A Bayesian approach for more reliable tail risk forecasts *

Dan Li †[a,b,c], Adam Clements[a,c] and Christopher Drovandi[b,c]

[a]School of Economics and Finance, Queensland University of Technology, Australia

[b]School of Mathematical Sciences, Queensland University of Technology, Australia

[c]QUT Centre for Data Science, Queensland University of Technology, Australia

Abstract

This paper demonstrates that existing quantile regression models used for jointly forecasting Value-at-Risk (VaR) and expected shortfall (ES) are sensitive to initial conditions. Given the importance of these measures in financial systems, this sensitivity is a critical issue. A new Bayesian quantile regression approach is proposed for estimating joint VaR and ES models. By treating the initial values as unknown parameters, sensitivity issues can be dealt with. Furthermore, new additive-type models are developed for the ES component that are more robust to initial conditions. A novel approach using the open-faced sandwich (OFS) method is proposed which improves uncertainty quantification in risk forecasts. Simulation and empirical results highlight the improvements in risk forecasts ensuing from the proposed methods.

Keywords— CAViaR, Value-at-Risk, Expected shortfall, Sequential Monte Carlo, Uncertainty quantification, Systemic risk

---

*We gratefully acknowledge the helpful feedback and valuable suggestions of the three anonymous reviewers and the journal editor. DL is supported by a scholarship from the QUT Centre for Data Science and a supervisor's top-up scholarship from the school of Mathematical Sciences of QUT. Computational resources used in this work were provided by QUT's High Performance Computing and Research Support Group (HPC).

†Corresponding author at: School of Economics and Finance, Queensland University of Technology, Australia. *E-mail addresses:* d33.li@qut.edu.au, d31.li@hdr.qut.edu.au (Dan Li), a.clements@qut.edu.au (Adam Clements), c.drovandi@qut.edu.au (Chris Drovandi).

1 Introduction

Prudential regulation of the financial system is based on measures of financial risk. Therefore accurate and reliable measures of financial risk are crucially important. Value-at-Risk (VaR) is the most commonly used risk measure over the years since its introduction in RiskMetrics (Morgan, 1995) due to its conceptual simplicity and its importance of setting regulatory capital requirements for financial institutions. VaR is a particular quantile of the conditional distribution of portfolio returns, and it measures the loss of a certain portfolio within a given period at a given confidence level (e.g. 95% or 99%). However, VaR does not indicate what the loss would be in the case that returns exceed the quantile, a quantity captured by Expected Shortfall (ES). ES is the expected loss conditional on returns beyond the VaR measure. The attractive properties of ES are discussed in Acerbi and Tasche (2002). Even though ES has been widely employed by financial institutions, it is not an elicitable measure, which means that there is no loss function that can be optimized uniquely by the true ES (Gneiting, 2011). Although ES is not elicitable, Fissler and Ziegel (2016) argued that VaR and ES are jointly elicitable. Taylor (2019) proposes new joint models of VaR and ES and a new loss function for estimating VaR and ES simultaneously.

For the ES component, we demonstrate that the joint models of Taylor (2019) are sensitive to the assumed starting values of the conditional VaR and ES series, particularly so for the additive-type model. Since unreliable risk measures can influence the judgment of financial decision-makers, such as less robust policymaking and capital allocation (see, for example, Danielsson et al., 2016 and Arismendi-Zambrano et al., 2022), it is crucial to address the sensitivity issue. Gerlach and Wang (2020) demonstrated, for some indices the additive-type model provides the best forecasting performance using realized measures. We show that the choice of the initial values can yield very different estimation and forecasting results for the additive-type model. We propose to address the sensitivity problem in two ways. Firstly, we argue that for both additive-type and multiplicative-type joint models, the sensitivity issue can be largely addressed by treating these initial values as unknown parameters and providing suitable prior settings accordingly. All parameters, including the initial values, are estimated in a Bayesian framework which we show produces more robust VaR and ES forecasts. Secondly, several new specifications of the additive-type model are developed in this work, which are less sensitive to the initial conditions than the original one of Taylor (2019). One of the newly proposed additive-type models is based on a link to a parametric GARCH model

employed in Gerlach and Wang (2020), which allows for a simulation study for the additive-type model. Previously, the simulation study of the joint VaR and ES model is restricted to the one with a multiplicative-type ES component according to Gerlach and Wang (2020).

The forecasting models of tail risk measures enable joint estimation of conditional quantile and ES based on an asymmetric Laplace (AL) working likelihood. We propose a Bayesian quantile regression approach to estimating the joint models. Besides largely addressing the sensitivity issues of the existing VaR and ES forecasting models, the Bayesian approach can provide uncertainty estimates of the unknown parameters based on the posterior samples, and thus the uncertainty associated with VaR and ES forecasts can be estimated. We explore the posterior coverage of prediction intervals for VaR and ES, which provides insights into the 'risks' associated with the tail risk measures. Estimating tail risk forecasting models via a Bayesian quantile regression approach is not new. Gerlach et al. (2011) adapted a Bayesian approach based on the skewed-Laplace distribution to estimate the Conditional Autoregressive VaR (CAViaR) models of Engle and Manganelli (2004). Gerlach and Wang (2020) employed an adaptive Bayesian Markov chain Monte Carlo method to estimate the joint VaR and ES models of Taylor (2019) and demonstrated the Bayesian approach outperformed the maximum likelihood method in terms of point estimation. To the best of our knowledge, however, the uncertainty estimates of the tail risk forecasting models produced from the Bayesian quantile regression approach have not been explored in the literature. Through the ubiquitous application of tail risk measures by financial institutions and regulators, quantification of model risk for VaR and ES forecasts is gaining attention (see, for example, Kerkhof et al., 2010; Alexander and Sarabia, 2012; Danielsson et al., 2016; Lazar and Zhang, 2019; and Farkas et al., 2020). As one source of model risk, an understanding of estimation uncertainty highlights the extent with which decision-makers can rely on tail risk measures is important. Kerkhof et al. (2010) and Farkas et al. (2020) addressed estimation risk by deriving confidence intervals for tail risk measures under the context of financial capital requirements. Patton et al. (2019) derived asymptotic distributions for the parameters of their proposed dynamic semiparametric models for VaR and ES. In this work, we focus on the quantification of parameter uncertainties via a Bayesian approach. We demonstrate that for VaR and ES forecasts, the corresponding uncertainty intervals under the adopted Bayesian quantile regression approach can be misleading, due to the misspecified AL working likelihood. To obtain more accurate uncertainty

estimates for the tail risk measures, an adjustment is made to the original posterior samples. A simulation study shows that significant improvements result from the adjustment.

VaR and ES not only quantify the risks taken by financial institutions individually, but they are also important elements in systemic risk modelling (see Silva et al., 2017 for a general overview of systemic risk models). Danielsson et al. (2016) provided empirical evidence of how the different values of VaR and ES estimates from different models affect systemic risk measures. They showed that the two popular systemic measures, conditional VaR (CoVaR) of Adrian and Brunnermeier (2011) and the marginal ES (MES) of Acharya et al. (2017), which fundamentally depend on VaR, can be influenced by the uncertainties in VaR estimates. In this work, we develop a procedure for jointly estimating the CoVaR and CoES (Co-Expected-Shortfall) systemic risk measures proposed by Adrian and Brunnermeier (2011) by following a similar approach taken for estimating joint VaR and ES models. We empirically illustrate that unreliable VaR forecasts resulting from the sensitivity issue can lead to unreliable CoVaR and CoES estimates. Such impact can be further investigated in other applications based on CoVaR and CoES, such as the case of measuring financial institutions' resiliency proposed in Gehrig and Iannino (2021).

The rest of the article is organized as follows. Section 2 reviews the existing joint VaR and ES models of Taylor (2019) and introduces the new additive-type models. The link between one of the proposed models to a parametric GARCH model is described in this section. In Section 3, we provide the framework of quantile regression and the working likelihood for estimating the joint VaR and ES models. An explanation of a Bayesian estimation and prediction approach for the quantile regression model is discussed in Section 4. The prior settings of the initial values of the conditional VaR and ES series are provided, and the method of posterior adjustment is also described in this section. In Section 5, we conduct a simulation study to investigate the accuracy of interval estimates of tail risk measures after employing the posterior adjustment method. Section 6 documents the empirical results and concluding comments are provided in Section 7.

2 Joint ES and VaR Models

2.1 Existing Models

Taylor (2019) proposed a semiparametric approach to model conditional VaR and ES jointly based on the equivalence between the quantile regression estimator and a Maximum Likelihood

Estimate (MLE) of an asymmetric Laplace (AL) density. For the VaR component, two specifications of the Conditional Autoregressive VaR (CAViaR) framework (Engle and Manganelli, 2004) were adopted. For the ES component, two models were proposed that ensure the corresponding two quantities do not cross (i.e. the absolute value of ES is always larger than VaR). The specifications of the two components are as follows:

$$\text{Symmetric Absolute Value (SAV):} \quad Q_t = \beta_0 + \beta_1 |r_{t-1}| + \beta_2 Q_{t-1}, \tag{1}$$

$$\text{Asymmetric Slope (AS):} \quad Q_t = \beta_0 + \beta_1 I(r_{t-1} > 0)|r_{t-1}| + \beta_2 I(r_{t-1} \leq 0)|r_{t-1}| + \beta_3 Q_{t-1}, \tag{2}$$

where Q_t is the conditional quantile of r_t, $I(\cdot)$ is the indicator function, and r_{t-1} is a financial return observed at $t-1$.

$$\text{Multiplicative:} \quad ES_t = (1 + \exp(\gamma_0))Q_t. \tag{3}$$

$$\text{Additive:} \quad ES_t = Q_t - x_t,$$

$$x_t = \begin{cases} \gamma_0 + \gamma_1(Q_{t-1} - r_{t-1}) + \gamma_2 x_{t-1} & \text{if } r_{t-1} \leq Q_{t-1} \\ x_{t-1} & \text{otherwise} \end{cases}, \tag{4}$$

where $\gamma_0 \geq 0$, $\gamma_1 \geq 0$, and $\gamma_2 \geq 0$ in Eq (4) ensure that ES and VaR do not cross. Specifically, four types of VaR-ES joint models are considered here, which are the combinations of the SAV CAViaR model and the Multiplicative ES model (SAV-Mult), the SAV CAViaR model and the Additive ES model (SAV-Add), the AS CAViaR model and the Multiplicative ES model (AS-Mult), and the AS CAViaR model and the Additive ES model (AS-Add).

