On the estimation of Value-at-Risk and Expected Shortfall at extreme levels

Journal of Commodity Markets 34 (2024) 100391

Contents lists available at ScienceDirect

Journal of Commodity Markets

journal homepage: www.elsevier.com/locate/jcomm

Regular article

On the estimation of Value-at-Risk and Expected Shortfall at extreme levels

Emese Lazar [a], Jingqi Pan [a], Shixuan Wang [b,*]

[a] ICMA Centre, Henley Business School, University of Reading, Reading, RG6 6DL, UK
[b] Department of Economics, University of Reading, Reading, RG6 6EL, UK

ARTICLE INFO

JEL classification:
C32
C53
G17
Q02

Keywords:
Risk models
Value-at-Risk
Expected Shortfall
Semiparametric model
Oil futures

ABSTRACT

The estimation of risk at extreme levels (such as 0.1%) can be crucial to capture the losses during market downturns, such as the global financial crisis and the COVID-19 market crash. For many existing models, it is challenging to estimate risk at extreme levels. In order to improve such estimation, we develop a framework to estimate Value-at-Risk and Expected Shortfall at an extreme level by extending the one-factor GAS model and the hybrid GAS/GARCH model to estimate Value-at-Risk and Expected Shortfall for two levels simultaneously, namely for an extreme level and for a more common level (such as 10%). Our simulation results indicate that the proposed models outperform the GAS model benchmarks in terms of in-sample and out-of-sample loss values, as well as backtest rejection rates. We apply the proposed models to oil futures (WTI, Brent, gas oil and heating oil) and compare them with a range of parametric, nonparametric, and semiparametric alternatives. The results show that our proposed models are generally superior to the alternatives.

1. Introduction

Many institutional decisions in financial risk management, such as those related to capital requirements, rely on good forecasts of conditional distributions of asset returns, with an emphasis on the left tails of these distributions. What keeps risk managers awake at night are not typical price fluctuations but unexpected downfalls of unusual magnitudes. The concern is that these may trigger systemic spirals that can cause big losses. Financial regulators are concerned with protecting the financial system against catastrophic events that could be a source of systemic risk. It is of interest to correctly measure risk at very small levels, but the small number of observations in the extreme tail of the returns' distribution constitutes a problem that such extreme returns occur very rarely. For daily returns, by definition, events that breach the 1% quantile occur about twice a year. Returns in the more extreme quantiles occur even more rarely, and our focus is the risk assessment of such events.

In this paper, we propose a framework to measure risk at extreme percentiles that extends two models of Patton et al. (2019) by simultaneously estimating risk at two different levels (an extreme level and an auxiliary level), by assuming a joint process that drives both sets of risk measures. The optimal choice of auxiliary level is a more common level (in the range of 2.5%–20%) which can be selected via time series cross-validation. We illustrate via simulations and commodities data that by simultaneously considering an auxiliary level, the risk estimates at the extreme levels outperform the alternatives in terms of loss values, and often in terms of backtest performance as well.

Value-at-Risk (VaR) is one of the most popular tail risk measures that is employed to assess and manage financial risk. VaR is an estimate of the quantile of the distribution of profit and losses and it can be measured at different levels. Due to its conceptual

* Corresponding author.
E-mail addresses: e.lazar@icmacentre.ac.uk (E. Lazar), jingqi.pan@pgr.reading.ac.uk (J. Pan), shixuan.wang@reading.ac.uk (S. Wang).

https://doi.org/10.1016/j.jcomm.2024.100391
Received 4 April 2023; Received in revised form 13 November 2023; Accepted 15 February 2024
Available online 22 February 2024

simplicity, VaR has become a popular risk measure of market risk. However, VaR ignores the shape and structure of the tail of the returns' distribution and is not a coherent risk measure (i.e. it is not "sub-additive") (Artzner et al., 1999). Expected Shortfall (ES) is a risk measure that has recently increased in popularity due to its favorable properties. It measures the expected value of the observations provided that they exceed VaR and is a coherent risk measure (Roccioletti, 2015). A transition from VaR at 1% level to ES at 2.5% level has been proposed by the Basel Committee on Banking Supervision (2013). However, the measurement of ES is inherently dependent on the value of the VaR estimate. As such, ES is not elicitable by itself, and only the (VaR, ES) tuple is elicitable (Ziegel, 2016).

It is well known that the volatility displayed by commodity market returns has often been high (Hung et al., 2008), as shown by the recent events related to COVID-19 and the ongoing international conflict between Russia and Ukraine. It has been documented that commodity asset returns are generally characterized by higher volatility than stock returns (Del Brio et al., 2020). Thus, it is vital to have special risk management tools for the commodity market, which are needed by market participants and policymakers. Specifically, market participants need to measure market risk at extreme levels in order to manage their portfolios. As for policymakers, they need to be aware of the risks faced by the economy, because extreme commodity price changes can have a big impact on the economy as a whole, as indicated by Sadorsky (1999) and others.

The literature on VaR and ES estimation is very rich. To measure risk at multiple levels, White et al. (2015) propose a vector autoregressive (VAR) framework to quantile models which extend the CAViaR model of Engle and Manganelli (2004) to multiple confidence levels. Following the results of Fissler and Ziegel (2016) that ES and VaR are jointly elicitable, Patton et al. (2019) present several novel models. Specifically, they propose four dynamic semiparametric models for VaR and ES, based on the generalized autoregressive score (GAS) framework (see Creal et al., 2013; Harvey, 2013). However, VaR and ES at the popular levels (e.g. 1%, 2.5%, and 5%) provide insufficient information about rare but drastic events such as the COVID-19 crisis. Also, copula models can be used to improve VaR predictions, such as in Li et al. (2022). Many papers ignore the possibility of multiple regimes in the risk models; one way to address this problem is by using Markov-switching models, as in Maciel (2021).

Researchers have devoted effort to estimate VaR and ES at extreme levels. There is no well-defined definition of extreme level for risk, but in the literature it is typically defined as at levels below 1%. Chavez-Demoulin et al. (2014) propose a nonparametric extension of the Peaks-Over-Threshold method from Extreme Value Theory (EVT) to estimate VaR and ES at 1% level. Hoga (2017) proposes tests to detect changes in extreme VaR at levels below 1% based on the Weissman estimator motivated by EVT. Danielsson and De Vries (1998) propose a semi-parametric method to assess the probability of extreme events for data with heavy tails and apply it for VaR at extreme levels such as 0.5%, 0.1%, and 0.005%. In the study of Kourouma et al. (2010), VaR and ES are estimated based on the EVT model using the Peaks-Over-Threshold method, and it is shown that this type of EVT model performs better during the 2008 financial crisis than the unconditional VaR models. Based on a GARCH-type volatility model with covariates, Hoga (2021) derives asymptotically valid forecast intervals for VaR and ES, which are proved to be adequate for extreme risk levels. The above papers all focus on VaR and ES estimations at extreme levels, but whilst they are based on EVT, our models forecast risk measures based on the GAS framework. There are several papers that improve on risk forecasts via forecast combinations, such as Taylor (2020) and Storti and Wang (2022), and the latter proposes forecast combinations of VaR models for various quantiles in order to compute ES. Our approach is, however, to use the information from a specific generic quantile to improve VaR and ES forecasts at an extreme level.

This paper makes three main contributions. First, from a methodology perspective, we propose an extension of two models (the one-factor GAS model and the hybrid GAS/GARCH model) of Patton et al. (2019) to be used for risk estimation at extreme levels, by simultaneously estimating VaR and ES at two different levels, namely at an extreme level and at an auxiliary level. Without relying on such an auxiliary level, the extreme risk measure will depend on a small number of observations in the extreme tail of the empirical distribution of the returns. Therefore, incorporating information on a more generic tail can help to improve the forecast of VaR and ES at extreme levels. We obtain parameter estimates that are more robust than the parameters of standard GAS models, as highlighted by our simulations. Second, from a practical perspective, we demonstrate how to employ time series cross-validation (TSCV) to select the optimal auxiliary level from a set of candidates in order to improve the forecast performance of the proposed models. The TSCV is a data-driven method that helps with the selection of the auxiliary level without relying on arbitrary judgment. Third, from an empirical perspective, we provide compelling evidence that our models outperform the alternatives in terms of the evaluation of VaR and ES forecasts in a forecasting exercise. Our empirical analysis is based on four oil futures and we find that the recent COVID-19 crisis period well illustrates the strengths of our models in terms of forecasting risk at extreme levels.

The rest of the paper is organized as follows. Section 2 discusses VaR and ES models including the four GAS models proposed by Patton et al. (2019) and introduces the proposed GAS models that simultaneously estimate VaR and ES at two levels. The simulation results regarding model performance are presented in Section 3. Section 4 presents the data used in our empirical study, the in-sample estimation results, and out-of-sample (OOS) forecast results. Section 5 presents robustness results based on a rolling window estimation. Section 6 concludes. An online Supplemental Appendix provides additional results.

