, \boldsymbol{\eta}, \boldsymbol{\beta}, \boldsymbol{\gamma}$ and $\boldsymbol{\Psi}$. The derivation of the EM algorithm is based on Proposition 3 of Petrella and Raponi (2019), properly extended to deal with the autoregressive structure of the quantile function $\boldsymbol{\mathcal{Q}}_t$ and the time dependency of the scale matrix $\boldsymbol{D}_t(\boldsymbol{\gamma})$.

Let $\boldsymbol{\Phi} = \{\boldsymbol{\omega}, \boldsymbol{\eta}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\Psi}\}$ denote the global set of parameters, and define $\hat{\boldsymbol{\Phi}} = \{\hat{\boldsymbol{\omega}}, \hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\Psi}}\}$ as the corresponding set of parameter estimates. For a given vector $\boldsymbol{\tau} = [\tau_1, \tau_2, ..., \tau_p]'$, the expected complete log-likelihood function (up to additive constants), given the observed data $\boldsymbol{Y}_t$ and the parameter estimates $\hat{\boldsymbol{\Phi}}$, is:

$$E\left[l_c(\boldsymbol{\Phi}|\boldsymbol{Y}_t, \hat{\boldsymbol{\Phi}})\right] = -\frac{1}{2}\sum_{t=1}^{T}\log|\boldsymbol{D}_t(\boldsymbol{\gamma})\tilde{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})\boldsymbol{D}_t(\boldsymbol{\gamma})| + \sum_{t=1}^{T}(\boldsymbol{Y}_t - \boldsymbol{\mathcal{Q}}_t)'\boldsymbol{D}_t(\boldsymbol{\gamma})^{-1}\tilde{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})^{-1}\tilde{\boldsymbol{\xi}} \tag{13}$$

$$-\frac{1}{2}\sum_{t=1}^{T} z_t(\boldsymbol{Y}_t - \boldsymbol{\mathcal{Q}}_t)'(\boldsymbol{D}_t(\boldsymbol{\gamma})\tilde{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})\boldsymbol{D}_t(\boldsymbol{\gamma}))^{-1}(\boldsymbol{Y}_t - \boldsymbol{\mathcal{Q}}_t) \tag{14}$$

$$-\frac{1}{2}\tilde{\boldsymbol{\xi}}'\tilde{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})^{-1}\tilde{\boldsymbol{\xi}}\sum_{t=1}^{T} u_t, \tag{15}$$

where

$$u_t = E[W_t|\boldsymbol{Y}_t, \hat{\boldsymbol{\Phi}}] = \left(\frac{\hat{\hat{m}}_t}{2+\hat{\hat{d}}}\right)^{\frac{1}{2}} \frac{K_{\nu+1}\left(\sqrt{(2+\hat{\hat{d}})\hat{\hat{m}}_t}\right)}{K_{\nu}\left(\sqrt{(2+\hat{\hat{d}})\hat{\hat{m}}_t}\right)} \tag{16}$$

$$z_t = E[W_t^{-1}|\boldsymbol{Y}_t, \hat{\boldsymbol{\Phi}}] = \left(\frac{2+\hat{\hat{d}}}{\hat{\hat{m}}_t}\right)^{\frac{1}{2}} \frac{K_{\nu+1}\left(\sqrt{(2+\hat{\hat{d}})\hat{\hat{m}}_t}\right)}{K_{\nu}\left(\sqrt{(2+\hat{\hat{d}})\hat{\hat{m}}_t}\right)} - \frac{2\nu}{\hat{\hat{m}}_t}, \tag{17}$$

with

$$\hat{\hat{m}}_t = (\boldsymbol{y}_t - \boldsymbol{\mathcal{Q}}_t)'(\boldsymbol{D}_t(\hat{\boldsymbol{\gamma}})\tilde{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\Psi}})\boldsymbol{D}_t(\hat{\boldsymbol{\gamma}}))^{-1}(\boldsymbol{y}_t - \boldsymbol{\mathcal{Q}}_t), \qquad \hat{\hat{d}} = \tilde{\boldsymbol{\xi}}'\tilde{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\Psi}})^{-1}\tilde{\boldsymbol{\xi}}, \tag{18}$$

and where W_t follows a standard exponential distribution.

For a given vector $\boldsymbol{\tau}$, the expected complete log-likelihood in (13)-(15) is then maximized with respect to the parameter set $\boldsymbol{\Phi}$, yielding the M-step updates $\hat{\boldsymbol{\Phi}} = \{\hat{\boldsymbol{\omega}}, \hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\Psi}}\}$. Notice that, unlike Petrella and Raponi (2019), closed-form solutions for $\hat{\boldsymbol{\omega}}, \hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$ do not exist, due to the autoregressive structure of the data and, therefore, numerical optimization is required. Updated estimates of $\tilde{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\Psi}})$ can instead be derived using the following expression:

$$\tilde{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\Psi}}) = \frac{1}{T}\sum_{t=1}^{T} z_t\boldsymbol{D}_t(\hat{\boldsymbol{\gamma}})^{-1}(\boldsymbol{Y}_t - \boldsymbol{\mathcal{Q}}_t)(\boldsymbol{Y}_t - \boldsymbol{\mathcal{Q}}_t)'\boldsymbol{D}_t(\hat{\boldsymbol{\gamma}})^{-1} \tag{19}$$

$$+\frac{1}{T}\sum_{t=1}^{T} u_t\tilde{\boldsymbol{\xi}}\tilde{\boldsymbol{\xi}}' - \frac{2}{T}\sum_{t=1}^{T}\boldsymbol{D}_t(\hat{\boldsymbol{\gamma}})^{-1}(\boldsymbol{Y}_t - \boldsymbol{\mathcal{Q}}_t)\tilde{\boldsymbol{\xi}}'. \tag{20}$$

Therefore, the EM algorithm can be implemented as follows:

E-step: Set the iteration number $h = 1$. Fix the vector $\boldsymbol{\tau}$ at the chosen quantile levels $\tau_1, ..., \tau_p$ of interest, and initialize the parameter set $\boldsymbol{\Phi} = \{\boldsymbol{\omega}, \boldsymbol{\eta}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\Psi}\}$. Then, given $\hat{\boldsymbol{\Phi}} = \hat{\boldsymbol{\Phi}}^{(h)} = \{\hat{\boldsymbol{\omega}}^{(h)}, \hat{\boldsymbol{\eta}}^{(h)}, \hat{\boldsymbol{\beta}}^{(h)}, \hat{\boldsymbol{\gamma}}^{(h)}, \hat{\boldsymbol{\Psi}}^{(h)}\}$, at each iteration h, calculate the weights:

$$\hat{u}_t^{(h)} = \left(\frac{\hat{\tilde{m}}_t^{(h)}}{2 + \hat{\tilde{d}}^{(h)}} \right)^{\frac{1}{2}} \frac{K_{\nu+1}\left(\sqrt{(2 + \hat{\tilde{d}}^{(h)})\hat{\tilde{m}}_t^{(h)}} \right)}{K_{\nu}\left(\sqrt{(2 + \hat{\tilde{d}}^{(h)})\hat{\tilde{m}}_t^{(h)}} \right)} \tag{21}$$

$$\hat{z}_t^{(h)} = \left(\frac{2 + \hat{\tilde{d}}^{(h)}}{\hat{\tilde{m}}_t^{(h)}} \right)^{\frac{1}{2}} \frac{K_{\nu+1}\left(\sqrt{(2 + \hat{\tilde{d}}^{(h)})\hat{\tilde{m}}_t^{(h)}} \right)}{K_{\nu}\left(\sqrt{(2 + \hat{\tilde{d}}^{(h)})\hat{\tilde{m}}_t^{(h)}} \right)} - \frac{2\nu}{\hat{\tilde{m}}_t^{(h)}} \tag{22}$$

where

$$\hat{\tilde{m}}_t^{(h)} = (\boldsymbol{y}_t - \boldsymbol{\mathcal{Q}}_t^{(h)})'(\boldsymbol{D}_t(\hat{\boldsymbol{\gamma}}^{(h)})\tilde{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\Psi}}^{(h)})\boldsymbol{D}_t(\hat{\boldsymbol{\gamma}}^{(h)}))^{-1}(\boldsymbol{y}_t - \boldsymbol{\mathcal{Q}}_t^{(h)}), \tag{23}$$

$$\hat{\tilde{d}}^{(h)} = \tilde{\boldsymbol{\xi}}'\tilde{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\Psi}}^{(h)})^{-1}\tilde{\boldsymbol{\xi}}. \tag{24}$$

M-step: Use the estimates $\hat{u}_t^{(h)}$ and $\hat{z}_t^{(h)}$ to maximize $E[l_c(\boldsymbol{\Phi}|\hat{\boldsymbol{\Phi}}^{(h)})]$ with respect to $\boldsymbol{\Phi}$, and obtain the updated set of parameter estimates $\hat{\boldsymbol{\Phi}}^{(h+1)}$.

The optimization procedure is iterated until convergence, that is, when the difference between the likelihood function evaluated at two consecutive iterations is smaller than 10^{-5}. We initialize the EM algorithm by providing the univariate parameter estimates of Taylor (2019) for each asset, while the initial value for the correlation matrix $\boldsymbol{\Psi}$ in (3) is calibrated using the empirical correlation matrix of the data. We fit the univariate models following the estimation procedure in Engle and Manganelli (2004) and Taylor (2019). In addition, we consider a strategy of multiple random starts with 100 different starting points to better explore the parameter space, and we retain the solution corresponding to the maximum likelihood value. This strategy prevents convergence issues and prevents the algorithm from being trapped in local maxima. From an algorithmic point of view, the EM method exploits the Nelder-Mead and Broyden-Fletcher-Goldfarb-Shanno (BFGS) optimization routines to obtain the updated estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, and it uses (19) to compute the updated estimate of $\tilde{\boldsymbol{\Sigma}}$.

The computational analysis was conducted using the `R` (version 4.0.2) software, where the functions to update $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\Sigma}}$ were coded with efficient C++ object-oriented programming.

The validity and performance of the proposed EM algorithm were also assessed using a simulation exercise (see Appendix B).

3 Portfolio construction

In this section, we approach the problem of portfolio allocation. Specifically, we construct the Skewness Mean-Variance (SMV) portfolio of Zhao et al. (2015), taking into account both the multivariate structure and the skewness of asset returns. Following Stolfi et al. (2018) and Zhao et al. (2015), we exploit an interesting property characterizing the MAL distribution in (3). We

show that any linear combination of its marginal components follows a univariate AL distribution whose parameters are a function of the MAL parameters in (3). Note that while the MAL density has thus far been regarded as a convenient tool for estimating the marginal quantiles, in this section, the MAL distribution is used as a data-driven assumption to describe the empirical characteristics of asset returns. As already stated in the Introduction, this choice has been positively accepted in the recent financial literature to detect the peakedness, fat-tails, and skewness of financial assets, overcoming the possible deficiencies of standard approaches relying, for example, on the Gaussian distribution assumption. We then evaluate the riskiness of the selected portfolio by calculating its corresponding VaR and ES using the results of Section 2.2.

3.1 Linear combinations of MAL components

Let us assume that $\boldsymbol{Y}_t$ is a p-dimensional random variable describing the joint dynamics of p variables at time t. Let us consider a linear combination (with weights to be determined) of each component of $\boldsymbol{Y}_t$. Then, the following proposition holds.

Proposition 1. *Let $\boldsymbol{Y}_t \sim MAL_p(\boldsymbol{\mu}_t, \boldsymbol{D}_t\tilde{\boldsymbol{\xi}}, \boldsymbol{D}_t\tilde{\boldsymbol{\Sigma}}\boldsymbol{D}_t)$, with density function defined in (3). Let $\boldsymbol{b}_t = (b_{t1}, ..., b_{tp})' \in \mathcal{R}^p$ be a vector of weights such that $\boldsymbol{b}_t \neq \boldsymbol{0}_p$, with $\boldsymbol{0}_p$ denoting a p-vector of zeros. Define the random variable $Y_t^{\boldsymbol{b}} = \sum_{j=1}^{p} b_{tj} Y_{tj}$. Then,*

$$Y_t^{\boldsymbol{b}} \sim AL(\mu_t^\star, \tau_t^\star, \delta_t^\star) \tag{25}$$

where

$$\mu_t^\star = \boldsymbol{b}_t'\boldsymbol{\mu}_t, \quad \tau_t^\star = \frac{1}{2}\left(1 - \frac{\boldsymbol{b}_t'\boldsymbol{D}_t\tilde{\boldsymbol{\xi}}}{\sqrt{2(\boldsymbol{b}_t'\boldsymbol{D}_t\tilde{\boldsymbol{\Sigma}}\boldsymbol{D}_t\boldsymbol{b}_t) + (\boldsymbol{b}_t'\boldsymbol{D}_t\tilde{\boldsymbol{\xi}})^2}}\right) \quad \text{and} \quad \delta_t^\star = \frac{(\boldsymbol{b}_t'\boldsymbol{D}_t\tilde{\boldsymbol{\Sigma}}\boldsymbol{D}_t\boldsymbol{b}_t)}{2\sqrt{2(\boldsymbol{b}_t'\boldsymbol{D}_t\tilde{\boldsymbol{\Sigma}}\boldsymbol{D}_t\boldsymbol{b}_t) + (\boldsymbol{b}_t'\boldsymbol{D}_t\tilde{\boldsymbol{\xi}})^2}}. \tag{26}$$

Proposition 1 brings out two main considerations. First, the distribution of $Y_t^{\boldsymbol{b}}$ is still AL, which greatly facilitates the computation of the VaR and ES in our context. Second, the parameters of $Y_t^{\boldsymbol{b}}$ are expressed as a function of the multivariate parameters $\boldsymbol{\mu}_t, \boldsymbol{D}_t$ and $\tilde{\boldsymbol{\Sigma}}$ of the MAL distribution in (3). This allows us to take into account the possible association among the marginal components of $\boldsymbol{Y}_t$ when choosing the allocation weights $\boldsymbol{b}_t$. In the next section, we exploit this property to retrieve the distribution of returns of a financial portfolio, whose optimal weights can be derived by solving a simple constrained optimization problem. Given the resulting optimal portfolio weights, we can then use the results of Section 2.2 to derive appropriate measures of the portfolio's VaR and ES.

3.2 The portfolio optimization problem

Assume that $\boldsymbol{Y}_t$ follows the distribution in (3). Given this specification, at each time t, investors may be interested in deriving a portfolio $Y_t^b = \sum_{j=1}^{p} b_{tj} Y_{tj}$ by investing a portion b_{tj} of their capital on the asset Y_{tj} so that $\sum_{j=1}^{p} b_{tj} = 1$. Then, in this setting, the result of Proposition 1 can be applied easily, yielding a portfolio with location, skewness and scale parameters equal to, respectively, $\mu_t^\star$, $\tau_t^\star$ and $\delta_t^\star$ as in (26).