We demonstrate that the estimation and prediction results of the joint VaR and ES model of Taylor (2019) can be sensitive to the choices of Q_1 and ES_1, especially for the additive-type models. We argue that this sensitivity can be largely addressed by treating the initial values as unknown parameters, whose posterior distributions can be approximated via the Bayesian approach introduced in Section 4. The details on Bayesian settings of the unknown parameters Q_1 and ES_1 are discussed later in Section 4.2. Furthermore, several new specifications of the additive-type model are proposed, which we demonstrate are less sensitive to the initial values of Q_1 and ES_1 compared to the SAV-Add and AS-Add models. Moreover, some of the new variations of the additive-type models can provide superior forecasting performance.

2.2 New Additive-type models

The development of a new additive model is motivated by exploiting the link between an additive-type VaR-ES joint model and a parametric GARCH-type model which allows a simulation study to be conducted. According to Gerlach and Wang (2020), the simulation study is restricted to the SAV-Mult model, as there is no direct link between the model with an additive ES component and a GARCH model. Our newly proposed additive-type ES component takes the form of:

$$\begin{aligned} \text{New Additive (NewAdd):} \quad & ES_t = Q_t - x_t, \\ & x_t = \gamma_0 + \gamma_1 |r_{t-1}| + \gamma_2 x_{t-1}, \end{aligned} \tag{5}$$

where $\gamma_0 \geq 0$, $\gamma_1 \geq 0$, and $\gamma_2 \geq 0$, as in the original additive-type model of Eq (4), to ensure that ES and VaR are not crossing. Constraining γ_2 to be equal to the autoregressive coefficient of the conditional quantile Q_{t-1} in Eq 1, we demonstrate here that the VaR and ES under the newly proposed additive-type model can be linked to the GARCH(1,1) model for the standard deviation with the following specifications:

$$\begin{aligned} r_t &= \sqrt{h_t}\epsilon_t, \\ \sqrt{h_t} &= a_0 + a_1 |r_{t-1}| + a_2 \sqrt{h_{t-1}}, \\ \epsilon_t &\overset{iid}{\sim} N(0,1), \end{aligned} \tag{6}$$

where h_t is the conditional variance of r_t and ϵ_t is the error term.

Gerlach and Wang (2020) showed that the true parameters of the SAV-Mult model can be obtained through a mapping from the above GARCH model to the SAV-Mult model. Specifically, $Q_t = \sqrt{h_t}\Phi^{-1}(\alpha)$, $ES_t = -\sqrt{h_t}\frac{\phi(\Phi^{-1}(\alpha))}{\alpha}$, where α is a chosen probability level of the quantile, $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal probability density function (PDF) and cumulative distribution function (CDF), respectively. Thus, the true parameters of the Q_t component model can be obtained as follows:

$$\beta_0 = a_0\Phi^{-1}(\alpha),\ \beta_1 = a_1\Phi^{-1}(\alpha),\ \beta_2 = a_2.$$

For the ES component of the SAV-Mult model, the true value of the parameter γ_0 can be calculated through the fixed ratio between ES_t and Q_t, $\frac{ES_t}{Q_t} = \frac{-\phi(\Phi^{-1}(\alpha))}{\alpha\Phi^{-1}(\alpha)} = 1 + \exp(\gamma_0)$.

For the new additive-type model of Eq (5), the additive difference x_t can be defined as

$$\begin{aligned} x_t &= Q_t - ES_t \\ &= \sqrt{h_t}\Phi^{-1}(\alpha) - \left(-\sqrt{h_t}\frac{\phi(\Phi^{-1}(\alpha))}{\alpha}\right) \\ &= \sqrt{h_t}C, \end{aligned} \tag{7}$$

where $C = \Phi^{-1}(\alpha) + \frac{\phi(\Phi^{-1}(\alpha))}{\alpha}$. Then $\sqrt{h_t}$ can be expressed in terms of x_t as $\sqrt{h_t} = x_t/C$. Substituting $\sqrt{h_t}$ in the GARCH model of Eq (6) with x_t/C, the difference x_t of the proposed model in Eq (5) can be written as:

$$x_t = a_0 C + a_1 C|r_{t-1}| + a_2 x_{t-1}, \tag{8}$$

where $a_2 = \beta_2$ which indicates that x_t in Eq 5, and Q_t modelled by the SAV CAViaR model, share the same coefficient for their autoregressive terms x_{t-1} and Q_{t-1}. Thus, the true parameter values of the proposed additive model can be obtained. It is worth noting that the true parameter values of the NewAdd model can only be identified when constraining γ_2 in Eq (5) to be equal to β_2 in Eq (1). We consider both the constrained and unconstrained γ_2 for the NewAdd model in the empirical analysis. We denote the model with constrained γ_2 as the NewAdd-C model, and the one with unconstrained γ_2 as the NewAdd-U model. As with the simple multiplicative-type model structure, one appealing property of the constrained model is simplicity, which may incur less estimation error. The constrained model can correctly capture the relationship between VaR and ES for some DGPs, such as the GARCH model in Eq 6. The unconstrained model provides more flexibility by allowing different dynamics between VaR and ES. However, the forecasting performance of the multiplicative-type model, which shares the same dynamics for VaR and ES, can outperform other more flexible models in terms of lower joint loss scores, as shown in Taylor (2019). Therefore, we choose the constrained model as one candidate model in the empirical study, and the results in Section 6 show that the constrained model can outperform the unconstrained model for some financial time series.

Furthermore, we extend the NewAdd-C and NewAdd-U models to take the same form as the AS CAViar model in Eq (2), so that the leverage effect can be captured. The expression for the

variation of the NewAdd model is given by

$$\text{NewAdd-AS:} \quad ES_t = Q_t - x_t,$$
$$x_t = \gamma_0 + \gamma_1 I(r_{t-1} > 0)|r_{t-1}| + \gamma_2 I(r_{t-1} \leq 0)|r_{t-1}| + \gamma_3 x_{t-1}, \tag{9}$$

where the parameters of x_t are non-negative. As with the NewAdd model, the parameter γ_3 can be constrained to be equal to the autoregressive coefficients of the CAViar models of (1) and (2). We denote the model with constrained γ_3 as the NewAdd-AS-C model, and the unconstrained version as the NewAdd-AS-U model.

It is worth mentioning that Gerlach and Wang (2020) proposed an additive model, called 'ES-X-CAViaR-X' model, with the use of realized measures in place of the absolute lagged returns in Eq (5). Even though the proposed model in Eq (5) has a similar structure to the ES-X-CAViaR-X model, they stated that there is no direct or exact link between an additive-type model and a parametric GARCH-type model. However, such a link can be identified for the new additive model in Eq (5) due to the constraints on the autoregressive parameter. The proposed model of Eq (5) is developed from the process of constructing a simulation design, but other merits of the new additive-type models have been identified. Besides the feasibility of linking the NewAdd model to the GARCH model as shown above, the proposed NewAdd model and its extensions are relatively insensitive to the choices of Q_1 and ES_1, which is demonstrated in the empirical analysis in Section 6.

The reduced sensitivity of the NewAdd model may be due to the independence between the difference x_t and conditional quantile Q_t. The additive difference x_t in the original additive-type model is driven by the difference between past conditional quantiles and returns, $Q_{t-1} - r_{t-1}$, and the dynamic of x_t is only updated when the lagged return exceeds the lagged quantile. However, x_t in the newly proposed model is directly driven by lagged returns, which mitigates its sensitivity to the initial value of conditional quantiles. Furthermore, as demonstrated in the empirical results in Section 6.1, among the proposed specifications of the additive model, the versions that share the same autoregressive coefficient with Q_t generally show lower sensitivity to initial conditions.

The forecasting performance of the ES-X-CAViaR-X model of Gerlach and Wang (2020), which incorporates the realized measures, is assessed together with other joint VaR and ES models. Under the ES-X-CAViaR-X model, Q_t and x_t follow symmetric processes. We extend the ES-X-CAViaR-

X model by developing asymmetries in both the Q_t and x_t processes. The extended models are defined as follows:

$$\begin{aligned}
\text{AS-ES-X-CAViaR-X-AS:} \quad Q_t &= \beta_0 + \beta_1 I(r_{t-1} > 0) X_{t-1} + \beta_2 I(r_{t-1} \leq 0) X_{t-1} + \beta_3 Q_{t-1}, \\
ES_t &= Q_t - x_t, \\
x_t &= \gamma_0 + \gamma_1 X_{t-1} + \gamma_2 I(r_{t-1} \leq 0) X_{t-1} + \gamma_3 x_{t-1}. \qquad (10) \\
\text{ES-X-CAViaR-X-AS:} \quad Q_t &= \beta_0 + \beta_1 X_{t-1} + \beta_2 Q_{t-1}, \\
ES_t &= Q_t - x_t, \\
x_t &= \gamma_0 + \gamma_1 X_{t-1} + \gamma_2 I(r_{t-1} \leq 0) X_{t-1} + \gamma_3 x_{t-1}, \qquad (11)
\end{aligned}$$

where X_{t-1} is a realized measure on day $t-1$. The improvements in the forecasting performance by allowing for asymmetries are observed in the empirical results shown in Section 6.1. The proposed asymmetric model is less sensitive compared with the additive-type model of Taylor (2019), and more consistent estimates can be obtained by treating the initial conditions as unknown parameters.

3 Quantile Regression with Asymmetric Laplace Likelihood

3.1 Quantile Regression Model

At a given quantile level $\alpha \in (0, 1)$, the quantile regression model of r_t is given by

$$r_t = Q_t + \epsilon_t = \boldsymbol{x}_t^\intercal \boldsymbol{\beta}_\alpha + \epsilon_t, \quad t = 1, \ldots, T, \qquad (12)$$

where Q_t is the α-th quantile of r_t, ϵ_t is the error term with zero mean and constant variance, $\boldsymbol{x}_t^\intercal$ is the transpose of a vector of covariates $\boldsymbol{x}_t$, and $\boldsymbol{\beta}_\alpha$ is the quantile parameter vector. Specifically, the covariates here are Q_{t-1} and lagged returns r_{t-1}. The α-th quantile, Q_t, can be based on the two CAViaR-type models, the SAV model and the AS model, in Eqs (1) and (2), respectively. According to Koenker and Machado (1999), the quantile regression estimator is equivalent to the maximum likelihood estimator of the AL density with the expression as shown below.