2. The augmented GAS model

2.1. Modeling VaR and ES

VaR provides banks and financial institutions with an estimate of the minimum loss level that occurs in the worst outcomes at a given level $\alpha \in (0, 1)$. Let $F_Y(\cdot|\Omega_{t-1})$ denote the cumulative distribution function of asset return Y_t over a time horizon (such as one day or one week) conditional on the information set Ω_{t-1}. Following Ziegel (2016), Nolde and Ziegel (2017), and Chen (2018), the VaR at level α at time t can be defined as:

$$VaR_t^\alpha = \inf\{Y_t | F_Y(Y_t|\Omega_{t-1}) \geq \alpha\}, \tag{1}$$

that is, VaR_t^α denotes the α-quantile of the underlying return distribution at time t. As such, VaR at level α can be written directly in terms of the inverse cumulative distribution function (Duffie and Pan, 1997):

$$VaR_t^\alpha = F_Y^{-1}(\alpha|\Omega_{t-1}). \tag{2}$$

ES measures the expectation of returns conditional on their value being less than VaR. ES is a coherent risk measure (Roccioletti, 2015) due its superior properties, and it has become increasingly popular in the risk management of banks and financial institutions. Recently, the Basel Committee on Banking Supervision (2013) proposed a transition from VaR at 1% level to ES at 2.5% level motivated by the global financial crisis in 2008. ES at level α at time t can be formally defined as (see Acerbi and Tasche, 2002):

$$ES_t^\alpha = \mathbb{E}[Y_t | Y_t \leq VaR_t^\alpha, \Omega_{t-1}]. \tag{3}$$

2.2. Generalized Autoregressive Score (GAS) framework

The application of the GAS framework for VaR and ES forecasting has been introduced by Patton et al. (2019). They propose the two-factor GAS model, the one-factor GAS model, the GARCH-FZ model, and the hybrid GAS/GARCH model to estimate VaR and ES jointly by minimizing the expectation of the VaR and ES joint loss function.[1] One of the most popular parameterizations of this joint loss function is the FZ0 loss function proposed by Fissler and Ziegel (2016), which has been further popularized by Patton et al. (2019). The FZ0 loss function is defined as:

$$L_{FZ0}(Y, v, e; \alpha) = -\frac{1}{\alpha e}\mathbf{1}\{Y \leq v\}(v - Y) + \frac{v}{e} + log(-e) - 1, \tag{4}$$

where Y is the return on the underlying asset, and v and e denote VaR and ES, respectively. $\mathbf{1}\{Y \leq v\}$ is an indicator function which returns 1 when $Y \leq v$ (i.e., the VaR is exceeded). Loss differences generated from the FZ0 loss function are homogeneous of degree zero. When $Y > v$, the returns do not affect the value of the loss. However, the loss value heavily relies on the returns when $Y \leq v$, with the parameter estimates being influenced by these extreme returns through the score. The parameters of the GAS models of Patton et al. (2019) are estimated by minimizing the loss function in Eq. (4). In the following, we briefly summarize their four model specifications.

2.2.1. The two-factor GAS model for ES and VaR

In the two-factor GAS model, the forecasts of VaR and ES are determined by the current value of VaR and ES and the forcing variable which is a function of the first order derivative and the Hessian of L_{FZ0}. The specification of the two-factor GAS model is shown below:

$$\begin{bmatrix} v_{t+1} \\ e_{t+1} \end{bmatrix} = \mathbf{W} + \mathbf{B}\begin{bmatrix} v_t \\ e_t \end{bmatrix} + \mathbf{A}\mathbf{H}_t^{-1}\nabla_t, \tag{5}$$

where $\mathbf{W}$ is a (2×1) vector and $\mathbf{B}$ and $\mathbf{A}$ are (2×1) matrices. The scoring function is given by:

$$\nabla_t \equiv \begin{bmatrix} \partial L_{FZ0}(Y_t, v_t, e_t; \alpha)/\partial v_t \\ \partial L_{FZ0}(Y_t, v_t, e_t; \alpha)/\partial e_t \end{bmatrix}, \tag{6}$$

and the scaling matrix $\mathbf{H}_t$ is the Hessian matrix:

$$\mathbf{H}_t = \begin{bmatrix} \frac{\partial^2 \mathbb{E}_{t-1}[L_{FZ0}(Y_t, {}_t, e_t;\alpha)]}{\partial v_t^2} & \frac{\partial^2 \mathbb{E}_{t-1}[L_{FZ0}(Y_t,v_t,e_t;\alpha)]}{\partial v_t \partial e_t} \\ . & \frac{\partial^2 \mathbb{E}_{t-1}[L_{FZ0}(Y_t,v_t,e_t;\alpha)]}{\partial e_t^2} \end{bmatrix}. \tag{7}$$

2.2.2. The one-factor GAS model for ES and VaR

The two-factor model allows ES and VaR to be updated as two separate, but correlated, processes. However, in the one-factor GAS model, VaR and ES are based on a time-varying risk measure κ_t (similar to the conditional variance process in the GARCH model). The one-factor GAS model is written as:

$$\begin{aligned} v_t &= a\ \exp\{\kappa_t\}, \\ e_t &= b\ \exp\{\kappa_t\},\ b < a < 0, \\ \kappa_t &= \omega + \beta\kappa_{t-1} + \gamma H_{t-1}^{-1} s_{t-1}, \end{aligned} \tag{8}$$

where the restriction $b < a < 0$ follows (Patton et al., 2019) and s_t is given by:

$$s_t \equiv \frac{\partial L_{FZ0}(Y_t, a\exp\{\kappa_t\}, b\exp\{\kappa_t\}; \alpha)}{\partial \kappa_t} = -\frac{1}{e_t}\left(\frac{1}{\alpha}\mathbf{1}\{Y_t \leq v_{t-1}\}Y_t - e_t\right), \tag{9}$$

[1] Fissler and Ziegel (2016) show that VaR and ES are jointly elicitable, while ES is not elicitable by itself (Gneiting, 2011).

and for simplicity, Patton et al. (2019) set the Hessian factor H_t as one. Thus, the one-factor GAS model for ES and VaR can be written as:

$$\kappa_t = \omega + \beta\kappa_{t-1} + \gamma \frac{1}{b\exp\{\kappa_{t-1}\}} \left(\frac{1}{\alpha} \mathbf{1}\{Y_{t-1} \le a\exp\{\kappa_{t-1}\}\} Y_{t-1} - b\exp\{\kappa_{t-1}\} \right). \tag{10}$$

2.2.3. The GARCH-FZ model for ES and VaR

Forecasting VaR and ES via a GARCH type model is one of the most prevailing ways to estimate risk measures, due to its parsimony. The GARCH-FZ model employs the framework of a GARCH model to generate VaR and ES, but the parameters of this model are estimated by minimizing the expectation of the loss function FZ0, instead of using (Q)MLE. The model is:

$$\begin{aligned} Y_t &= \mu_t + \sigma_t\,\eta_t, \quad \eta_t \sim iid\ F_\eta(0,1), \\ \sigma_t^2 &= \omega + \beta\sigma_{t-1}^2 + \gamma Y_{t-1}^2, \end{aligned} \tag{11}$$

where σ_t^2 is the conditional variance which follows a GARCH(1, 1) process. In terms of VaR and ES, the dynamic structure is analogous to the one-factor GAS model shown above:

$$\begin{aligned} v_t &= a\,\sigma_t, \quad \text{where } a = F_\eta^{-1}(\alpha), \\ e_t &= b\,\sigma_t, \quad \text{where } b = \mathbb{E}\,[\eta_t|\eta_t \le a], \end{aligned} \tag{12}$$

where η_t is the standardized residual.

2.2.4. The hybrid GAS/GARCH model for ES and VaR

In the hybrid GAS/GARCH model, the process κ_t in the one-factor GAS model and the volatility σ_t in the GARCH model both contribute to the dynamics of VaR and ES. Thus, as a combination of both models, the hybrid GAS/GARCH model is specified as:

$$\begin{aligned} Y_t &= \exp\{\kappa_t\}\eta_t, \ \eta_t \sim iid\ F_\eta(0,1), \\ v_t &= a\ \exp\{\kappa_t\}, \\ e_t &= b\ \exp\{\kappa_t\}, \ b < a < 0, \\ \kappa_t &= \omega + \beta\kappa_{t-1} + \gamma \frac{1}{e_{t-1}} \left(\frac{1}{\alpha} \mathbf{1}\{Y_{t-1} \le, v_{t-1}\} Y_{t-1} - e_{t-1} \right) + \delta \log|Y_{t-1}|, \end{aligned} \tag{13}$$

where κ_t is the log-volatility which is affected by Y_{t-1} in terms of the logarithm of absolute return rather than the square of return.

We now turn our attention to modeling risk at extreme levels within a GAS framework. According to Eqs. (10) and (13), κ_t depends on s_{t-1} (the first order derivative of the FZ0 loss function, driven mostly by the indicator function in Eq. (4)) and κ_{t-1}. Fig. 1 presents the κ_t and s_t processes for the one-factor GAS and hybrid GAS/GARCH models for two different levels, estimated from WTI crude futures prices. In general, κ_t remains mostly unchanged at extreme level $\alpha = 0.1\%$, and so the VaR and ES at time t are largely unaffected by the small changes in κ_t. For a less extreme level ($\alpha = 5\%$, for example), κ_t and s_t are more dynamic, being influenced by the returns in the tail of the distribution (see the last four figures in Fig. 1). VaR and ES at a higher level can use past information more efficiently. Therefore, in order to improve on the estimation of GAS models of Patton et al. (2019) for extreme levels, we propose the augmented GAS models, which are introduced in the following section.

2.3. The augmented GAS models for ES and VaR

In this section, we propose to enhance two dynamic semi-parametric models, which are the one-factor GAS model and the hybrid GAS/GARCH model of Patton et al. (2019), to improve the forecasts of risk measures at an extreme level. We achieve this by simultaneously modeling VaR and ES at two different levels, an extreme level α_1 (such as 0.1%) and an auxiliary level α_2 (a more common level in the range of 2.5%–20%).[2] In this setup, the same κ_t process drives the risk estimates VaR and ES for both levels. As such, we introduce two augmented GAS models, namely the augmented GAS one-factor model (we label it as A-GAS-1F) and the augmented hybrid model (we label it as A-Hybrid). These two models are jointly labeled as A-GAS models. We denote the VaR and ES at the extreme level of interest α_1 as $v_{1,t}$ and $e_{1,t}$, and at the auxiliary level α_2 as $v_{2,t}$ and $e_{2,t}$. Also, we investigate the backtesting performance of the $v_{1,t}$ and $e_{1,t}$ forecasts because these are at the level of interest. In the following, we elaborate on the details of the proposed models.

[2] An additional auxiliary level α_3 could also be considered at higher computational cost. However, in our preliminary analysis, this does not provide considerable improvements.

