Time-Varying and High-Dimensional Tail Risk Management based on the Multivariate Asymmetric Laplace Distribution

Yaming Yang*
School of Statistics, Tianjin University of Finance and Economics,
Tianjin 300222, China.
and
Peilin Yang †
Stanford University Graduate School of Business and Barcelona School of Economics.
and
Xiaowei Chen ‡
School of Finance, Nankai University, Tianjin 300350, China

May 20, 2024

Abstract

We develop a new multi-factor algorithm with time-varying factor loadings based on the Multivariate Asymmetric Laplace (MAL) distribution and provide the theoretical guarantee, facilitating the forecast of Expected Shortfall of multiple financial assets simultaneously. We generalize the traditional joint quantile regression framework to a time-varying setting, which allows us to capture the dynamic tail interdependence among assets. The proposed methodology is highly flexible and computationally tractable. The location-scale mixture representation of the MAL distribution allows estimation of the model parameters utilizing expectation–maximization algorithm, and the factor structure makes the model scalable to high dimensions. Applying this new statistical model to a panel of 50 stocks from 11 sectors of S&P 500 index over 2015-2022, we show that our model performs well on out-of-sample forecasting of Value-at-Risk and Expected Shortfall.

Keywords: Risk Management; Expected Shortfall; Generalized Autoregressive Score; Factor model

*Email: ymyang09@163.com.
†Email: peilin.yang@upf.edu.
‡*Corresponding author: chenx@nankai.edu.cn.*

1 Introduction

Time-varying tail risk measurement matters in risk management of high-dimensional financial system (see discussions in important researches Engle (2004), Engle et al. (2008), Shephard (2020), Basak & Shapiro (2001), Tobias & Brunnermeier (2016)), but it becomes challenging when markets alter significantly during financial turmoil. Since the new risk measurement promoted by Basel II in 1996, many tail risk management measures, such as Value at Risk (VaR) and Expected Shortfall (ES), have been widely used. VaR is the negated $\alpha-$quantile of the distribution that measures the maximum potential loss of an asset or a portfolio with a given probability over a predefined time horizon. ES is the expected loss given an exceeding of VaR. Recent management practice and theoretical works widely accept the appealing properties of ES. Wang & Zitikis (2021) build the economic foundation and rationale for using ES. Du & Escanciano (2017) develop a simple tool for backtesting ES, and Garcia et al. (2007) elicit proper conditioning information for a decentralized risk management system based on VaR. Alexander & Baptista (2004) consider a trade-off between ES with VaR and delineate the mean-VaR efficient set. Capponi & Rubtsov (2022) consider a framework that combines future growth and tail risk management. Admittedly, these works argue the importance of using VaR/ES to predict and monitor the tail risk index. According to these indexes, there still needs to be more specific approaches for managing each asset and the corresponding portfolio time-varying risk. This article aims to construct a syncretic tool of time-varying tail risk management tool from another perspective.

The major contributions of this article concerning portfolio optimization are twofold: One is we establish a quantile risk management framework under high-dimensional condition. Our framework directly evaluates portfolio risk at a specific quantile compared to the traditional Mean-Variance method. Instead of following the canonical quantile estimation of each asset and evaluating the nature of the portfolio, we set up likelihood framework quantile risk measure to steer clear of the difficulty of the traditional methods without analytical estimation results. The vantage point of our approach derives from the elegant property of Multivariate Asymmetric Laplace (MAL) distribution (see the seminal work of this field Kotz et al. (2001*b*), Kotz et al. (2001*a*) and Yu & Moyeed (2001)). In contrast, using traditional quantile estimation and then combining it with mean-variance analysis is no longer applicable because the pinball function (used in quantile estimation) has no differentiable property. Secondly, we consider factor loadings by combining two aspects, grouping and time-varying together. Grouping nature is widespread in risk management problems, e.g., market risk related to all assets and industrial risk factors affect a part of

assets, and time-varying nature is embodied in situations where risk index varies during different turmoil periods such as catastrophes, pandemics, and war. In this paper, we develop a multi-factor model to estimate dynamic tail interdependence for a large bunch of financial asset returns.

The method mentioned above affords a sketch of the framework. Still, more is needed for real-world practice because there are many ways to model/estimate the parameters that describe the property mean, variance, and skewness. To illustrate the multiple-factor model, we make the following assumptions according to our practice in portfolio management. One is that individual assets can be grouped according to the sectors, and each of these groups is subject to one common factor and one group-specific factor. This assumption is common in the financial world, (see Sharpe (1964) and Fama & French (1993)). They furnish a spectrum of evidence of the market and specific factors for modeling high dimensions into some essential dimensions. Our work will model the time-varying correlation matrix of MAL distribution through this factor model. Thus, this induces another assumption of factors: time-varying property, which is also very important in practice. Creal & Tsay (2015) and Oh & Patton (2017) generate the varying factor loadings using the Generalized Autoregressive Score (GAS) model, linking the likelihood function and factor variation. These two assumptions lay down the basis of our research work.

In the empirical part, we consider a panel of 50 daily stock returns across 11 sectors of the S&P 500 index from January 17, 2015 to December 30, 2022. Firstly, we estimate the parameters of the proposed factor models using the whole sample of 2011 observations with different model specifications. Secondly, we conduct out-of-sample forecasts for VaR/ES from January 05, 2016 to December 30, 2022. To evaluate the forecast performance, we use different scoring rules to conduct backtesting, where we compare our out-of-sample forecasts. Finally, we solve the portfolio optimization problems adopting the Skewness Mean-Variance (SMV) stretegy of Merlo et al. (2021). We compare the portfolio performances against five alternative models. The empirical analysis shows that our approach always performs well in terms of forecasting marginal VaR and ES. Graphically, the degree of tail interdependence tends to strengthen during financial turmoil, as one might expect. The confidence set of the model encompasses the portfolios constructed using our proposed models.

The article is organized as follows. In section 2, we introduce a multi-factor model with dynamic loadings, which can be embeded in the MAL distribution to capture the time-varying tail interdependence for a large bunch of assets. Section 3 presents the VaR and ES estimation model based on MAL distribution. In section 4, we discuss the procedure of

the parameter estimation. We provide a suitable EM-algorithm to estimate the parameters and use the variance target approach to reduce the computational burden. Section 5 and section 6 discuss the simulation and empirical application and evaluate the performance of the modeling approach, respectively. Section 7 is the application of our method: portfolio risk management.

2 Time-Varying Factor Loadings

Modeling VaR and ES for a large collection of asset returns while accounting for possible time-varying dependence is challenging. The modeling methodology should be flexible and scalable to high dimensions, while also being parsimonious to avoid the "curse of dimensionality". To strike this balance, we propose to model the time-varying dependence among multiple returns by using a factor structure that endowed with score-driven parameter dynamics. The factor structure can be embeded in the MAL distribution in the light of the location-scale mixture representation of the MAL distribution proposed in Kotz et al. (2001*a*). We start by introducing the construction of factor model in this section.

Consider the MAL distribution $\boldsymbol{Y}_t = [Y_{1t}, ..., Y_{Nt}]'$ by the following structure[1].

$$Y_{it} = \mu_{it} + \delta_i \xi_i W_t + \delta_i \varsigma_i \sqrt{W_t}(\tilde{\boldsymbol{\lambda}}'_{it} \boldsymbol{z}_t + \sigma_{it}\varepsilon_{it}), \quad i = 1, \ldots, N, \tag{1}$$

where $W_t \sim \mathcal{E}(1)$ follows standard exponential distribution, $\boldsymbol{z}_t$ is a $k \times 1$ vector of common factors, and $\boldsymbol{z}_t \sim \boldsymbol{N}_k(\boldsymbol{0}_k, \boldsymbol{I}_k)$ follows k-variate standard normal distribution, $\tilde{\boldsymbol{\lambda}}_{it}$ is a $k \times 1$ vector of scaled factor loadings, ε_{it} is an idiosyncratic variable, $\varepsilon_{it} \sim N(0,1)$, $\boldsymbol{z}_t \perp \varepsilon_{it}$ and $\varepsilon_{it} \perp \varepsilon_{jt}$, $\forall i \neq j$. Apparently $\tilde{\boldsymbol{\lambda}}'_{it}\boldsymbol{z}_t + \sigma_{it}\varepsilon_{it}$ is a linearly composition of i.i.d. random variables. [2] Define the vector $\tilde{\boldsymbol{\lambda}}_{it}$ and scalar σ_{it} as

$$\tilde{\boldsymbol{\lambda}}_{it} = \frac{\boldsymbol{\lambda}_{it}}{\sqrt{1 + \boldsymbol{\lambda}'_{it}\boldsymbol{\lambda}_{it}}}, \qquad \sigma^2_{it} = \frac{1}{1 + \boldsymbol{\lambda}'_{it}\boldsymbol{\lambda}_{it}},$$

for an unrestricted $k \times 1$ vector $\boldsymbol{\lambda}_{it}$ to maintain the unit variance. Given this structure, the correlation matrix of $[\tilde{\boldsymbol{\lambda}}'_{1t}\boldsymbol{z}_t + \sigma_{1t}\varepsilon_{1t}, \ldots, \tilde{\boldsymbol{\lambda}}'_{Nt}\boldsymbol{z}_t + \sigma_{Nt}\varepsilon_{Nt}]'$ is $\boldsymbol{R}_t = \tilde{\boldsymbol{L}}'_t\tilde{\boldsymbol{L}}_t + \boldsymbol{S}_t$, where $\tilde{\boldsymbol{L}}_t = [\tilde{\boldsymbol{\lambda}}_{1t}, \ldots, \tilde{\boldsymbol{\lambda}}_{Nt}]$ and $\boldsymbol{S}_t = diag(\sigma^2_{1t}, \ldots, \sigma^2_{Nt})$. This decomposition allows for the dependence structures to be determined flexibly and dynamically by multiple factors while also giving rise to a closed-form likelihood expression.