3.2 Working Likelihood

The parameters of the quantile regression model can be estimated using a working likelihood based on the following AL density:

$$f(r_t) = \frac{\alpha(1-\alpha)}{\sigma} \exp(-(r_t - Q_t)(\alpha - I(r_t \leq Q_t))/\sigma), \qquad (13)$$

where σ is a scale parameter, and Q_t is the quantile with a chosen probability level α. Taylor (2019) introduced a conditional AL scale σ_t and linked the time-varying scale σ_t to conditional ES estimation. The conditional ES can be expressed in terms of σ_t as $ES_t = -\sigma_t/\alpha$, with the assumption of a zero expected return of an asset. After incorporating the conditional ES into Eq (13), the resulting AL density is:

$$f(r_t) = \frac{\alpha - 1}{ES_t} \exp\left(\frac{(r_t - Q_t)(\alpha - I(r_t \leq Q_t))}{\alpha ES_t}\right). \tag{14}$$

Based on the AL density of Eq (14), the associated AL working likelihood for the observed data $\boldsymbol{r} = \{r_t, t = 1, \ldots, T\}$ can be written as:

$$f(\boldsymbol{r}|\boldsymbol{\theta}) = \prod_{t=1}^{T} f(r_t), \tag{15}$$

where $\boldsymbol{\theta}$ is a set of parameters of the VaR-ES joint models believed to have generated the conditional quantiles $\mathbf{Q} = \{Q_t, t = 1, \ldots, T\}$ and the corresponding expected shortfall $\mathbf{ES} = \{ES_t, t = 1, \ldots, T\}$ of $\boldsymbol{r}$.

The observations $\boldsymbol{r}$ are not assumed to follow an AL distribution, whereas the AL working likelihood allows us to conduct statistical inference within the quantile regression. Then the VaR and ES joint models can be estimated with maximum likelihood based on the AL density in Eq (14). Besides the likelihood maximization approach adopted by Taylor (2019), the parameters of the proposed models can also be estimated through Bayesian approaches. Yu and Moyeed (2001) introduced a Bayesian quantile regression approach based on the AL likelihood. The posterior of the parameters can be generated via the Bayesian approach, as explained in Section 4, with the use of the AL working likelihood of Eq (15), where Q_t can be modeled by Eq (1) or (2) and ES_t can be modeled by Eq (3), (4), (5) or (9). Gerlach and Wang (2020) have employed an adaptive Markov chain Monte Carlo (MCMC) method for estimation. In this work, we utilize sequential Monte Carlo (SMC, Chopin, 2002, Del Moral et al., 2006) methods for estimation and prediction. SMC is chosen here as it is more efficient for generating a full predictive density for VaR and ES forecasts by reusing the intermediate samples created by the data-annealing SMC approach.

4 Bayesian Inference

Given the working AL likelihood of Eq (15), the Bayesian quantile regression approach (e.g., Yu and Moyeed, 2001; Yu and Stander, 2007) can be employed, with the use of SMC (e.g., Chopin,

2002) to sample from the joint posterior of the parameters of the proposed models. As a by-product, the marginal posterior distribution of each parameter can be obtained. The posterior distributions of parameters can easily produce interval estimates of interest, and the uncertainty estimates of VaR and ES forecasts can be obtained based on the parameter posterior samples.

4.1 Bayesian Estimation

4.1.1 Bayesian Statistics

Based on the working AL likelihood function of Eq (15) for an observed sample $\boldsymbol{r}$, the posterior of parameters $\boldsymbol{\theta}$ is given by

$$\pi(\boldsymbol{\theta}|\boldsymbol{r}) = \frac{f(\boldsymbol{r}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{Z}, \tag{16}$$

$$\propto f(\boldsymbol{r}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}), \tag{17}$$

where $\pi(\boldsymbol{\theta})$ is the prior and $Z = f(\boldsymbol{r}) = \int_{\boldsymbol{\theta}} f(\boldsymbol{r}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}$ is the normalizing constant which is not required by most algorithms for sampling the posterior. Since it is infeasible to directly sample from the posterior $\pi(\boldsymbol{\theta}|\boldsymbol{r})$ of the proposed models, the SMC sampling method is adopted.

4.1.2 Sequential Monte Carlo

In SMC, N weighted samples (called particles), $\{W_t^i, \boldsymbol{\theta}_t^i\}_{i=1}^N$, which represent a sequence of distributions $\pi_t(\boldsymbol{\theta}|\boldsymbol{r})$ for $t = 0, \ldots, T$, need to be constructed. The sequence starts from a simple distribution that is easy to sample from and ultimately reaches the target posterior distribution. The sequence of distributions can be constructed using either the likelihood annealing method or the data-annealing method. The sequence formed by the likelihood annealing approach, with the prior as the initial distribution, takes the form

$$\pi_t(\boldsymbol{\theta}|\boldsymbol{r}) \propto f(\boldsymbol{r}|\boldsymbol{\theta})^{\gamma_t}\pi(\boldsymbol{\theta}),$$

where γ_t is a power parameter, and $0 = \gamma_1 \leq \cdots \leq \gamma_t \leq \cdots \leq \gamma_T = 1$. The final distribution in the sequence is the posterior $\pi_T(\boldsymbol{\theta}|\boldsymbol{r}) \equiv \pi(\boldsymbol{\theta}|\boldsymbol{r})$. Under the data-annealing strategy, the sequence of distributions can be constructed as

$$\pi_t(\boldsymbol{\theta}|r_{1:t}) \propto f(r_{1:t}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}), \tag{18}$$

where $r_{1:t}$ represents the observations available up to current time t. Li et al. (2021) argued that the intermediate distributions $\pi_t(\boldsymbol{\theta}|r_{1:t})$ constructed by the data-annealing method can be efficiently

reused for making predictions for GARCH-type models. As demonstrated in Section 4.4, such efficiency brought from the data-annealing approach can also be useful in predicting VaR and ES.

Three main steps are iteratively applied in SMC to traverse the particles, which are re-weighting, resampling, and moving (or mutation). The re-weighting step re-weights the sample for target $t-1$, $\left\{W_{t-1}^i, \boldsymbol{\theta}_{t-1}^i\right\}_{i=1}^N$, to generate a new set of particles for target t. The new unnormalized weights are given by

$$w_t^i = W_{t-1}^i \frac{\pi_t(\boldsymbol{\theta}_{t-1}^i | r_{1:t})}{\pi_{t-1}(\boldsymbol{\theta}_{t-1}^i | r_{1:t-1})},$$

for $i = 1, \ldots, N$. The normalized weights can be calculated as $W_t^i = \frac{w_t^i}{\sum_{k=1}^N w_t^k}$, and the new particles remain unchanged as $\boldsymbol{\theta}_t^i = \boldsymbol{\theta}_{t-1}^i$. The re-weighting step usually reduces the effective sample size (ESS), which measures the quality of the particle set, so the particles need to be resampled to raise the ESS back to a desired level. The resampling process resamples the particle values proportional to their weights, where particles with relatively high weights are duplicated, and those with relatively small weights are dropped. Since the resampled particle set may contain duplicates of those particles with relatively high weights before resampling, a moving step is required to diversify the resampled particles. A moving step usually employs MCMC kernels of invariant distribution π_t to perturb the particles. Here, a Metropolis-Hastings (Hastings, 1970) kernel is adopted, whose parameters can be adapted online in SMC (see, for example, Salomone et al., 2018; Bon et al., 2021).

4.2 Settings of Unknown Parameters Q_1 and ES_1

As stated earlier, the joint models of Taylor (2019) are found to be sensitive to the initial values assumed for Q_1 and ES_1. In the literature, Q_1 and ES_1 are initialized to the empirical α-quantile of the early part of a financial time series and the corresponding mean of returns that exceed the α-quantile, respectively. However, we show that different empirical initializations can lead to different posterior distributions of the unknown parameters, and hence different predictive distributions of VaR and ES forecasts may be generated.

We treat these initial values as unknown parameters to address this issue, and then the posterior distributions of these parameters are generated through the Bayesian approach described above. To undertake the Bayesian quantile regression, prior distributions need to be assigned to these

unknown parameters. The priors of Q_1 and ES_1 are chosen to be uniformly distributed as:

$$Q_1 \sim \mathcal{U}(\mathrm{L_Q}, 0), \text{ and } ES_1 \sim \mathcal{U}(\mathrm{L_E}, 0),$$

where the upper bounds are zeros, and the choices of the lower bounds $\mathrm{L_Q}$ and $\mathrm{L_E}$ rely on the probability levels of tail risks. Since the difference between Q_1 and ES_1 is generally not known in advance, we suggest that to set the lower bounds of Q_1 and ES_1 at the same level for the additive-type models. Moreover, we require Q_1 to be larger than ES_1. This condition can be achieved by abandoning the draws from the prior of ES_1, whose values are larger than the draws from the prior of Q_1.

In order to test if the joint models are sensitive to the choice of prior distribution, different types of distributions other than the uniform distribution are adopted for priors. The exponential, gamma and log-normal distributions are also considered. In a similar way to the uniform case, we use negative values of realizations that are generated from the three distributions as prior samples since Q_1 and ES_1 are always negative. To simulate values from the three distributions, we set their mean parameter equal to absolute values of the empirical quantile and ES of the sample to be analyzed. We suggest not setting too small standard deviation parameters for the priors as the information about the dispersion is limited. For the sake of consistency, other prior distributions mostly cover the range of the corresponding uniform prior and allocate lower densities for the values out of the bounds of the uniform prior. It is worth noting that the multiplicative-type models only require Q_1 due to the models' specifications of the ES component. Therefore, the settings of the difference between Q_1 and ES_1 can be ignored for the multiplicative-type models. The empirical analysis illustrates the issue and show how this issue can be largely addressed.