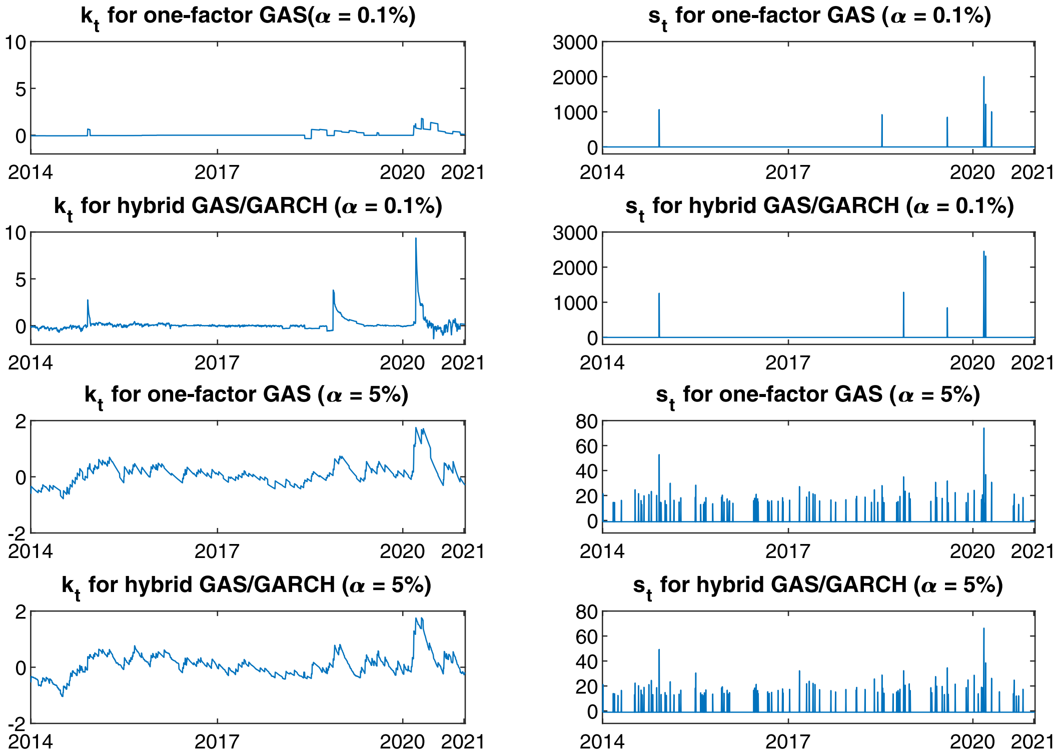

Fig. 1. This figure presents the κ_t and s_t processes of the GAS one-factor (left) and Hybrid (right) models for $\alpha = 0.1\%$ and $\alpha = 5\%$, estimated for the WTI futures prices from Jan 2014 to Jan 2021, with the model parameters re-estimated every 30 trading days using a rolling window of 1805 observations (7 years).

2.3.1. The augmented GAS one-factor model for ES and VaR

Under the GAS framework, the VaR and ES processes linearly depend on κ_t. Similarly, in the A-GAS-1F model, $v_{1,t}$, $e_{1,t}$, $v_{2,t}$ and $e_{2,t}$ are all driven by κ_t, which on the other hand depends on its lagged values (κ_{t-1}) and the score at the auxiliary level α_2. The model can be defined as:

$$
\begin{aligned}
&v_{1,t} = a_1 \exp\{\kappa_t\},\ \ e_{1,t} = b_1 \exp\{\kappa_t\},\\
&v_{2,t} = a_2 \exp\{\kappa_t\},\ \ e_{2,t} = b_2 \exp\{\kappa_t\},\\
&\kappa_t = \omega + \beta\kappa_{t-1} + \gamma s_{t-1,\alpha_2},\\
&s_{t,\alpha_2} \equiv \frac{\partial L_{FZ0}(Y_t, a_2 \exp\{\kappa_t\}, b_2 \exp\{\kappa_t\}; \alpha_2)}{\partial \kappa_t} = -\frac{1}{e_{2,t}}\left(\frac{1}{\alpha_2}\mathbf{1}\{Y_t \le v_{2,t}\}Y_t - e_{2,t}\right),
\end{aligned}
\tag{14}
$$

where $v_{1,t}$ and $e_{1,t}$ are the VaR and ES at the extreme level α_1 and $v_{2,t}$ and $e_{2,t}$ are the VaR and ES at the auxiliary level α_2. The score s_{t,α_2} only depends on α_2, being the first order derivative of the FZ0 loss function for the auxiliary level α_2. It is possible to also consider the s_{t,α_1} process as a second forcing variable in the κ_t process. However, given that α_1 is an extreme level, s_{t,α_1} is a predominantly flat process without much variation, which would make the corresponding parameter in the κ_t process difficult to estimate.[3]

2.3.2. The augmented hybrid GAS/GARCH model for ES and VaR

Extending the hybrid GAS/GARCH model of Patton et al. (2019), we propose the augmented hybrid GAS/GARCH model (labeled A-Hybrid) which uses an auxiliary level of risk α_2, given by:

$$
\begin{aligned}
&Y_t = \exp\{\kappa_t\}\eta_t,\quad \eta_t \sim iid\ F_\eta(0,1),\\
&v_{1,t} = a_1 \exp\{\kappa_t\},\ \ e_{1,t} = b_1 \exp\{\kappa_t\},\\
&v_{2,t} = a_2 \exp\{\kappa_t\},\ \ e_{2,t} = b_2 \exp\{\kappa_t\},\\
&\kappa_t = \omega + \beta\kappa_{t-1} + \gamma\left(\frac{1}{e_{2,t-1}}\left(\frac{1}{\alpha_2}\mathbf{1}\{Y_{t-1} \le v_{2,t-1}\}Y_{t-1} - e_{2,t-1}\right)\right) + \delta \log|Y_{t-1}|,
\end{aligned}
\tag{15}
$$

[3] In our preliminary experiments, considering the s_{t,α_1} process does not lead to improvements in the results.

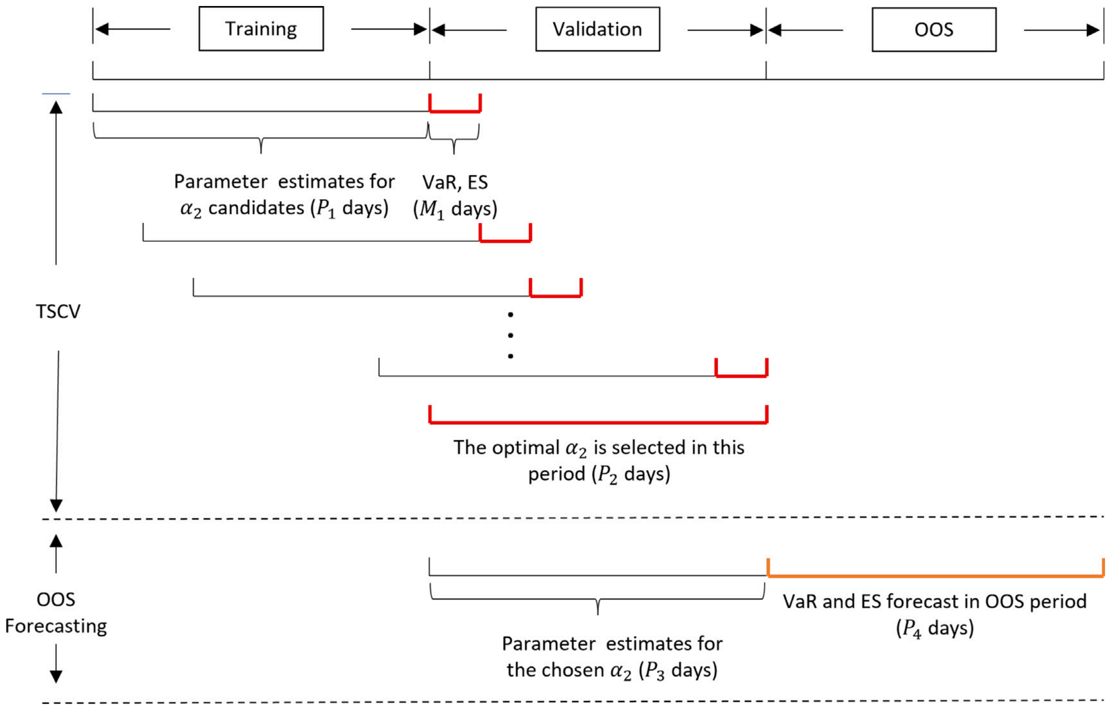

Fig. 2. This diagram presents the TSCV procedure for selecting the best auxiliary level α_2 for the A-GAS model. Also, it illustrates the forecasting procedure for VaR and ES at the extreme level α_1 in the OOS period. The parameters estimated from the training set (P_1 days) are used to forecast the VaR and ES in the first M_1 days in the validation period (P_2 days). Then, the parameters are re-estimated using information from day $1+M_1$ to day P_1+M_1 to generate the VaR and ES for the following M_1 days. After $\lfloor P_2/M_1 \rfloor$ repetitions, VaR and ES are forecasted for the validation period. The α_2 value with the lowest average FZ0 loss value over the validation period is selected as the optimal α_2 that is used in the OOS period. To obtain VaR and ES forecasts over the OOS period (P_4 days), the parameters of the A-GAS model with the optimal α_2 value are estimated using data from the last P_3 days prior to the OOS period.

where the log-volatility κ_t is the same as in the hybrid GAS/GARCH model but is based on α_2.[4]

2.3.3. Parameter estimation

In the augmented models, the forecasts of risk measures at the extreme level α_1 consider the losses at the auxiliary risk level α_2. Thus, the VaR and ES at α_1 are obtained by minimizing the joint loss function which is the sum of both FZ0 loss functions, at both α_1 and α_2 levels. Let $\mathfrak{L}_{FZ0}$ be the sum of the FZ0 loss functions, defined as:

$$\mathfrak{L}_{FZ0}(Y, v_1, v_2, e_1, e_2; \alpha_1, \alpha_2) = \underbrace{L_{FZ0}(Y, v_1, e_1; \alpha_1)}_{L_{FZ0}(\alpha_1)} + \underbrace{L_{FZ0}(Y, v_2, e_2; \alpha_2)}_{L_{FZ0}(\alpha_2)}, \tag{16}$$

where the $L_{FZ0}(Y, v_i, e_i; \alpha_i)$ is the FZ0 loss function for α_i as given by Patton et al. (2019) in their Eq. (4). By minimizing the expectation of $\mathfrak{L}_{FZ0}(Y, v_1, v_2, e_1, e_2; \alpha_1, \alpha_2)$, the model parameters are estimated via:

$$\hat{\theta}_T = \arg\min_{\theta} \frac{1}{T} \sum_{t=1}^{T} \mathfrak{L}_{FZ0}(Y_t, v_{1,t}, v_{2,t}, e_{1,t}, e_{2,t}; \alpha_1, \alpha_2), \tag{17}$$

where $v_{i,t}$ and $e_{i,t}$ are the VaR and ES forecasts at time t, obtained with the information set available at time $t-1$, at two risk levels α_i, $i = 1$ and 2. Before evaluating the performance of the augmented models, an essential consideration is the selection of the hyper-parameter, the auxiliary level α_2. To find an optimized α_2, we propose to use time series cross-validation, which is discussed in the following section.