Typically, in risk management applications, the skewness parameter is fixed a priori by the researcher at a certain level (constant over time) $\tau_t^\star = \tilde{\tau}$, as it essentially measures the overall riskiness of a financial product (a portfolio, in our case). Therefore, once we estimate the time-varying MAL parameters from the quantile regression model in (2), for a fixed level of risk $\tilde{\tau}$, the investor's portfolio decision is based on the solution of the selected portfolio strategy. As stated above, to obtain the optimal portfolio allocation, we adopt the SMV strategy of Zhao et al. (2015), which seeks to minimize the portfolio variance and at the same time control for the skewness of the asset returns. Formally, according to Proposition 1, the SMV portfolio solves the following constrained optimization problem:

$$\underset{\boldsymbol{b}_t \in \mathcal{R}^p}{\operatorname{argmin}} \quad \boldsymbol{b}_t' \boldsymbol{D}_t \tilde{\boldsymbol{\Sigma}} \boldsymbol{D}_t \boldsymbol{b}_t \tag{27a}$$

$$\text{s.t.} \quad \tau_t^\star = \tilde{\tau}, \quad \forall t \tag{27b}$$

$$\boldsymbol{b}_t' \mathbf{1}_p = 1 \tag{27c}$$

where $\tilde{\boldsymbol{\Sigma}}$ was introduced in (3) and accounts for the covariance matrix of the returns, while $\boldsymbol{b}_t$ denotes the portfolio's weights at time t held by the investor over the period $[t, t+1)$.

From an empirical point of view, the constraint in (27b) implies that the portfolio weights must be adjusted at each holding period to guarantee that the VaR of the portfolio has a constant level $\tilde{\tau}$, namely, $\mathbb{P}(Y_t^{\boldsymbol{b}} < \mu_t^\star | \mathcal{F}_{t-1}) = \tilde{\tau}$. Once we obtain the optimal portfolio weights for the period $[t, t+1)$, we can compute the conditional portfolio's VaR and ES at level $\tilde{\tau}$ by simply applying the result in (8) to the univariate case.

As explained above, since the parameters $\mu_t^\star, \tau_t^\star$ and $\delta_t^\star$ depend on the parameter estimates of the MAL distribution, information on the dependence structure and on the empirical characteristics embedded in the data is channeled through these estimates into the portfolio's VaR and ES forecasts. This motivates our approach even further, since it can offer an operative and useful tool to help investors and asset managers in deriving optimal portfolio allocations and, at the same time, monitoring multiple VaR and ES jointly.

4 Assessment of VaR and ES forecasts

To assess the performance of VaR and ES predictions jointly, we introduce a new backtesting procedure, based on the multivariate approach discussed in Section 2.

Backtesting techniques are based on quantitative tests that scrutinize model performance in terms of accuracy and precision with respect to a defined criterion. Existing approaches, however, rely on tests that analyze VaR and ES predictions separately; i.e., they focus only on the individual evaluation of one risk measure or the other. VaR evaluation is typically based on coverage tests, which measure the percentage of times that the returns exceed the estimated VaR at a chosen probability level τ (see, e.g., the unconditional coverage (LR_{uc}) test of Kupiec (1995), the conditional coverage (LR_{cc}) test of Christoffersen (1998) and the Dynamic Quantile (DQ) test of Engle and Manganelli (2004)).

To evaluate ES forecasts, the backtesting analysis becomes more complicated since ES is not an elicitable measure (Gneiting, 2011) and therefore suitable scoring functions cannot be determined (Taylor, 2019). The test of McNeil and Frey (2000) is commonly used in this context, which is based on the discrepancy between the observed return and the ES forecast for the periods in which the return exceeds the VaR forecast. Another suitable option is the backtesting procedure of Du and Escanciano (2017), which is based on the Unconditional ES (U_{ES}) and Conditional ES (C_{ES}) tests.

However, since ES relies on observations exceeding the VaR, it is clear that assessment of ES forecasts cannot be independent of the predicted VaR values. This, together with the fact that ES is not elicitable, motivates the introduction of a scoring function for jointly evaluating VaR and ES forecasts. Based on the characterization of consistent scoring functions introduced by Fissler and Ziegel (2016) and Nolde et al. (2017), several scoring rules have been proposed in the literature for the univariate setting (see, e.g., Patton et al. (2019), Fissler et al. (2015) and Taylor (2019)).

In what follows, we provide a new scoring rule that can be used in a multivariate setting to jointly evaluate VaR and ES forecasts of multiple (and possibly correlated) financial assets. To provide support for our proposal of estimating multiple VaR and ES by maximizing the MAL likelihood, we define a new scoring function (S_{MAL}) using the negative of the MAL log score:

$$\begin{aligned} S_{MAL}\left(\boldsymbol{\mathcal{Q}}_t, \boldsymbol{ES}_t, \boldsymbol{y}_t; \tilde{\boldsymbol{\Sigma}}, \boldsymbol{\tau}\right) = \frac{1}{2}\log\left(|\tilde{\boldsymbol{\Sigma}}|\right) + \log\left(|(\boldsymbol{\tau}\boldsymbol{ES}_t') \circ \mathbf{I}_p|\right) - \frac{\nu}{2}\log\left(\frac{\tilde{m}_t}{2+\tilde{d}}\right) \\ +(\boldsymbol{y}_t - \boldsymbol{\mathcal{Q}}_t)'\left((\boldsymbol{\tau}\boldsymbol{ES}_t') \circ \mathbf{I}_p\right)^{-1}\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\boldsymbol{\xi}} - \log\left(K_\nu\left(\sqrt{(2+\tilde{d})\tilde{m}_t}\right)\right) \end{aligned} \tag{28}$$

where $\circ$ denotes the Hadamard product and $\mathbf{I}_p$ represents the identity matrix of order p.

Notice that when p is equal to 1, the S_{MAL} in (28) reduces to the AL log score of Taylor (2019). When $p > 1$, the loss function S_{MAL} allows us to (i) perform a joint assessment of the pairs (VaR, ES) specific to each asset and, at the same time, (ii) control for the existing correlation among

returns.

5 Empirical study

In this section, we apply the methodology presented in Sections 2 and 3 to real data in order to evaluate and compare the empirical implications with those obtained by using a univariate framework. Specifically, we follow Taylor (2019) and use the weekly returns of the FTSE 100, NIKKEI 225, and S&P 500 stock market indices from April 26, 1985, to February 01, 2021. Using a rolling window exercise, we estimate the one-week-ahead VaR and ES forecasts implied by the CAViaR specifications described in Section 2.2, and we select the most desirable model using the Diebold and Mariano (2002) test. In a second empirical exercise, we aggregate the market indices to form a financial portfolio and determine its optimal allocation weights by solving the optimization problem described in Section 3.2. We finally compute and assess the resulting portfolio's conditional VaR and ES for the out-of-sample period, which consists of the last 368 observations of the sample.

5.1 Data

Our sample is collected from Bloomberg, and it consists of 1868 weekly returns for each of the three stock indices. The main summary statistics are displayed in Table 1 below, providing evidence of the well-known stylized facts on fat tails, high kurtosis and serial and cross-sectional correlation that typically characterize financial assets. Moreover, all series exhibit a negative skewness, the Jarque-Bera test significantly rejects the normality assumption, the Ljung-Box test indicates the presence of serial correlation and the Augmented Dickey-Fuller test supports the hypothesis of the absence of unit roots. These results clearly motivate us to consider a quantile regression approach as an investigative tool.

5.2 Out-of-sample VaR and ES forecasting

Using the approach introduced in Section 2, in this section, we derive a joint estimation of VaR and ES for the three stock market indices described above. Specifically, we estimate the out-of-sample series of VaR and ES by considering the three different specifications in (4), (5) and (6), with both the multiplicative factor in (10) and the AR formulation in (11)-(12) for the ES component. Moreover, since we are concerned with the downside risk, we evaluate the out-of-sample forecasts at three different probability levels, namely, $\boldsymbol{\tau} = [0.1, 0.1, 0.1], \boldsymbol{\tau} = [0.05, 0.05, 0.05]$ and $\boldsymbol{\tau} = [0.01, 0.01, 0.01]$.

The first objective is to assess the performance of the CAViaR specifications using the proposed multivariate framework. We start by evaluating the VaR forecasts using the conventional LR_{uc}, LR_{cc} and DQ tests, while we perform the U_{ES} and C_{ES} tests of Du and Escanciano (2017) to

evaluate the ES predictions. The results are shown in Table 2, where Panel A refers to the case of the ES vector modeled as in (10) and Panel B refers to the AR specification in (11)-(12). Looking at the VaR forecasts, in both panels, for all three indices and for all three quantile levels, we find that the CAViaR-AS specification is always successfully backtested at the 5% significance level. The results are less clear for the other two CAViaR specifications. The same results are confirmed when evaluating the ES predictions, as the CAViaR-AS specification again yields outstanding performances for all three indices and for all three quantile levels.

To jointly evaluate the VaR and ES forecasts associated with each stock market index, in addition to the results of the coverage tests, Table 3 reports the values of the loss functions S_{FZN} of Nolde et al. (2017) and S_{FZ0} of Patton et al. (2019) averaged over the out-of-sample period, where:

$$S_{FZN}(\mathcal{Q}_t, ES_t, y_t) = (\mathbf{1}_{(y_t<\mathcal{Q}_t)} - \tau)\frac{\mathcal{Q}_t}{2\tau\sqrt{-ES_t}} - \frac{1}{2\sqrt{-ES_t}}(\mathbf{1}_{(y_t<\mathcal{Q}_t)}\frac{y_t}{\tau} - ES_t) + \sqrt{-ES_t} \quad (29)$$

and

$$S_{FZ0}(\mathcal{Q}_t, ES_t, y_t) = \frac{1}{\tau ES_t}\mathbf{1}_{(y_t<\mathcal{Q}_t)}(y_t - \mathcal{Q}_t) + \frac{\mathcal{Q}_t}{ES_t} + \log(-ES_t) - 1. \quad (30)$$

The losses in (29) and (30) belong to the class of scoring rules proposed in Nolde et al. (2017) and Patton et al. (2019) and have the additional advantage of generating loss differences (between competing forecasts) that are homogeneous of degree 1/2 and zero, respectively.

Overall, the results show that both the CAViaR-AS and CAViaR-IG dynamics are associated with smaller losses compared to the CAViaR-SAV model, except for the case of $\boldsymbol{\tau} = [0.1, 0.1, 0.1]$. Moreover, in line with Taylor (2019), we find evidence of a better forecasting performance when using the constant multiplicative factor $(1 + e^{\gamma_{0j}})$ to model the ES parameter (Panel A) compared to the AR dynamics (Panel B).

Finally, to reinforce our analysis, we evaluate the forecasting performance of the three competing CAViaR models using the scoring function in (28). Specifically, at each time t, and for the specified level $\boldsymbol{\tau}$, we define by $S^{(j)}_{MAL_t}(\boldsymbol{\tau})$ the scoring function associated with model j, and we denote the difference between the scoring functions of model i and model j by $\Delta^{(i,j)}_{MAL,t} = S^{(i)}_{MAL_t}(\boldsymbol{\tau}) - S^{(j)}_{MAL_t}(\boldsymbol{\tau})$, where $i, j = 1, 2, 3$. We then test for the null hypothesis that $\mathbb{E}[\Delta^{(i,j)}_{MAL,t}] = 0$ against $\mathbb{E}[\Delta^{(i,j)}_{MAL,t}] < 0$ using the Diebold and Mariano (2002) test for all the pairs of models i and j. If the null hypothesis is rejected, then the forecasts delivered by model i are more accurate than those of model j, and therefore model i is preferable to model j. The results of the test, together with the corresponding p-values, are reported in Table 4. The table clearly shows that the CAViaR-AS specification outperforms both the CAViaR-IG and CAViaR-SAV models at all the three quantile levels and for both the adopted ES formulations of constant multiplicative factor (Panel A) and AR dynamics (Panel B). Therefore, to select the best performing model between the two CAViaR-AS in Panels A and B, we again apply the Diebold and Mariano (2002) test to the two competing CAViaR-AS specifications. The results are reported in Table 5, and they suggest that

the CAViaR-AS model with the ES specified as a constant multiple of the VaR provides the most accurate predictions at all three quantile levels. This is in line with Taylor (2019), who also found that the same specification not only produces the smallest losses but also delivers the most accurate predictions compared with all the other competing CAViaR dynamics[2]. These results corroborate the fact that accounting for asymmetries in the autoregressive process of a given quantile improves the model's forecasting ability (see, e.g., Engle and Manganelli (2004), Xiliang and Xi (2009), Taylor (2005) and Laporta et al. (2018)).

To show the advantages and the different implications of our approach, we compare our results with those obtained by considering each asset separately, as if we ignored their possible dependence structure. Specifically, the three CAViaR models are estimated individually for each stock market index using the univariate approach of Taylor (2019). To assess the performance of the three models and to combine the individual forecasts of the three indices in a single value, we use the sum of the three corresponding AL log scores (see Taylor (2019)) as a consistent scoring rule. That is, at each time t, and for each model j, we define the following scoring function:

$$S^{(j)}_{AL_t}(\boldsymbol{\tau}) = \sum_{p=1}^{3} S^{(j)}_{AL_{p,t}}(\tau_p) \tag{31}$$

where $S^{(j)}_{AL_{p,t}}(\tau_p)$ denotes the AL log-score of Taylor (2019), corresponding to model j and asset p:

$$S^{(j)}_{AL_{p,t}}(\tau_p) = -\log\left(\frac{\tau_p - 1}{ES^{(j)}_{p,t}}\right) - \frac{\left(y_{p,t} - \mathcal{Q}^{(j)}_{p,t}\right)\left(\tau_p - \mathbf{1}_{(y_{p,t} < \mathcal{Q}^{(j)}_{p,t})}\right)}{\tau_p ES^{(j)}_{p,t}}. \tag{32}$$

As explained in Frongillo and Kash (2015), summing the three AL scoring functions would produce a consistent scoring rule in this case, since each function $S^{(j)}_{AL_{p,t}}(\tau_p)$ elicits the pair (VaR, ES) for the corresponding p-th asset (see Fissler and Ziegel (2016) and Taylor (2019)).
Then, as before, we define the difference between the scoring functions of model i and model j by $\Delta^{(i,j)}_{AL,t} = S^{(i)}_{AL_t}(\boldsymbol{\tau}) - S^{(j)}_{AL_t}(\boldsymbol{\tau})$ and apply the Diebold and Mariano (2002) test to look for the best model in terms of forecasting accuracy. The results are reported in Table 6 (Panels A and B). As shown in the table, the conclusion of the test is now less clear and does not provide any significant evidence in favor of a particular model. This is one of the primary advantages of our approach, as we are able to identify a clear hierarchy among competing models.