4.3 Posterior Adjustment

Generally, the working AL likelihood is not the true likelihood for generating data (Yang et al., 2016). The assumption of the independent and identically distributed (i.i.d.) error term ϵ_t is unlikely to hold with financial time series data. The posterior from the misspecified working likelihood may yield wider or narrower credible intervals than the ones from the true likelihood. Sriram et al. (2013) showed that the misspecified AL likelihood can still lead to consistent posterior results for the model parameters. However, this does not imply that the uncertainty interval estimates constructed from the posterior are automatically valid (Yang et al., 2016). Therefore,

the Bayesian quantile regression based on an AL likelihood cannot be considered a truly Bayesian procedure (Benoit and Van den Poel, 2017), and we need to make corrections while constructing the posterior distributions of parameters and the corresponding predictions. The issue of uncertainty quantification has been considered in Liu et al. (2020), who applied a Bayesian quantile regression framework to analyze mass spectrometry datasets. As far as we know, the posterior coverage issue, which we aim to explore here, has not been considered for the quantile regression models of VaR and ES in the literature.

To construct asymptotically valid credible intervals of the unknown parameters, and hence the prediction intervals of the VaR and ES forecasts, we adopt an adjustment for the posterior samples proposed by Shaby (2014), called the open-faced sandwich (OFS) adjustment. Due to model misspecification, the asymptotic covariance matrix of the posterior distribution is different from that of the asymptotic distribution of the Bayesian point estimator (Chernozhukov and Hong, 2003). The ultimate goal of the OFS adjustment is to match the limiting distribution of posterior samples with the asymptotic distribution of the Bayesian point estimates, which are typically the posterior mean or median, so that the uncertainty estimates of parameters can be calibrated accordingly.

Denote the N weighted samples $\{W_t^i, \boldsymbol{\theta}_t^i\}_{i=1}^N$ as the SMC posterior samples of the unknown parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma})$ based on observations $r_{1:t}$, where $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ represent parameters of the VaR component model and the ES component model, respectively. The OFS adjustment to the weighted samples requires an estimator $\hat{\boldsymbol{\Omega}}$ of the matrix $\boldsymbol{\Omega} = \mathbf{H}^{-1}\mathbf{P}^{1/2}\mathbf{H}^{1/2}$, where $\mathbf{H}^{-1}$ is the covariance matrix of the limiting posterior distribution, and $\mathbf{J}^{-1} = \mathbf{H}^{-1}\mathbf{P}\mathbf{H}^{-1}$ is the covariance matrix of the asymptotic distribution of the Bayes estimator, $\hat{\boldsymbol{\theta}}$. Both $\mathbf{H}$ and $\mathbf{P}$ need to be estimated to obtain the estimator $\hat{\boldsymbol{\Omega}}$. The matrix $\mathbf{H}^{-1}$ can be estimated by the sample covariance matrix of the SMC samples, denoted as $\hat{\mathbf{H}}^{-1}$. The matrix $\mathbf{P}$ is given by $\mathbf{P} = \mathbb{E}_{\boldsymbol{\theta}}\left[\nabla_{\boldsymbol{\theta}}\ell(r_{1:t}|\boldsymbol{\theta})\left(\nabla_{\boldsymbol{\theta}}\ell(r_{1:t}|\boldsymbol{\theta})\right)^{\intercal}\right]$, where $\nabla_{\boldsymbol{\theta}}\ell(r_{1:t}|\boldsymbol{\theta})$ is the gradient of the log likelihood evaluated at the true parameter $\boldsymbol{\theta}$. An estimate of the matrix $\mathbf{P}$ is given by

$$\hat{\mathbf{P}} = \sum_{j=1}^{t} \nabla\ell(r_j|\hat{\boldsymbol{\theta}})\left(\nabla\ell(r_j|\hat{\boldsymbol{\theta}})\right)^{\intercal}, \tag{19}$$

where $\hat{\boldsymbol{\theta}}$ is the Bayes estimator. The adjusted samples can be obtained by pre-multiplying the

unadjusted samples by $\hat{\boldsymbol{\Omega}}$ as follows

$$\tilde{\boldsymbol{\theta}}_t^i = \hat{\boldsymbol{\theta}} + \hat{\boldsymbol{\Omega}} \left(\boldsymbol{\theta}_t^i - \hat{\boldsymbol{\theta}} \right), \tag{20}$$

where $\tilde{\boldsymbol{\theta}}_t^i$ is the adjusted sample, and $\boldsymbol{\theta}_t^i$ is the unadjusted sample. The covariance matrix of the adjusted posterior samples $\left\{ W_t^i, \tilde{\boldsymbol{\theta}}_t^i \right\}_{i=1}^N$ is $\mathbf{J}^{-1}$ as desired.

It is relatively easy to estimate $\mathbf{H}^{-1}$ for the VaR-ES joint models, but some difficulties arise in computing the estimated matrix $\hat{\mathbf{P}}$ for the SAV-Add and AS-Add models when treating Q_1 and ES_1 as unknown parameters. For the two additive-type models, the partial derivatives of the log likelihood $\ell(r_{1:t}|\hat{\boldsymbol{\theta}})$ with respect to the parameters Q_1 and ES_1 need to be calculated recursively. The details of the recursive approach to find $\nabla \ell(r_{1:t}|\hat{\boldsymbol{\theta}})$ are provided in the Appendix A. The AL log likelihood of each data point r_t for the VaR-ES joint models evaluated at the point estimates $\hat{\boldsymbol{\theta}}$ is given by

$$\ell(r_t|\hat{\boldsymbol{\theta}}) = \begin{cases} \log(\alpha - 1) - \log(ES_t) + \frac{(\alpha-1)(r_t - Q_t)}{\alpha ES_t}, & \text{if } r_t \leq Q_t \\ \log(\alpha - 1) - \log(ES_t) + \frac{\alpha(r_t - Q_t)}{\alpha ES_t}, & \text{otherwise} \end{cases}. \tag{21}$$

To choose the correct expression of the above log likelihood, the value of Q_t needs to be compared with r_t. By substituting the Bayesian point estimate $\hat{\boldsymbol{\theta}}$ into the CAViaR-type models, the point estimate of Q_t can be obtained. Given the estimates of $\mathbf{P}$ and $\mathbf{H}$, the adjusted posterior samples $\left\{ W_t^i, \tilde{\boldsymbol{\theta}}_t^i \right\}_{i=1}^N$ can be obtained by following Eq (20), and thus Bayesian inference can be carried out based on the adjusted samples. The improvements of the interval estimates of the adjusted distributions is demonstrated in the simulation study.

4.4 Bayesian Prediction

Under the Bayesian approach, the posterior predictive distribution for VaR and ES forecasts can be provided, which account for the uncertainty associated with the unknown parameters. The one-step-ahead posterior predictions of the conditional quantile $\hat{Q}_{t+1}$ and expected shortfall $\hat{ES}_{t+1}$ are considered here, given their past values and observations up to time t.

As stated, it is relatively easy to produce the posterior predictive distribution by data-annealing SMC. Under the data-annealing approach, the intermediate samples from the posterior $\pi(\boldsymbol{\theta}|r_{1:t})$ based on different subsets of observations $r_{1:t}$ are ready to be reused to make predictions, which means there is no need to re-estimate the model again with different sample sizes. The posterior

predictive distribution of $\hat{Q}_{t+1}$ based on the adjusted parameters $\tilde{\boldsymbol{\theta}}$ takes the following form:

$$p\left(\hat{Q}_{t+1}|r_{1:t}, Q_{1:t}\right) = \int_{\tilde{\boldsymbol{\theta}}} p\left(\hat{Q}_{t+1}|r_{1:t}, Q_{1:t}, \tilde{\boldsymbol{\theta}}\right) \pi_t(\tilde{\boldsymbol{\theta}}|r_{1:t}, Q_{1:t})\, \mathrm{d}\tilde{\boldsymbol{\theta}}.$$

Based on the adjusted weighted samples $\left\{W_t^i, \tilde{\boldsymbol{\theta}}_t^i\right\}_{i=1}^N$ from the posterior $\pi_t(\tilde{\boldsymbol{\theta}}|r_{1:t})$, the posterior predictive distribution of $\hat{Q}_{t+1}$ and $\hat{ES}_{t+1}$ can be approximated by $\left\{W_t^i, \hat{Q}_{t+1}^i\right\}_{i=1}^N$ and $\left\{W_t^i, \hat{ES}_{t+1}^i\right\}_{i=1}^N$, respectively. Given the values of $r_{1:t}$, $Q_{1:t}$, and $\tilde{\boldsymbol{\theta}}_t^i$ for $i = 1, \ldots, N$, $\hat{Q}_{t+1}^i$ and $\hat{ES}_{t+1}^i$ are specified according to the VaR-ES forecasting models described in Section 2. From the predictive distribution, one-step-ahead forecasts and the associated uncertainty interval estimates can be easily obtained.

We use a family of joint loss functions developed by Fissler and Ziegel (2016) to jointly evaluate VaR and ES forecasts from the VaR-ES models. As explained in Fissler and Ziegel (2016), the joint loss functions taking the forms in Appendix B are strictly consistent, which means that the loss function can be minimized uniquely by true VaR and ES. Three specifications of the loss scores adopted in Taylor (2019) are considered here and provided in Appendix B. The weighted samples from the posterior predictive distributions of VaR and ES forecasts are substituted into the three loss functions, and we use the corresponding posterior median of the loss scores to undertake the evaluation.

5 Simulation Study

In the simulation study, we evaluate the performance of the 95% credible intervals of the unknown parameters and the 95% prediction intervals of the one-step ahead VaR and ES forecasts. To the best of our knowledge, the uncertainties derived from a Bayesian approach for the jointly estimated VaR and ES forecasts have not been explored in the literature, so the unadjusted interval estimates from the Bayesian quantile regression are treated as the benchmark available before making any adjustments.

As stated in Section 2.2, the true parameters of the SAV-Mult model and the SAV-NewAdd-C model are known if simulating from the GARCH model of Eq (6). We simulated 40 return series from the GARCH model with the following parameters that are used in Gerlach and Wang (2020):

$$a_0 = 0.02, \quad a_1 = 0.10, \quad a_2 = 0.85.$$

Each simulated return series has a sample size of $T = 2000$. We set $\alpha = 1\%$ in this study and

the corresponding true values of the parameters of the SAV-Mult model and the SAV-NewAdd-C model are given in Table 1 and 2, respectively.