2.3.4. Time series cross-validation

For the above augmented models, α_2 is the hyper-parameter to be determined. Cross-validation has been introduced as a method to help choose the best hyper-parameters for models in general (see, for example, Hart, 1994 for a description of this methodology). However, for time series, this method cannot be used in its classic form. Thus, we apply a special version of cross-validation that is suitable for time series applications, proposed by Hyndman and Athanasopoulos (2018). Within this procedure, a series of validation

[4] For the same reason, we choose to let κ_t only depend on risk measures at the auxiliary level α_2.

Table 1
The average loss values of the Augmented GAS models.

α_2	A-GAS-1F			A-Hybrid		
	Validation loss	OOS Forecast loss	Loss reduction	Validation loss	OOS Forecast loss	Loss reduction
2.5%	1.813	1.805	24.4%	1.813	1.796	15.8%
5%	1.783	1.773	25.7%	1.781	1.778	16.7%
7.5%	1.772	1.765	26.1%	1.794	1.773	16.9%
10%	1.802	1.763	26.2%	1.793	1.773	16.9%
12.5%	1.778	1.767	26.0%	1.803	1.778	16.6%
15%	1.767	1.769	25.9%	1.813	1.784	16.4%
17.5%	1.812	1.766	26.1%	1.820	1.789	16.2%
20%	1.786	1.762	26.2%	1.805	1.792	16.0%
TSCV	1.618	1.702	28.7%	1.644	1.719	19.4%
GAS	–	2.388	–	–	2.134	–
GJR-G-Est	–	1.691	–	–	1.691	–
GJR-G-True	–	1.675	–	–	1.675	–

Note: This table presents the average loss values over the validation and OOS forecasting period obtained from 1000 replications of VaR and ES estimations at the level $\alpha_1 = 0.1\%$, assuming a GJR-GARCH-SKT as the DGP. The full sample size is $T = 9000$, with 3000 as the training sample, 3000 as the validation sample, and 3000 as the OOS sample. The first 8 rows correspond to the A-GAS models with different α_2 values, whilst row 9 corresponds to the A-GAS models with TSCV. Row 10 corresponds to the GAS model of Patton et al. (2019). The last two rows are for the GJR-GARCH-SKT, where the GJR-G-Est corresponds to the GJR-GARCH-SKT model with the parameters re-estimated in every simulation path, and GJR-G-True corresponds to the GJR-GARCH-SKT model with the true parameters of the DGP. The loss values of the one-factor models are located in the left panel and loss values for the hybrid models are in the right panel. The columns "Validation loss" and "OOS forecast loss" present the average loss value of the augmented GAS models during the validation period and the OOS forecasting period, based on the FZ0 loss function. The Column "Loss reduction" presents the relative reduction in the loss value of the risk measures obtained by the augmented models compared to their corresponding GAS model.

sets are formed, each consisting of an equal-weighted segment of the time series observations. The corresponding training set consists of observations that occurred before the validation set. Therefore, no future information is used when making forecasts for the validation sets. One small change that we make is to use the rolling-windows in our emprical study, rather than the expanding-windows as proposed by Hyndman and Athanasopoulos (2018). This is due to the consideration of possible structural breaks, which could lead to large forecasting errors and result in the model being unreliable for forecasting. The TSCV procedure[5] in our empirical study is illustrated in Fig. 2.

3. Simulation study

In this section, we investigate the performance of the A-GAS models and compare it with that of the GAS models of Patton et al. (2019) via Monte Carlo simulations. To measure model performance, we use loss values and a range of backtests. In the following, we consider an extreme level as $\alpha_1 = 0.1\%$. We choose the data generation process (DGP) as the GJR-GARCH (1,1) model with skewed *t* distribution (we label it as GJR-GARCH-SKT) which considers the leverage effect, and it is among the most suitable model for volatility and VaR forecasting (Liu and Hung, 2010).[6] Specifically, the DGP is:

$$\begin{aligned} &Y_t = \sigma_t \eta_t,\ \eta_t \sim iid\ F_\eta(0,1), \\ &\sigma_t^2 = \omega + \gamma Y_{t-1}^2 + \delta \mathbf{1}\{Y_{t-1} < 0\} Y_{t-1}^2 + \beta \sigma_{t-1}^2, \end{aligned} \tag{18}$$

where the parameter values of the DGP are set to be $(\omega, \gamma, \delta, \beta) = (0.0225, 0.0065, 0.1779, 0.8835)$, and the error term η_t follows a skewed *t* distribution of Hansen (1994) with degrees of freedom $\nu = 7.5269$ and skewness parameter $\lambda = -0.1455$.[7]

In the specification of the A-GAS model, we consider a variety of values for α_2, specifically $\{2.5\%, 5\%, 7.5\%, 10\%, 12.5\%, 15\%, 17.5\%, 20\%\}$ is the set of possible values for the auxiliary level.[8] The simulation is based on 1000 replications. The whole sample size $T = 9000$ is divided into three equal segments, specifically training period, validation period and the OOS period. The first 3000 days of the sample are the training period used to obtain the parameter estimates; the next 3000 days constitute the validation period for the selection of α_2; and the last 3000 days represent the OOS period for the purpose of final forecasting evaluation. To reduce the computational cost of the simulation, VaR and ES forecasts in the validation period are generated in a fixed-window manner, which is based on the fact that there is no structural break in the DGP and thus the rolling-windows are not needed. TSCV selects the optimum α_2 obtained via minimizing the average loss in the validation period in each replication, and this α_2 will be used over the entire OOS period. We also report the results of individual α_2 for the purpose of comparison.

Table 1 presents the FZ0 loss values and loss reductions of the A-GAS models at the extreme level $\alpha_1 = 0.1\%$ for different α_2 values. If α_2 is chosen by TSCV, the two A-GAS models outperform the classical alternative models. The loss reduction is defined as the relative reduction in the loss value of the risk measures obtained by the augmented models compared to their corresponding

5 We use a fixed window in the simulation setting because there is no structural break in the our simulation set-up, and this can largely reduce the computational cost. This can be regarded as a special case when $M_1 = P_2$ in Fig. 2.

6 We also performed a simulation analysis using a Markov-switching process as DGP. The results are reported in the Supplemental Appendix.

7 All parameter values are obtained via fitting the model to the de-meaned returns on the S&P 500 from 1 January 2000 to 8 January 2021.

8 Other values of α_2 can also be considered. Due to the computational cost, we restrict α_2 to these eight choices.

Table 2
Backtest rejection rates.

α_2	A-GAS-1F			A-Hybrid		
	UC	CC	BS	UC	CC	BS
2.5%	35.8%	22.7%	29.9%	37.4%	25.6%	25.4%
5%	36.1%	23.6%	28.7%	38.0%	25.9%	28.5%
7.5%	35.8%	23.0%	29.6%	36.5%	26.1%	30.6%
10%	35.7%	23.6%	29.8%	37.8%	26.6%	29.2%
12.5%	35.0%	23.6%	28.9%	37.6%	28.1%	31.1%
15%	36.6%	24.6%	29.5%	37.2%	26.7%	32.6%
17.5%	36.1%	23.7%	28.2%	37.5%	27.4%	33.3%
20%	34.4%	22.0%	27.0%	37.1%	26.7%	32.5%
TSCV	27.9%	15.7%	22.9%	31.5%	21.9%	30.3%
GAS	47.8%	37.4%	40.0%	47.9%	38.3%	34.0%
GJR-G-Est	13.2%	7.7%	20.9%	13.2%	7.7%	20.9%
GJR-G-True	8.0%	5.7%	21.0%	8.0%	5.7%	21.0%

Note: This table presents the backtest rejection rates obtained from 1000 replications, indicating the frequency of backtest rejections at 5% significance level. The DGP used in the simulation is the GJR-GARCH-SKT with full sample size $T = 9000$, and the risk level used to compute VaR and ES is $\alpha_1 = 0.1\%$. The first 8 rows correspond to the A-GAS models with different α_2 values, whilst row 9 corresponds to the A-GAS models with TSCV. Row 10 corresponds to the GAS model of Patton et al. (2019). The last two rows are for the GJR-GARCH-SKT, where the GJR-G-Est corresponds to the GJR-GARCH-SKT model with the parameters re-estimated in every simulation path, and GJR-G-True corresponds to the GJR-GARCH-SKT model with the true parameters of the DGP. Columns 2–4 and columns 5–7 present the rejection rates for the Unconditional Coverage (UC), Conditional Coverage (CC) and Bootstrapping backtest (BS), respectively.

Table 3
The relative loss reduction of A-GAS models for various α_1 levels.