Akin to much related research on factor structure, we give a specific structure of matrix $\tilde{\boldsymbol{L}}$. N assets can be split into G groups according to an observed characteristic, such

[1] We will elaborate on this MAL in Section 3.

[2] In equation (4), it is combination of N variables, but we use dimension reduction method such as empirically industrial categories or PCA methods.

as sector, industry, firm size, risk attributes, etc. We assume that all assets within a specific group have identical factor loading. Hereafter, we consider the corresponding factor structures in Oh & Patton (2023) and Opschoor et al. (2021). The factor structure involves $G+1$ elements, including one common factor shared by all groups and G group-specific factors shared exclusively by one certain group. The loading matrix of the structure is defined as follows

$$\tilde{\boldsymbol{L}}_t' = \begin{pmatrix} \tilde{\lambda}_{M,1,t} & \tilde{\lambda}_{1,1,t} & 0 & 0 & \cdots & 0 \\ \tilde{\lambda}_{M,2,t} & 0 & \tilde{\lambda}_{2,2,t} & 0 & \cdots & 0 \\ \tilde{\lambda}_{M,3,t} & 0 & 0 & \tilde{\lambda}_{3,3,t} & 0 & 0 \\ \vdots & \vdots & \vdots & 0 & \ddots & 0 \\ \tilde{\lambda}_{M,G,t} & 0 & 0 & 0 & 0 & \tilde{\lambda}_{G,G,t} \end{pmatrix} \otimes \boldsymbol{\iota}_g \tag{2}$$

where $\otimes$ denotes the Kronecker product, $\boldsymbol{\iota}_g$ is an identical vector. The length of $\boldsymbol{\iota}_g$ depends on the number of elements in each group. Consequently, the matrix has $2G$ distinct factor loadings. The first column vector in the loading matrix is supposed to be interpreted as the loadings of each group on a common factor. The group-specific factor loadings are arranged diagonally. They only affect internal specific group correlation. The factor structure embedded in the loading matrix $\tilde{\boldsymbol{L}}_t$ implies that $\boldsymbol{R}_t$ exhibits a block structure too.

Given loading matrix in (2), to incorporate the time variation of the factor loadings, we make the following assumptions.

Assumption 1 *The factor loading $\lambda_{M,g,t}$ is related to a score function and the last period of factor loading $\lambda_{M,g,t-1}$ (same for $\lambda_{g,g,t}$)*

$$\lambda_{M,g,t+1} = \omega_g^M + A^M s_{M,g,t} + B^M \lambda_{M,g,t}, \qquad for \quad g = 1, \ldots, G,$$
$$\lambda_{g,g,t+1} = \omega_g^C + A^C s_{C,g,t} + B^C \lambda_{g,g,t}, \qquad for \quad g = 1, \ldots, G.$$

where $\boldsymbol{s_t} = [s_{M,1,t}, ..., s_{M,G,t}, s_{C,1,t}, ., , , s_{C,G,t}]'$ is the score vector of the (expected) log-likelihood function. To keep the model parsimonious, we impose that A^M, B^M, A^C and B^C are irrelative to g, that is, these terms are fixed for any $g = 1, \ldots, G$ in both common factor and group-specific factor loading dynamics. While the intercept vector ω_g^M and ω_g^C, for $g = 1, \ldots, G$ are different across different groups. Thus, $\boldsymbol{\Theta} = [\omega_1^M, \ldots, \omega_G^M, \omega_1^C, \ldots, \omega_G^C, A^M, B^M, A^C, B^C]$ is the parametric vector determining the factor loading dynamics.

Assumption (1) is also called the generalized autoregressive score model (GAS model see Creal et al. (2013)). It provides a time-varying relationship of factor loadings. This structure describes the distribution evolution between $\boldsymbol{Y}_t$ and $\boldsymbol{Y}_{t+1}$. There are many

ways to model the time-varying factor loadings, such as GARCH, ACD, and ACI. Here, we choose to use the likelihood function score because it provides an approximation of this change. Note that a vital model specification is about the choice of the scaling function of the scores. We employ the unit scaling for the score following Opschoor et al. (2021) and Oh & Patton (2023) to reduce the computational burden. That is, $\boldsymbol{s}_t = \partial E\{\ell_c(\boldsymbol{Y}_t; \boldsymbol{R}_t)\}/\partial \boldsymbol{\eta}_t$, where $E\{\ell_c(\boldsymbol{Y}_t)\}$ represents the expected log-likelihood function of $\boldsymbol{Y}_t$, and $\boldsymbol{\eta}_t = [\lambda_{M,1,t}, \ldots, \lambda_{M,g,t}, \lambda_{1,1,t}, \ldots, \lambda_{g,g,t}]'$. [3]

The GAS framework mentioned above implies $2G + 4$ parameters governing the GAS dynamics. Computational challenge may arise when G becomes large. To stave off the difficulty, we will use the variance target approach to reduce the computational burden and build a suitable EM algorithm to estimate the parameters of the conditional joint distribution. This is explained in detail in Section 4. All proofs of proposition and theorems are presented in the Appendix.

3 Quantile and ES estimation based on MAL distribution

We have provided the specifications of the possible interdependence of tails simultaneously. The focal point of this section is to discuss the premise of estimation. Let $Q_{Y_{it}}(\tau_i \mid \mathcal{F}_{t-1})$ denote the quantile function of the i-th component of $\boldsymbol{Y}_t$, conditional on information set $\mathcal{F}_{t-1}$, for $i = 1, \ldots, N$, and $t = 1, \ldots, T$, respectively. We decompose $\boldsymbol{Y}_t$ into quantile functions and error terms

$$\boldsymbol{Y}_t = \boldsymbol{\mu}_t + \boldsymbol{\epsilon}_t, \quad t = 1, 2, \ldots, T, \tag{3}$$

where $\boldsymbol{\mu}_t = Q_{\boldsymbol{Y}_t}(\boldsymbol{\tau} \mid \mathcal{F}_{t-1})$ is the conditional quantile function (at fixed levels $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_N)'$, respectively), $\boldsymbol{\epsilon}_t$ denotes an $N \times 1$ vector of error terms, of which the i-th marginal τ_i-quantile is restricted to equal to zero.

For the decomposition in (3), we then consider a dynamic generalization of the MAL distribution proposed by Kotz et al. (2001*a*). Specifically, we consider the multivariate Asymmetric Laplace distribution MAL$(\boldsymbol{\mu}_t, \boldsymbol{D\xi}, \boldsymbol{D\Sigma}_t\boldsymbol{D})$ incorporating time-varying correlation matrix, with density function

$$\begin{aligned} &f_{\boldsymbol{Y}_t}(\boldsymbol{Y}_t \mid \boldsymbol{\mu}_t, \boldsymbol{D\xi}, \boldsymbol{D\Sigma}_t\boldsymbol{D}, \mathcal{F}_{t-1}) \\ &\quad = \frac{2\exp\left\{(\boldsymbol{Y}_t - \boldsymbol{\mu}_t)'\boldsymbol{D}^{-1}{\boldsymbol{\Sigma}_t}^{-1}\boldsymbol{\xi}\right\}}{(2\pi)^{N/2}\left|\boldsymbol{D\Sigma}_t\boldsymbol{D}\right|^{1/2}} \left(\frac{m_t}{2 + d_t}\right)^{\nu/2} K_\nu\left(\sqrt{(2 + d_t)m_t}\right), \end{aligned} \tag{4}$$

where $\boldsymbol{\mu}_t$ is the shift parameter vector, we shall refer to the MAL as the Centralized MAL(CMAL) distribution if $\boldsymbol{\mu}_t = \boldsymbol{0}$, indicating that each component density are forced

[3] More calculation of score function is presented in appendix.

to be jointed at the same origin, $\boldsymbol{D\xi}$ is an $N \times 1$ scale parameter vector, with $\boldsymbol{D} = \text{diag}\,[\delta_1, \delta_2, \ldots, \delta_N]$, $\delta_i > 0$, and $\boldsymbol{\xi} = [\xi_1, \xi_2, \ldots, \xi_N]'$. $\boldsymbol{\Sigma}_t$ is an $N \times N$ positive definite matrix such that $\boldsymbol{\Sigma}_t = \boldsymbol{\Lambda R}_t \boldsymbol{\Lambda}$, with $\boldsymbol{R}_t$ being a time-varying correlation matrix and $\boldsymbol{\Lambda} = diag[\varsigma_1, \ldots, \varsigma_N]$. Moreover, $m_t = (\boldsymbol{Y}_t - \boldsymbol{\mu}_t)'(\boldsymbol{D\Sigma}_t\boldsymbol{D})^{-1}(\boldsymbol{Y}_t - \boldsymbol{\mu}_t)$, $d_t = \boldsymbol{\xi}'\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\xi}$, and K_ν denotes the Modified Bessel Function of the third kind with index parameter $\nu = (2-N)/2$.

It follows from Kotz et al. (2001*a*) that $\boldsymbol{Y}_t$ admits a shifted location-scale mixture of multivariate normal distribution, having the representation

$$\boldsymbol{Y}_t = \boldsymbol{\mu}_t + \boldsymbol{D\xi}W_t + \sqrt{W_t}\boldsymbol{D}{\boldsymbol{\Sigma}_t}^{1/2}\boldsymbol{Z}_t. \tag{5}$$

where $\boldsymbol{Z}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_N)$ denotes an N-variate standard normal distribution, and $W_t \sim \mathcal{E}(1)$ is a standard exponential random variable, with $\boldsymbol{Z}_t$ being independent of W_t. We thus have $E\boldsymbol{Y}_t = \boldsymbol{\mu}_t + \boldsymbol{D\xi}$, and the dynamic covariance matrix of error terms $\boldsymbol{\epsilon}_t$ depends on $\boldsymbol{R}_t$ through the relationship $Cov(\boldsymbol{\epsilon}_t) = \boldsymbol{D}(\boldsymbol{\Lambda R}_t\boldsymbol{\Lambda} + \boldsymbol{\xi\xi}')\boldsymbol{D}$. Specifically, $\boldsymbol{R}_t$ can be interpreted as the time-varying correlation matrix of the subordinated Gaussian process. However, estimating high-dimensional correlation matrix $\boldsymbol{R}_t$ is intrinsically challenging. By using the factor model introduced in Section 2, the representation in (1) can then be embedded in the MAL distribution.