A second question of interest concerns the "efficiency gain" of the multiple approach compared to the univariate one. In this sense, we would like to test whether taking into account the association

[2]To further justify this choice, we also compare the CAViaR-AS model with the Quantile AutoRegression of Koenker and Xiao (2006). Specifically, we estimate the regression model of Petrella and Raponi (2019) using the lagged returns (at lag 1) of each asset as the covariates. Comparing these two models would allow one to evaluate the potential contribution of assuming a CAViaR specification in the quantile dynamics. According to the coverage tests and the scoring functions defined in (29) and (30), we still find that the performance of the CAViaR-AS specifications is better.

structure among the market indices would provide us with better predictions in terms of VaR and ES. To do this, we use the backtesting procedure to identify the most "efficient" model, that is, the model producing the best forecasts according to the Diebold and Mariano (2002) test. To measure the efficiency gain, we analyze the difference, if any, in the predictive accuracy between the forecast (VaR, ES) produced by our multivariate approach and the univariate ones. Therefore, for a given CAViaR specification, we test for the difference between the scores obtained with the scoring rule in (31) and those obtained with (28). The null hypothesis is that, on average, the difference is not statistically different from zero, i.e., that the two approaches have the same forecasting performance. The alternative hypothesis is that the difference is smaller than zero, i.e., that the multivariate approach delivers significantly better predictions (smaller losses).

Table 7 shows the resulting test statistics and the corresponding p-values for each of the possible pairs of competing models. Interestingly, for all the considered risk levels and for all three CAViaR specifications, we are always able to reject the null hypothesis at the 5% level, providing evidence of the efficiency gain of our proposed joint approach[3].

To offer a graphical intuition of supporting the results, Figure 1 shows the time series of the difference between the scoring function in (31) that is consistent with the univariate approach and the scoring function proposed in (28) over the whole out-of-sample period and for the three considered CAViaR specifications, with the ES modeled as a multiple of the VaR. The left graph in Figure 1 refers to the case of $\boldsymbol{\tau} = [0.1, 0.1, 0.1]$, the center graph displays the case of $\boldsymbol{\tau} = [0.05, 0.05, 0.05]$, and the right plot displays $\boldsymbol{\tau} = [0.01, 0.01, 0.01]$. The black line represents the difference of the two scoring functions obtained by using the CAViaR-SAV specification in (4), while the red and blue lines refer to the CAViaR-AS and CAViaR-IG dynamics, respectively. The efficiency gain of the multivariate approach clearly emerges from these pictures. Indeed, for all the considered risk levels, and regardless of the dynamic specification of the quantiles, the difference between the two approaches is almost always positive. This confirms the idea that the losses associated with the univariate model can be very large if the dependence structure of the data is not accounted for.

In Figure 2, we display the series of out-of-sample forecasts of the VaR and ES for each of the three stock indices, which are estimated by assuming the selected CAViaR-AS specification in both the univariate and joint approaches. The VaR predictions obtained with the univariate approach of Taylor (2019) are represented by a dotted blue line, while the VaR estimates produced by our joint approach are depicted by the solid red line. The estimated ES is represented by the dotted green line (using the univariate approach) and the solid orange line (using the joint approach). The left panels of Figure 2 refer to $\boldsymbol{\tau} = [0.1, 0.1, 0.1]$, the center panels refer to $\boldsymbol{\tau} = [0.05, 0.05, 0.05]$

[3]As a further robustness check, we also considered the best CAViaR specification for each asset and then applied the Diebold and Mariano (2002) test. The proposed joint method still appears to be more efficient than the univariate method of Taylor (2019), so our findings are unchanged.

and the right ones refer to $\boldsymbol{\tau} = [0.01, 0.01, 0.01]$. The gray dots denote the original return series of each stock index. In all the cases, the estimates of VaR and ES produced by our joint approach lie below the corresponding values obtained with the univariate setting, suggesting that our proposed method can lead to more conservative results.

Finally, to obtain a more intuitive representation of the relationship between the estimated VaR and ES over time and across quantile levels, Figure 3 displays the absolute difference between the out-of-sample VaR and ES forecasts for each of the three stock indices. The plots in the first row are obtained by assuming the CAViaR-AS specification with the ES modeled as in (10), while the plots in the second row refer to the case of the CAViaR-AS specification with the ES following the dynamic in (11)-(12) at the $\boldsymbol{\tau} = [0.1, 0.1, 0.1]$ (left column), $\boldsymbol{\tau} = [0.05, 0.05, 0.05]$ (center column) and $\boldsymbol{\tau} = [0.01, 0.01, 0.01]$ (right column) quantile levels. The blue, red and orange lines refer to the FTSE 100, NIKKEI 225 and S&P 500 stock market indices, respectively, while the gray bands correspond to the main recession periods and to various economic and financial crises that have occurred since 2014. As one can reasonably expect, the difference follows the overall market volatility. Indeed, we find that the difference between the estimated risk measures is typically smaller in calm periods and larger in periods of turbulent markets, with more pronounced upward spikes when the AR dynamics for the ES are used. High volatility is also clearly evident in correspondence and in the aftermath of major economic and financial crises, such as, for example, the Chinese stock market crash at the start of 2016, the Brexit in 2018 and the outbreak of the COVID-19 pandemic in 2020.

Based on these considerations, in the next section, we consider the CAViaR-AS specification in (5) with the ES expressed in (10) to implement the portfolio optimization procedure.

5.3 Out-of-sample portfolio VaR and ES forecasting

In this section, we use the three stock market indices FTSE 100, NIKKEI 225 and S&P 500 to build a SMV portfolio that delivers a certain fixed level of risk $\tilde{\tau}$. The optimal allocation weights are determined by solving the optimization problem described in Section 3.2 using the parameter estimates provided by the CAViaR-AS specification in Section 5.2.

We evaluate the benefits of our approach by considering alternative strategies. First, we use the estimation method of Zhao et al. (2015), where the covariance matrix $\tilde{\boldsymbol{\Sigma}}$ in (27a) is estimated using the sample variance and the sample mean of the return series. We call this strategy Moment-SMV. Second, we evaluate the classic MV of Markowitz (1952). In this case, we model the conditional covariance of the asset returns using several well-known autoregressive dynamics, i.e., the multivariate GARCH Dynamic Conditional Correlation model (Engle 2002) under both the multivariate Normal (MV-G-DCC-N) and Student-t (MV-G-DCC-t) error distributions and the asymmetric Dynamic Conditional Correlation model with multivariate Normal (MV-G-aDCC-N)

and Student-t (MV-G-aDCC-t) errors. Moreover, since the MV strategy can often be inadequate in controlling for asymmetric risk-averse agents, we also consider the above strategies under the multivariate skew Normal (SN) and multivariate skew Student-t (St) distributions of Bauwens and Laurent (2005) as further competing strategies, which we denote by MV-G-DCC-SN, MV-G-DCC-St, MV-G-aDCC-SN and MV-G-aDCC-St, respectively. Then, for each model, we forecast the one-week-ahead conditional covariance matrix and plug it into the portfolio optimization problem.

We jointly estimate the VaR and ES of the resulting portfolios and analyze their out-of-sample performance using the last 368 returns of the sample. The backtesting results for the considered strategies are shown in Table 8, where we report the AL log-score of Taylor (2019) together with the S_{FZN} and S_{FZ0} loss functions in (29) and (30) for the joint evaluation of the pair (VaR, ES). The results clearly show that our approach stands out compared to the other strategies. Indeed, the strategies based on the multivariate Normal and t- distributions and their skewed counterparts produce highly volatile VaR forecasts and suffer from larger average losses over the out-of-sample period. On the other hand, the SMV and Moment-SMV models deliver better performance gains over the MV portfolios, with the SMV being preferred at the three VaR levels, especially in the most extreme case of $\tau = 0.01$. This gain may be traced back to the higher efficiency in the estimation procedure based on the ML approach proposed in Section 2.3. It is worth noting that these conclusions remain valid even when we use the S_{FZ0} and S_{FZN} scoring rules, which do not directly depend on the AL likelihood function.

From a financial viewpoint, in Table 8, we also evaluate the risk-adjusted returns of the competing portfolios, measured by the Sharpe Ratio (SR) and the Herfindahl-Hirchman Index of weights concentration (HHI). We find that the SMV strategy delivers the portfolio with the highest SR and the least concentrated portfolios at both $\tau = 0.1$ and $\tau = 0.05$, with average HHIs of 0.558 and 0.557, respectively. On the other hand, when $\tau = 0.01$, the MV strategy seems to yield the portfolios with the highest SRs, while the Mom-SMV strategy produces the lowest degree of weights concentration.

A graphical representation of the SMV portfolio weights and their evolution over time is provided in Figure 4. In each plot of the figure, the blue line denotes the allocation weights assigned to the FTSE 100, the red line refers to the NIKKEI 225, and the allocation weights of the S&P 500 are displayed in orange. The left panel concerns $\tau = 0.1$, the center one concerns $\tau = 0.05$, and on the right-hand side, we plot the results for $\tau = 0.01$. The SMV strategy tends to invest mainly in the FTSE 100 and the S&P 500, while it tends to hold only a small short position on the NIKKEI 225. It is also interesting that the portfolio weights exhibit the highest volatility in periods of high uncertainty in the market, such as, for example at the end of 2016 and during the ongoing COVID-19 pandemic. Finally, we evaluate the evolution of the wealth generated by the portfolios at the three risk levels during the out-of-sample period. Figure 5 highlights a positive

trend for all quantile levels $\tau = 0.1$ (violet line), $\tau = 0.05$ (green line) and $\tau = 0.01$ (yellow line) from 2014 to 2015 and from 2016 until the outbreak of COVID-19 at the beginning of 2020.

6 Discussion and conclusions

This paper proposes a dynamic joint quantile regression model for estimating the VaR and ES of multiple financial assets in one step, extending the univariate framework of Taylor (2019). To implement the methodology, we suggest a likelihood-based approach based on the MAL density proposed in Petrella and Raponi (2019), generalized to the case of time-varying parameters. This offers a powerful tool to model the dynamics of multiple VaR and ES jointly. Indeed, the location parameter of the MAL density represents the vector of the VaRs, while the ES can be expressed as a simple function of the density scale parameter.

We show that our approach can offer several important advantages, both theoretical and practical. First, it provides a significant gain in terms of estimation efficiency, as it allows us to estimate multiple VaR and ES in just one step. Second, it can significantly improve the forecast accuracy, since it accounts for the dependence structure among financial assets, which cannot be detected by univariate methods. These results are also confirmed empirically. Indeed, using three stock market indices as in Taylor (2019), we estimate the pairs (VaR, ES) for each of the three assets and evaluate the forecasts using a new scoring function based on the MAL density, which allows us to account for the dependence structure among the considered assets at each point in time. The forecasts of VaR and ES are compared with those obtained by the univariate approach of Taylor (2019), i.e., by considering the three stock market indices separately, as they are independent of each other. We find a significant gain in terms of the forecasting accuracy using the proposed multivariate framework, leading to more reliable risk measure estimates.

Following Zhao et al. (2015), we also exploit the properties of the time-varying MAL distribution to derive a new portfolio optimization method, where the optimal allocation weights are adjusted at each holding period to ensure that the portfolio meets a predetermined level of risk. Empirically, we find that our optimization method produces a portfolio with less concentrated allocation weights and a higher Sharpe Ratio than other existing strategies.

Several extensions and generalizations could be analyzed, leaving space for future research. Although we focused only on CAViaR models, one could consider other VaR-based models in the quantile regression framework or specify different ES dynamics where the factor $(1 + e^{\gamma_{tj}})$ varies over time according to an autoregressive process for γ_{tj}. Another interesting research problem concerns the evaluation of the portfolio performance when a larger set of indices is considered, which may help us in providing an empirical ranking based on the VaR and ES forecasts. In this case, a penalized approach, as used, for instance, by Petrella and Raponi (2019), could be

adopted to deal with the curse of dimensionality, improve estimation, achieve greater parsimony and conduct a variable selection procedure. Finally, other portfolio strategies can be implemented, where the choice of the weights may be motivated by other practical considerations or regulatory restrictions.

Acknowledgements

This project has received funding from the postdoctoral fellowships programme Beatriu de Pínos, funded by the Secretary of Universities and research (Government of Catalonia) and by the Horizon 2020 programme of research and innovation of the European union under the Marie Sktodowska-Curie grant agreement No 801370.

Index	Mean	Median	SD	Skewness	Kurtosis	J-B	L-B	ADF
FTSE 100	0.086	0.234	2.396	-1.456	14.717	**17517.036**	**62.674**	**-19.729**
NIKKEI 225	0.046	0.237	2.946	-0.748	6.421	**3383.250**	**175.024**	**-19.159**
S&P 500	0.163	0.320	2.338	-0.947	7.367	**4503.301**	**403.851**	**-20.325**
Correlation matrix								
	FTSE 100	NIKKEI 225	S&P 500					
FTSE 100	1							
NIKKEI 225	0.510	1						
S&P 500	0.709	0.501	1					

Table 1: Summary statistics of the weekly returns of the three indices for the entire sample from April 26, 1985, to February 01, 2021. The test statistics are displayed in boldface when the null hypothesis is rejected at the 1% significance level. J-B, L-B and ADF denote the Jarque-Bera test, the Ljung-Box test on squared returns with 4 lags and the Augmented Dickey-Fuller unit root test with 4 lags, respectively.