$\alpha_2 \backslash \alpha_1$	A-GAS-1F						A-Hybrid					
	0.1%	0.25%	0.5%	0.75%	1%	2.5%	0.1%	0.25%	0.5%	0.75%	1%	2.5%
2.5%	24.39%	16.60%	11.88%	8.89%	6.55%	N/A	15.81%	13.54%	7.50%	4.39%	3.11%	N/A
5%	25.73%	18.65%	14.00%	11.03%	9.12%	2.61%	16.68%	14.94%	8.73%	5.95%	4.66%	1.54%
7.5%	26.07%	19.35%	14.87%	11.84%	9.91%	3.70%	16.90%	15.29%	9.45%	6.66%	5.22%	2.12%
10%	26.15%	19.74%	15.17%	12.39%	10.34%	4.16%	16.91%	15.34%	9.81%	7.04%	5.71%	2.51%
12.5%	25.99%	20.20%	15.55%	12.65%	10.54%	4.39%	16.65%	15.70%	9.92%	7.12%	5.85%	2.72%
15%	25.90%	19.88%	15.53%	12.66%	10.60%	**4.52%**	16.38%	15.68%	9.99%	7.30%	5.97%	2.78%
17.5%	26.05%	20.15%	15.51%	12.68%	10.54%	4.49%	16.17%	15.61%	9.94%	7.22%	5.97%	2.78%
20%	26.22%	20.01%	15.56%	12.65%	10.52%	4.43%	16.03%	15.51%	10.05%	7.16%	5.88%	**2.79%**
TSCV	**28.70%**	**20.58%**	**15.76%**	**12.73%**	**10.65%**	4.43%	**19.42%**	**16.67%**	**10.39%**	**7.44%**	**6.10%**	2.75%

Note: This table presents the relative loss reduction of the A-GAS models as compared to their corresponding GAS models, based on 1000 replications of VaR and ES estimations for a GJR-GARCH-SKT DGP with full sample size $T = 9000$. Loss reduction values for the one-factor GAS models are on the left and for the Hybrid GAS/GARCH model are on the right. The first column presents the α_2 candidates from 2.5% up to 20% and the last row reports the values obtained by applying TSCV in the selection of α_2. The rest of the columns present the relative loss reduction for different target levels (α_1). In each column, the α_2 with the highest reduction is highlighted in bold. When $\alpha_1 \geq \alpha_2$, the loss values are reported as N/A.

GAS model. The loss reduction of the augmented GAS one-factor model and the hybrid model are approximately 26% and 16%. The A-GAS models with TSCV are found to be the best GAS-type models, with loss values of 1.702 and 1.719 for the A-GAS-1F and the A-Hybrid model, equivalent to a 28.7% and 19.4% loss reduction, respectively. We further compare the loss obtained by the augmented models with the loss value from the "true" VaR and ES calculated from the DGP (when no model risk is present, labeled as GJR-G-True), as well as with the loss values obtained by estimating the DGP model (obtained when only parameter estimation risk is present, labeled as GJR-G-Est). The results suggest that the augmented models estimated via TSCV lead to risk values that have losses very close to the true loss values, complimenting the accuracy of the risk forecasts.

Next, we compare the augmented GAS models with the alternatives in terms of backtests of the risk forecasts. Three backtests are considered. First, we implement the unconditional coverage (UC) test proposed by Kupiec (1995) which uses the proportion of failures as its main tool to evaluate VaR. Second, the conditional coverage (CC) test proposed by Christoffersen (1998) is considered, and this test addresses the clustering of failures. Third, to evaluate the ES forecasts, we employ the bootstrap (BS) test of McNeil and Frey (2000), which focuses on the discrepancies between the observed returns and the ES forecasts for the periods in which the return exceeds the VaR forecast. We calculate the rate that the null is rejected at 5% level, and we call this the Rejection Rate. This is reported over the OOS period.[9]

Table 2 presents the backtest rejection rate at the extreme level $\alpha_1 = 0.1\%$ for VaR and ES of the two A-GAS models, the GAS model, the true model with true parameter labeled as GJR-G-True, and the true model with estimated parameter labeled as GJR-G-Est. In general, the A-GAS models outperform the GAS model. The A-GAS models with TSCV provide the lowest backtest rejection rates except for the BS backtest results of the A-Hybrid model. On the other hand, the A-GAS-1F model with TSCV has the best

[9] More backtest results are reported in the online Supplemental Appendix.

Table 4
Summary statistics and parameter estimates.

	WTI	Brent	GO	HO
Panel A: Summary statistics				
Mean (annualized)	6.301	3.765	3.445	3.959
Std. dev. (annualized)	42.077	36.969	34.178	36.895
Skewness	0.037	−0.590*	−0.254*	−0.653*
Kurtosis	20.406*	14.411*	8.017*	10.005*
VaR ($\alpha = 0.1\%$)	−12.970	−11.066	−11.483	−15.758
VaR ($\alpha = 0.25\%$)	−10.806	−9.688	−8.610	−9.225
VaR ($\alpha = 0.5\%$)	−9.109	−7.562	−6.757	−8.173
VaR ($\alpha = 1\%$)	−7.089	−6.392	−5.633	−6.217
ES ($\alpha = 0.1\%$)	−22.800	−19.257	−14.882	−19.487
ES ($\alpha = 0.25\%$)	−15.974	−13.685	−11.724	−14.663
ES ($\alpha = 0.5\%$)	−12.952	−10.986	−9.741	−11.652
ES ($\alpha = 1\%$)	−10.460	−8.982	−7.851	−9.360
Panel B: Parameter estimates				
ω	0.064	0.036	0.014	0.029
	(0.016)	(0.012)	(0.006)	(0.010)
β	0.914	0.926	0.951	0.936
	(0.008)	(0.008)	(0.007)	(0.010)
γ	0.076	0.069	0.047	0.060
	(0.009)	(0.008)	(0.007)	(0.009)
ν	7.583	6.985	8.098	7.029
	(0.714)	(0.658)	(0.866)	(0.659)
λ	−0.094	−0.075	−0.058	−0.030
	(0.019)	(0.017)	(0.019)	(0.022)

Note: Summary statistics and parameter estimates for the four futures return series, over the full sample period from January 2000 to January 2021. Panel A reports the annualized mean and standard deviation of the returns expressed in percentages, the skewness, kurtosis, as well as the sample VaR and ES estimates for four different values of risk level α. Panel B presents the estimated parameters of the GARCH (1,1) model with skewed t distributed errors, with the standard errors in parentheses.

* Denotes values of skewness (kurtosis) significantly different from zero (3) at 5% level.

performance in the BS backtest. The GJR-GARCH-SKT model performs best in terms of backtest rejection rate, which is as expected, because this is the DGP model used for the simulation. It is important to note that using the true DGP model is only possible in a simulation setup, whilst in practice the true DGP is unknown.

Additionally, we explore model performance for different values of α_1. Table 3 presents the relative loss reduction obtained by the A-GAS models compared to the GAS models, for different extreme levels of α_1. We consider augmented models for various values of α_1 and α_2. Specifically, $\alpha_1 \in \{0.1\%, 0.25\%, 0.5\%, 0.75\%, 1\%, 2.5\%\}$ and $\alpha_2 \in \{2.5\%, 5\%, 7.5\%, 10\%, 12.5\%, 15\%, 17.5\%, 20\%\}$. We set the restriction that α_1 must be lower than α_2. As α_1 decreases, the loss reduction obtained via the A-GAS model as compared to the corresponding GAS model increases for all α_2 candidates. Also, we find that the A-GAS models with TSCV have the greatest improvement in terms of loss values. The greatest reduction is obtained for $\alpha_1 = 0.1\%$ and it is about 15% and 20% in relative terms for the A-GAS-1F and A-Hybrid models, respectively.

Overall, based on the results of the above simulation studies, incorporating information from an auxiliary level improves the performance of the GAS models when α_1 is extremely small. Also, TSCV is shown to be highly effective to choose the hyper-parameter α_2.

4. Empirical study

4.1. Data description

To evaluate the empirical forecast performance of the proposed models, we study daily returns from four oil futures, the WTI crude oil, Brent crude oil, Gas oil (GO) and Heating oil (HO). These series are representative of the behavior of the returns in the commodity futures markets.[10] The sample period is between 1 January 2000 and 8 January 2021.[11]

[10] Commodity markets are considered to be highly volatile (Del Brio et al., 2020). The proposed risk models can be applied in other markets, such as equity markets. However, due to the highly volatile nature of commodity returns, the issue of estimating risk at extreme levels is most imperative in these markets, which motivated our empirical investigation.

[11] Our data source is Refinitiv Eikon and the Refinitiv Identification Code (RIC) for these four oil futures are: CLc1 (WTI crude oil), LCOc1 (Brent crude oil), LGOc1 (Gas oil), and HOc1 (Heating oil). To ensure continuity in the data, we remove the days with negative prices, market-specific non-trading days and zero returns from each return series.

Table 5
Average losses during the validation period for different values of α_2.

α_2	A-GAS-1F				A-Hybrid			
	WTI	Brent	GO	HO	WTI	Brent	GO	HO
2.5%	2.422	2.486	2.091	2.520	2.494	2.622	1.915	2.667
5%	2.379	2.632	2.196	2.531	2.523	2.425	1.925	2.641
7.5%	2.505	2.509	2.044	2.534	2.493	**2.350**	2.040	2.580
10%	2.293	2.578	2.101	**2.288**	2.502	2.467	1.969	2.466
12.5%	2.286	**2.414**	**1.885**	2.440	2.399	2.414	1.919	2.485
15%	**2.266**	2.463	1.973	2.342	2.465	2.420	**1.895**	2.427
17.5%	2.268	2.524	1.975	2.425	**2.343**	2.476	1.981	2.470
20%	2.285	2.436	1.991	2.452	2.437	2.527	2.037	**2.424**

Note: This table presents the average loss values of VaR and ES at the extreme level $\alpha_1 = 0.1\%$ in the validation period for eight α_2 values, estimated for the return series of four oil futures from January 2007 to December 2013. The left panel indicates the average FZ0 loss value for the A-GAS-1F model whilst the values for the A-Hybrid model are presented in the right panel. The lowest value in each column is in bold, and the corresponding α_2 is selected for the OOS period.

Table 6
Parameter estimates of the GAS and A-GAS models.