The general dimension reduction approach using the above mentioned factor structure is originally put forward in the context of copula models (see Creal & Tsay (2015) and Oh & Patton (2017)). As mentioned in Oh & Patton (2017), the joint distribution of $X_{it} = \tilde{\boldsymbol{\lambda}}_{it}'\boldsymbol{z}_t + \sigma_{it}\varepsilon_{it}$ is available in closed-form for only a few particular combinations of choice of $F_{\boldsymbol{z}}$ and F_ε. The most common example being where the distribution of $F_{\boldsymbol{z}}$ and F_ε are both Gaussian, in which case X_{it} is also Gaussian. Several recent works in the context of constructing high-dimensional copula models, like student's t copula and skew t copula, are taking advantage of this dimension reduction approach. Instead of applying the factor structure to copula models, we consider a direct application of the factor structure on the observed variables $\boldsymbol{Y}_t$. Our model retains the advantage that having a decomposition of distribution in closed form. In addition, it can also be used to capture asymmetric dependence.

There is only one more step to do: ξ_i and ς_i are still unknowns scale/skew parameters for fulfilling condition $P(Y_{it} < Q_{Y_{it}}(\tau_i)|\mathcal{F}_{t-1}) = \tau_i$. Therefore, we need to have the following proposition.

Proposition 1 *Let $Y_{it}, i = 1, ..., N$ be marginal components of a N-dimensional random variable $\boldsymbol{Y}_t$, of which the density function is defined in* (4)*, and let $\boldsymbol{\tau} = (\tau_1, ..., \tau_N)$ be a $N \times 1$ vector, such that $\tau_i \in (0, 1)$, for all i. Then $P\left(Y_{it} < Q_{Y_{it}}(\tau_i)|\mathcal{F}_{t-1}\right) = \tau_i$ for*

$i = 1, ..., N$, *if and only if* $\xi_j = (1 - 2\tau_j) / \{\tau_j (1 - \tau_j)\}$, $\varsigma_i^2 = 2/\{\tau_i(1 - \tau_i)\}$.

The constraints in $\boldsymbol{\xi}$ and $\boldsymbol{\Lambda}$ represent essential conditions to ensure that the specification in (4) holds and is also imposed to guarantee model identifiability. It is easy to prove that, when such constraints are satisfied, the marginal distribution of each component of MAL distribution in (4) follows an univariate Asymmetric Laplace distribution. For a given τ_i and $\delta_i > 0$, minimizing the objective function of the quantile regression is equivalent to the maximum likelihood based on the AL density.

In order to estimate conditional quantile for each margin, various dynamic models can be used. We impose CAViaR models for the quantile regression function. The CAViaR proposed by Engle & Manganelli (2004) specifies the evolution of the corresponding quantile at the τ_j-th level over time using autoregressive process. In line with previous research, the CAViaR specifications that we considered in the subsequent analysis are as follows

$$Q_{Y_{it}} (\tau_i \mid \mathcal{F}_{t-1}) = \beta_{1i} + \beta_{2i} Q_{Y_{it-1}} (\tau_i \mid \mathcal{F}_{t-2}) + \beta_{3i} \mid Y_{it-1} \mid, \tag{6a}$$

$$Q_{Y_{it}} (\tau_i \mid \mathcal{F}_{t-1}) = \beta_{1i} + \beta_{2i} Q_{Y_{it-1}} (\tau_i \mid \mathcal{F}_{t-2}) + \beta_{3i} Y_{it-1}^{+} + \beta_{4i} Y_{it-1}^{-}. \tag{6b}$$

Equation (6a) is called Symmetric Absolute Value (SAV) specification and equation (6b) is called Asymmetric Slope (AS) specification. $\boldsymbol{\beta} = [\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N]'$, with $\boldsymbol{\beta}_i = [\beta_{1i}, \ldots, \beta_{3i}]$ in SAV specification and $\boldsymbol{\beta}_i = [\beta_{1i}, \ldots, \beta_{4i}]$ in AS specification are parameters to be estimated.

Notice that, the MAL distribution should not be viewed as a parametric assumption on the asset returns but rather offer a working likelihood function to estimate the parameters in a quantile regression framework. Although there is no assumption on the tail distribution, Bassett et al. (2004) provided an interest link between ES and quantile regression, namely that ES can be expreesed by quantile as follow

$$ES_t = \mathbb{E}[Y_t] - \frac{\mathbb{E}[(Y_t - \mathcal{Q}_{Y_t})(\tau - \mathbb{1}(Y_t \leq \mathcal{Q}_{Y_t}))]}{\tau}. \tag{7}$$

Following Merlo et al. (2021), we can express the ES in terms of the conditional mean and the AL scale parameters $\boldsymbol{\delta}$. Recall that each marginal of the MAL follows the univaraite AL distribution with scale parameter equals to $\delta_i = \mathbb{E}[(Y_i - \mathcal{Q}_{Y_i})(\tau_i - \mathbb{1}(Y_i \leq \mathcal{Q}_{Y_i})]$. The estimator of marginal ES in expression (7), can, therefore, takes the form

$$ES_{it} = \mathbb{E}[Y_{it}] - \frac{\delta_i}{\tau_i}, \quad t = 1, \ldots, T, \quad i = 1 \ldots, N.$$

For the estimation of ES, following Bassett et al. (2004), the marginal expection of $\mathbb{E}[Y_{it}]$ can be evaluated empirically using the sample mean $\bar{Y}_i$ of Y_{it}. As shown before, we can estimate VaR for all marginal assets using CAViaR model, i.e, the shift parameter

vectors $\boldsymbol{\mu}_t$ has already been determined, then we can use maximum likelihood approach to estimate scale parameters δ_i, $i = 1, \ldots, N$, i.e., $\boldsymbol{D}$ in (4) along with the conditional correlation matrix $\boldsymbol{R_t}$ in the centralized MAL distribution in order to estimate ES, while accounting for tail dependence. Since ES and VaR tend to vary together to some extent, the ES estimates are required not to cross the corresponding VaR estimates. It means that given a small quantile probability τ, the estimates of ES should be lower than the estimates of VaR. [4]

4 EM Estimation Algorithm

With our focus on the estimation of multiple GAS parameters, we extend the EM algorithm by exploiting the location-scale mixture of multivariate Gaussian laws, in which the dependence structure of the subordinated Gaussian process is expressed by a factor structure that endowed with score-driven parameter dynamics. Our extension enables us to capture time-varying dependencies which can be important for measuring tail risk of returns. Let $\boldsymbol{D}_{(\boldsymbol{\delta})}$ and $\boldsymbol{\Sigma}_{t(\boldsymbol{\Theta})}$ denote the scale and correlation matrix to be estimated in the second step to clarify that these two matrix depend on the unknown parameters $\boldsymbol{\delta}$ and $\boldsymbol{\Theta}$, respectively.

Proposition 2 *For a given vector $\boldsymbol{\tau} = [\tau_1, ..., \tau_N]'$, the observed data $\boldsymbol{Y}_t$, $t = 1, ..., T$, and a set of parameter estimates $\hat{\boldsymbol{\Phi}} = \{\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\Theta}}\}$. The expected likelihood function of linear representation is*

$$\begin{aligned} \mathrm{E}\left\{\ell_c\left(\boldsymbol{\Phi} \mid \boldsymbol{Y}_t, \hat{\boldsymbol{\mu}}_t, \hat{\boldsymbol{\Phi}}\right)\right\} = & -\frac{1}{2}\sum_{t=1}^{T}\log\left|\boldsymbol{D}_{(\boldsymbol{\delta})}\boldsymbol{\Sigma}_{t(\boldsymbol{\Theta})}\boldsymbol{D}_{(\boldsymbol{\delta})}\right| + \sum_{t=1}^{T}\left(\boldsymbol{Y}_t - \hat{\boldsymbol{\mu}}_t\right)'\boldsymbol{D}_{(\boldsymbol{\delta})}^{-1}\boldsymbol{\Sigma}_{t(\boldsymbol{\Theta})}^{-1}\boldsymbol{\xi} \\ & - \frac{1}{2}\sum_{t=1}^{T}\hat{z}_t\left(\boldsymbol{Y}_t - \hat{\boldsymbol{\mu}}_t\right)'\left(\boldsymbol{D}_{(\boldsymbol{\delta})}\boldsymbol{\Sigma}_{t(\boldsymbol{\Theta})}\boldsymbol{D}_{(\boldsymbol{\delta})}\right)^{-1}\left(\boldsymbol{Y}_t - \hat{\boldsymbol{\mu}}_t\right) \\ & - \frac{1}{2}\sum_{t=1}^{T}\hat{u}_t\boldsymbol{\xi}'\boldsymbol{\Sigma}_{t(\boldsymbol{\Theta})}^{-1}\boldsymbol{\xi} + const \end{aligned} \tag{8}$$

where const contains any components that do not depend on $\boldsymbol{D}$ and $\boldsymbol{\Sigma}_t$, and

$$\hat{u}_t = \left(\frac{\hat{m}_t}{2+\hat{d}_t}\right)^{1/2}\frac{K_{v+1}\left\{\sqrt{(2+\hat{d}_t)\hat{m}_t}\right\}}{K_v\left\{\sqrt{(2+\hat{d}_t)\hat{m}_t}\right\}}, \quad \hat{z}_t = \left(\frac{2+\hat{d}_t}{\hat{m}_t}\right)^{1/2}\frac{K_{v+1}\left\{\sqrt{(2+\hat{d}_t)\hat{m}_t}\right\}}{K_v\left\{\sqrt{(2+\hat{d}_t)\hat{m}_t}\right\}} - \frac{2v}{\hat{m}_t}$$

with

$$\hat{m}_t = \left(\boldsymbol{Y}_t - \hat{\boldsymbol{\mu}}_t\right)'\left(\hat{\boldsymbol{D}}_{(\hat{\boldsymbol{\delta}})}\hat{\boldsymbol{\Sigma}}_{t(\hat{\boldsymbol{\Theta}})}\hat{\boldsymbol{D}}_{(\hat{\boldsymbol{\delta}})}\right)^{-1}\left(\boldsymbol{Y}_t - \hat{\boldsymbol{\mu}}_t\right), \quad \hat{d}_t = \boldsymbol{\xi}'\hat{\boldsymbol{\Sigma}}_{t(\hat{\boldsymbol{\Theta}})}^{-1}\boldsymbol{\xi}.$$

[4] In our empirical study, we impose a nonlinear constraint on the maximum likelihood estimation to ensure that VaR and ES estimates do not cross.