τ	[0.1, 0.1, 0.1]					[0.05, 0.05, 0.05]					[0.01, 0.01, 0.01]				
	LR_{uc}	LR_{cc}	DQ	U_{ES}	C_{ES}	LR_{uc}	LR_{cc}	DQ	U_{ES}	C_{ES}	LR_{uc}	LR_{cc}	DQ	U_{ES}	C_{ES}
Panel A: Multiplicative factor for the ES															
SAV															
FTSE 100	**0.043**	**0.455**	**7.194**	**1.467**	**7.735**	**2.658**	**3.469**	**8.288**	**0.128**	**6.711**	**0.756**	**1.111**	**6.350**	5.063	9.991
NIKKEI 225	5.571	**5.838**	10.463	**0.022**	**6.553**	**1.851**	**2.806**	**7.047**	**0.543**	11.086	**2.857**	**4.510**	11.047	**−1.280**	**0.913**
S&P 500	**0.450**	**0.804**	**9.431**	**1.604**	12.776	**1.203**	**3.890**	**8.917**	**0.931**	12.748	4.890	**5.815**	30.388	**−1.229**	9.568
AS															
FTSE 100	**1.067**	**1.210**	**6.641**	**0.663**	**8.773**	**1.203**	**2.314**	**6.266**	**0.434**	**5.243**	**1.096**	**4.184**	**3.351**	**−1.562**	**3.428**
NIKKEI 225	**3.571**	**4.932**	**1.064**	**0.261**	**8.043**	**2.658**	**3.469**	**8.094**	**0.579**	**9.193**	**1.747**	**2.019**	**8.332**	**−1.490**	**2.241**
S&P 500	**0.724**	**0.961**	**7.585**	**1.428**	**8.660**	**0.704**	**2.942**	**7.101**	**1.091**	**8.038**	**0.414**	**2.022**	**6.519**	**1.439**	**0.013**
IG															
FTSE 100	**0.450**	**0.456**	**5.705**	**1.011**	9.926	**2.658**	**3.469**	**8.381**	**0.263**	**6.421**	4.261	**4.399**	9.853	**1.670**	**6.078**
NIKKEI 225	6.547	6.717	11.814	**−0.295**	**8.787**	**1.851**	**2.806**	**6.954**	**0.597**	**9.478**	**1.747**	**2.019**	**5.738**	**−1.195**	**6.915**
S&P 500	**0.724**	**0.961**	10.492	**1.298**	12.164	**1.851**	**5.047**	11.778	**0.579**	**7.730**	**0.952**	**1.309**	**6.009**	**1.699**	**3.150**
Panel B: AR formulation for the ES															
SAV															
FTSE 100	4.589	**4.714**	9.787	**0.184**	**4.671**	**2.587**	**3.490**	11.336	**−1.699**	**2.256**	**1.055**	**2.922**	**9.309**	−3.699	**2.256**
NIKKEI 225	5.329	**5.431**	10.613	**−1.570**	**7.718**	**1.547**	**2.297**	**7.295**	**−1.330**	**2.575**	**2.070**	**4.164**	12.106	4.330	11.575
S&P 500	6.840	8.057	12.670	**0.152**	12.066	**1.476**	**3.516**	**2.591**	1.969	**0.315**	4.020	**5.144**	15.297	5.869	**0.315**
AS															
FTSE 100	**3.441**	**3.337**	**5.771**	**−1.594**	**1.128**	**2.249**	**5.271**	**3.573**	**1.470**	**7.960**	**3.182**	**2.913**	**3.248**	**−1.470**	**7.960**
NIKKEI 225	**3.651**	**5.168**	**6.650**	**−1.646**	**1.631**	**2.251**	**4.451**	**7.473**	**−1.350**	**2.542**	**0.177**	**4.879**	**3.313**	**1.350**	**2.542**
S&P 500	**2.250**	**3.371**	**1.745**	**0.523**	**0.008**	**1.535**	**3.321**	**6.457**	**1.103**	**6.304**	**2.922**	**3.042**	**6.036**	**−1.103**	**6.304**
IG															
FTSE 100	5.214	**1.282**	**9.456**	**0.251**	**7.350**	**2.284**	**3.146**	**8.951**	**1.553**	**5.164**	**3.036**	**3.935**	**9.082**	**1.840**	**4.991**
NIKKEI 225	12.095	6.845	**6.904**	**1.285**	9.997	**1.228**	**3.180**	**6.852**	**1.151**	**4.656**	**2.135**	**3.106**	**6.277**	**1.704**	**2.610**
S&P 500	9.859	**1.564**	**6.456**	**1.732**	**4.639**	**1.731**	**4.584**	10.327	**−1.703**	**1.925**	**1.041**	**5.188**	15.358	**−1.898**	**0.013**

Table 2: Marginal out-of-sample VaR and ES forecast evaluation using the joint approach with the multiplicative factor in (10) (Panel A) and the AR formulation in (11)-(12) (Panel B) for the ES. At the 5% significance level, the critical values of LR_{uc} and LR_{cc} are 3.84 and 5.99, respectively. The U_{ES} is rejected if the test statistic is greater (in absolute value) than 1.96. Finally, the DQ test uses lagged violations at lag 4 while the C_{ES} test considers the first 4 lagged autocorrelations, and the critical value for both is 9.49. The test statistics are displayed in boldface when the null hypotheses are not rejected at the 5% significance level.

$\boldsymbol{\tau}$	$[0.1, 0.1, 0.1]$		$[0.05, 0.05, 0.05]$		$[0.01, 0.01, 0.01]$	
	S_{FZ0}	S_{FZN}	S_{FZ0}	S_{FZN}	S_{FZ0}	S_{FZN}
Panel A: Multiplicative factor for the ES						
SAV						
FTSE 100	1.398	2.661	1.795	3.221	2.109	3.399
NIKKEI 225	1.545	3.091	2.051	3.666	2.231	3.678
S&P 500	1.368	2.552	1.753	3.126	1.907	3.215
AS						
FTSE 100	1.371	2.578	1.750	2.833	1.954	3.254
NIKKEI 225	1.585	2.890	1.949	3.155	2.111	3.636
S&P 500	1.346	2.431	1.706	2.690	1.867	3.142
IG						
FTSE 100	1.383	2.636	1.765	3.000	1.946	3.320
NIKKEI 225	1.539	3.025	2.023	3.387	2.196	3.689
S&P 500	1.357	2.535	1.715	2.943	1.897	3.257
Panel B: AR formulation for the ES						
SAV						
FTSE 100	1.414	2.878	2.116	2.362	4.866	5.740
NIKKEI 225	1.766	2.082	2.241	3.055	5.220	4.779
S&P 500	1.406	2.767	1.940	3.177	4.929	5.241
AS						
FTSE 100	1.496	1.845	1.985	2.325	6.998	5.009
NIKKEI 225	1.714	2.059	2.315	3.731	7.233	4.120
S&P 500	1.478	2.756	2.048	2.944	6.929	4.934
IG						
FTSE 100	1.411	1.849	2.112	2.800	3.201	5.025
NIKKEI 225	1.671	1.993	2.305	3.076	3.432	5.629
S&P 500	1.370	2.764	2.031	3.007	3.164	4.983

Table 3: Marginal out-of-sample VaR and ES forecast evaluation based on the average losses using the scoring functions in (29) and (30) for the joint approach with the multiplicative factor in (10) (Panel A) and the AR formulation in (11)-(12) (Panel B) for the ES.

$\boldsymbol{\tau}$	$[0.1, 0.1, 0.1]$			$[0.05, 0.05, 0.05]$			$[0.01, 0.01, 0.01]$		
	CAViaR-SAV	CAViaR-AS	CAViaR-IG	CAViaR-SAV	CAViaR-AS	CAViaR-IG	CAViaR-SAV	CAViaR-AS	CAViaR-IG
Panel A: Multiplicative factor for the ES									
CAViaR-SAV	-	3.169	2.020	-	4.090	4.630	-	3.832	3.777
	-	(0.999)	(0.978)	-	(1.000)	(1.000)	-	(1.000)	(1.000)
CAViaR-AS	-3.169	-	-2.575	-4.090	-	-2.394	-3.832	-	-1.931
	(0.001)	-	(0.005)	(0.000)	-	(0.009)	(0.000)	-	(0.027)
CAViaR-IG	-2.020	2.575	-	-4.630	2.394	-	-3.777	1.931	-
	(0.022)	(0.995)	-	(0.000)	(0.991)	-	(0.000)	(0.973)	-
Panel B: AR formulation for the ES									
CAViaR-SAV	-	2.635	1.919	-	3.688	4.114	-	3.927	4.728
	-	(0.996)	(0.972)	-	(1.000)	(1.000)	-	(1.000)	(1.000)
CAViaR-AS	-2.635	-	-2.320	-3.688	-	6.090	-3.927	-	6.273
	(0.004)	-	(0.010)	(0.000)	-	(1.000)	(0.000)	-	(1.000)
CAViaR-IG	-1.919	2.320	-	-4.114	-6.090	-	-4.728	-6.273	-
	(0.028)	(0.990)	-	(0.000)	(0.000)	-	(0.000)	(0.000)	-

Table 4: Test statistics and p-values (in parentheses) of the Diebold and Mariano (2002) pairwise test between competing CAViaR models in predicting one-week-ahead returns using the joint approach with the multiplicative factor in (10) (Panel A) and the AR formulation in (11)-(12) (Panel B) for the ES. In each panel, the null hypothesis is that on average, the forecasts obtained with model i are not statistically different from those obtained with model j using the multivariate scoring rule $S^{(j)}_{MAL_t}(\boldsymbol{\tau})$ in (28).

$\boldsymbol{\tau}$	$[0.1, 0.1, 0.1]$	$[0.05, 0.05, 0.05]$	$[0.01, 0.01, 0.01]$
	AR formulation		
	CAViaR-AS	CAViaR-AS	CAViaR-AS
Multiplicative factor			
CAViaR-AS	-1.997	-6.273	-5.029
	(0.023)	(0.000)	(0.000)

Table 5: Test statistics and p-values (in parentheses) of the Diebold and Mariano (2002) pairwise test between the CAViaR-AS specifications using the joint approach with the constant multiplicative factor in (10) and the AR formulation in (11)-(12) for the ES component in predicting one-week-ahead returns. The null hypothesis is that the two approaches have the same forecasting performance.

$\boldsymbol{\tau}$	$[0.1, 0.1, 0.1]$			$[0.05, 0.05, 0.05]$			$[0.01, 0.01, 0.01]$		
	CAViaR-SAV	CAViaR-AS	CAViaR-IG	CAViaR-SAV	CAViaR-AS	CAViaR-IG	CAViaR-SAV	CAViaR-AS	CAViaR-IG
Panel A: Multiplicative factor for the ES									
CAViaR-SAV	-	1.383	0.971	-	1.905	0.938	-	0.920	0.274
	-	(0.916)	(0.834)	-	(0.971)	(0.826)	-	(0.821)	(0.608)
CAViaR-AS	-1.383	-	-1.030	-1.905	-	-1.936	-0.920	-	-0.975
	(0.084)	-	(0.152)	(0.029)	-	(0.027)	(0.179)	-	(0.165)
CAViaR-IG	-0.971	1.030	-	-0.938	1.936	-	-0.274	0.975	-
	(0.166)	(0.848)	-	(0.174)	(0.973)	-	(0.392)	(0.835)	-
Panel B: AR formulation for the ES									
CAViaR-SAV	-	0.179	0.273	-	0.473	0.111	-	0.188	-0.740
	-	(0.571)	(0.608)	-	(0.682)	(0.544)	-	(0.575)	(0.230)
CAViaR-AS	-0.179	-	-0.003	-0.473	-	-0.537	-0.188	-	-0.520
	(0.429)	-	(0.499)	(0.318)	-	(0.296)	(0.425)	-	(0.302)
CAViaR-IG	-0.273	0.003	-	-0.111	0.537	-	0.740	0.520	-
	(0.392)	(0.501)	-	(0.456)	(0.704)	-	(0.770)	(0.698)	-

Table 6: Test statistics and p-values (in parentheses) of the Diebold and Mariano (2002) pairwise test between competing CAViaR models in predicting one-week-ahead returns using the univariate approach of Taylor (2019) with the multiplicative factor in (10) (Panel A) and the AR formulation in (11)-(12) (Panel B) for the ES. In each panel, the null hypothesis is that, on average the forecasts obtained with model i are not statistically different from those obtained with model j using the $S^{(j)}_{AL_t}(\boldsymbol{\tau})$ scoring rule in (31).

τ	$[0.1, 0.1, 0.1]$			$[0.05, 0.05, 0.05]$			$[0.01, 0.01, 0.01]$		
	Univariate approach								
	CAViaR-SAV	CAViaR-AS	CAViaR-IG	CAViaR-SAV	CAViaR-AS	CAViaR-IG	CAViaR-SAV	CAViaR-AS	CAViaR-IG
Joint approach									
Panel A: Multiplicative factor for the ES									
CAViaR-SAV	-4.368	-4.432	-4.488	-2.966	-2.748	-2.890	-3.166	-3.348	-3.290
	(0.000)	(0.000)	(0.000)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)
CAViaR-AS	-4.386	-4.561	-4.601	-3.108	-2.973	-3.127	-3.066	-2.748	-3.190
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)
CAViaR-IG	-4.389	-4.523	-4.573	-3.063	-2.880	-3.029	-2.656	-2.048	-2.690
	(0.000)	(0.000)	(0.000)	(0.001)	(0.002)	(0.001)	(0.004)	(0.020)	(0.005)
Panel B: AR formulation for the ES									
CAViaR-SAV	-8.913	-8.622	-8.764	-6.290	-6.104	-6.118	-4.467	-4.859	-4.433
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
CAViaR-AS	-8.536	-8.278	-8.417	-6.413	-6.472	-6.306	-4.856	-5.353	-4.856
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
CAViaR-IG	-8.737	-8.440	-8.613	-6.530	-6.467	-6.394	-4.894	-5.336	-4.852
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Table 7: Test statistics and p-values (in parentheses) of the Diebold and Mariano (2002) pairwise test between the competing joint and univariate approaches in predicting one-week-ahead returns with the multiplicative factor in (10) (Panel A) and the AR formulation in (11)-(12) (Panel B) for the ES. The null hypothesis is that the two approaches have the same forecasting performance.