	Panel A: WTI				Panel B: Brent			
	GAS-1F	Hybrid	A-GAS-1F	A-Hybrid	GAS-1F	Hybrid	A-GAS-1F	A-Hybrid
β	0.986	0.810	0.971	0.870	0.942	0.785	0.955	0.927
	(0.004)	(0.015)	(0.028)	(0.019)	(0.03)	(0.020)	(0.005)	(0.026)
γ	0.002	0.000	0.068	0.077	0.002	0.000	0.051	0.056
	(0.015)	(0.000)	(0.168)	(0.237)	(0.007)	(0.000)	(0.096)	(0.015)
δ	–	0.058	–	0.064	–	0.074	–	0.039
	–	(0.009)	–	(0.009)	–	(0.038)	–	(0.013)
a_1	−6.271	−8.950	−5.198	−7.443	−9.689	−8.246	−7.082	−9.393
	(6.031)	(0.621)	(5.340)	(1.910)	(1.360)	(1.299)	(1.689)	(0.159)
b_1	−6.727	−9.745	−5.711	−7.444	−9.935	−8.362	−8.911	−14.490
	(4.485)	(0.580)	(5.354)	(1.881)	(1.372)	(1.819)	(1.671)	(0.127)
a_2	–	−1.191	−1.596	–	–	−1.952	−3.662	
	–	–	(14.142)	(5.662)	–	–	(6.905)	(2.001)
b_2	–	–	−2.388	−2.991	–	–	−3.237	−7.8646
	–	–	(14.268)	(5.533)	–	–	(6.866)	(3.097)
Ave. loss	2.274	2.287	2.058	2.057	2.276	2.199	2.152	2.127
	Panel C: GO				Panel D: HO			
	GAS-1F	Hybrid	A-GAS-1F	A-Hybrid	GAS-1F	Hybrid	A-GAS-1F	A-Hybrid
β	0.963	0.903	0.983	0.830	0.989	0.999	0.981	0.957
	(0.003)	(0.002)	(0.001)	(0.030)	(0.002)	(0.000)	(0.042)	(0.105)
γ	0.005	0.001	0.013	0.014	0.000	0.000	0.034	0.079
	(0.004)	(0.000)	(0.004)	(0.066)	(0.003)	(0.000)	(0.008)	(0.103)
δ	–	0.068	–	0.100	–	0.004	–	0.038
	–	(0.000)	–	(0.002)	–	(0.000)	–	(0.019)
a_1	−7.004	−5.878	−6.711	−5.943	−8.739	−7.823	−7.165	−7.823
	(0.731)	(0.743)	(0.221)	(0.326)	(2.156)	(2.126)	(0.243)	(2.550)
b_1	−8.660	−6.555	−6.737	−7.409	−8.966	−7.823	−10.360	−7.916
	(0.849)	(0.68)	(0.217)	(0.353)	(2.314)	(2.149)	(0.138)	(3.154)
a_2	–	–	−2.171	−1.773	–	–	−2.891	−1.403
	–	–	(0.644)	(1.305)	–	–	(0.410)	(8.840)
b_2	–	–	−2.966	−3.261	–	–	−4.231	−2.798
	–	–	(0.684)	(1.299)	–	–	(0.777)	(10.033)
Ave. loss	1.907	1.676	1.828	1.732	2.192	2.015	2.072	2.034

Note: This table presents parameter estimates and standard errors (in parentheses) for two GAS models and two A-GAS models used to forecast VaR and ES at the extreme level $\alpha_1 = 0.1\%$ for four oil futures series over the in-sample period from January 2007 to December 2013. For each return series, the first two columns in each panel present the parameter estimates for the one-factor GAS model and the Hybrid GAS/GARCH model, and the following two columns indicate the parameter values for the one-factor A-GAS model and the A-Hybrid GAS/GARCH model, respectively. The last row of each panel presents the average in-sample FZ0 loss for the four return series.

Table 4 presents the summary statistics of these four series over the full sample period. All return series exhibit substantial kurtosis of between 8 and 20. The table also shows the sample VaR and ES for four levels: 0.1%, 0.25%, 0.5% and 1%, and Panel

Table 7
Out-of-sample backtest performance.

	UC test (VaR) p-values				CC test (VaR) p-values				BS test (ES) p-values			
Panel A: Jan 2014 to Jan 2021												
	WTI	Brent	GO	HO	WTI	Brent	GO	HO	WTI	Brent	GO	HO
RW-500	0.012	0.001	0.010	0.003	0.002	0.000	0.002	0.001	0.032	0.000	0.008	0.030
RW-1000	0.012	**0.051**	0.040	0.013	0.002	0.004	0.004	0.002	**0.108**	**0.138**	0.014	0.048
RW-1500	**0.148**	0.014	0.000	**0.150**	0.007	0.002	0.000	**0.352**	**0.110**	**0.374**	**0.104**	**0.102**
CF-500	**0.396**	**0.160**	**0.366**	**0.399**	**0.694**	**0.369**	0.007	**0.697**	**0.284**	**0.570**	**0.288**	**0.410**
CF-1000	**0.859**	**0.512**	**0.132**	**0.399**	**0.982**	**0.806**	**0.319**	**0.697**	**0.490**	0.000	**0.794**	**0.292**
CF-1500	**0.532**	**0.512**	**0.366**	**0.529**	**0.822**	**0.806**	**0.661**	**0.820**	0.000	0.000	**0.338**	0.000
GARCH-N	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.028	0.008
GARCH-SKT	0.000	0.001	**0.132**	0.003	0.000	0.000	0.006	0.012	**0.236**	**0.516**	**0.140**	**0.054**
GARCH-EDF	0.000	0.000	0.010	0.000	0.000	0.000	0.002	0.000	**0.498**	**0.694**	**0.196**	**0.740**
GJR-GARCH-SKT	0.003	0.003	**0.366**	0.001	0.001	0.001	0.007	0.003	**0.488**	**0.606**	**0.492**	**0.098**
GARCH-FZ	0.000	0.000	0.010	0.003	0.000	0.000	0.037	0.012	**0.294**	**0.674**	**0.146**	0.012
GAS-1F	**0.396**	**0.887**	**0.366**	0.047	0.007	**0.988**	**0.661**	**0.137**	**0.472**	**0.474**	**0.784**	**0.146**
Hybrid	0.046	**0.051**	**0.132**	0.001	**0.136**	0.004	**0.319**	0.000	**0.140**	**0.146**	**0.078**	0.006
GAS-2F	**0.148**	**0.417**	**0.132**	0.047	0.007	0.007	**0.319**	**0.137**	**0.368**	**0.594**	**0.706**	**0.084**
A-GAS-1F	0.000	**0.051**	0.040	0.013	0.000	0.004	0.004	0.044	**0.062**	0.028	0.044	**0.550**
A-Hybrid	0.001	**0.051**	0.000	0.013	0.003	**0.147**	0.000	0.044	**0.810**	0.046	**0.952**	0.028
Panel B: Jan 2020 to Jan 2021												
	UC test (VaR) p-values				CC test (VaR) p-values				BS test (ES) p-values			
	WTI	Brent	GO	HO	WTI	Brent	GO	HO	WTI	Brent	GO	HO
RW-500	0.029	0.003	0.002	0.002	0.002	0.001	0.001	0.001	**0.526**	0.078	**0.074**	**0.094**
RW-1000	0.000	0.003	0.002	0.002	0.000	0.001	0.001	0.001	**0.094**	0.072	**0.596**	**0.080**
RW-1500	0.000	0.000	0.000	0.030	0.000	0.000	0.000	**0.093**	**0.118**	**0.420**	**0.108**	**0.482**
CF-500	**0.262**	**0.275**	0.028	**0.268**	**0.531**	**0.548**	0.002	**0.539**	0.000	0.000	**0.538**	0.000
CF-1000	**0.262**	**0.275**	0.028	**0.268**	**0.531**	**0.548**	**0.089**	**0.539**	0.000	0.000	**0.478**	0.000
CF-1500	**0.262**	**0.275**	0.028	**0.472**	**0.531**	**0.548**	**0.089**	0.000	0.000	0.000	**0.508**	0.000
GARCH-N	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.001	**0.076**	**0.052**	**0.138**	**0.074**
GARCH-SKT	0.002	0.000	0.002	0.030	0.001	0.000	0.001	**0.093**	**0.506**	**0.528**	**0.276**	**0.498**
GARCH-EDF	0.000	0.000	0.002	0.002	0.000	0.000	0.001	0.001	**0.530**	**0.592**	**0.088**	**0.484**
GJR-GARCH-SKT	0.029	0.003	0.028	0.030	0.002	0.001	0.002	**0.093**	**0.456**	**0.544**	**0.552**	**0.556**
GARCH-FZ	0.000	0.000	0.028	0.002	0.000	0.000	**0.089**	0.009	**0.184**	**0.556**	**0.494**	**0.402**
GAS-1F	0.029	0.031	**0.260**	0.030	0.002	**0.097**	**0.528**	**0.093**	**0.508**	**0.520**	0.000	**0.480**
Hybrid	0.002	0.003	0.028	0.002	0.009	0.001	**0.089**	0.001	**0.070**	**0.082**	**0.498**	**0.092**
GAS-2F	0.002	0.003	0.028	0.030	0.001	0.001	**0.089**	**0.093**	**0.322**	**0.570**	**0.504**	**0.492**
A-GAS-1F	0.002	0.003	0.002	0.030	0.009	0.001	0.001	**0.093**	**0.516**	**0.604**	**0.156**	**0.488**
A-Hybrid	0.029	**0.275**	0.002	**0.472**	**0.090**	**0.548**	0.001	0.000	**0.510**	0.000	**0.362**	0.000

Note: This table presents the p-values of two VaR backtests and an ES backtest for four oil futures, over the whole OOS period (Panel A) and the COVID-19 period (Panel B) for 16 risk forecasting models at level $\alpha_1 = 0.1\%$. Columns 2–5 and 6–9 present the results for the Unconditional Coverage (UC) and the Conditional Coverage (CC) backtest for the evaluation of VaR. The last 4 Columns present the results of the Bootstrapping (BS) backtest for the evaluation of ES. Values greater than 0.05 (indicating no evidence against optimality at 5% significance level) are in bold.

B presents the estimated parameters of the GARCH (1,1) model with a skewed t distribution, fitted to the de-meaned returns, with the model defined as:

$$\begin{aligned} Y_t &= \sigma_t \eta_t, \ \eta_t \sim iid \text{ Skew } t(0, 1, \nu, \lambda), \\ \sigma_t^2 &= \omega + \beta \sigma_{t-1}^2 + \gamma Y_{t-1}^2. \end{aligned} \tag{19}$$

The full sample is divided into a training period (January 2000 to December 2006), a validation period (January 2007 to December 2013), and an OOS forecasting period (January 2014 to January 2021). In the validation period, we employ the TSCV introduced in Section 2.3.4 with eight candidates of $\alpha_2 \in \{2.5\%, 5\%, 7.5\%, 10\%, 12.5\%, 15\%, 17.5\%, 20\%\}$, and the α_2 value which provides the lowest loss value is selected as the optimal α_2 for forecasting OOS. Then we produce risk forecasts for $\alpha_1 = 0.1\%$ for the OOS period, and the forecasting performance is evaluated in the OOS period.