To estimate the model parameters, we use a two-step approach. First, we estimate the parameters $\boldsymbol{\mu}_t$ of each of the margin separately to obtain the estimation of the $\boldsymbol{\tau}$-th quantile of the components of $\boldsymbol{Y_t}$. Second, we estimate $\boldsymbol{D}$ and $\boldsymbol{\Sigma}_t$ conditional on the marginal estimates of the shift parameters $\boldsymbol{\mu}_t$. For a given vector $\boldsymbol{\tau}$ and the estimates of marginal quantiles $\hat{\boldsymbol{\mu}}_t$, the estimation of $\boldsymbol{\Phi}$ can be obtained by maximizing the expected complete log-likelihood function in (8) with respect to the parameter set $\boldsymbol{\Phi}$. Since the scale matrix is static while the factor loadings in correlation matrix follow a GAS dynamic structure, closed-form solutions of $\boldsymbol{\Theta}$ do not exist. Therefore, the estimation of $\boldsymbol{\Theta}$ is obtained through numerical optimization. The estimation of $\boldsymbol{\delta}$ can then be obtained by solving the first-order condition

$$\sum_{t=1}^{T}(\boldsymbol{Y}_t-\hat{\boldsymbol{\mu}}_t)\boldsymbol{\xi}^{\top}\hat{\boldsymbol{\Sigma}}^{-1}_{t(\hat{\boldsymbol{\Theta}})}+n\boldsymbol{D}_{(\boldsymbol{\delta})}-\sum_{t=1}^{T}\hat{z}_t(\boldsymbol{Y}_t-\hat{\boldsymbol{\mu}}_t)(\boldsymbol{Y}_t-\hat{\boldsymbol{\mu}}_t)\prime\boldsymbol{D}^{-1}_{(\boldsymbol{\delta})}\hat{\boldsymbol{\Sigma}}^{-1}_{t(\hat{\boldsymbol{\Theta}})}=\boldsymbol{0}_{N\times N}.$$

Therefore, the EM algorithm can be summarized in the following steps:

E-step: Given $\hat{\boldsymbol{\Phi}}^{(h)}=\{\hat{\boldsymbol{\delta}}^{(h)},\hat{\boldsymbol{\Theta}}^{(h)}\}$, at each iteration h, compute the weights

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

$$\hat{z}_t^{(h)}=\left(\frac{2+\hat{d}_t^{(h)}}{\hat{m}_t^{(h)}}\right)^{\frac{1}{2}}\frac{K_{v+1}\left(\sqrt{\left(2+\hat{d}_t^{(h)}\right)\hat{m}_t^{(h)}}\right)}{K_{v}\left(\sqrt{\left(2+\hat{d}_t^{(h)}\right)\hat{m}_t^{(h)}}\right)}-\frac{2v}{\hat{m}_t^{(h)}},$$

where

$$\hat{m}_t^{(h)}=(\boldsymbol{Y}_t-\hat{\boldsymbol{\mu}}_t)'(\hat{\boldsymbol{D}}_{(\hat{\boldsymbol{\delta}}^{(h)})}\hat{\boldsymbol{\Sigma}}_{t(\hat{\boldsymbol{\Theta}}^{(h)})}\hat{\boldsymbol{D}}_{(\hat{\boldsymbol{\delta}}^{(h)})})^{-1}(\boldsymbol{Y}_t-\hat{\boldsymbol{\mu}}_t),$$

$$\hat{d}_t^{(h)}=\boldsymbol{\xi}'\hat{\boldsymbol{\Sigma}}^{-1}_{t(\hat{\boldsymbol{\Theta}}^{(h)})}\boldsymbol{\xi}.$$

M-step: Use the estimates $\hat{u}_t^{(h)}$ and $\hat{z}_t^{(h)}$ to get the updated parameter set $\hat{\boldsymbol{\Phi}}^{(h+1)}$ by maximizing $\mathrm{E}\left\{\ell_c\left(\boldsymbol{\Phi}\mid\boldsymbol{Y}_t,\hat{\boldsymbol{\mu}}_t,\hat{\boldsymbol{\Phi}}\right)\right\}$ with respect to $\boldsymbol{\Phi}$. Specifically, we estimate parameter set $\hat{\boldsymbol{\Theta}}^{(h+1)}$ using the variance target approach described later in this section. While $\hat{\boldsymbol{D}}^{(h+1)}_{(\boldsymbol{\delta})}$ is obtained through numerical optimization by solving the following equation

$$\sum_{t=1}^{T}(\boldsymbol{Y}_t-\hat{\boldsymbol{\mu}})\boldsymbol{\xi}'\hat{\boldsymbol{\Sigma}}^{-1}_{t(\hat{\boldsymbol{\Theta}}^{(h+1)})}+n\boldsymbol{D}_{(\boldsymbol{\delta})}-\sum_{t=1}^{T}\hat{z}_t^{(h)}(\boldsymbol{Y}_t-\hat{\boldsymbol{\mu}})(\boldsymbol{Y}_t-\hat{\boldsymbol{\mu}})'\boldsymbol{D}_{(\boldsymbol{\delta})}{}^{-1}\hat{\boldsymbol{\Sigma}}^{-1}_{(\hat{\boldsymbol{\Theta}}^{(h+1)})}=\boldsymbol{0}_{N\times N}.$$

The iteration between *E-step* and *M-step* continues until convergence, that is, when the ℓ_2 norm distance between the parameters estimated involving two successive evaluations is smaller than 10^{-3}. We initialize $\hat{\boldsymbol{D}}^{(0)}_{(\boldsymbol{\delta})}$ by conducting MLE of AL distribution for

each marginal asset return. While the calibration of the initial value of parameter set $\hat{\boldsymbol{\Theta}}^{(0)}$ is derived from the MAL distribution random number generator introduced in Kotz et al. (2001*a*) using all the observed data. Since MAL random numbers $\boldsymbol{Y}_t = [Y_{1t}, ..., Y_{Nt}]'$ can be generated by a standard exponential distributed random number W_t and a set of standard normal distributed random number $\boldsymbol{Z}_t = [Z_{1t}, ..., Z_{Nt}]'$, along with prespecified MAL parameters $\boldsymbol{\mu}_t, \boldsymbol{D}, \boldsymbol{\xi}$ and $\boldsymbol{\Lambda}$. For given estimates of $\hat{\boldsymbol{\mu}}_t$ and $\hat{\boldsymbol{D}}$, for each $t = 1, ..., T$, we can firstly empirically approximate the subordinated exponential random number by $\hat{W}_t = \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{Y}_{it} - \hat{\boldsymbol{\mu}}_{it})/\hat{\delta}_i\xi_i$, since the mean of $\boldsymbol{Z}_t$ is close to zero when the dimension of $\boldsymbol{Z}_t$ is relatively large. Then, $(\boldsymbol{Y_t} - \hat{\boldsymbol{\mu}}_t - \hat{\boldsymbol{D}}\boldsymbol{\xi}\hat{W}_t)/\sqrt{\hat{W}_t}$ can be seen as a set of random number generated from $\boldsymbol{N}(\boldsymbol{0}, \hat{\boldsymbol{D}}\boldsymbol{\Sigma_t}\hat{\boldsymbol{D}})$. Then, for a given time-invariate $\hat{\boldsymbol{D}}$ and $\boldsymbol{\Lambda}$, we can obtain an estimate of the unconditional correlation matrix $\bar{\boldsymbol{R}}$. With $\bar{\boldsymbol{R}}$ in hand, following the general estimation procedure of variance target introduced below, we can obtain $\hat{\boldsymbol{\Theta}}^{(0)}$ by minimizing expression (9).