Table 1: Coverage rates of interval estimates of the SAV-Mult model

		95% Coverage Rate	
	True	Unadjusted	Adjusted
β_0	-0.047	34.5%	**45.8%**
β_1	-0.233	38.6%	**70.3%**
β_2	0.850	39.5%	**59.1%**
γ_0	-1.926	80.9%	**96.1%**
$\text{VaR}_{1001:2000}$	-	38.5%	**52.5%**
$\text{ES}_{1001:2000}$	-	56.1%	**90.5%**
1% Hit Rate	-	1.06%	1.02%

Both the multiplicative- and the additive-type models are fitted to the 40 simulated datasets, and we use $N = 5000$ particles in SMC to approximate posterior distributions. We calculated one-step ahead VaR and ES forecasts for the final 1000 returns r_t of each dataset from $t = 1001$ to $t = 2000$ based on the returns $r_{1:t-1}$ up to time $t - 1$. The coverage rate, which is the frequency of the true values that fall within the interval estimates, of the 95% credible interval of the parameter estimates is calculated based on 1000 in-sample periods for each dataset. Based on the estimates of posterior predictive distributions, the coverage rate of the 95% prediction interval of the one-step ahead forecasts is calculated for the final 1000 data points of each dataset. The average values of the coverage rates over the 40 datasets are provided in Table 1 and 2.

Table 2: Coverage rates of interval estimates of the SAV-NewAdd-C model

		95% Coverage Rate	
	True	Unadjusted	Adjusted
β_0	-0.047	29.0%	**41.3%**
β_1	-0.233	43.3%	**73.8%**
$\beta_2 = \gamma_2$	0.850	34.2%	**56.4%**
γ_0	0.007	**89.8%**	70.2%
γ_1	0.034	**97.0%**	73.9%
$\text{VaR}_{1001:2000}$	-	37.4%	**55.9%**
$\text{ES}_{1001:2000}$	-	61.0%	**62.8%**
1% Hit Rate	-	1.06%	1.04%

The 1% hit rate in Table 1 and 2 refers to the percentage of observations falling below the

1% quantile estimates, where the posterior median values of Q_t are used as the point quantile estimates. Both the multiplicative-type and additive-type models provide reasonable hit rates. As the adjustments mainly change the curvature of the posterior, the median point estimates do not change much. Although the SAV-Mult and SAV-NewAdd-C models have the same VaR structure, their results for VaR may differ due to the restriction on β_2 in the SAV-NewAdd-C model. We can see that from Table 1 and 2, both the adjusted and unadjusted uncertainties of VaR related parameters and one-step ahead forecasts for the two models are slightly different. The original uncertainties of the quantile and ES forecasts and most of the parameter estimates of the two models are underestimated, which leads to low coverage rates before the adjustment. It is obvious that most of the coverage rates after adjustment are more accurate (i.e. closer to 95%) than the unadjusted ones, especially for the ES forecasts from the SAV-Mult model. However, the adjusted coverage rates are not perfect due to the imperfect estimates of the matrix $\mathbf{\Omega}$ required by the OFS adjustment. The approach of estimating matrix $\mathbf{P}$, which is a component of $\mathbf{\Omega}$, is designed for the case that multiple independent replicates can be produced from a stochastic process. However, in our situation, only a single return series with T observations can be generated per draw. Our adjusted results produce significantly more accurate uncertainty quantification than the unadjusted ones, not only for the parameters but importantly for the VaR and ES forecasts. However, our results also show that more research is required to improve the accuracy of the uncertainty quantification further.

Furthermore, we suggest that the true parameters of these models can also be approximated via the method of ordinary least squares for an overdetermined system (Anton and Rorres, 2013). Suppose that a linear system $\mathbf{Ax} = \mathbf{b}$ has m equations and n unknown variables, which implies that $\mathbf{A}$ is an $m \times n$ matrix, $\mathbf{x}$ is a $n \times 1$ vector and $\mathbf{b}$ is a $m \times 1$ vector. If the number of equations m is greater than the number of unknown variables n, such a system is referred to as an overdetermined system. An approximate solution to the overdetermined system is given by

$$\mathbf{x} = (\mathbf{A}^\intercal \mathbf{A})^{-1} \mathbf{A}^\intercal \mathbf{b}. \tag{22}$$

For the above simulation settings, since we know the true values of Q_t and ES_t, a system of equations can be constructed for Q_t and ES_t with the corresponding unknown parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. Then the unknown parameters can be approximated by following Eq (22). We have used the

overdetermined system method to calibrate the true values of the unknown parameters obtained by mapping from the GARCH model to the joint VaR-ES model of this simulation settings, and there is approximately no difference in the results from the two different methods.

6 Empirical Results

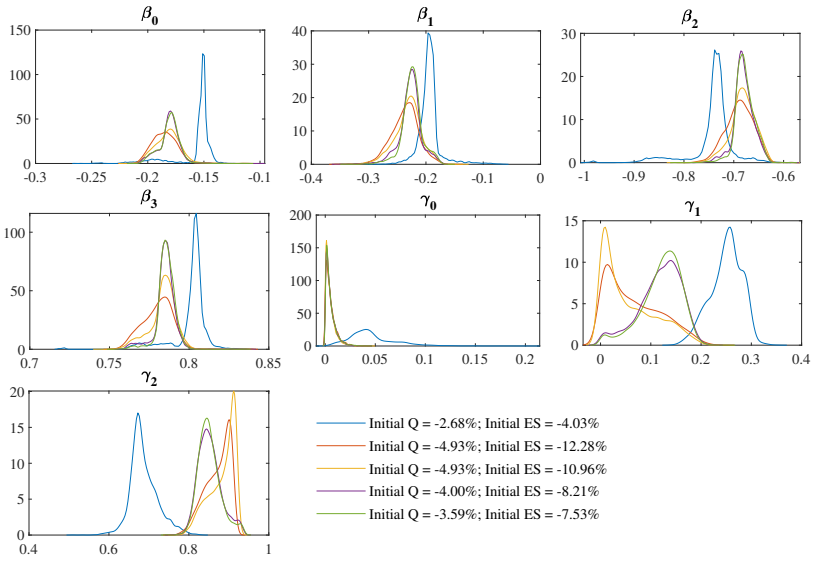
Daily S&P 500 returns are used for the following empirical analysis. The sample to be examined spans 7 April 1986 - 2 September 2021, containing 8927 observations in total. The dataset covers the crash of 1987, the Global Financial Crisis, and the 2020 stock market crash, which allows estimation and prediction results under different market conditions. To investigate if the sensitivity issue is affected by the length of the estimation period, an expanding window approach is employed for estimating each model. The starting in-sample size is 2000, and 250 out-of-sample forecasts are generated based on the in-sample estimation. Then the second estimation results are based on 2250 in-sample observations, and another 250 forecasts are generated. By expanding the in-sample period in such a way, each model was estimated 28 times for the dataset, where the final estimation in-sample includes 8750 observations. In total, 6927 out-of-sample forecasts are evaluated.

In addition to the S&P 500 data, another five financial time series of different asset classes and markets are analyzed. Data for IBM, NIKKEI 225, NASDAQ and FTSE 100 indices consisting of 5000 daily returns ending on 31 December 2021 are used. We also include the returns of USO ETF (US Oil Fund) from 10 April 2006 to 17 December 2021. These five datasets allow us to analyze the ES-X-CAViaR-X model and its proposed extensions, given that the up-to-date high-frequency data used for generating realized measures are available. The five-minute prices were downloaded from Thomson Reuters Datascope. We use realized volatility (RV) that is computed by aggregating five-minute high-frequency returns for a trading day. As with the S&P 500, the expanding window approach is adopted to estimate each candidate model, and the starting in-sample size is 1000.

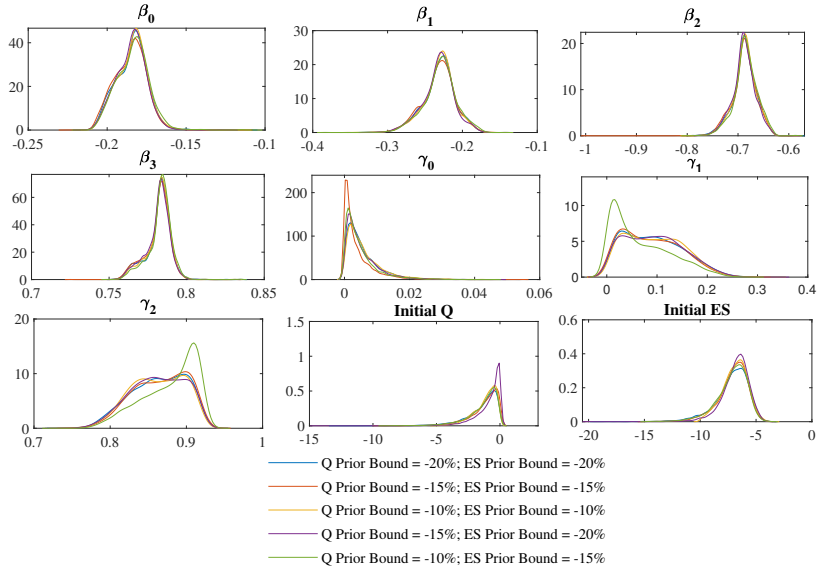
6.1 Sensitivity of the joint VaR and ES models

6.1.1 S&P 500 Sample

The sensitivity to initial conditions can be visually illustrated by the posterior distributions of the parameters of the joint models. In this section, we present the estimation results based on the initial 2250 observations. The sensitivity problem is severe enough to be easily visible as



(a) AS-Add model with predetermined values of Q_1 and ES_1



(b) AS-Add model with Q_1 and ES_1 treated as unknown parameters

Figure 1: Marginal posterior distributions of the parameters of the AS-Add model by using SMC.

shown in Figure 1a. The estimation results based on the in-sample size of 2250 also imply that the sensitivity issue may not be mitigated with a long estimation period. Figure 1a presents the posterior distributions of the parameters of the AS-Add model at $\alpha = 1\%$ probability level with different choices for the initial values of Q_1 and ES_1. The values of Q_1 and ES_1 are initialized at the empirical α-quantile and the corresponding empirical ES of the first 300, 400, 500, 800, and 1000 observations, leading to the initial values as shown in the following figures. Strongly overlapping marginal posterior distributions in the figures indicate more consistent results. We can see that with

different fixed initial values, the posterior distributions can behave very differently. This issue can be largely addressed by treating the initial Q and ES as unknown parameters, as shown in Figure 1b. The priors of Q_1 and ES_1 are set to be uniformly distributed over a range of $(L, 0)$ for different levels of tail risks, where L is the lower bound of the uniform distribution. Where applicable, the choices of different lower bounds are given in the figures of marginal posterior distributions. In addition to the uniform prior, the posteriors based on exponential, gamma, and log-normal priors are presented in Figure 2. With different prior distributions for the initial parameters, consistent marginal posteriors of parameters other than Q_1 and ES_1 can still be produced.