Portfolio	Mean	SD	S_{FZ0}	S_{FZN}	S_{AL}	SR	HHI
Panel A: $\boldsymbol{\tau} = [0.1, 0.1, 0.1]$							
SMV	-2.716	0.883	1.317	1.549	2.473	0.225	0.558
Mom-SMV	-2.785	0.859	1.405	1.592	2.486	0.221	0.563
MV-G-DCC-N	-2.519	1.175	1.399	1.621	2.516	0.192	0.680
MV-G-DCC-t	-2.391	1.107	1.380	1.911	2.491	0.198	0.662
MV-G-aDCC-N	-2.520	1.174	1.399	1.622	2.491	0.195	0.697
MV-G-aDCC-t	-2.392	1.107	1.379	1.912	2.487	0.193	0.673
MV-G-DCC-SN	-3.198	1.483	1.419	1.876	2.511	0.187	0.693
MV-G-DCC-St	-3.582	1.674	1.448	2.216	2.544	0.197	0.676
MV-G-aDCC-SN	-3.199	1.482	1.418	1.876	2.511	0.188	0.691
MV-G-aDCC-St	-3.582	1.633	1.459	2.208	2.544	0.200	0.678
Panel B: $\boldsymbol{\tau} = [0.05, 0.05, 0.05]$							
SMV	-4.016	1.240	1.563	1.612	2.529	0.204	0.557
Mom-SMV	-3.759	1.244	1.592	1.615	2.615	0.189	0.562
MV-G-DCC-N	-3.232	1.508	1.696	1.624	2.735	0.192	0.680
MV-G-DCC-t	-3.174	1.476	1.655	2.024	2.696	0.198	0.662
MV-G-aDCC-N	-3.234	1.507	1.695	1.625	2.735	0.195	0.697
MV-G-aDCC-t	-3.176	1.476	1.655	2.025	2.696	0.193	0.673
MV-G-DCC-SN	-3.812	1.770	1.666	1.868	2.706	0.187	0.693
MV-G-DCC-St	-4.429	2.082	1.668	2.336	2.711	0.197	0.676
MV-G-aDCC-SN	-3.809	1.765	1.664	1.871	2.705	0.188	0.691
MV-G-aDCC-St	-4.430	2.027	1.679	2.327	2.722	0.200	0.678
Panel C: $\boldsymbol{\tau} = [0.01, 0.01, 0.01]$							
SMV	-5.227	1.936	1.989	1.656	3.031	0.154	0.533
Mom-SMV	-5.346	2.036	2.018	1.756	3.043	0.141	0.527
MV-G-DCC-N	-4.570	2.132	2.523	2.201	3.524	0.192	0.680
MV-G-DCC-t	-4.913	2.312	2.263	2.073	3.266	0.198	0.662
MV-G-aDCC-N	-4.572	2.132	2.523	2.202	3.522	0.195	0.697
MV-G-aDCC-t	-4.914	2.311	2.263	2.074	3.266	0.193	0.673
MV-G-DCC-SN	-5.004	2.325	2.406	1.865	3.407	0.187	0.693
MV-G-DCC-St	-6.405	3.054	2.125	2.534	3.129	0.197	0.676
MV-G-aDCC-SN	-5.009	2.336	2.395	1.877	3.396	0.188	0.691
MV-G-aDCC-St	-6.395	2.948	2.145	2.503	3.149	0.200	0.678

Table 8: Evaluation of the out-of-sample forecasts of the portfolios VaR and ES. Mean and SD report the average and standard deviation of the portfolio VaR. S_{FZ0}, S_{FZN} and S_{AL} show the average losses using the scoring functions of Patton et al. (2019), Nolde et al. (2017) and Taylor (2019) in (30), (29) and (32), respectively. SR and HHI denote the portfolio Sharpe Ratio and the averaged Herfindahl-Hirschman Index.

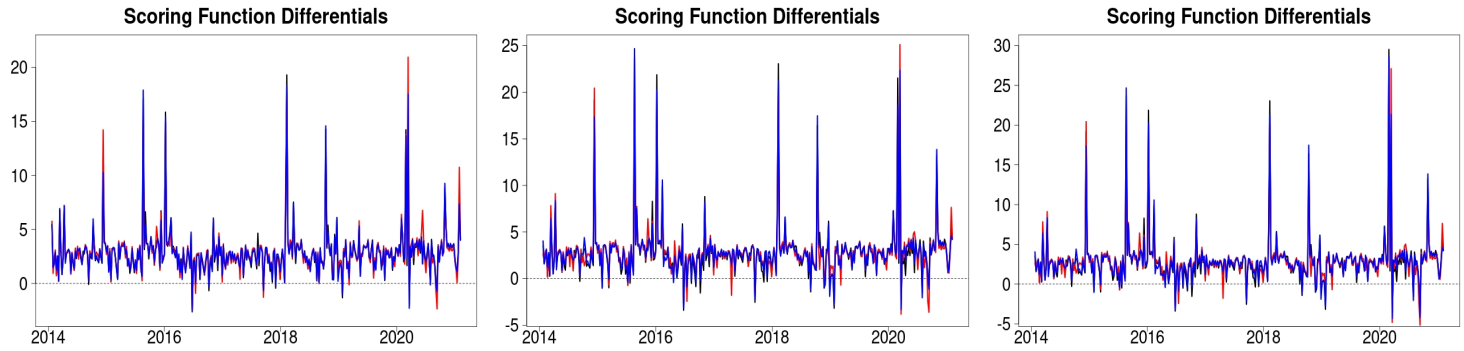

Figure 1: Scoring function differentials, $S_{AL_t}(\boldsymbol{\tau}) - S_{MAL_t}(\boldsymbol{\tau})$, between the $S_{AL_t}(\boldsymbol{\tau})$ loss in (31) of the univariate approach of Taylor (2019) and the $S_{MAL_t}(\boldsymbol{\tau})$ loss in (28) for the joint method, over the out-of-sample period at $\boldsymbol{\tau} = [0.1, 0.1, 0.1]$ (left plot), $\boldsymbol{\tau} = [0.05, 0.05, 0.05]$ (center plot) and $\boldsymbol{\tau} = [0.01, 0.01, 0.01]$ (right plot) for the CAViaR-SAV (black), CAViaR-AS (red) and CAViaR-IG (blue) specifications, with the ES modeled as in (10).

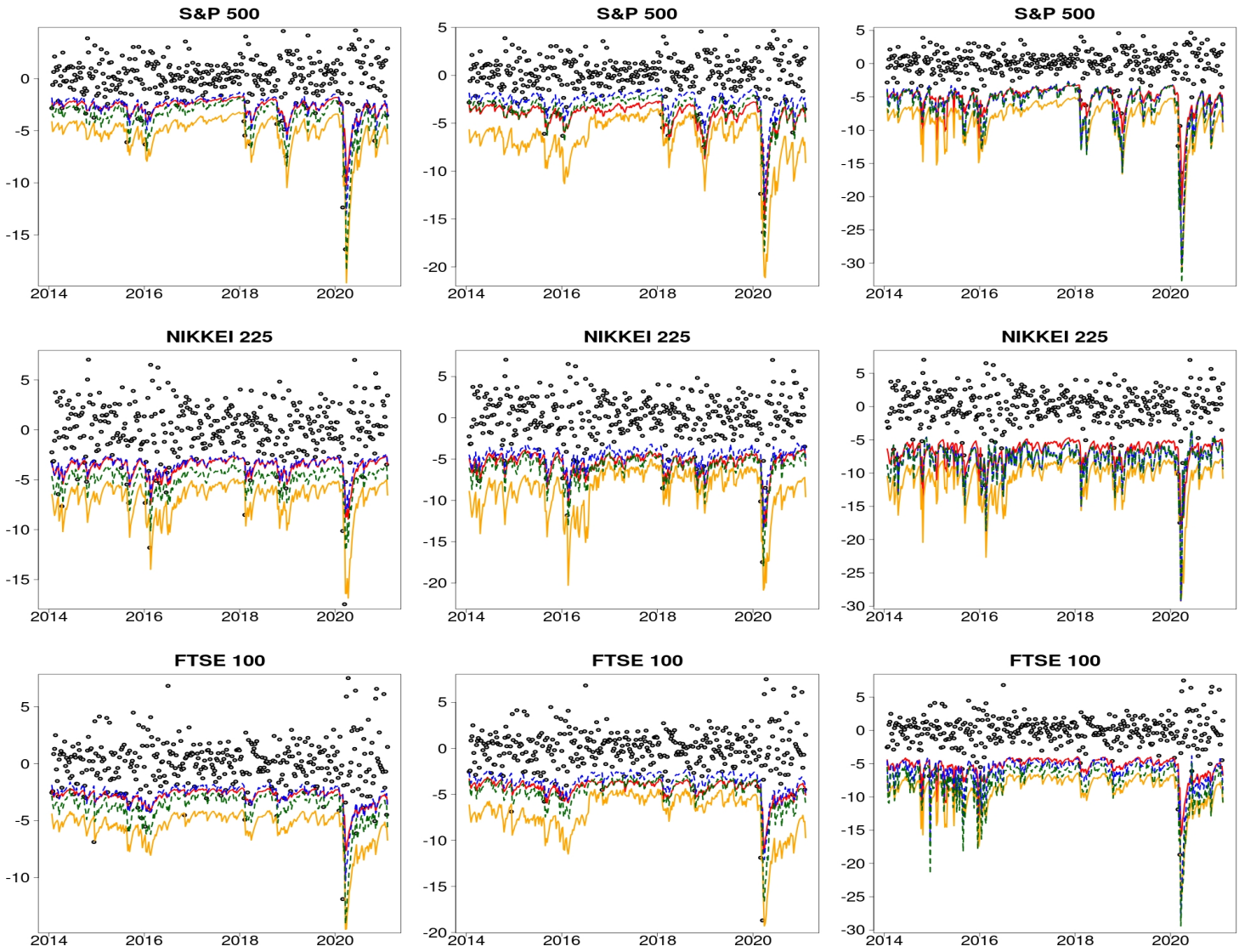

Figure 2: Out-of-sample forecasts of VaR and ES for the three stock indices, estimated with the CAViaR-AS specification using both the univariate and joint approaches, with the ES modeled as in (10). The dotted blue and the solid red lines refer to the VaR predictions, estimated with the univariate and the multiple approach, respectively. The estimated ES is represented by the dotted green line (for the univariate method of Taylor (2019)) and the solid orange line (for the multivariate approach). The left panels refer to $\boldsymbol{\tau} = [0.1, 0.1, 0.1]$, the center panels refer to $\boldsymbol{\tau} = [0.05, 0.05, 0.05]$ and the case of $\boldsymbol{\tau} = [0.01, 0.01, 0.01]$ is displayed in the right panels. The gray dots represent the observed weekly returns for the considered stock index.

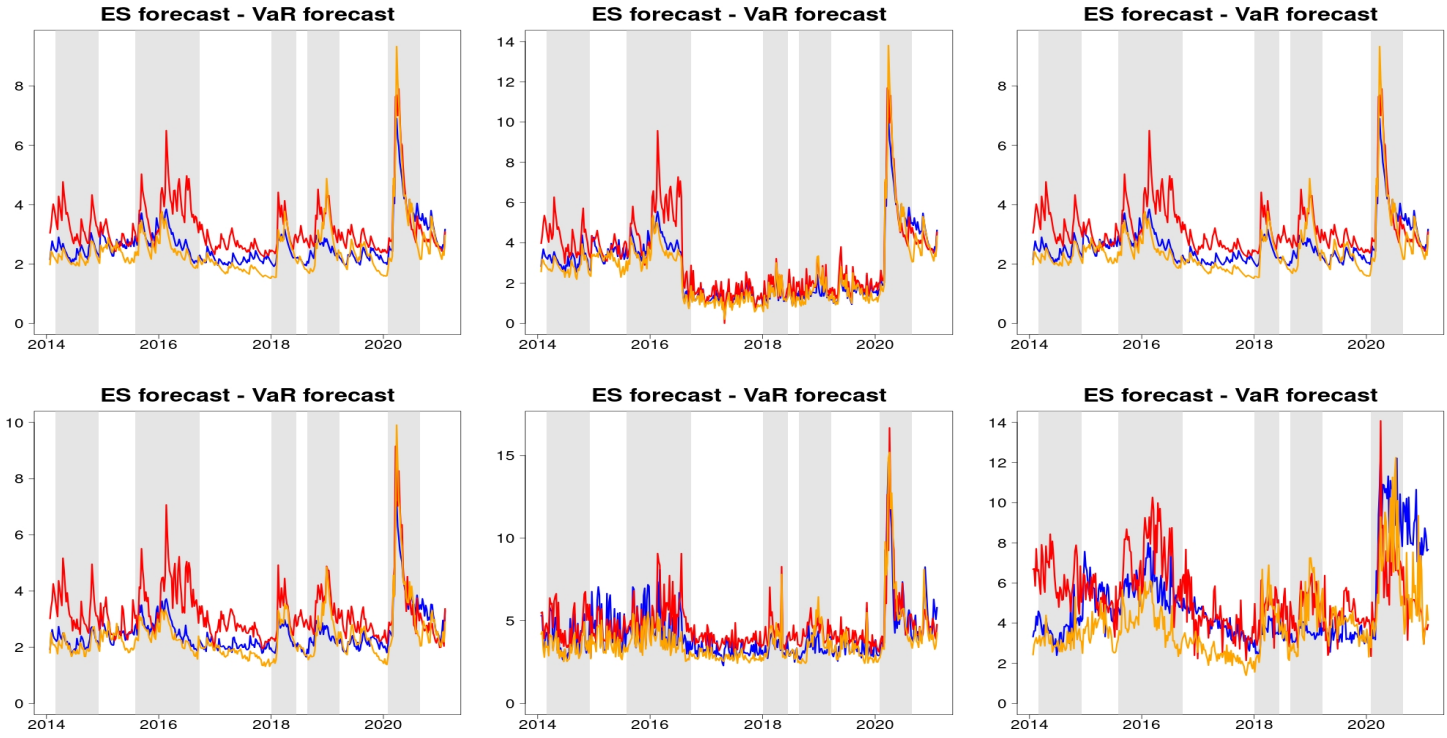

Figure 3: Absolute difference between the out-of-sample VaR and ES forecasts for the three stock indices, estimated with the CAViaR-AS specification and using the ES modeled both as in (10) (first row) and with the AR specification of (11)-(12) (second row), at the $\boldsymbol{\tau} = [0.1, 0.1, 0.1]$ (left column), $\boldsymbol{\tau} = [0.05, 0.05, 0.05]$ (center column) and $\boldsymbol{\tau} = [0.01, 0.01, 0.01]$ (right column) quantile levels. The blue, red and orange lines refer to the FTSE 100, NIKKEI 225 and S&P 500 stock market indices, respectively. The gray bands correspond to the recession dates and to various economic and financial crises that occurred in 2014,03-2015,02; 2015,07-2016,09; 2018,01-2018,06; 2018,08-2019,03; and 2020,02-2020,03.