Variance-Targeting As the parameter set $\boldsymbol{\Theta}$ is still large, we adopt the "variance targeting" type estimator proposed by Oh & Patton (2018) to estimate the parameters in the GAS dynamics. Variance Targeting is a two-step procedure that can dramatically reduce the computational burden. Assume that the conditional mean of the factor loading vector satisfy $\bar{\boldsymbol{\eta}} = \boldsymbol{\omega} + \boldsymbol{B} \cdot \bar{\boldsymbol{\eta}}$, where $\bar{\boldsymbol{\eta}} = \mathbb{E}[\boldsymbol{\eta}_t]$, $\boldsymbol{\omega} = [\omega_1^M, \ldots, \omega_G^M, \omega_1^C, \ldots, \omega_G^C]$ and $\boldsymbol{B} = diag(B^M, \ldots, B^M, B^C, \ldots, B^C) \in \mathbb{R}^{2G\times 2G}$. In the first step, we match $\bar{\boldsymbol{\eta}}$ to the empirical intra and between-group correlations using moment estimation. Specifically, we consider the scaled loading matrix with each element replaced by its unconditional expectation

$$\overline{\boldsymbol{L}}' = \begin{pmatrix} \frac{\mathbb{E}[\lambda_{M,1,t}]}{v_1} & \frac{\mathbb{E}[\lambda_{1,1,t}]}{v_1} & 0 & \cdots & 0 \\ \frac{\mathbb{E}[\lambda_{M,2,t}]}{v_2} & 0 & \frac{\mathbb{E}[\lambda_{2,2,t}]}{v_2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\mathbb{E}[\lambda_{M,G,t}]}{v_G} & 0 & 0 & \cdots & \frac{\mathbb{E}[\lambda_{G,G,t}]}{v_G} \end{pmatrix},$$

where $v_i \equiv \sqrt{1 + \mathbb{E}\left[\lambda_{M,i,t}\right]^2 + \mathbb{E}\left[\lambda_{i,i,t}\right]^2}$. Then the correlation matrix of intra and between-group implied by model is $\bar{\boldsymbol{L}}'\bar{\boldsymbol{L}}$. The corresponding empirical unconditional intra and between group correlation matrix based on empirical unconditioanl correlation matrix $\bar{\boldsymbol{R}}$ is denoted by $\bar{\boldsymbol{\Omega}} \in \mathbb{R}^{G\times G}$ where the g-th diagonal element of $\bar{\boldsymbol{\Omega}}$ is the average of the correlations between all variables belonging to group g, and the off-diagonal element $\bar{\Omega}_{ij}$ is the average of the correlations between all pairs of variables from group i and j. The moment estimation is conducted by minimizing the following function

$$\hat{\bar{\boldsymbol{\eta}}} = \arg\min_{\bar{\boldsymbol{\eta}}}[vech\left(\bar{\boldsymbol{\Omega}} - \overline{\boldsymbol{L}}'\overline{\boldsymbol{L}}\right)]^T[vech\left(\bar{\boldsymbol{\Omega}} - \overline{\boldsymbol{L}}'\overline{\boldsymbol{L}}\right)]. \tag{9}$$

In the second step, with $\hat{\bar{\boldsymbol{\eta}}}$ estimated in the first step, $\boldsymbol{\omega}$ can be replace by $(\boldsymbol{I}_{2G} - \boldsymbol{B})\hat{\bar{\boldsymbol{\eta}}}$. Thus only $\boldsymbol{\Theta} = [A^M, B^M, A^C, B^C]$ are left to be estimated numerically.

5 Simulation

In this section, we conduct a simulation to study the finite sample properties of the maximum likelihood estimation of our model. We consider both of the choices of CAViaR specifications and investigate parameter estimation accuracy at different probability levels. We simulate $N = 50$ dimensional time series of length $T = 250$, with $G = 10$ equally sized groups, i.e., each group includes $N/G = 5$ individual asset returns. The sample size corresponds to the data dimensions in our empirical study. Time-varying location parameter $\boldsymbol{\mu}_t$ is generated according to the two CAViaR specifications in (6a) and (6b). The dynamic factor loadings evolve according to the GAS model introduced in Section 2. In addition, we also need to generate a diagonal matrix $\boldsymbol{D}$ to construct the scale parameter vector in the MAL distribution. To keep the simulation model parsimonious, we impose that $\omega_g^M = \omega^M$. For the GAS dynamic part, we set $[A^M, A^C, B^M, B^C] = [0.01, 0.98, 0, 015, 0, 92]$, $\omega^M = 0.03$ and let ω_g^C be randomly generated in the interval $[0.03, 0.06]$. For the parameters in MAL density, we set the components in $\boldsymbol{D}$ are randomly generated in the interval $[0.03, 0.1]$. The parameters in the CAViaR models are as follows

$$SAV: \quad [\beta_1, \beta_2, \beta_3] = [0.3, 0.9, 0.1].$$
$$AS: \quad [\beta_1, \beta_2, \beta_3, \beta_4] = [0.2, 0.85, 0.2, -0.3].$$

For each of the CAViaR specifications, we carry out 100 Monte Carlo replications and evaluate the accuracy of parameter estimation with two different quantile vectors $\boldsymbol{\tau} = [\tau_1, \ldots, \tau_{50}]'$, respectively $\tau_i = 0.01, 0.05$, for $i = 1, \ldots, 50$.

Table 1 presents the mean and the standard deviation of parameter estimates in CAViaR models and $[A^M, A^C, B^M, B^C, \omega^M]$ in the GAS model. For δ_i, $i = 1, \ldots, 50$ and ω_g^C, $g = 1, \ldots, 10$, we only report results for the Mean Absolute Error (MAE) and Mean Percentage Error (MPE), averaged across the 100 simulations. We find that all parameters are estimated near their true values. Overall, we conclude that the parameters of the MAL factor model with GAS dynamic loadings can be accurately estimated if the model is correctly specified.

Table 1: Summary of parameter estimation results of Monte Carlo simulation.

	SAV					AS				
	0.01			0.05		0.01			0.05	
Panel A: $[A^M, A^C, B^M, B^C, \omega^M]$										
	Ture	Mean	Std	Mean	Std	Ture	Mean	Std	Mean	Std
β_1	0.300	0.328	0.069	0.338	0.067	0.200	0.222	0.052	0.205	0.024
β_2	0.900	0.887	0.009	0.876	0.011	0.850	0.840	0.016	0.839	0.023
β_3	0.100	0.112	0.006	0.125	0.009	0.200	0.200	0.011	0.205	0.015
β_4						-0.300	-0.286	0.026	-0.290	0.023
A^M	0.010	0.011	0.007	0.009	0.004	0.010	0.011	0.006	0.009	0.004
B^M	0.980	0.971	0.006	0.979	0.003	0.980	0.972	0.007	0.979	0.003
A^C	0.015	0.015	0.010	0.016	0.007	0.015	0.018	0.011	0.009	0.008
B^C	0.920	0.919	0.027	0.914	0.020	0.920	0.921	0.035	0.917	0.017
ω^M	0.030	0.026	0.005	0.031	0.004	0.030	0.027	0.005	0.031	0.004
Panel B: MAE and MPE (both in %) for δ_i, $i = 1, \ldots, 50$ and ω_g, $g = 1, \ldots, 10$										
	MAE	MPE	MAE	MPE		MAE	MPE	MAE	MPE	
$\boldsymbol{\delta}$	0.083	0.565	0.261	0.506		0.068	0.265	0.098	0.511	
$\boldsymbol{\omega^C}$	1.623	3.410	0.423	0.735		1.554	3.059	0.190	0.315	

NOTE: This table reports the parameter estimates of Monte Carlo simulation. Results are based on 100 Monte Carlo replications.

Table 2: Summary statistics of the returns of 50 equities for the entire sample.

	Mean	5%	25%	50%	75%	95%
Mean	0.020	-0.002	0.014	0.019	0.027	0.039
Std	0.776	0.517	0.683	0.745	0.862	1.137
Skewness	-0.368	-1.513	-0.511	-0.241	-0.057	0.206
Kurtosis	15.947	8.037	11.442	13.771	17.884	34.302
Pearson correlation	0.438	0.265	0.359	0.436	0.506	0.616

NOTE: This table presents summary statistics of the returns of 50 equities for the entire sample from January 17, 2015 to December 30, 2022. The column presents the mean and quantile of the measures listed in the rows.

6 Empirical Application

6.1 Data and summary statistics

In this section, we apply the multi-factor MAL model proposed in the previous section to daily stock returns from January 17, 2015 to December 30, 2022, a total of $T = 2011$ trade days. We investigate 50 randomly chosen constituents of the S&P 500 index covering all 11 sectors.[5] The proportion of the number of institutions in each group is roughly equal to the proportion of the total number of institutions in each sector of the S&P 500 index. In our sample, "Information Technology and Industries" form the largest group with 7 firms, followed by "Consumer Discretionary, Energy, and Financial" with 6 firms. The other 6 groups include 3 firms. Table 2 presents the main summary statistics of the first four moments across the 50 assets and the summary of 1225 pairwise correlations of the daily returns. Ljung-Box tests for serial correlation at 4 lags found that 49 out of 50 asset returns reject the null hypothesis at a 5% significant level, indicating the time series does contain a significant autocorrelation. Jarque-Bera test shows that all 50 series reject the null hypothesis of normal distribution at the 5% level. Further, augmented Dickey-Fuller tests reject the null hypothesis of a unit root at a 5% level for all 50 series. The above results underline the skewed, fat-tailed, serial and cross-sectional correlated, and non-Gaussian natures of stock returns.

[5] A list of those firms is provided in the appendix.

Table 3: The GAS parameter estimation results of MLE.

	SAV				AS			
τ	0.05		0.01		0.05		0.01	
	Estimates	SE	Estimates	SE	Estimates	SE	Estimates	SE
A^M	0.007	0.002	0.009	0.000	0.004	0.010	0.005	0.000
B^M	0.990	0.008	0.947	0.004	0.995	0.003	0.979	0.002
A^C	0.011	0.001	0.026	0.000	0.010	0.001	0.007	0.000
B^C	0.988	0.001	0.947	0.003	0.989	0.002	0.949	0.002

NOTE: This table reports the GAS parameter estimation results, applied to daily return of 50 stocks using the full sample of $T = 2011$ observations. Standard errors are provided based on the sandwich robust covariance matrix estimator.

6.2 Full-Sample Estimation and Out-of-Sample Forecast

In this section, we use the two-step method described in Section 2 to estimate the parameters of the score-driven factor MAL models with two CAViaR specifications at both probability levels using the full sample of 2011 observations. Table 3 shows the estimates of $[A^M, A^C, B^M, B^C]$ with different model specifications.