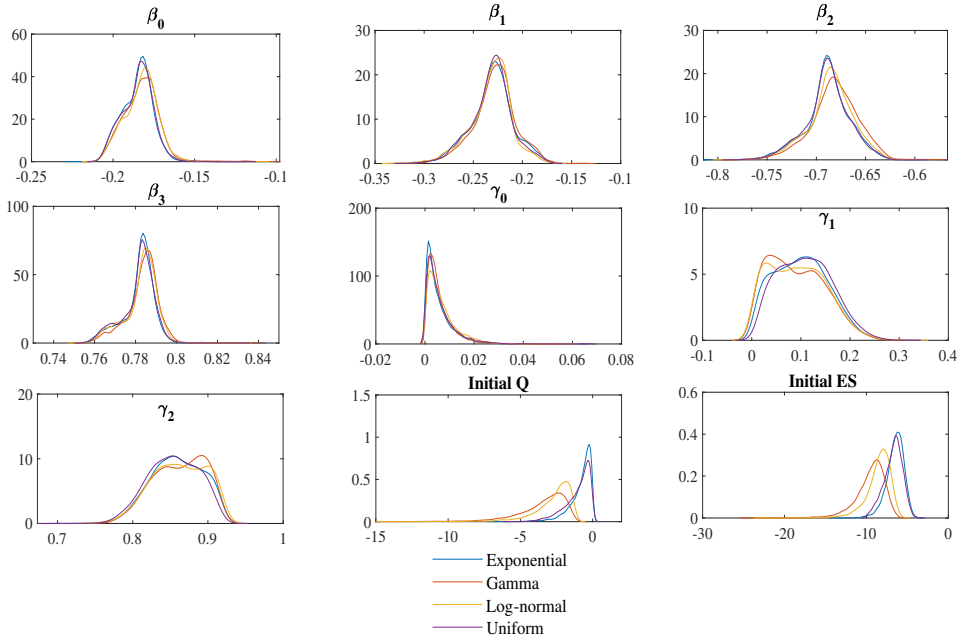


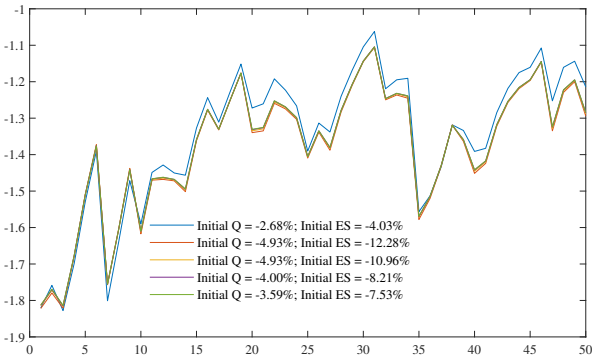
Figure 2: Marginal posterior distributions of the parameters of the AS-Add model with Q_1 and ES_1 treated as unknown parameters by using different prior distributions.

The marginal posterior distributions for the AS-Mult model and other additive-type models are reported in Appendix C. From Figure 18a, we can see that the AS-Mult model also demonstrates some degree of sensitivity to Q_1, but less so than the AS-Add model. By treating Q_1 as an unknown parameter, as shown in Figure 18b, the resulting marginal posteriors based on different priors of Q_1 are more consistent than the ones in Figure 18a. The NewAdd model of Eq (5) is much less sensitive to Q_1 and ES_1 compared to the AS-Add model. The posterior distributions of the parameters of the AS-NewAdd-C model, a constrained version of Eq (5), are shown in Figure 19a. With the same choices of initial values of the AS-Add model, these posteriors are less sensitive

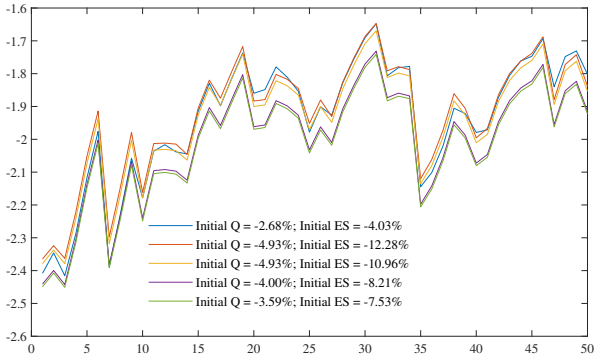
to different values of Q_1 and ES_1. Nevertheless, the AS-NewAdd-C model still shows sensitivity to some degree, which also can be addressed by treating the initial values as unknown parameters as demonstrated in Figure 19b. Furthermore, the sensitivity of other variations of the proposed additive-type models of Eq (5) and (9) are also less sensitive to the initial conditions compared to the original additive-type model, leading to more consistent posteriors where initial conditions are treated as unknown parameters. The marginal posterior distributions for these variations of the additive-type model are presented in Appendix C. In addition, the benefit of the reduced sensitivity will be demonstrated later by checking the standard deviations of each model's forecast evaluation results based on different initial conditions. Regarding the range of the uniform priors for Q_1 and ES_1, we suggest setting their lower bounds at the same level as it is difficult to know the difference between the initial quantile and ES in advance.

Moreover, for the original additive-type model, different posterior distributions caused by the various choices of Q_1 and ES_1 can influence the prediction of tail risks. Figures 3a and 3b present the first 50 prediction results of quantile and ES based on the posterior distributions in Figures 1a. It can be seen that different posteriors of parameters can lead to different forecasting results, especially for the predicted ES. The sensitivity can be largely reduced in the case that the initial values of Q and ES are treated as unknown parameters. Even with different choices of priors, the posterior distributions and the corresponding forecasts of quantile and ES are similar, as shown in Figures 3c and 3d. In addition, for the proposed additive-type model of the constrained version of Eq (5), the forecasts resulting from using different initial values of the quantile and ES are relatively consistent as shown in Figures 3e and 3f.

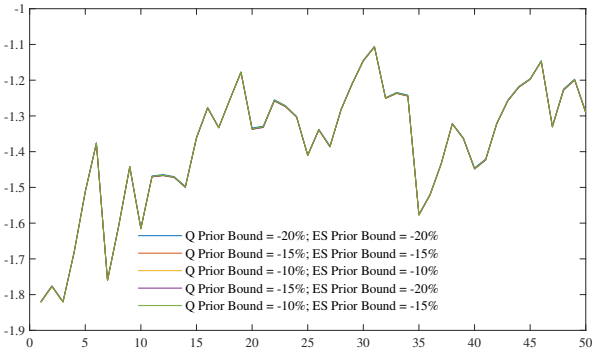
To illustrate the economic size of the sensitivity effect more clearly, we investigate the forecasting results based on the parameter estimation with the initial 5250 observations. Figure 4a shows 100 one-step-ahead forecasts of ES from 21 Jun 2007 to 09 Nov 2017. We can see that different initial values lead to different forecasting results for this period during the GFC. With an arbitrary choice of initial conditions, a financial institution's economic capital may not be efficiently allocated as the amount required to hold for risk coverage is not accurately forecast. For instance, on 1 Oct 2007, the highest estimated ES is 10.43% larger than the lowest value. In this case, different values of the initial inputs can result in wasted regulatory capital or insufficient capital for covering risk. Especially for the highly volatile period, optimization of capital allocation plays a vital role. The



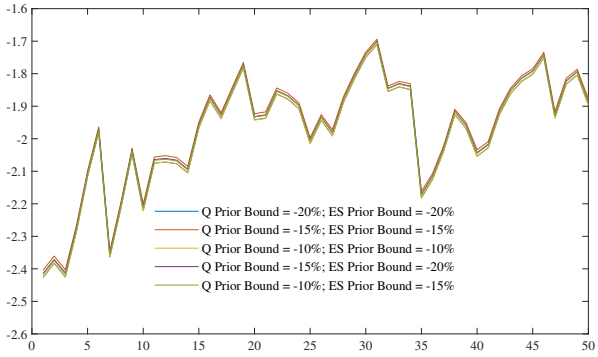
(a) $Q_{2251:2300}$: AS-Add model with choices of Q_1 & ES_1



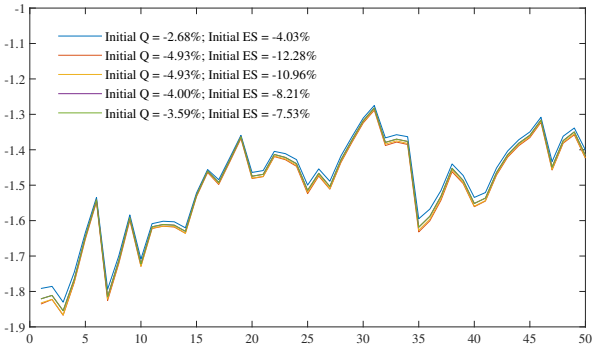
(b) $ES_{2251:2300}$: AS-Add model with choices of Q_1 & ES_1



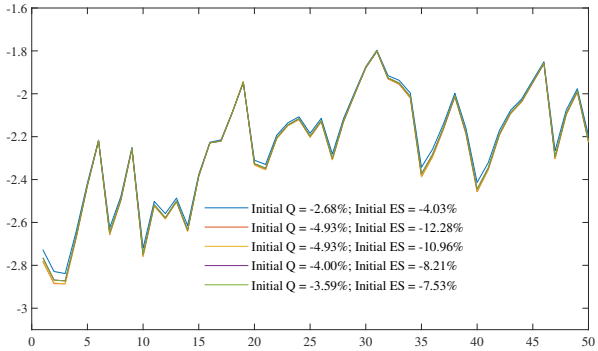
(c) $Q_{2251:2300}$: AS-Add model with parameters Q_1 & ES_1