4.2. Estimation results

Table 5 presents the loss values for VaR and ES at the extreme level $\alpha_1 = 0.1\%$ for different values of α_2 for the four series considered.[12] The optimal value of α_2 is found in the range from 10% to 15% for the A-GAS-1F model, while for the A-Hybrid

[12] We use $M_1 = 30$ days to reduce computational time.

Table 8
Out-of-sample losses and loss rankings.

	Average loss				Loss ranking				
Panel A: Jan 2014 to Jan 2021									
	WTI	Brent	GO	HO	WTI	Brent	GO	HO	Average
RW-500	4.315	4.056	3.869	5.031	15	14	16	15	15
RW-1000	3.908	3.803	3.197	4.569	11	13	13	12	12.25
RW-1500	3.911	4.116	3.314	3.514	12	15	14	5	11.5
CF-500	3.599	2.994	2.998	3.474	9	6	11	3	7.25
CF-1000	3.272	3.035	2.781	3.870	6	7	9	11	8.25
CF-1500	3.310	3.175	2.730	3.575	7	9	8	7	7.75
GARCH-N	4.912	4.770	3.526	5.131	16	16	15	16	15.75
GARCH-SKT	3.335	2.990	2.631	3.703	8	5	6	8	6.75
GARCH-EDF	3.688	3.425	2.836	3.804	10	10	10	10	10
GJR-GARCH-SKT	*2.886*	2.736	**2.333**	3.558	2	3	1	6	*3*
GARCH-FZ	4.127	3.498	2.428	4.697	14	11	3	13	10.25
GAS-1F	**2.665**	2.890	2.472	3.509	1	4	4	4	3.25
Hybrid	3.986	3.147	2.568	4.970	13	8	5	14	10
GAS-2F	3.055	3.572	2.648	*3.448*	5	12	7	2	6.5
A-GAS-1F	2.997	**2.271**	*2.404*	**2.966**	4	1	2	1	**2**
A-Hybrid	2.911	*2.523*	3.061	3.717	3	2	12	9	6.5
Panel B: Jan 2020 to Jan 2021									
	WTI	Brent	GO	HO	WTI	Brent	GO	HO	Average
RW-500	10.824	11.153	9.277	9.962	11	12	16	14	13.25
RW-1000	12.517	12.562	7.458	11.196	14	14	13	16	14.25
RW-1500	13.743	14.715	8.654	5.689	16	16	15	6	13.25
CF-500	6.232	5.715	4.693	5.188	5	3	7	4	4.75
CF-1000	7.930	7.155	4.315	6.547	10	7	4	7	7
CF-1500	7.678	7.593	4.437	3.339	9	8	6	3	6.5
GARCH-N	11.561	12.780	8.324	9.890	12	15	14	13	13.5
GARCH-SKT	6.624	7.105	4.765	6.802	6	6	8	8	7
GARCH-EDF	7.639	8.416	5.939	7.011	8	9	12	9	9.5
GJR-GARCH-SKT	4.871	5.868	*3.372*	5.657	4	4	2	5	3.75
GARCH-FZ	12.183	9.066	**2.613**	10.237	13	11	1	15	10
GAS-1F	4.449	6.730	4.364	8.340	3	5	5	12	6.25
Hybrid	13.683	8.995	5.391	8.048	15	10	9	10	11
GAS-2F	7.115	11.617	5.898	8.265	7	13	11	11	10.5
A-GAS-1F	*3.699*	*3.065*	4.248	*2.887*	2	2	3	2	**2.25**
A-Hybrid	**2.713**	**2.925**	5.544	**2.330**	1	1	10	1	*3.25*

Note: This table presents the average losses and loss rankings (with the best performing model ranked 1 and the worst ranked 16) based on average FZ0 losses, for VaR and ES forecasts at level $\alpha_1 = 0.1\%$ of four oil futures, over the OOS period (Panel A) and the COVID-19 period (Panel B). Columns 2–5 present the average FZ0 losses, with the lowest (second lowest) in each column shown in bold (italics). Columns 6–9 present the loss rankings. The last column presents the average rank across the four series, with the best (second best) model shown in bold (italics).

model we find that α_2 is above 15%, except for $\alpha_2 = 7.5\%$ for Brent. Table 6 presents the estimated parameters together with their standard errors for the GAS and A-GAS models with α_2 chosen by TSCV over the validation period. Both A-GAS models estimated on all four energy commodity futures demonstrate a higher value of γ compared to the original GAS models. This implies that the estimated VaR and ES of the A-GAS models at the extreme level α_1 are influenced by the value of VaR and ES at the auxiliary level α_2. A higher value of γ in the A-GAS models indicates that the κ_t process (and the VaR and ES processes) is more reactive to the forcing variable s_{t-1}. The parameters a_2 and b_2 in the A-GAS models are also reasonable, ensuring that the VaR and ES at the auxiliary level are lower than at the extreme level. Notably, the loss values of the A-GAS-1F model are lower than those of the GAS-1F across all series considered. Furthermore, the A-Hybrid model reports lower losses than the Hybrid model for both WTI and Brent.

4.3. Out-of-sample results

We now turn to the OOS forecast performance of the A-GAS models at the extreme level $\alpha_1 = 0.1\%$, as compared to a total of fourteen alternative models. Six non-parametric models are considered as benchmarks, including the traditional rolling window methods with window lengths of 500, 1000 and 1500 trading days and rolling window methods based on the Cornish–Fisher expansion (Cornish and Fisher, 1938), with the same window lengths as the first three models. Four prevailing GARCH models are also considered as benchmarks, namely, the GARCH model with normal distribution (GARCH-N), GARCH with skewed t distribution

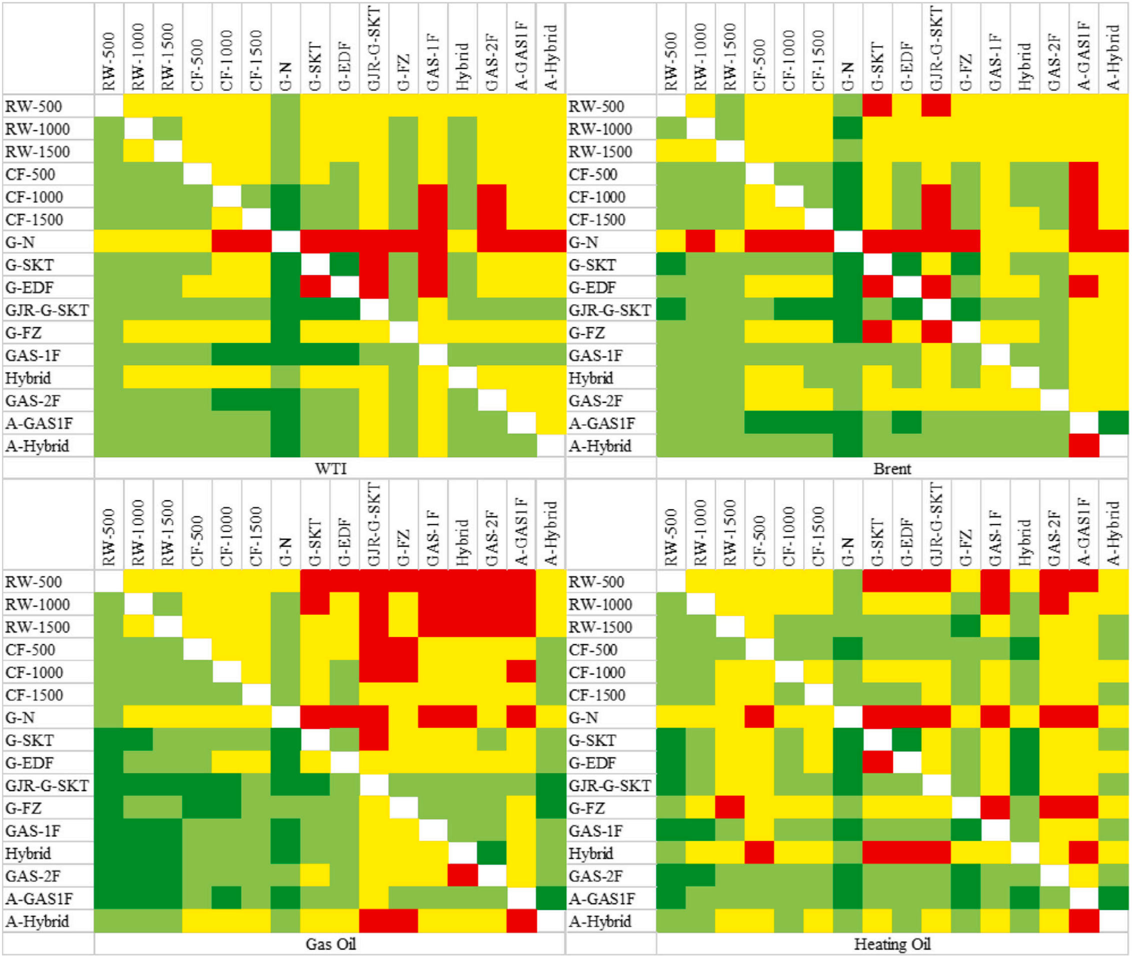

Fig. 3. Diebold–Mariano (DM) test results comparing the FZ0 losses over the OOS period from January 2014 to January 2021, for 16 models across four oil futures, comparing risk forecasts at level $\alpha_1 = 0.1\%$. Dark green (red) blocks mean that the row model has significantly lower (greater) average loss than the column model at 10% significance level; light green (yellow) blocks mean that the row model has lower (greater) average loss than the column model, but the difference is not significant.