As shown in Table 3, for both of the CAViaR specifications, the values of B^M and B^C are close to 1, representing a strong persistence of the time-varying factor loadings. This result is similar to other applications of GAS models (see Creal et al. (2013), Oh & Patton (2018)).

Sequentially, we derive an out-of-sample estimation of VaR and ES for the 50 asset returns. We use a rolling window of 250 observations. [6] The out-of-sample period consists of 352 forecasts from January 05, 2016 to December 31, 2022, which includes several periods of market turmoil. The parameters in the MAL distribution and the GAS model are reestimated after every 5 observations (or roughly 1 calendar week). Using the rolling window exercise, we construct a one-step-ahead forecast.

We first conduct VaR backtesting based on three coverage tests, namely the unconditional coverage (LR_{uc}) test of Kupiec et al. (1995), the conditional coverage (LR_{cc}) test of Christoffersen (1998) and the Dynamic Quantile (DQ) test of Engle & Manganelli (2004). Here, we include four lags in the regression of our test. Table 4 summarizes the results of the three tests on the 50 time series. We find that CAViaR-SAV specification outperforms the

[6] We use similar evaluation procedures and time window as Opschoor et al. (2021).

Table 4: Summary of marginal out-of-sample VaR forecast evaluation results.

	# H_0 is not rejected at 5%				# H_0 is not rejected at 1%			
$\boldsymbol{\tau}$	0.05		0.01		0.05		0.01	
	SAV	AS	SAV	AS	SAV	AS	SAV	AS
LR_{uc}	50	44	33	24	44	38	16	5
LR_{cc}	49	48	34	28	48	43	22	14
DQ	43	37	15	5	38	30	9	4

NOTE: This table reports the number of assets of which the null hypothesis is not rejected by LR_{uc}, LR_{cc} and DQ tests at the 5% and 1% significant levels.

CAViaR-AS specification at both the quantile levels. This result can probably be explained by the fact that the S&P 500 constitutes what we considered characterized by a sudden spur of volatility and roughly coincides with the out-of-sample period during which time several systemic events occurred, such as Brexit in 2018, the COVID-19 pandemic outbreak in 2020 and the implementation of the U.S. dollar interest rate hike policy in 2022.

Next, we evaluate the forecast performance of out-of-sample ES. Because ES is not an elicitable risk measure (see Fissler & Ziegel (2016) for more detailed discussions)[7], there is an absence of scoring functions for ES. We conduct the backtesting of ES by using the bootstrap test proposed by Mcneil & Frey (2000) based on $n = 10000$ samples. The test focuses on the size of the discrepancy between the asset returns and the ES forecasts when quantile violation occurs. To test the hypothesis of the zero mean, a bootstrap test is conducted to avoid the distribution assumption about the exceedance residuals. We compare the multi-factor models with dynamic loadings against two other alternatives. In the first case, the forecasts of ES are obtained separately from each marginal AL distribution, where we completely disregard the interrelation among assets. In the second case, we assume the correlation matrix time-invariant, that is, the difference between this case and the proposed factor model is that the procedure of updating $\boldsymbol{\Theta}$ in the $M - step$ can be omitted, and the empirical unconditional correlation matrix $\bar{\boldsymbol{R}}$ is used in the out-of-sample forecasting instead of $\hat{\boldsymbol{R}}_{t(\boldsymbol{\Theta})}$.

Panel A of Table 5 summarized the p-value results of 50 assets for a two-sided bootstrap test of the hypothesis that the exceedance residuals of out-of-sample estimations have zero mean against the alternative that the mean is not equal to zero. Since the systemic overes-

[7]In more technical terms, a risk measure is considered elicitable if it can be consistently estimated based on observations or forecasts.

Table 5: Summary of marginal out-of-sample ES forecast evaluation using the bootstrap test.

	SAV						AS					
	0.05			0.01			0.05			0.01		
	Indep	static corr	dynamic corr	Indep	static corr	dynamic corr	Indep	static corr	dynamic corr	Indep	static corr	dynamic corr
Panel A: Summary of p-values for the two-sided bootstrap test												
Min	0.448	0.447	0.447	0.449	0.428	0.437	0.457	0.457	0.453	0.448	0.447	0.461
25% quantile	0.474	0.475	0.481	0.481	0.478	0.477	0.483	0.483	0.491	0.471	0.474	0.484
Median	0.492	0.493	0.508	0.509	0.551	0.497	0.555	0.556	0.593	0.488	0.521	0.514
75% quantile	0.595	0.640	0.585	0.663	0.731	0.598	0.684	0.684	0.747	0.611	0.581	0.691
Max	0.982	0.977	0.971	0.923	0.961	0.938	0.993	0.993	0.993	0.959	0.988	0.989
# H_0 is rejected at 5% level	0	0	0	0	0	0	0	0	0	0	0	0
# H_0 is rejected at 1% level	0	0	0	0	0	0	0	0	0	0	0	0
Panel B: Summary of p-values for the one-sided bootstrap test												
Min	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010
25% quantile	0.000	0.000	0.001	0.000	0.005	0.469	0.019	0.019	0.044	0.000	0.003	0.467
Median	0.005	0.014	0.023	0.034	0.132	0.482	0.141	0.141	0.242	0.006	0.037	0.487
75% quantile	0.135	0.179	0.279	0.304	0.441	0.498	0.469	0.469	0.474	0.188	0.472	0.501
Max	0.519	0.519	0.550	0.526	0.538	0.535	0.527	0.527	0.535	0.537	0.537	0.529
# H_0 is rejected at 5% level	34	30	27	26	21	2	18	18	14	34	26	3
# H_0 is rejected at 1% level	28	23	22	21	15	2	10	10	5	26	20	0

NOTE: This table reports the results of ES evaluated using the bootstrap test for zero mean exceedance residuals. Panel A and Panel B summarized the p-values and the number of assets that are rejected at the 5% and 1% confidence levels for the two-sided and one-sided bootstrap test, respectively.

timate of ES is more prone to be accompanied by serious consequences, we also conduct a one-sided test against the alternative hypothesis that the residuals have a mean less than zero. The summary of the results of the one-sided test is presented in Panel B of Table 5. Both the SAV and AS models survive the two-sided bootstrap test even at the 1% confidence level at both quantile levels, indicating that the empirical version of residuals in the event of quantile violation is not significantly different from zero. While the result of the one-sided test shows that the possible overestimation of ES may exist in the forecast of some asset returns. Results in Table 5 shows that SAV specification provide a better fit than AS specification at the 1% quantile level, whereas the AS specification performs better at the 5% quantile level. We also find that the models considering subordinated correlations fit considerably better than their independent counterparts; our multi-factor models with dynamic loadings perform best, yielding the least number of assets that are rejected by the one-sided test. This result is robust to using both CAViaR specifications at both quantile levels.

Scoring Function Evaluation In what follows, to reinforce our analysis, we implement scoring functions to evaluate VaR and ES forecasts jointly. Specifically, we define the scoring functions of model j and k by $\Delta_{it}^{(j,k)} = S_{it}^{(j)} - S_{it}^{(k)}$, where $S_{it}^{(j)}$ is the scoring function value of asset i associated with model j at time t. We then use the test proposed by Diebold & Mariano (2002) (DM) to test the hypothesis that $\mathbb{E}\left[\Delta_{MAL,it}^{(j,k)}\right] = 0$ against $\mathbb{E}\left[\Delta_{MAL,it}^{(j,k)}\right] < 0$. Consider two scoring functions, respectively, S_{FZN} in Nolde & Ziegel (2017) and S_{FZ0} in Patton et al. (2019), where

$$S_{FZN}\left(\mathcal{Q}_t, ES_{it}, Y_{it}\right) = \left(\mathbf{1}_{(Y_{it}<\mathcal{Q}_{it})} - \tau\right)\frac{\mathcal{Q}_{it}}{2\tau\sqrt{-ES_{it}}} - \frac{1}{2\sqrt{-ES_{it}}}\left(\mathbf{1}_{(Y_{it}<\mathcal{Q}_{it})}\frac{Y_{it}}{\tau} - ES_{it}\right) + \sqrt{-ES_{it}},$$

$$S_{FZ0}\left(\mathcal{Q}_{it}, ES_{it}, Y_{it}\right) = \frac{1}{\tau ES_{it}}\mathbf{1}_{(Y_{it}<\mathcal{Q}_{it})}(Y_{it} - \mathcal{Q}_{it}) + \frac{\mathcal{Q}_{it}}{ES_{it}} + \log(-ES_{it}) - 1.$$

The test statistics, together with the corresponding p-values of DM test, are reported in Table 6. The same results are confirmed when jointly evaluating the performance of the out-of-sample VaR and ES forecasts, as the dynamic factor model again yields outstanding performances for both the CAViaR models at both the quantile levels. We then conduct the test between two CAViaR specifications both with their correlations modeled by dynamic factor structure. The joint test results of VaR and ES using scoring rules are presented in Table 7. It shows that the SAV model provide a significantly better fit than the AS model in the joint test of VaR and ES at both of the quantile levels.