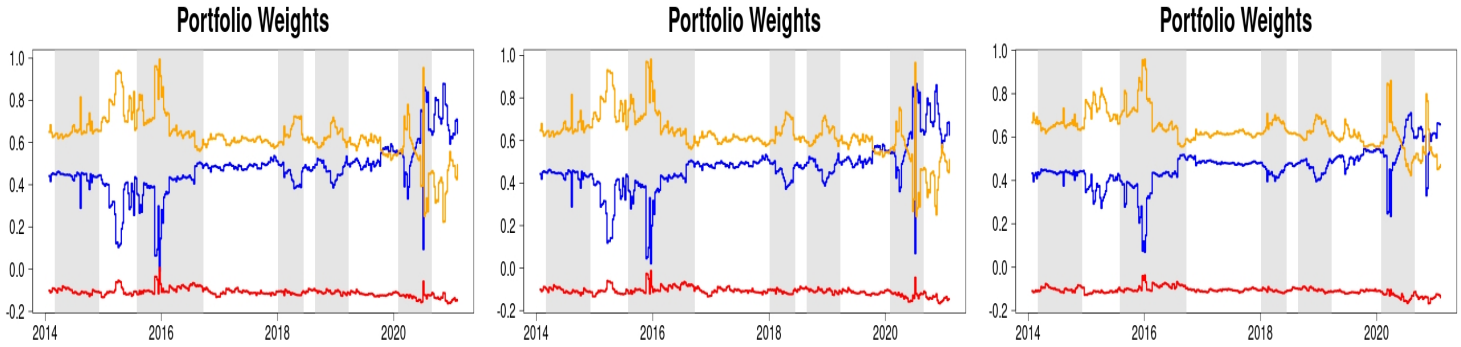

Figure 4: Optimal portfolio weights path over the out-of-sample period computed using the selected CAViaR-AS model at $\tau = 0.1$ (left panel), $\tau = 0.05$ (central panel) and $\tau = 0.01$ (right panel). The optimal portfolio weights comprise the FTSE 100 (blue), NIKKEI 225 (red) and S&P 500 (orange) stock market indices. The gray bands correspond to the recession dates and to various economic and financial crises that occurred in 2014,03-2015,02; 2015,07-2016,09; 2018,01-2018,06; 2018,08-2019,03; and 2020,02-2020,03.

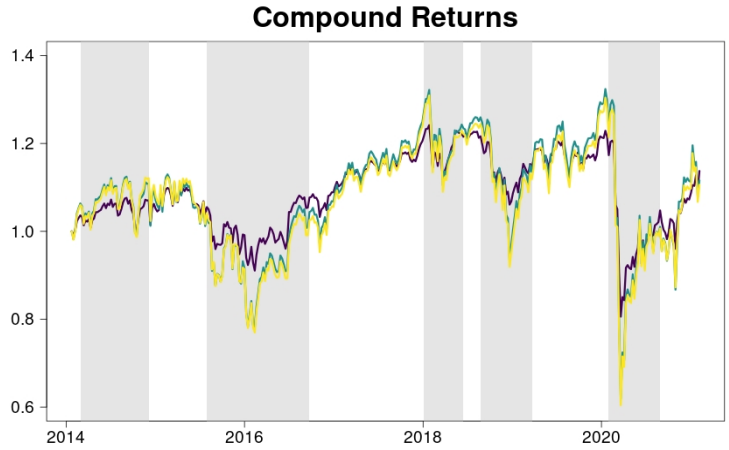

Figure 5: Compound returns over the out-of-sample period computed using the selected CAViaR-AS model at $\tau = 0.1$ (violet), $\tau = 0.05$ (green) and $\tau = 0.01$ (yellow). The gray bands correspond to the recession dates and to various economic and financial crises that occurred in 2014,03-2015,02; 2015,07-2016,09; 2018,01-2018,06; 2018,08-2019,03; and 2020,02-2020,03.

Appendix A. Proof of Proposition 1

As stated in Petrella and Raponi (2019), the $\text{MAL}_p(\boldsymbol{\mu}, \boldsymbol{D}\tilde{\boldsymbol{\xi}}, \boldsymbol{D}\tilde{\boldsymbol{\Sigma}}\boldsymbol{D})$ in (3) can be written as a location-scale mixture, having the following representation:

$$\boldsymbol{Y} = \boldsymbol{\mu} + \boldsymbol{D}\tilde{\boldsymbol{\xi}}W + \sqrt{W}\boldsymbol{D}\tilde{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{Z} \tag{33}$$

where $\tilde{\boldsymbol{\xi}} = [\tilde{\xi}_1, \tilde{\xi}_2, ..., \tilde{\xi}_p]'$, having generic element $\tilde{\xi}_j = \frac{1-2\tau_j}{\tau_j(1-\tau_j)}$, $j = 1, ..., p$. $\tilde{\boldsymbol{\Sigma}}$ is a $p \times p$ positive definite matrix such that $\tilde{\boldsymbol{\Sigma}} = \tilde{\boldsymbol{\Lambda}}\boldsymbol{\Psi}\tilde{\boldsymbol{\Lambda}}$, with $\boldsymbol{\Psi}$ being a correlation matrix and $\tilde{\boldsymbol{\Lambda}} = \text{diag}[\tilde{\sigma}_1, \tilde{\sigma}_1, ..., \tilde{\sigma}_p]$, with $\tilde{\sigma}_j^2 = \frac{2}{\tau_j(1-\tau_j)}$, $j = 1, ..., p$. Finally, $\boldsymbol{Z} \sim \mathcal{N}_p(\boldsymbol{0}_p, \boldsymbol{I}_p)$ denotes a p-variate standard Normal distribution and $W \sim \text{Exp}(1)$ has a standard Exponential distribution, with $\boldsymbol{Z}$ being independent of W. Notice that, under the constraints imposed on $\tilde{\boldsymbol{\xi}}$ and $\tilde{\boldsymbol{\Lambda}}$, the representation in (33) implies that:

$$\boldsymbol{Y} \mid W = w \sim \mathcal{N}_p(\boldsymbol{\mu} + \boldsymbol{D}\tilde{\boldsymbol{\xi}}w, w\boldsymbol{D}\tilde{\boldsymbol{\Sigma}}\boldsymbol{D}). \tag{34}$$

Let $\phi_{\boldsymbol{Y}}(\boldsymbol{t})$ denote the characteristic function of $\boldsymbol{Y}$, with $\boldsymbol{t} \in \mathcal{R}^p$. Using the result in (33), it follows that:

$$\phi_{\boldsymbol{Y}}(\boldsymbol{t}) = \mathbb{E}_W\left[\mathbb{E}_Y[e^{i\boldsymbol{t}'\boldsymbol{Y}} \mid W = w]\right] = \int_0^\infty \mathbb{E}_Z[e^{i\boldsymbol{t}'(\boldsymbol{\mu}+\boldsymbol{D}\tilde{\boldsymbol{\xi}}w+\sqrt{w}\boldsymbol{D}\tilde{\boldsymbol{\Sigma}}^{\frac{1}{2}}\boldsymbol{Z})} \mid W = w]e^{-w}dw. \tag{35}$$

Now, using the conditional distribution of $\boldsymbol{Y}$ given W in (34), we have that:

$$\mathbb{E}_Z[e^{i\boldsymbol{t}'(\boldsymbol{\mu}+\boldsymbol{D}\tilde{\boldsymbol{\xi}}w+\sqrt{w}\boldsymbol{D}\tilde{\boldsymbol{\Sigma}}^{\frac{1}{2}}\boldsymbol{Z})} \mid W = w] = e^{i\boldsymbol{t}'\boldsymbol{\mu}+i\boldsymbol{t}'\boldsymbol{D}\tilde{\boldsymbol{\xi}}w-\frac{w}{2}\boldsymbol{t}'\boldsymbol{D}\tilde{\boldsymbol{\Sigma}}\boldsymbol{D}\boldsymbol{t}}.$$

Substituting this result into (35) yields:

$$\phi_{\boldsymbol{Y}}(\boldsymbol{t}) = e^{i\boldsymbol{t}'\boldsymbol{\mu}} \int_0^\infty e^{-w(1+\frac{1}{2}\boldsymbol{t}'\boldsymbol{D}\tilde{\boldsymbol{\Sigma}}\boldsymbol{D}\boldsymbol{t}-i\boldsymbol{t}'\boldsymbol{D}\tilde{\boldsymbol{\xi}})}dw. \tag{36}$$

Finally, integrating over W, we obtain:

$$\phi_{\boldsymbol{Y}}(\boldsymbol{t}) = e^{i\boldsymbol{t}'\boldsymbol{\mu}}\left(1 + \frac{1}{2}\boldsymbol{t}'\boldsymbol{D}\tilde{\boldsymbol{\Sigma}}\boldsymbol{D}\boldsymbol{t} - i\boldsymbol{t}'\boldsymbol{D}\tilde{\boldsymbol{\xi}}\right)^{-1}. \tag{37}$$

Now, let $\boldsymbol{b} = (b_1, ..., b_p)' \in \mathcal{R}^p$ be a $p \times 1$ vector such that $\boldsymbol{b} \neq \boldsymbol{0}_p$ and consider a new random variable $Y^{\boldsymbol{b}} = \sum_{j=1}^p b_j Y_j = \boldsymbol{b}'\boldsymbol{Y}$, having characteristic function $\phi_{Y^{\boldsymbol{b}}}(z)$, with $z \in \mathcal{R}$. Notice that $Y^{\boldsymbol{b}}$ is a linear transformation of the marginals $Y_1, \ldots, Y_p$. Therefore, the relation $\phi_{Y^{\boldsymbol{b}}}(z) = \phi_{\boldsymbol{b}'\boldsymbol{Y}}(z) = \phi_{\boldsymbol{Y}}(\boldsymbol{b}z)$ holds, since:

$$\phi_{Y^{\boldsymbol{b}}}(z) = e^{iz\boldsymbol{b}'\boldsymbol{\mu}}\left(1 + \frac{1}{2}z^2\boldsymbol{b}'\boldsymbol{D}\tilde{\boldsymbol{\Sigma}}\boldsymbol{D}\boldsymbol{b} - iz\boldsymbol{b}'\boldsymbol{D}\tilde{\boldsymbol{\xi}}\right)^{-1}. \tag{38}$$

The characteristic function in (38) resembles the characteristic function of the AL univariate distribution discussed in Yu and Moyeed (2001) and Kozumi and Kobayashi (2011) where $\mu^\star, \tau^\star, \delta^\star$ are the scale, skewness and scale parameters, respectively. Therefore, the characteristic function of $Y^{\boldsymbol{b}}$ in (38) can be rewritten as the characteristic function of a univariate AL distribution with parameters:

$$\mu^\star = \boldsymbol{b}'\boldsymbol{\mu}, \quad \tau^\star = \frac{1}{2}\left(1 - \frac{\boldsymbol{b}'\boldsymbol{D}\tilde{\boldsymbol{\xi}}}{\sqrt{2(\boldsymbol{b}'\boldsymbol{D}\tilde{\boldsymbol{\Sigma}}\boldsymbol{D}\boldsymbol{b}) + (\boldsymbol{b}'\boldsymbol{D}\tilde{\boldsymbol{\xi}})^2}}\right) \quad \text{and} \quad \delta^\star = \frac{(\boldsymbol{b}'\boldsymbol{D}\tilde{\boldsymbol{\Sigma}}\boldsymbol{D}\boldsymbol{b})}{2\sqrt{2(\boldsymbol{b}'\boldsymbol{D}\tilde{\boldsymbol{\Sigma}}\boldsymbol{D}\boldsymbol{b}) + (\boldsymbol{b}'\boldsymbol{D}\tilde{\boldsymbol{\xi}})^2}}. \tag{39}$$

□

In conclusion, we obtain that $Y^{\boldsymbol{b}} \sim \text{AL}(\mu^\star, \tau^\star, \delta^\star)$ and $\mathbb{P}(Y^{\boldsymbol{b}} < \mu^\star) = \tau^\star$. Consequently, the parameter $\tau^\star$ controls the probability assigned to each side of $Y^{\boldsymbol{b}}$ and $\mu^\star$ is the corresponding quantile at level $\tau^\star$. Notice that the denominator $2(\boldsymbol{b'D\tilde{\Sigma}Db}) + (\boldsymbol{b'D\tilde{\xi}})^2$ in (39) is well defined on the positive real line since $(\boldsymbol{b'D\tilde{\xi}})^2 \geq 0$ and $2(\boldsymbol{b'D\tilde{\Sigma}Db}) > 0$ because $\tilde{\boldsymbol{\Sigma}}$ is a positive definite matrix. Furthermore, when $\boldsymbol{\tau} = [0.5, 0.5, ..., 0.5]$, we have that $\boldsymbol{b'D\tilde{\xi}} = 0$ and (39) simplifies to $\tau^\star = 0.5$ and $\delta^\star = \frac{\sqrt{\boldsymbol{b'D\tilde{\Sigma}Db}}}{2\sqrt{2}}$, which implies that the distribution of $Y^{\boldsymbol{b}}$ is symmetric around $\mu^\star$.

Appendix B. Simulation study

In this Appendix we conduct a simulation study to evaluate the finite sample properties of the proposed method and its ability to jointly estimate the pair (VaR, ES) for multiple correlated assets. This simulation exercise addresses the following issues. First, we consider different distributional choices for the error term to investigate the behavior of the model in the presence of non-Gaussian errors. Second, we evaluate the bias and accuracy of the ML estimators when the interest of the research is focused upon the lower tails of the distribution. Finally, we inspect the impact of dimensionality on both the estimated parameters and the computational burden of the optimization routine.

In the first experiment, we consider a sample size of $T = 1500$ and set $p = 3$. The observations are simulated using the following data generating process:

$$\boldsymbol{Y}_t = \mathcal{Q}_{\boldsymbol{Y}_t}(\boldsymbol{\tau}|\mathcal{F}_{t-1}) + \boldsymbol{\epsilon}_t, \qquad t = 1, 2, ..., T, \tag{40}$$

where $\mathcal{Q}_{\boldsymbol{Y}_t}(\boldsymbol{\tau}|\mathcal{F}_{t-1})$ is generated according to the three different CAViaR specifications described in (4)-(6). For the ES component, we adopted both the multiplicative factor specification in (10) and the AR formulation in (11)-(12). Following Petrella and Raponi (2019), two different simulation scenarios are considered for the error terms $\boldsymbol{\epsilon}_t$ in (40):

(i) a multivariate Normal distribution ($\mathcal{N}_3$) with zero mean and variance-covariance matrix equal to $\boldsymbol{D}_t\tilde{\boldsymbol{\Sigma}}\boldsymbol{D}_t$, that is $\boldsymbol{\epsilon}_{it} \sim \mathcal{N}_3(\boldsymbol{0}, \boldsymbol{D}_t\tilde{\boldsymbol{\Sigma}}\boldsymbol{D}_t)$;

(ii) a multivariate Student-t distribution ($\mathcal{T}_3$) with 5 degrees of freedom, scale parameter $\boldsymbol{D}_t\tilde{\boldsymbol{\Sigma}}\boldsymbol{D}_t$ and non centrality parameter $\boldsymbol{D}_t\tilde{\boldsymbol{\xi}}$, that is, $\boldsymbol{\epsilon}_{it} \sim \mathcal{T}_3(5, \boldsymbol{D}_t\tilde{\boldsymbol{\xi}}, \boldsymbol{D}_t\tilde{\boldsymbol{\Sigma}}\boldsymbol{D}_t)$.