(d) $ES_{2251:2300}$: AS-Add model with parameters Q_1 & ES_1



(e) $Q_{2251:2300}$: AS-NewAdd-C model with choices of Q_1 & ES_1

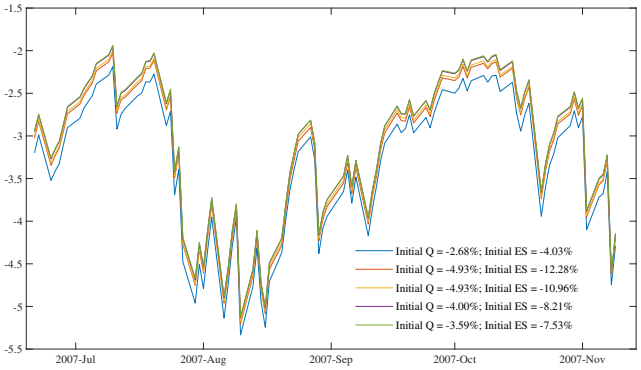


(f) $ES_{2251:2300}$: AS-NewAdd-C model with choices of Q_1 & ES_1

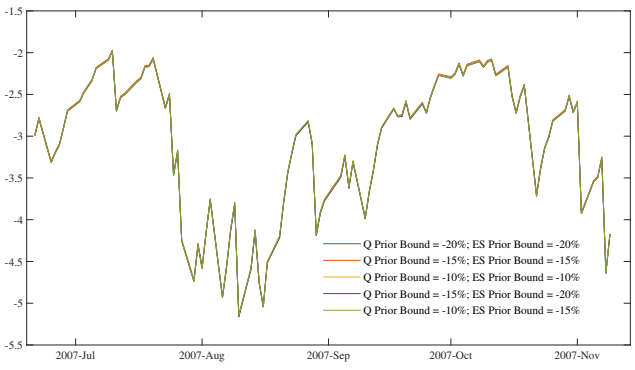
Figure 3: Posterior median of the first 50 out-of-sample quantile and ES forecasts

proposed approach largely addresses the sensitivity in forecasting results, as shown in Figure 4b, minimizing the chance of inefficiently allocating capital against risky asset portions.

Figures 5 and 6 illustrate the standard deviations of the loss scores from the three loss functions described in Section 4.4.2 for the original additive-type models, the proposed additive-type models, and the multiplicative-type models. By treating Q_1 and ES_1 as unknown parameters, the loss scores for all the candidate models are relatively consistent, and have lower standard deviations for different selections of priors most of the time, stabilizing the performance of the forecasting



(a) ES forecasts: AS-Add model with choices of Q_1 & ES_1



(b) ES forecasts: AS-Add model with parameters Q_1 & ES_1

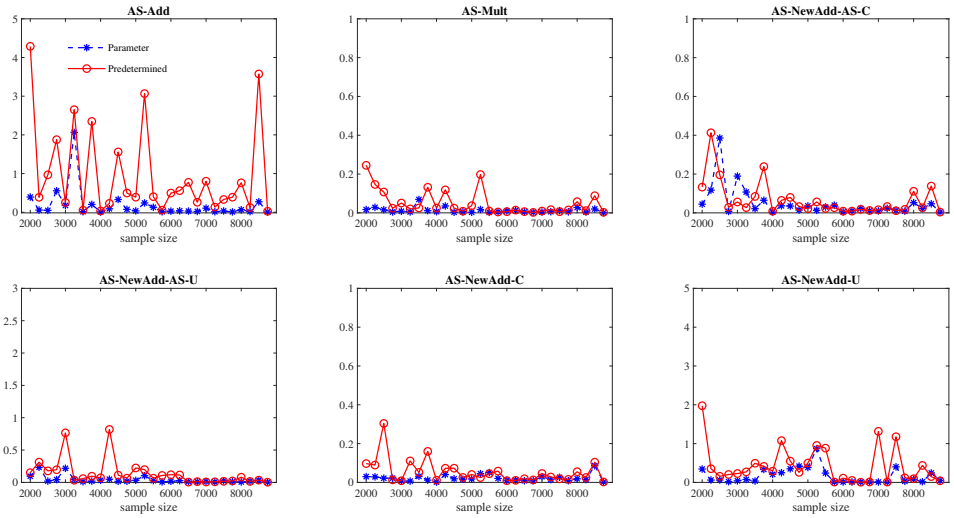
Figure 4: Posterior median of 100 out-of-sample ES forecasts

scheme. In contrast, by not treating the initial conditions as parameters, the loss scores are more variable and have larger standard deviations with different initial values, especially for the AS-Add and SAV-Add models. It is worth emphasizing that the loss scores are computed based on the posterior predictive distributions of VaR and ES. Therefore, loss scores with larger standard deviations indicate higher sensitivity in the forecasts of VaR and ES. Moreover, the sensitivity issue is prone to occur even with a large estimation sample. For instance, as shown in Figure 5, the AS-Add model can be highly sensitive to the choices of Q_1 and ES_1 even when the in-sample size is 8000 and 8500. Overall, the proposed variations of the additive-type models, whose standard deviations of the loss scores are relatively low, are less sensitive to the initial conditions compared with the original additive-type model.

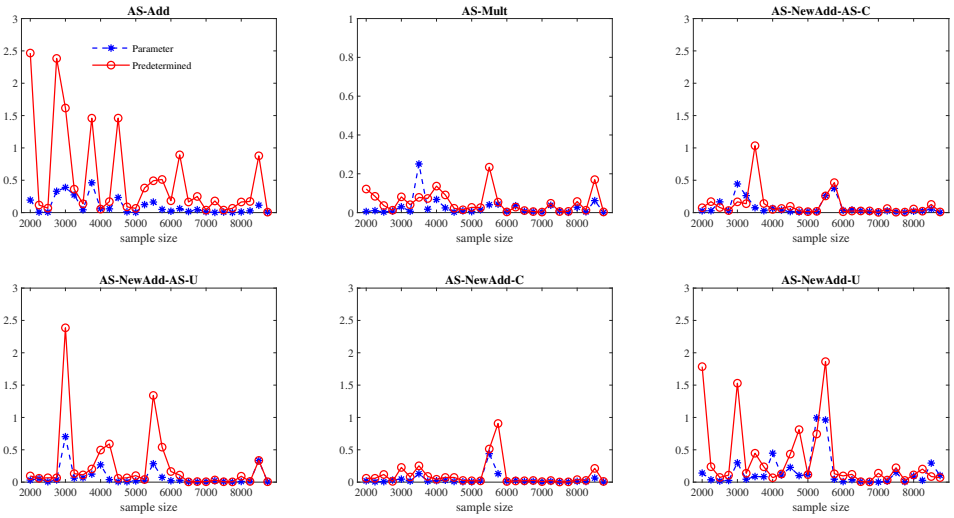
Table 3: 1% VaR and 1% ES Joint Loss Scores

Model	Out-of-sample Loss Score (Median)		
	L1	L2	L3
AS-Add	4740	1670	15421
AS-NewAdd-C	4715	1652	15317
AS-NewAdd-U	4722	1657	15346
AS-NewAdd-AS-C	4718	**1643**	**15312**
AS-NewAdd-AS-U	**4703**	1664	15320
AS-Mult	4720	1652	15326
	L1	L2	L3
SAV-Add	4751	1712	15567
SAV-NewAdd-C	4741	1696	15504
SAV-NewAdd-U	4745	1782	15571
SAV-NewAdd-AS-C	4730	1689*	15463*
SAV-NewAdd-AS-U	4726*	1783	15511
SAV-Mult	4741	1700	15508

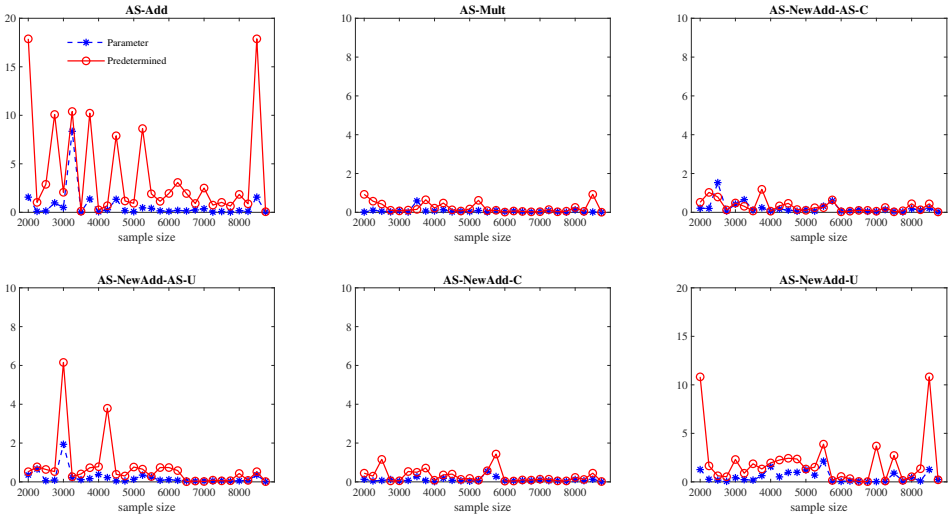
Notes:
L1: Loss score from the joint loss function of Eq (33); L2: Loss score from the joint loss function of Eq (34); L3: Loss score from the joint loss function of Eq (35).
Bold indicates lowest values of loss score for a model with an Asymmetric Slope-type VaR component.
* indicates lowest values of loss score for a model with a Symmetric Absolute Value-type VaR component.