(GARCH-SKT), GARCH with empirical distribution function (GARCH-EDF) in which case VaR and ES are estimated from the sample VaR and ES of the estimated standardized residuals obtained from the GARCH model, and the GJR-GARCH with skewed t distribution (GJR-GARCH-SKT). We next consider four models introduced by Patton et al. (2019): the two-factor GAS model (GAS-2F), the one-factor GAS model (GAS-1F), the GARCH model using FZ loss minimization (GARCH-FZ), and the hybrid GAS/GARCH model (Hybrid). Finally, we consider the two proposed augmented GAS models, the A-GAS-1F model and the A-hybrid model. We estimate the parametric and semiparametric models using the first 7 years starting with 2007 as the in-sample period, and retain the parameter estimates to build forecasts for the OOS period.

Table 7 presents the p-values of the VaR and ES backtests in the OOS period (from January 2014 to January 2021) and over the COVID-19 period (from January 2020 to January 2021) for 16 models and for four oil futures. As before, we forecast risk at the level of $\alpha_1 = 0.1\%$. We find that, over the whole OOS period, the augmented models can pass the VaR backtests (UC and CC) for the Brent series. Considering the ES backtest (BS), our models provide reasonable backtest results for the time series of WTI, GO, and HO. However, when we consider only the COVID-19 period, the models proposed by Patton et al. (2019) cannot pass the UC test anymore, and our models experience a significant improvement during this period; this is especially true for the A-Hybrid model and the VaR backtests.

To provide a more robust evaluation, we also employ a comparison based on the FZ0 loss function to assess the performance of the models considered. Table 8 presents the average losses and the ranks of the models based on loss values at the level $\alpha_1 = 0.1\%$ over the OOS period and the COVID-19 period. Over the whole OOS period, the A-GAS-1F model provides the lowest average loss for Brent and HO. Even though the A-GAS-1F model is not the best-performing model for all four futures, its average ranking is the best among all the models. During the COVID-19 period, which contains a high concentration of extreme losses, the A-Hybrid model has the lowest average loss for all futures except for GO series. The A-GAS-1F model provides stable performance during the COVID-19 period, with the best average ranking overall.

While average losses are a useful tool to consider forecast performance out-the-sample, they do not provide information on the significance of the loss differences between models. Fig. 3 presents the results of the Diebold–Mariano (DM) test that performs pairwise model comparisons based on loss differences over the OOS period, with the null hypothesis that the row model and the column model have equal loss values. The A-GAS-1F model has superior performance for all series, especially for Brent and HO, outperforming all alternative models considered.

5. Robustness check

In practice, time series are often characterized by the presence of structural breaks in the fitted models. In Section 4, the forecasts of VaR and ES in the OOS are based on the model parameters estimated in the in-sample period, without updating. In this section, we perform a robustness check by updating the parameter estimates using rolling-windows for the forecasts of VaR and ES in the OOS. Specifically, each model is re-estimated every 30 trading days using a window length of 1805 observations (7 years), with the first window starting from January 2007. The same OOS period is used to evaluate VaR and ES estimates.[13] We find that the results are similar to the ones reported in earlier sections of this paper. Over the OOS period, the loss ranks of the A-GAS models slightly decrease, but the A-GAS-1F model has the best performance overall for most of the oil futures series considered. When considering the COVID-19 period in isolation, the A-GAS models, based on rolling window estimation, show the same superior performance as before.

6. Conclusion

This paper introduces augmented versions of the GAS models that jointly estimate risk at an extreme level and an auxiliary level, with the purpose to improve on the forecasts of VaR and ES at an extreme level. By using TSCV to select the optimal auxiliary level, we document an improvement in the risk forecasts both in-sample and during the OOS periods considered. Our simulation study also highlights this improvement in terms of the forecast loss and the backtest rejection rates. We employ the proposed A-GAS models to forecast the VaR and ES of four oil futures over the period from January 2000 to January 2021. We compare these with forecasts made by fourteen alternative models, and we implement several backtests to compare their performance. The main finding is that VaR and ES forecasts obtained from the A-GAS models outperform the risk forecasts based on popular GARCH models or historical simulations, and they also lead to improved loss values compared with the original GAS models for three out of four future series considered. The A-GAS models perform even better during the COVID-19 period which is characterized by extreme losses. As such, the proposed augmented versions of popular GAS risk models can provide improved risk forecasts at extreme levels by utilizing the information from prevailing risk levels without considering exogenous information. Applications of these models to study the risk of other asset classes would be of future interest. Additionally, the proposed framework of estimating risk at extreme levels can be extended to more than two risk level or by considering alternative risk models.

Funding disclosure

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Emese Lazar: Conceptualization, Methodology, Writing – review & editing. **Jingqi Pan:** Conceptualization, Formal analysis, Methodology, Software, Writing – original draft. **Shixuan Wang:** Conceptualization, Methodology, Writing – review & editing.

Declaration of competing interest

None

Data availability

The data that support the findings of this study is openly available from Refinitiv DataStream.

Appendix A. Supplemental material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jcomm.2024.100391.

[13] The online Supplementary Appendix presents the results for the robustness check.

References

Acerbi, C., Tasche, D., 2002. Expected shortfall: a natural coherent alternative to value at risk. Econ. Notes 31 (2), 379–388.
Artzner, P., Delbaen, F., Eber, J.-M., Heath, D., 1999. Coherent measures of risk. Math. Finance 9 (3), 203–228.
Basel Committee on Banking Supervision, 2013. Fundamental review of the trading book: A revised market risk framework. Available at http://www.bis.org/publ/bcbs265.pdf. (Accessed September 2022).
Chavez-Demoulin, V., Embrechts, P., Sardy, S., 2014. Extreme-quantile tracking for financial time series. J. Econometrics 181 (1), 44–52.
Chen, J., 2018. On exactitude in financial regulation: Value-at-risk, expected shortfall, and expectiles. Risks 6 (2).
Christoffersen, P.F., 1998. Evaluating interval forecasts. Internat. Econom. Rev. 39 (4), 841–862.
Cornish, E.A., Fisher, R.A., 1938. Moments and cumulants in the specification of distributions. Rev. Inst. Int. Statist. 5 (4), 307–320.
Creal, D., Koopman, S.J., Lucas, A., 2013. Generalized autoregressive score models with applications. J. Appl. Econometrics 28 (5), 777–795.
Danielsson, J., De Vries, C.G., 1998. Beyond the Sample: Extreme Quantile and Probability Estimation. Rotterdam Tinbergen Institute (Preprint).
Del Brio, E.B., Mora-Valencia, A., Perote, J., 2020. Risk quantification for commodity ETFs: Backtesting value-at-risk and expected shortfall. Int. Rev. Financ. Anal. 70, 101163.
Duffie, D., Pan, J., 1997. An overview of Value at Risk. J. Derivatives 4 (3), 7–49.
Engle, R.F., Manganelli, S., 2004. CAViaR: Conditional autoregressive value at risk by regression quantiles. J. Bus. Econom. Statist. 22 (4), 367–381.
Fissler, T., Ziegel, J.F., 2016. Higher order elicitability and Osband's principle. Ann. Statist. 44 (4), 1680–1707.
Gneiting, T., 2011. Making and evaluating point forecasts. J. Amer. Statist. Assoc. 106 (494), 746–762.
Hansen, B.E., 1994. Autoregressive conditional density estimation. Internat. Econom. Rev. 35 (3), 705–730.
Hart, J.D., 1994. Automated kernel smoothing of dependent data by using time series cross-validation. J. R. Stat. Soc. Ser. B Stat. Methodol. 56 (3), 529–542.
Harvey, A.C., 2013. Dynamic Models for Volatility and Heavy Tails: With Applications To Financial and Economic Time Series, vol. 52, Cambridge University Press.
Hoga, Y., 2017. Testing for changes in (extreme) VaR. Econom. J. 20 (1), 23–51.
Hoga, Y., 2021. The uncertainty in extreme risk forecasts from covariate-augmented volatility models. Int. J. Forecast. 37 (2), 675–686.
Hung, J.-C., Lee, M.-C., Liu, H.-C., 2008. Estimation of value-at-risk for energy commodities via fat-tailed GARCH models. Energy Econ. 30 (3), 1173–1191.
Hyndman, R.J., Athanasopoulos, G., 2018. Forecasting: Principles and Practice. OTexts.
Kourouma, L., Dupre, D., Sanfilippo, G., Taramasco, O., 2010. Extreme value at risk and expected shortfall during financial crisis. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1744091, (Accessed on September 2022).
Kupiec, P., 1995. Techniques for verifying the accuracy of risk measurement models. J. Derivatives 3 (2), 73–84.
Li, H., Liu, Z., Wang, S., 2022. Vines climbing higher: Risk management for commodity futures markets using a regular vine copula approach. Int. J. Finance Econ. 27 (2), 2438–2457.
Liu, H.-C., Hung, J.-C., 2010. Forecasting S&P100 stock index volatility: The role of volatility asymmetry and distributional assumption in GARCH models. Expert Syst. Appl. 37 (7), 4928–4934.
Maciel, L., 2021. Cryptocurrencies value-at-risk and expected shortfall: Do regime-switching volatility models improve forecasting? Int. J. Finance Econ. 26 (3), 4840–4855.
McNeil, A.J., Frey, R., 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. J. Empir. Financ. 7 (3–4), 271–300.
Nolde, N., Ziegel, J.F., 2017. Elicitability and backtesting: Perspectives for banking regulation. Ann. Appl. Stat. 11 (4), 1833–1874.
Patton, A.J., Ziegel, J.F., Chen, R., 2019. Dynamic semiparametric models for expected shortfall (and value-at-risk). J. Econometrics 211 (2), 388–413.
Roccioletti, S., 2015. Backtesting Value At Risk and Expected Shortfall. Springer.
Sadorsky, P., 1999. Oil price shocks and stock market activity. Energy Econ. 21 (5), 449–469.
Storti, G., Wang, C., 2022. Nonparametric expected shortfall forecasting incorporating weighted quantiles. Int. J. Forecast. 38 (1), 224–239.
Taylor, J.W., 2020. Forecast combinations for value at risk and expected shortfall. Int. J. Forecast. 36 (2), 428–441.
White, H., Kim, T.-H., Manganelli, S., 2015. VAR for VaR: Measuring tail dependence using multivariate regression quantiles. J. Econometrics 187 (1), 169–188.
Ziegel, J.F., 2016. Coherence and elicitability. Math. Finance 26 (4), 901–918.