The true values of the CAViaR model and the ES dynamics are calibrated using the real data in the empirical application. Specifically, we set

$$\boldsymbol{\omega} = [-0.20, -0.12, -0.24]', \quad \boldsymbol{\eta} = [0.85, 0.70, 0.60]', \quad \boldsymbol{\beta}_1 = [-0.10, -0.05, -0.20]', \quad \boldsymbol{\beta}_2 = [0.05, 0.10, 0.20]'$$

and

$$\boldsymbol{\gamma}_0 = [-1.1, -1.5, -1.3]', \quad \boldsymbol{\gamma}_1 = [0.05, 0.10, 0.02]', \quad \boldsymbol{\gamma}_2 = [0.12, 0.05, 0.20]' \quad \text{and} \quad \boldsymbol{\gamma}_3 = [0.80, 0.70, 0.60]'.$$

For the CAViaR-IG dynamic, each element of the vector $\boldsymbol{\omega}$ is considered in terms of absolute value to guarantee that the autoregressive process in (6) is well-defined. Finally, we set $\boldsymbol{\Psi} = \begin{bmatrix} 1 & 0.3 & 0.7 \\ 0.3 & 1 & 0.5 \\ 0.7 & 0.5 & 1 \end{bmatrix}$.

Since we are interested in evaluating the downside risk, we analyze three different quantile vectors, namely, $\boldsymbol{\tau} = [0.1, 0.1, 0.1]$, $\boldsymbol{\tau} = [0.05, 0.05, 0.05]$ and $\boldsymbol{\tau} = [0.01, 0.01, 0.01]$. For each model, we

carry out $B = 250$ Monte Carlo replications and report the percentage relative bias (Bias%) and the Root Mean Square Error (RMSE) averaged across the B simulations. Tables 9 and 10 report the results for the parameters $\boldsymbol{\omega}, \boldsymbol{\eta}$ and $\boldsymbol{\beta}$ of the three CAViaR specifications. As can be noted, our estimation method is able to recover the true CAViaR specifications under both the $\mathcal{N}_3$ and $\mathcal{T}_3$ scenarios and for both the considered ES dynamics. Indeed, both the Bias% and the RMSE remain reasonably small under all the different scenarios even though, as expected, their values tend to increase slightly as the quantile level becomes more extreme (due to the reduced information available at the tails of the distribution) and when we consider a heavy-tailed distribution ($\mathcal{T}_3$ scenario). To computationally evaluate the speed of convergence of the EM algorithm, in the last row of each panel, we also report the median number of iterations and CPU Time (in seconds) required by the implemented `R` code using an Intel Xeon E5-2609 2.40GHz processor. The running times range from 9.613 seconds for the simplest SAV specification with a constant multiplicative factor specified for the ES component to 47.242 seconds for the most complex AS model with an autoregressive ES component, confirming the practical feasibility of our optimization algorithm.

Finally, to evaluate the impact of dimensionality on the optimization routine, we considered the simulation experiment in (40) with $T = 1500$ and $\boldsymbol{\tau} = [0.1, 0.1, ..., 0.1]$ and let $p = \{3, 5, 10, 12\}$ grow. Specifically, for each value of p, Tables 11 and 12 report the Bias% and RMSE of the parameter $\theta = || \boldsymbol{\theta} ||$ over 100 Monte Carlo replications, where $\boldsymbol{\theta} = [\boldsymbol{\omega}, \boldsymbol{\eta}, \boldsymbol{\beta}]'$ and $|| \cdot ||$ denote the ℓ_1 norm of $\boldsymbol{\theta}$. As before, we also report the median number of iterations and CPU Time (in seconds) needed to fit the model. As one can easily see, the Bias% remains relatively constant regardless of the value of p, even though, as expected, the RMSE increases slightly with the dimensionality of the problem.

CAViaR	SAV			AS			IG		
$\boldsymbol{\tau}$	$[0.1, 0.1, 0.1]$	$[0.05, 0.05, 0.05]$	$[0.01, 0.01, 0.01]$	$[0.1, 0.1, 0.1]$	$[0.05, 0.05, 0.05]$	$[0.01, 0.01, 0.01]$	$[0.1, 0.1, 0.1]$	$[0.05, 0.05, 0.05]$	$[0.01, 0.01, 0.01]$
Panel A: $\boldsymbol{\epsilon}_{it} \sim \mathcal{N}_3(\mathbf{0}, \boldsymbol{D}_t\tilde{\boldsymbol{\Sigma}}\boldsymbol{D}_t)$									
ω_1	−0.120 (0.027)	0.758 (0.047)	3.874 (0.149)	−1.621 (0.036)	−2.048 (0.050)	−3.472 (0.151)	−1.494 (0.031)	2.520 (0.044)	3.749 (0.156)
ω_2	−1.020 (0.037)	−3.119 (0.052)	2.057 (0.145)	−1.265 (0.030)	−3.738 (0.054)	−4.086 (0.169)	−1.334 (0.036)	1.558 (0.053)	3.622 (0.142)
ω_3	−1.371 (0.028)	−2.245 (0.053)	3.900 (0.162)	−1.454 (0.035)	−2.617 (0.063)	−3.238 (0.147)	−1.195 (0.033)	2.751 (0.048)	3.964 (0.148)
η_1	1.184 (0.025)	1.337 (0.057)	−1.528 (0.153)	2.241 (0.030)	2.379 (0.049)	−3.185 (0.155)	0.548 (0.048)	−2.690 (0.059)	−2.941 (0.159)
η_2	1.228 (0.035)	1.753 (0.041)	−2.537 (0.143)	−0.179 (0.034)	−1.855 (0.062)	1.992 (0.199)	1.794 (0.039)	−2.314 (0.041)	−3.570 (0.144)
η_3	1.560 (0.027)	2.472 (0.036)	−1.132 (0.174)	1.830 (0.041)	2.991 (0.062)	4.230 (0.178)	1.981 (0.031)	2.220 (0.054)	2.781 (0.153)
β_{11}	−1.573 (0.031)	−2.688 (0.038)	3.971 (0.155)	−1.409 (0.035)	−3.308 (0.056)	−3.366 (0.168)	−0.141 (0.040)	2.704 (0.041)	3.497 (0.162)
β_{12}	−0.841 (0.033)	−1.521 (0.060)	2.459 (0.140)	−0.517 (0.031)	−0.797 (0.045)	1.871 (0.159)	1.427 (0.045)	1.981 (0.062)	2.082 (0.185)
β_{13}	−1.799 (0.045)	−2.066 (0.057)	3.211 (0.165)	−1.439 (0.039)	−1.937 (0.055)	3.540 (0.167)	1.899 (0.041)	2.853 (0.073)	3.198 (0.159)
β_{21}				1.862 (0.033)	2.008 (0.046)	5.335 (0.147)			
β_{22}				−0.187 (0.039)	1.705 (0.054)	4.059 (0.161)			
β_{23}				0.634 (0.039)	1.089 (0.049)	3.899 (0.158)			
Iterations	6	8	12	8	14	27	6	9	14
CPU Time	9.613	11.824	17.090	11.498	22.830	33.933	9.669	12.816	18.159
Panel B: $\boldsymbol{\epsilon}_{it} \sim \mathcal{T}_3(5, \boldsymbol{D}_t\tilde{\boldsymbol{\xi}}, \boldsymbol{D}_t\tilde{\boldsymbol{\Sigma}}\boldsymbol{D}_t)$									
ω_1	−1.196 (0.039)	−2.393 (0.063)	2.996 (0.156)	−0.196 (0.049)	−2.898 (0.051)	−2.393 (0.168)	−1.291 (0.029)	1.996 (0.056)	3.560 (0.163)
ω_2	−1.833 (0.045)	−1.572 (0.059)	1.385 (0.162)	−2.833 (0.035)	−2.380 (0.052)	−4.572 (0.179)	−1.533 (0.034)	1.385 (0.052)	2.608 (0.160)
ω_3	−1.397 (0.036)	−1.297 (0.041)	3.736 (0.168)	−2.397 (0.046)	−2.120 (0.061)	−4.297 (0.151)	−1.558 (0.032)	1.736 (0.068)	1.825 (0.169)
η_1	2.699 (0.051)	0.292 (0.067)	−2.097 (0.170)	2.699 (0.041)	2.266 (0.057)	2.292 (0.167)	1.132 (0.027)	−2.097 (0.050)	−1.940 (0.164)
η_2	1.050 (0.048)	1.849 (0.068)	−4.367 (0.177)	0.050 (0.038)	1.421 (0.047)	1.849 (0.198)	1.805 (0.037)	−2.367 (0.047)	−3.034 (0.153)
η_3	1.508 (0.042)	2.455 (0.061)	3.581 (0.167)	1.508 (0.032)	1.783 (0.049)	3.455 (0.151)	1.335 (0.036)	1.581 (0.057)	2.243 (0.140)
β_{11}	−1.718 (0.030)	−0.402 (0.050)	2.688 (0.154)	−2.718 (0.049)	−3.440 (0.068)	−3.702 (0.160)	−1.908 (0.036)	2.688 (0.054)	1.954 (0.162)
β_{12}	−1.615 (0.048)	−0.402 (0.052)	3.189 (0.163)	−1.615 (0.038)	−2.691 (0.064)	−2.402 (0.147)	−0.940 (0.034)	1.189 (0.053)	2.514 (0.189)
β_{13}	−2.557 (0.037)	1.942 (0.049)	3.087 (0.180)	−2.157 (0.047)	−2.542 (0.060)	1.942 (0.199)	1.769 (0.041)	2.087 (0.080)	3.941 (0.170)
β_{21}				1.371 (0.059)	1.057 (0.056)	3.782 (0.170)			
β_{22}				−0.803 (0.040)	−1.690 (0.052)	1.166 (0.166)			
β_{23}				0.489 (0.039)	1.607 (0.065)	2.210 (0.163)			
Iterations	7	10	15	11	16	28	7	11	16
CPU Time	9.617	12.799	18.141	11.691	22.983	34.631	9.628	13.812	18.385

Table 9: Bias% and RMSE (in brackets) of the point estimates for $\boldsymbol{\omega}, \boldsymbol{\eta}$ and $\boldsymbol{\beta}$ of the three CAViaR specifications in (4)-(6) with the ES modeled as a multiple of VaR as in (10), under the $\mathcal{N}_3$ and $\mathcal{T}_3$ scenarios. The last two rows of each panel show the median number of iterations and CPU Time (in seconds) required to fit the model using a single run of the EM algorithm.

CAViaR	SAV			AS			IG		
$\boldsymbol{\tau}$	$[0.1, 0.1, 0.1]$	$[0.05, 0.05, 0.05]$	$[0.01, 0.01, 0.01]$	$[0.1, 0.1, 0.1]$	$[0.05, 0.05, 0.05]$	$[0.01, 0.01, 0.01]$	$[0.1, 0.1, 0.1]$	$[0.05, 0.05, 0.05]$	$[0.01, 0.01, 0.01]$
Panel A: $\boldsymbol{\epsilon}_{it} \sim \mathcal{N}_3(\mathbf{0}, \boldsymbol{D}_t\tilde{\boldsymbol{\Sigma}}\boldsymbol{D}_t)$									
ω_1	−0.688 (0.037)	−2.193 (0.094)	−3.853 (0.155)	−2.266 (0.060)	2.609 (0.156)	2.245 (0.196)	0.505 (0.059)	−2.726 (0.075)	2.522 (0.150)
ω_2	−0.413 (0.037)	−2.936 (0.053)	−3.462 (0.165)	−1.189 (0.065)	2.139 (0.055)	3.807 (0.191)	−1.986 (0.077)	−2.309 (0.084)	3.130 (0.162)
ω_3	−1.796 (0.055)	−1.046 (0.063)	−3.785 (0.145)	−1.949 (0.057)	1.749 (0.073)	2.424 (0.165)	−1.162 (0.064)	−4.810 (0.121)	3.644 (0.153)
η_1	1.691 (0.037)	2.940 (0.061)	2.999 (0.140)	0.642 (0.061)	−3.533 (0.067)	−2.804 (0.199)	1.430 (0.040)	−2.116 (0.074)	−3.672 (0.162)
η_2	1.884 (0.043)	−2.023 (0.050)	−2.133 (0.177)	−1.217 (0.052)	−2.763 (0.091)	−4.777 (0.140)	−1.813 (0.054)	−1.854 (0.119)	−3.088 (0.155)
η_3	1.018 (0.065)	−1.174 (0.057)	3.017 (0.162)	−3.500 (0.065)	−2.376 (0.045)	−2.197 (0.150)	1.158 (0.081)	−2.133 (0.136)	−2.316 (0.167)
β_{11}	−1.850 (0.048)	−3.514 (0.045)	3.345 (0.133)	1.850 (0.064)	1.395 (0.053)	4.856 (0.176)	−2.688 (0.060)	1.699 (0.102)	2.803 (0.189)
β_{12}	−1.609 (0.041)	−2.701 (0.057)	2.465 (0.186)	2.602 (0.066)	3.041 (0.091)	3.230 (0.141)	−2.357 (0.074)	−2.640 (0.128)	2.921 (0.182)
β_{13}	−1.224 (0.038)	−1.291 (0.045)	2.808 (0.164)	1.659 (0.063)	2.407 (0.067)	4.289 (0.119)	−1.928 (0.065)	−2.191 (0.117)	4.053 (0.143)
β_{21}				1.320 (0.058)	−2.763 (0.141)	−2.013 (0.193)			
β_{22}				1.599 (0.069)	−2.484 (0.121)	−3.276 (0.162)			
β_{23}				1.082 (0.060)	−2.179 (0.118)	3.936 (0.174)			
Iterations	12	14	19	15	23	35	10	15	18
CPU Time	13.791	20.739	25.738	16.112	22.956	45.946	13.796	20.781	25.784
Panel B: $\boldsymbol{\epsilon}_{it} \sim \mathcal{T}_3(5, \boldsymbol{D}_t\tilde{\boldsymbol{\xi}}, \boldsymbol{D}_t\tilde{\boldsymbol{\Sigma}}\boldsymbol{D}_t)$									
ω_1	−1.687 (0.049)	−3.265 (0.121)	−3.792 (0.151)	−2.793 (0.037)	2.309 (0.126)	3.497 (0.215)	−1.006 (0.038)	−2.278 (0.117)	3.580 (0.142)
ω_2	−2.891 (0.078)	−2.992 (0.084)	−4.143 (0.189)	−1.244 (0.048)	4.041 (0.121)	4.667 (0.223)	−2.308 (0.069)	−3.344 (0.102)	2.875 (0.127)
ω_3	−1.169 (0.087)	−2.719 (0.128)	−4.852 (0.156)	−1.529 (0.070)	1.858 (0.093)	2.045 (0.216)	−1.845 (0.044)	−2.442 (0.125)	4.644 (0.194)
η_1	2.834 (0.038)	3.721 (0.096)	−3.089 (0.157)	1.081 (0.057)	2.490 (0.171)	−3.874 (0.190)	1.595 (0.051)	2.387 (0.093)	−4.524 (0.146)
η_2	−1.345 (0.062)	−2.788 (0.166)	−4.423 (0.154)	−1.449 (0.063)	−2.036 (0.176)	−3.810 (0.215)	0.404 (0.059)	−2.002 (0.086)	−4.258 (0.189)
η_3	1.642 (0.067)	−2.089 (0.094)	−2.026 (0.144)	−1.104 (0.073)	2.716 (0.096)	−2.981 (0.221)	2.186 (0.057)	3.929 (0.106)	2.565 (0.192)
β_{11}	−2.171 (0.046)	−3.313 (0.091)	2.090 (0.132)	0.920 (0.066)	3.032 (0.133)	3.917 (0.186)	−1.893 (0.040)	−3.994 (0.138)	3.613 (0.161)
β_{12}	−2.212 (0.050)	−3.605 (0.112)	1.394 (0.146)	1.261 (0.082)	2.959 (0.090)	3.883 (0.197)	−1.525 (0.047)	−2.143 (0.104)	3.346 (0.181)
β_{13}	−1.536 (0.042)	−2.111 (0.085)	3.773 (0.129)	1.971 (0.082)	3.189 (0.145)	2.862 (0.184)	−2.807 (0.054)	−3.592 (0.092)	4.842 (0.164)
β_{21}				1.171 (0.051)	1.847 (0.137)	3.866 (0.177)			
β_{22}				1.229 (0.067)	2.621 (0.154)	3.195 (0.196)			
β_{23}				1.456 (0.063)	2.158 (0.133)	3.535 (0.187)			
Iterations	13	14	20	16	24	37	11	16	20
CPU Time	13.797	20.760	25.790	16.783	22.176	47.242	13.794	20.752	26.766