Table 6: The statistics and p-values of the Diebold & Mariano (2002) (DM) test for one-step ahead forecasts

	0.05						0.01					
	Indep		static corr		dynamic corr		Indep		static corr		dynamic corr	
	FZN	FZ0	FZN	FZ0	FZN	FZ0	FZN	FZ0	FZN	FZ0	FZN	FZ0
Panel A: CAViaR-SAV specification for VaR												
Indep	-	-	3.454	3.514	3.860	2.149	-	-	5.376	5.264	4.388	4.316
	-	-	(1.000)	(1.000)	(1.000)	(0.984)	-	-	(1.000)	(1.000)	(1.000)	(1.000)
static corr	-3.454	-3.514	-	-	2.665	0.664	-5.376	-5.264	-	-	3.043	2.047
	(0.000)	(0.000)	-	-	(0.996)	(0.747)	(0.000)	(0.000)	-	-	(0.999)	(0.980)
dynamic corr	-3.860	-2.149	-2.665	-0.664	-	-	-4.388	-4.316	-3.043	-2.047	-	-
	(0.000)	(0.016)	(0.004)	(0.253)	-	-	(0.000)	(0.000)	(0.001)	(0.020)	-	-
Panel B: CAViaR-AS specification for VaR												
Indep	-	-	3.513	3.472	2.594	1.935	-	-	5.413	5.304	3.912	4.305
	-	-	(1.000)	(1.000)	(0.995)	(0.974)	-	-	(1.000)	(1.000)	(1.000)	(1.000)
static corr	-3.513	-3.472	-	-	2.482	1.830	-5.413	-5.304	-	-	1.792	1.072
	(0.000)	(0.000)	-	-	(0.994)	(0.966)	(0.000)	(0.000)	-	-	(0.963)	(0.858)
dynamic corr	-2.594	-1.935	-2.482	-1.830	-	-	-3.912	-4.305	-1.792	-1.072	-	-
	(0.005)	(0.027)	(0.007)	(0.034)	-	-	(0.000)	(0.000)	(0.037)	(0.142)	-	-

NOTE: This table reports the statistics and p-values of the Diebold & Mariano (2002) (DM) test for one-step ahead forecasts among three model settings. The null hypothesis is that the model in the jth row has the same forecasting performance as the model in the kth column regarding the mean of score values using certain scoring rules.

Table 7: The joint test results of VaR and ES using scoring rules

	AS-dynamic corr			
	0.05		0.01	
	FZN	FZ0	FZN	FZ0
SAV-dynamic corr	-2.779	-2.171	-1.731	-1.915
	(0.003)	(0.015)	(0.042)	(0.028)

7 Portfolio Tail Risk Management

Following Zhao et al. (2015) and Merlo et al. (2021), we evaluate the optimal weights by using the Skewness Mean-Variance (SMV) model, which is built to accommodate fat-tail and skewness characteristics of financial assets. This choice is reasonably applicable when the returns are not Gaussian distributed. As shown in Kotz et al. (2001*b*), a linear combination of the margins of a MAL variables follows a univariate asymmetric Laplace distribution. Thus, for the sake of the optimal portfolio selection in the context of MAL distribution, it suffices to obtain the allocation weight vector to minimize the portfolio variance at a prespecified probability level where the portfolio is a linear combination of marginal components of MAL distribution. It enlightens us to rewrite the additive property of MAL and the portfolio strategy to adapt to our model.

Proposition 3 *Let* $\boldsymbol{Y}_t \sim MAL(\boldsymbol{\mu}_t; \boldsymbol{D\xi}, \boldsymbol{D\Sigma}_t\boldsymbol{D})$*, with density function defined in* (4)*. Let* $\boldsymbol{b}_t = (b_{1t}, \ldots, b_{Nt})' \in \mathbb{R}^N$ *be a vector of weights such that* $\sum_{i=1}^N b_{it} = 1$*. Then the portfolio* $Y_t^b = \sum_{i=1}^N b_{it} Y_{it}$ *follows univariate AL distribution with* $Y_t^b \sim AL(\mu_t^\star, \tau_t^\star, \delta_t^\star)$*, where*

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

Based on aforementioned proposition, formally, the portfolio construction is obtained by solving the following optimization problem

$$\begin{aligned} \underset{\boldsymbol{b}_t \in \mathcal{R}^N}{\operatorname{argmin}} \quad & \boldsymbol{b}_t'\boldsymbol{D\Sigma}_t\boldsymbol{D}\boldsymbol{b}_t \\ \text{s.t. } & \tau_t^\star = \tilde{\tau}, \quad \forall t \\ & \boldsymbol{b}_t'\mathbf{1}_N = 1. \end{aligned} \tag{10}$$

Given the constraints in (10), the portfolio allocation weights must be adjusted to guarantee the portfolio VaR has a constant probability level $\tilde{\tau}$ in each holding period, namely $P\left(Y_t^b < \mu_t^\star \mid \mathcal{F}_{t-1}\right) = \tilde{\tau}$. Such an approach provides a useful tool for customizing asset

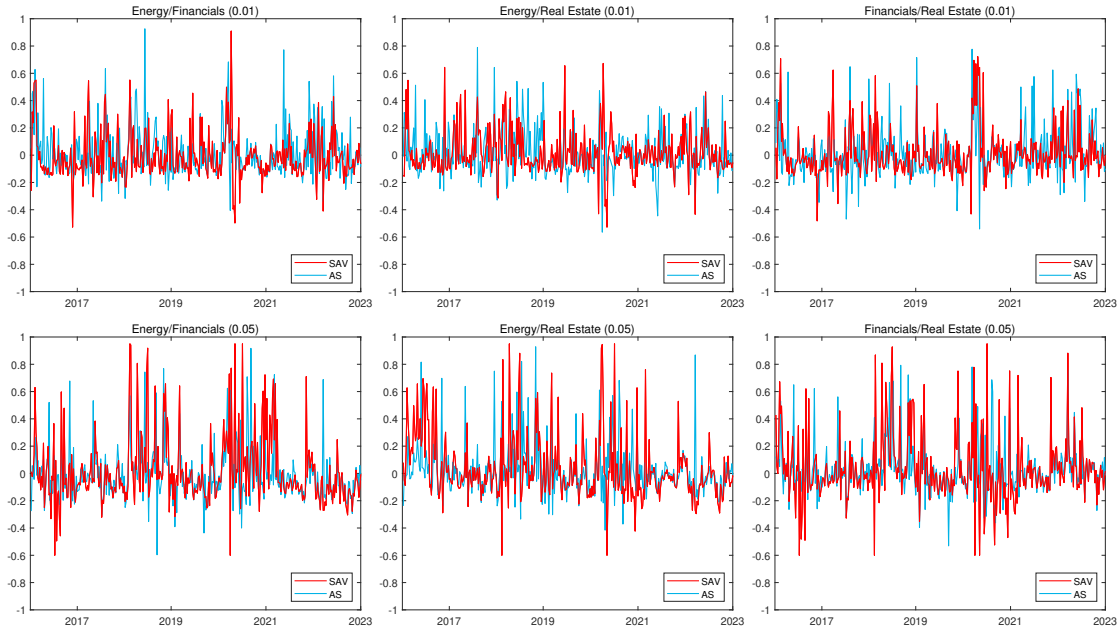

Figure 1: Out-of-sample forecast examples of within and between sector tail correlations.

allocation strategies. In addition, this optimization model will also facilitate a direct comparison of portfolio VaR and ES in out-of-sample forecasts with other benchmark models.

Now, we use the dynamic multi-factor model presented above to construct an optimal portfolio. The optimal allocation weights of 50 assets are determined by solving the SMV portfolio strategy. Figure 1 shows an example of between-sector dynamic correlations. For clarity, each panel presents the correlations between two sectors with both of the two specifications. In each panel of Figure 1, the red line denotes the correlations between certain two sectors with SAV specification, and the blue line denotes the correlations with AS specification. We find that the intercorrelation among assets tends to be reinforced during periods of high uncertainty in the market, which includes Brexit in 2018, the breakout of the COVID-19 pandemic in 2020, and the implementation of the U.S. dollar interest rate hike policy in 2022. As one can reasonably expect, financial variables are more correlated during periods of market turmoil.

To prove the advantage and practicality of our approach, we proceed to build portfolios using Mean-Variance (MV) strategy in Markowitz (1952) as a comparison. In this case, we use the GARCH model with time-varying covariance and student t innovations to model the daily stock returns. To ensure fairness between both classes of models, we model the conditional covariance in the GARCH model using several benchmark models that also apply autoregressive expressions in modeling correlation matrix. We estimate the one-step-ahead conditional covariance matrix for each model and plug it into the MV optimization problem. The benchmark models consist of three popular multivariate GARCH

(MGARCH) models: Dynamic Conditional Correlation (DCC) model (Engle (2002)), Dynamic Equicorrelation (DECO) and Block DECO models of Engle & Kelly (2012). Since the above three MGARCHs require repetitive numerical calculation of the inverse of covariance, which makes it computationally cumbersome, and the curse of dimensional effect is more pronounced. Therefore, we also consider two *score-driven* factor copula models proposed by Opschoor et al. (2021) to model the time-varying dependencies, respectively, the (Multi-Factor) MF and (Multi-Factor Lower-Triangular) MFLT models. These two multi-factor copula models are embedded with closed-form likelihood expression and are highly flexible to high dimensions. The MF and MFLT models are the best among all candidate factor models in terms of the statistical fitting in Opschoor et al. (2021)'s research. Based on the parameter estimation of our multi-factor MAL models, we adjust the portfolio of $N = 50$ returns dynamically.

Table 8 shows the summarized results, where we report the mean and standard deviations of VaR and ES of portfolios during the out-of-sample period. We also report the score function results of the portfolio VaR and ES using the considered strategies. For all the scoring rules, models that deliver lower values are preferred. Since a substantial number of competing models are built to construct the portfolio, we apply the Model Confidence Set (MCS) procedure in Hansen et al. (2011) to be the selector of models, which involves a sequence of tests that permit the construction of a set of "superior" models with a given level of confidence. Alongside evaluating forecast accuracy, we also calculate two relevant quantities to assess the forecasting performance of the different models from an economic perspective, including portfolio concentration (CO) and portfolio turnover (TO). These indicators are commonly used to evaluate and compare portfolios. CO at time t is defined as $CO_t = \sum_{i=1}^{N} w_{i,t|t-1}^2$, where $w_{i,t|t-1}$ is the optimal weight for the $i-$th asset at time t. TO measures the absolute change of weights adjustment from t to $t+1$. Turnover at time t is defined as $TO_t = \sum_{i=1}^{N} \left| w_{i,t+1|t} - w_{i,t|t-1} \right|$.

Table 8 embodies that our approach yields the highest portfolio ES overall. From a financial viewpoint, the portfolio that delivers higher ES usually posses higher control on tail risks. Regarding forecasting volatility, our approach also performs the best, delivering the lowest standard deviation for both VaR and ES over the out-of-sample period, which means that the portfolio tail risk of our approach is less susceptible to market fluctuations. This can be traced back to the difference in the train of thought to calculate VaR and ES. Our approach focuses on the cross-sectional tail interdependency at a certain probability level. Thus, the tail risk control effect of our approach is better that obtained by the benchmark models in which the VaR and ES are estimated based on conditional mean and

covariance. Moreover, the portfolio based on the SMV strategy is constructed with the aim of minimizing the volatility of the portfolio VaR and ES, which is different from the MV strategy, of which goal is to minimize the volatility of the whole predictive density. As for the evaluation of portfolio VaR and ES forecasts, all the models are included in the MCS using usual significance levels, like 5% or 10%, but our approach delivers the best performance gains over the other approaches based on GARCH models applying MV optimal strategy in the joint backtesting of the pair (VaR, ES), yielding the lowest values both of S_{FZN} and S_{FZ0} functions at both of the probability levels.

From the economic perspective, a model that produces a lower average CO generally implies less extreme portfolio weights. The concentration of our model belongs to the medium level, while our model produces the lowest portfolio turnover. A portfolio strategy that produces lower TO implies in general less transaction costs.

Table 8: The results on the SMV strategy with factor MAL model and MV strategy

	VaR (mean)	VaR (std)	ES (mean)	ES (std)	S_{FZN}	S_{FZ0}	CO	TO
Panel A: $\boldsymbol{\tau} = [0.05, \ldots, 0.05]$								
SMV-SAV	-0.421	0.155	-0.560	0.146	1.072(1.000)	0.131(1.000)	0.326	**0.979**
SMV-AS	-0.453	0.160	-0.550	0.160	1.097(1.000)	0.160(1.000)	0.239	1.221
MF	-0.455	0.248	-0.692	0.343	1.145(0.999)	0.349(0.959)	0.261	2.202
MFLT	-0.445	0.205	-0.681	0.293	1.154(0.951)	0.256(1.000)	0.245	2.107
DCC	-0.383	0.155	-0.649	0.233	1.175(0.701)	0.390(0.679)	0.328	2.054
DECO	-0.388	0.171	-0.666	0.263	1.157(0.997)	0.371(0.960)	**0.183**	1.003
BL-DECO	-0.380	0.166	-0.652	0.254	1.142(0.999)	0.307(0.997)	0.197	1.198
Panel B: $\boldsymbol{\tau} = [0.01, \ldots, 0.01]$								
SMV-SAV	-0.648	0.175	-0.740	0.200	1.567(1.000)	0.880(1.000)	0.230	**0.982**
SMV-AS	-0.612	0.198	-0.696	0.185	1.557(1.000)	0.857(1.000)	0.241	1.038
MF	-0.710	0.384	-0.947	0.486	1.724(0.979)	1.369(0.917)	0.261	2.202
MFLT	-0.697	0.327	-0.935	0.430	1.645(1.000)	0.964(1.000)	0.245	2.107
DCC	-0.685	0.279	-1.027	0.420	1.843(0.368)	1.527(0.378)	0.328	2.054
DECO	-0.698	0.307	-1.053	0.460	1.604(1.000)	1.159(1.000)	**0.183**	1.003
BL-DECO	-0.683	0.297	-1.031	0.447	1.689(0.995)	1.205(0.992)	0.197	1.198

NOTE: This table reports results on the SMV strategy with factor MAL model and MV strategy with 5 benchmark GARCH models. For each model, we present the mean and standard deviation of the portfolio VaR and ES and the average scoring values of the portfolio during the out-of-sample periods. We also present the p-value associated with the model confidence set (MCS) of Hansen et al. (2011). At last, we report the mean of the concentration (CO) and the portfolio turnover (TO). Numbers in bold are the lowest portfolio concentration and portfolio turnover.

8 Conclusion

Motivated by the growing interest in measures of extreme events, this article proposes a flexible and parsimonious model for time-varying high-dimensional measures of tail risks. We suggest a likelihood-based approach based on the MAL density to obtain the estimation of marginal VaR and ES, and then combine MAL distribution with a score-driven factor model to obtain a measure of dynamic tail interdependence. Parameters can be estimated straightforwardly by using EM-algorithm and a two-step approach for GAS dynamics. By combining with the SMV strategy, our proposed multi-factor MAL model can be applied in portfolio construction up to a dimension of 50 or more. In the application scenario, our models always outperform benchmark models in terms of forecast accuracy of marginal VaR and ES. Moreover, our optimization methods produce portfolios with the highest ES and the lowest VaR and ES volatility, representing a better control of tail risk compared with benchmark models. Our model can also provide insights into systemic risk. Greater tail interdependence forecasts by our model obviously associate with downward risk of the whole financial market.

Acknowledgement

The first author Yaming Yang gratefully acknowledge the project National Social Science Fund of China (Grant No. 21CTJ016). Peilin Yang was supported by the European Research Council (grant number: 882499). Please send correspondence to Xiaowei Chen, School of Finance, Nankai University, 300350, China; Email: chenx@nankai.edu.cn. On behalf of all authors, the corresponding author (Xiaowei Chen) states that there is no conflict of interest.

References

Alexander, G. J. & Baptista, A. M. (2004), 'A comparison of var and cvar constraints on portfolio selection with the mean-variance model', *Management Science* **50**(9), 1261–1273.

Basak, S. & Shapiro, A. (2001), 'Value-at-risk-based risk management: optimal policies and asset prices', *The Review of Financial Studies* **14**(2), 371–405.

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

Capponi, A. & Rubtsov, A. (2022), 'Systemic risk-driven portfolio selection', *Operations Research* **70**(3), 1598–1612.

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

Creal, D. D. & Tsay, R. S. (2015), 'High dimensional dynamic stochastic copula models', *Journal of Econometrics* **189**(2), 335–345.

Creal, D., Koopman, S. J. & Lucas, A. (2013), 'Generalized autoregressive score models with applications', *Journal of Applied Econometrics* **28**(5), 777–795.

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

Du, Z. & Escanciano, J. C. (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. (2004), 'Risk and volatility: Econometric models and financial practice', *American Economic Review* **94**(3), 405–420.

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

Engle, R. & Kelly, B. (2012), 'Dynamic equicorrelation', *Journal of Business & Economic Statistics* **30**(2), 212–228.

Engle, R., Shephard, N. & Sheppard, K. (2008), 'Fitting vast dimensional time-varying covariance models.'.

Fama, E. F. & French, K. R. (1993), 'Common risk factors in the returns on stocks and bonds', *Journal of Financial Economics* **33**(1), 3–56.

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

Garcia, R., Renault, É. & Tsafack, G. (2007), 'Proper conditioning for coherent var in portfolio management', *Management Science* **53**(3), 483–494.

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

Kotz, S., Kozubowski, T. J. & Podgórski, K. (2001*a*), Asymmetric multivariate laplace distribution, *in* 'The Laplace Distribution and Generalizations', Springer, pp. 239–272.

Kotz, S., Kozubowski, T. & Podgórski, K. (2001*b*), *The Laplace Distribution and Generalizations: a Revisit with Applications to Communications, Economics, Engineering, and Finance*, number 183, Springer Science & Business Media.

Kupiec, P. H. et al. (1995), *Techniques for Verifying the Accuracy of Risk Measurement Models*, Vol. 95, Division of Research and Statistics, Division of Monetary Affairs, Federal ....

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

Mcneil, A. J. & Frey, R. (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.

Merlo, L., Petrella, L. & Raponi, V. (2021), 'Forecasting var and es using a joint quantile regression and its implications in portfolio allocation', *Journal of Banking & Finance* **133**, 106248.

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

Oh, D. H. & Patton, A. J. (2017), 'Modeling dependence in high dimensions with factor copulas', *Journal of Business & Economic Statistics* **35**(1), 139–154.

Oh, D. H. & Patton, A. J. (2018), 'Time-varying systemic risk: Evidence from a dynamic copula model of cds spreads', *Journal of Business & Economic Statistics* **36**(2), 181–195.

Oh, D. H. & Patton, A. J. (2023), 'Dynamic factor copula models with estimated cluster assignments', *Journal of Econometrics* **237**(2), 105374.

Opschoor, A., Lucas, A., Barra, I. & Van Dijk, D. (2021), 'Closed-form multi-factor copula models with observation-driven dynamic factor loadings', *Journal of Business & Economic Statistics* **39**(4), 1066–1079.

Patton, A. J., Ziegel, J. F. & Chen, R. (2019), 'Dynamic semiparametric models for expected shortfall (and value-at-risk)', *Journal of Econometrics* **211**(2), 388–413.

Sharpe, W. F. (1964), 'Capital asset prices: A theory of market equilibrium under conditions of risk', *The Journal of Finance* **19**(3), 425–442.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1964.tb02865.x*

Shephard, N. (2020), Statistical aspects of arch and stochastic volatility, *in* 'Time Series Models', Chapman and Hall/CRC, pp. 1–68.

Tobias, A. & Brunnermeier, M. K. (2016), 'Covar', *The American Economic Review* **106**(7), 1705.

Wang, R. & Zitikis, R. (2021), 'An axiomatic foundation for the expected shortfall', *Management Science* **67**(3), 1413–1429.

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

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