Table 10: Bias% and RMSE (in brackets) of the point estimates for $\boldsymbol{\omega}, \boldsymbol{\eta}$ and $\boldsymbol{\beta}$ of the three CAViaR specifications in (4)-(6) with the AR process for the ES as in (11)-(12) under the $\mathcal{N}_3$ and $\mathcal{T}_3$ scenarios. The last two rows of each panel show the median number of iterations and CPU Time (in seconds) required to fit the model using a single run of the EM algorithm.

	$p=3$			$p=5$			$p=10$			$p=12$		
CAViaR	SAV	AS	IG	SAV	AS	IG	SAV	AS	IG	SAV	AS	IG
Panel A: $\boldsymbol{\epsilon}_{it} \sim \mathcal{N}_3(\mathbf{0}, \boldsymbol{D}_t\tilde{\boldsymbol{\Sigma}}\boldsymbol{D}_t)$												
Bias%	0.754	0.672	0.855	0.697	0.742	0.748	0.803	0.729	0.764	0.869	0.680	0.654
RMSE	0.083	0.088	0.075	0.118	0.390	0.272	0.299	0.267	0.315	0.351	0.391	0.341
Iterations	6	9	6	12	14	11	18	27	22	47	55	48
CPU Time	9.812	11.746	9.756	46.207	60.240	47.248	88.980	102.840	98.040	126.540	137.880	120.660
Panel B: $\boldsymbol{\epsilon}_{it} \sim \mathcal{T}_3(5, \boldsymbol{D}_t\tilde{\boldsymbol{\xi}}, \boldsymbol{D}_t\tilde{\boldsymbol{\Sigma}}\boldsymbol{D}_t)$												
Bias%	0.837	0.809	0.785	0.793	0.884	0.895	0.734	0.831	0.702	0.893	0.708	0.929
RMSE	0.095	0.085	0.083	0.207	0.375	0.321	0.372	0.396	0.380	0.362	0.379	0.442
Iterations	6	9	7	11	16	12	18	29	23	48	57	50
CPU Time	9.783	11.742	9.776	47.900	64.440	48.944	84.720	98.840	96.760	130.200	164.100	155.460

Table 11: Bias% and RMSE of the point estimates of $\theta = \| \boldsymbol{\theta} \|$, where $\boldsymbol{\theta} = [\boldsymbol{\omega}, \boldsymbol{\eta}, \boldsymbol{\beta}]'$, for different values of p and for the three CAViaR specifications in (4)-(6), with the ES modeled as a multiple of VaR as in (10). Panels A and B refer to the $\mathcal{N}_3$ and $\mathcal{T}_3$ scenarios, respectively, where the last two rows of each panel show the median number of iterations and CPU Time (in seconds) required to fit the model using a single run of the EM algorithm.

	$p=3$			$p=5$			$p=10$			$p=12$		
CAViaR	SAV	AS	IG	SAV	AS	IG	SAV	AS	IG	SAV	AS	IG
Panel A: $\boldsymbol{\epsilon}_{it} \sim \mathcal{N}_3(\mathbf{0}, \boldsymbol{D}_t\tilde{\boldsymbol{\Sigma}}\boldsymbol{D}_t)$												
Bias%	0.786	0.791	0.773	0.731	0.865	0.671	0.912	0.713	0.792	0.803	0.753	0.800
RMSE	0.140	0.149	0.147	0.265	0.279	0.232	0.243	0.322	0.223	0.364	0.358	0.437
Iterations	12	13	11	23	24	21	38	37	35	67	75	68
CPU Time	12.617	15.799	13.141	82.140	89.400	85.400	109.680	168.240	118.320	196.620	303.360	196.540
Panel B: $\boldsymbol{\epsilon}_{it} \sim \mathcal{T}_3(5, \boldsymbol{D}_t\tilde{\boldsymbol{\xi}}, \boldsymbol{D}_t\tilde{\boldsymbol{\Sigma}}\boldsymbol{D}_t)$												
Bias%	0.722	0.729	0.742	0.704	0.783	0.639	0.737	0.765	0.762	0.828	0.734	0.633
RMSE	0.152	0.180	0.152	0.256	0.297	0.302	0.315	0.356	0.283	0.393	0.425	0.395
Iterations	12	14	12	25	23	23	38	39	43	68	77	70
CPU Time	12.681	16.824	13.090	89.880	94.920	83.652	106.260	168.360	117.960	177.420	314.840	258.620

Table 12: Bias% and RMSE of the point estimates of $\theta = || \boldsymbol{\theta} ||$, where $\boldsymbol{\theta} = [\boldsymbol{\omega}, \boldsymbol{\eta}, \boldsymbol{\beta}]'$, for different values of p and for the three CAViaR specifications in (4)-(6), with the AR process for the ES as in (11)-(12). Panels A and B refer to the $\mathcal{N}_3$ and $\mathcal{T}_3$ scenarios, respectively, where the last two rows of each panel show the median number of iterations and CPU Time (in seconds) required to fit the model using a single run of the EM algorithm.

References

Acerbi, C. and D. Tasche (2002). On the coherence of Expected Shortfall. *Journal of Banking & Finance 26*(7), 1487–1503.

Alexander, G. J. and A. M. Baptista (2008). Active portfolio management with benchmarking: Adding a Value-at-Risk constraint. *Journal of Economic Dynamics and Control 32*(3), 779 – 820.

Arditti, F. D. (1971). Another look at mutual fund performance. *Journal of Financial and Quantitative Analysis 6*(3), 909–912.

Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath (1997). Thinking coherently. *Risk 10*(11), 68–72.

Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath (1999). Coherent measures of risk. *Mathematical Finance 9*(3), 203–228.

Barone-Adesi, G. (1985). Arbitrage equilibrium with skewed asset returns. *Journal of Financial and Quantitative Analysis 20*(3), 299–313.

Bassett, G. W., R. Koenker, and G. Kordas (2004). Pessimistic portfolio allocation and choquet expected utility. *Journal of Financial Econometrics 2*(4), 477–492.

Baur, D. G. (2013). The structure and degree of dependence: A quantile regression approach. *Journal of Banking & Finance 37*(3), 786–798.

Bauwens, L. and S. Laurent (2005). A new class of multivariate skew densities, with application to generalized autoregressive conditional heteroscedasticity models. *Journal of Business & Economic Statistics 23*(3), 346–354.

Bernardi, M., G. Gayraud, L. Petrella, et al. (2015). Bayesian tail risk interdependence using quantile regression. *Bayesian Analysis 10*(3), 553–603.

Bonaccolto, G., M. Caporin, and S. Paterlini (2019). Decomposing and backtesting a flexible specification for CoVaR. *Journal of Banking & Finance 108*, 105659.

Bu, D., Y. Liao, J. Shi, and H. Peng (2019). Dynamic Expected Shortfall: A spectral decomposition of tail risk across time horizons. *Journal of Economic Dynamics and Control 108*, 103753.

Cai, Z. and X. Wang (2008). Nonparametric estimation of conditional VaR and Expected Shortfall. *Journal of Econometrics 147*(1), 120–130.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review 39*(4), 841–862.

Diebold, F. X. and R. S. Mariano (2002). Comparing predictive accuracy. *Journal of Business & Economic Statistics 20*(1), 134–144.

Du, Z. and J. C. Escanciano (2017). Backtesting Expected Shortfall: accounting for tail risk. *Management Science 63*(4), 940–958.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics 20*(3), 339–350.

Engle, R. F. and S. Manganelli (2004). CAViaR: conditional autoregressive Value at Risk by regression quantiles. *Journal of Business & Economic Statistics 22*(4), 367–381.

Engle, R. F. and V. K. Ng (1993). Measuring and testing the impact of news on volatility. *The Journal of Finance 48*(5), 1749–1778.

Fissler, T. and J. F. Ziegel (2016). Higher order elicitability and Osband's principle. *The Annals of Statistics 44*(4), 1680–1707.

Fissler, T., J. F. Ziegel, and T. Gneiting (2015). Expected Shortfall is jointly elicitable with Value at Risk - implications for backtesting. *arXiv preprint arXiv:1507.00244*.

Friend, I. and R. Westerfield (1980). Co-skewness and capital asset pricing. *The Journal of Finance 35*(4), 897–913.

Frongillo, R. and I. A. Kash (2015). Vector-Valued property elicitation. In *Conference on Learning Theory*, pp. 710–727.

Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association 106*(494), 746–762.

Gourieroux, C., W. Liu, and G. Liu (2012). Converting Tail-VaR to VaR: An econometric study. *Journal of Financial Econometrics 10*(2), 233–264.

Koenker, R. and Z. Xiao (2006). Quantile autoregression. *Journal of the American Statistical Association 101*(475), 980–990.

Konno, H. and K.-i. Suzuki (1995). A mean-variance-skewness portfolio optimization model. *Journal of the Operations Research Society of Japan 38*(2), 173–187.

Kotz, S., T. Kozubowski, and K. Podgorski (2012). *The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance*. Springer Science & Business Media.

Kozumi, H. and G. Kobayashi (2011). Gibbs sampling methods for bayesian quantile regression. *Journal of Statistical Computation and Simulation 81*(11), 1565–1578.

Kraus, A. and R. H. Litzenberger (1976). Skewness preference and the valuation of risk assets. *The Journal of Finance 31*(4), 1085–1100.

Kraus, D. and C. Czado (2017). D-vine copula based quantile regression. *Computational Statistics & Data Analysis 110*, 1–18.

Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives 3*(2), 73–84.

Laporta, A. G., L. Merlo, and L. Petrella (2018). Selection of Value at Risk models for energy commodities. *Energy Economics 74*, 628–643.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance 7*(1), 77–91.

McNeil, A. J. and R. Frey (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *Journal of Empirical Finance 7*(3-4), 271–300.

Mittnik, S. and S. T. Rachev (1991). Alternative multivariate stable distributions and their applications to financial modeling. In *Stable processes and related topics*, pp. 107–119. Springer.

Nolde, N., J. F. Ziegel, et al. (2017). Elicitability and backtesting: Perspectives for banking regulation. *The Annals of Applied Statistics 11*(4), 1833–1874.

Paolella, M. S. (2015). Multivariate asset return prediction with mixture models. *The European Journal of Finance 21*(13-14), 1214–1252.

Patton, A. J., J. F. Ziegel, and R. Chen (2019). Dynamic semiparametric models for Expected Shortfall (and Value-at-Risk). *Journal of Econometrics 211*(2), 388–413.

Petrella, L. and V. Raponi (2019). Joint estimation of conditional quantiles in multivariate linear regression models with an application to financial distress. *Journal of Multivariate Analysis 173*, 70–84.

Rockafellar, R. T. and S. Uryasev (2000). Optimization of conditional Value-at-Risk. *Journal of Risk 2*(3), 21–42.

Shi, Y., C. T. Ng, and K.-F. C. Yiu (2018). Portfolio selection based on asymmetric Laplace distribution, coherent risk measure, and expectation-maximization estimation. *Quantitative Finance and Economics 2*(4), 776–797.

Stolfi, P., M. Bernardi, and L. Petrella (2018). The sparse method of simulated quantiles: an application to portfolio optimization. *Statistica Neerlandica 72*(3), 375–398.

Taylor, J. W. (2005). Generating volatility forecasts from Value at Risk estimates. *Management Science 51*(5), 712–725.

Taylor, J. W. (2008). Estimating Value at Risk and Expected Shortfall using expectiles. *Journal of Financial Econometrics 6*(2), 231–252.

Taylor, J. W. (2019). Forecasting Value at Risk and Expected Shortfall using a semiparametric approach based on the asymmetric Laplace distribution. *Journal of Business & Economic Statistics 37*(1), 121–133.

White, H., T.-H. Kim, and S. Manganelli (2015). VAR for VaR: Measuring tail dependence using multivariate regression quantiles. *Journal of Econometrics 187*(1), 169–188.

Xiliang, Z. and Z. Xi (2009). Estimation of Value-at-Risk for energy commodities via CAViaR model. In *Cutting-Edge Research Topics on Multiple Criteria Decision Making*, pp. 429–437. Springer.

Yamai, Y. and T. Yoshiba (2005). Value-at-Risk versus Expected Shortfall: A practical perspective. *Journal of Banking & Finance 29*(4), 997–1015.

Yiu, K. (2004). Optimal portfolios under a Value-at-Risk constraint. *Journal of Economic Dynamics and Control 28*(7), 1317 – 1334.

Yu, K. and R. A. Moyeed (2001). Bayesian quantile regression. *Statistics & Probability Letters 54*(4), 437–447.

Zhao, S., Q. Lu, L. Han, Y. Liu, and F. Hu (2015). A mean-CVaR-skewness portfolio optimization model based on asymmetric Laplace distribution. *Annals of Operations Research 226*(1), 727–739.

Zhu, D. and J. W. Galbraith (2011). Modeling and forecasting Expected Shortfall with the generalized asymmetric Student-t and asymmetric exponential power distributions. *Journal of Empirical Finance 18*(4), 765–778.