(a) Sample standard deviations of the loss score from the joint loss function of Eq (33)

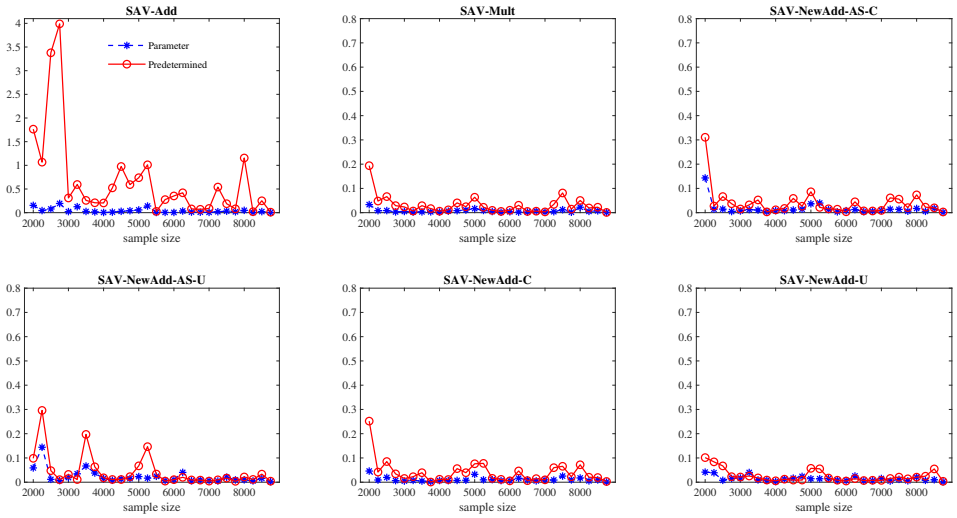


(b) Sample standard deviations of the loss score from the joint loss function of Eq (34)

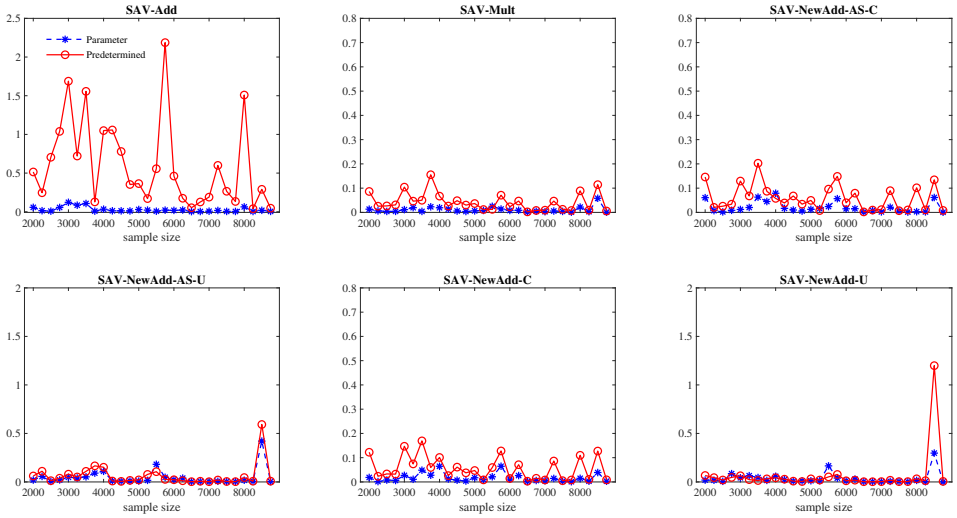


(c) Sample standard deviations of the loss score from the joint loss function of Eq (35)

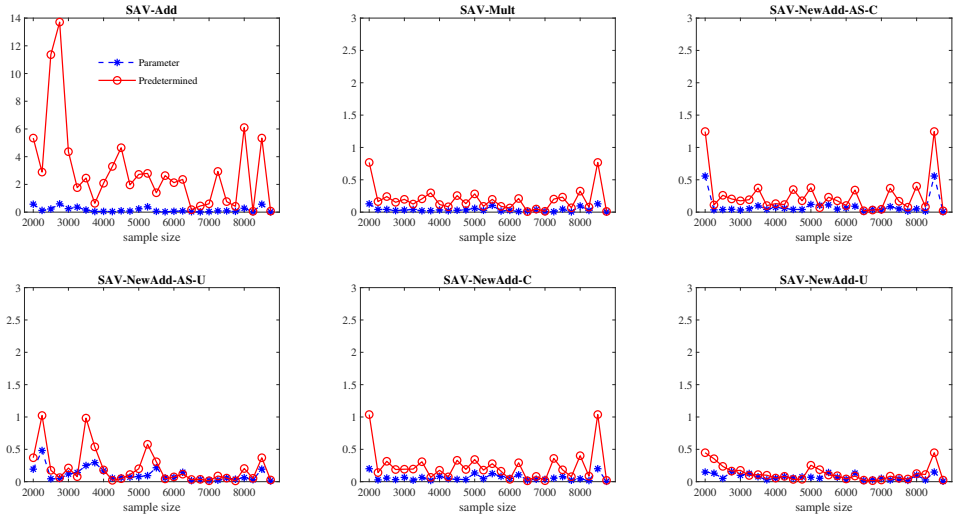
Figure 5: Sample standard deviations of joint loss scores for 1% VaR and 1% ES produced from models with an asymmetric VaR component based on different in-sample sizes



(a) Sample standard deviations of the loss score from the joint loss function of Eq (33)



(b) Sample standard deviations of the loss score from the joint loss function of Eq (34)



(c) Sample standard deviations of the loss score from the joint loss function of Eq (35)

Figure 6: Sample standard deviations of joint loss scores for 1% VaR and 1% ES produced from models with a symmetric VaR component based on different in-sample sizes

Table 3 provides the loss scores for jointly evaluating the out-of-sample forecasting performance of the candidate models estimated by treating initial conditions as unknown parameters. The presented loss scores are the averages over the ones based on different selections of priors of Q_1 and ES_1. As indicated by the joint loss scores L2 and L3, the proposed AS-NewAdd-AS-C model provides the lowest losses, performing best for the S&P 500 dataset. Regarding the joint loss score L1, the AS-NewAdd-AS-U model outperforms other candidates. Moreover, the joint VaR-ES models whose VaR component is modelled by the AS CAViaR model, which can capture the leverage effect, is always advantageous over the one modelled by the SAV CAViaR model as shown in Table 3. It is worth mentioning that regardless of whether the VaR component is asymmetric or symmetric, the models that perform best are from our newly proposed additive-type models.

Table 4: 1% VaR Coverage Test p-values

Model	UC	CC	DQ
AS-Add	0.005	0.015	0.045
AS-NewAdd-C	**0.085**	**0.155**	**0.125**
AS-NewAdd-U	0.039	0.037	0.009
AS-NewAdd-AS-C	**0.250**	**0.309**	**0.281**
AS-NewAdd-AS-U	**0.167**	**0.244**	0.000
AS-Mult	**0.085**	**0.061**	0.010
SAV-Add	**0.167**	**0.090**	0.000
SAV-NewAdd-C	**0.358**	**0.122**	0.000
SAV-NewAdd-U	**0.250**	**0.108**	0.000
SAV-NewAdd-AS-C	**0.571**	**0.133**	0.000
SAV-NewAdd-AS-U	**0.167**	0.023	0.000
SAV-Mult	**0.423**	**0.128**	0.000

Notes:
All tests are conducted at the 5% significance level.
Bold indicates the model is not rejected by the test at a 5% level.

In addition to the joint loss functions, the performance of VaR forecasts is also examined via backtesting procedures. Three standard tests are considered here: the unconditional coverage (UC) test of Kupiec (1995), the conditional coverage test of Christoffersen (1998) and the dynamic conditional quantile (DQ) test of Engle and Manganelli (2004). From the p-values of the backtests reported in Table 4, the results of the UC and CC tests are reasonable for most of the models, where the AS-Add model and the unconstrained AS-NewAdd model are rejected all the time. For the DQ test, which is more powerful in rejecting a misspecified model than the other two tests, most of the models are rejected except the constrained AS-NewAdd model and its asymmetric version. The performance of ES forecasts is also assessed through backtests. The ES residual test of McNeil and Frey (2000) is adopted, which tests if the average difference between the returns

exceeding VaR and ES forecasts is approximately zero. The regression-based ES backtest of Bayer and Dimitriadis (2022) is also conducted. The regression backtest regresses returns on the ES forecasts and an intercept, and then it tests if the intercept is equal to 0 and the slope coefficient is equal to 1. From Table 5, we can see that only the original additive-type models are rejected by the ES residual test. It is worth mentioning that the adjusted results are not used here since the sensitivity issue is irrelevant to the uncertainty adjustment. The main purpose of this section is to demonstrate the sensitivity issue, and illustrate how the issue is largely addressed by the proposed approach. Moreover, as discussed in Section 5, since the point estimates would not change much with the uncertainty adjustment, the loss scores and backtesting results are reported based on unadjusted posterior samples.

Table 5: 1% ES Backtest Rejection Count

Model	ES Residual	ES Regression
AS-Add	1	0
AS-NewAdd-C	0	0
AS-NewAdd-U	0	0
AS-NewAdd-AS-C	0	0
AS-NewAdd-AS-U	0	0
AS-Mult	0	0
SAV-Add	1	0
SAV-NewAdd-C	0	0
SAV-NewAdd-U	0	0
SAV-NewAdd-AS-C	0	0
SAV-NewAdd-AS-U	0	0
SAV-Mult	0	0

Notes:
All tests are conducted at the 5% significance level.
1 indicates the model is rejected by the test at a 5% level.

6.1.2 Out-of-sample VaR and ES Forecasts for Other Market Indices and Financial Assets

The comprehensiveness of the empirical analysis is extended by analyzing different financial datasets. The ES-X-CAViaR-X model of Gerlach and Wang (2020) and its extended specifications in Eq (10) and (11) are included, where RV is used as the realized measure. Table 6 provides the AL loss score ratios of candidate models to that of the AS-Add model to assess VaR and ES jointly. The AL loss score of Eq (35) is just the negative of the AL log-likelihood which is a consistent scoring function. We can see that the AS-ES-X-CAViaR-X-AS model, an extension of the ES-X-CAViaR-X model of Gerlach and Wang (2020) results in the lowest joint loss scores for most

Table 6: 1% VaR and 1% ES AL Loss Score Ratio

Model	Oil	IBM	NIKKEI	FTSE	NASDAQ
AS-Add	1.000	1.000	1.000	1.000	1.000
AS-Mult	**0.998**	0.999	0.996	1.000	1.003
AS-NewAdd-C	**0.998**	0.999	0.996	0.999	0.995
AS-NewAdd-U	0.999	1.000	**0.993**	0.999	1.003
AS-NewAdd-AS-C	**0.998**	0.999	0.996	0.999	0.992
AS-NewAdd-AS-U	0.999	1.000	0.996	0.999	1.000
AS-ES-X-CAViaR-X-AS	1.020	**0.972**	0.999	**0.993**	**0.945**
SAV-Add	1.010	1.000	1.007	1.013	1.008
SAV-Mult	1.008	1.001	1.006	1.013	1.021
SAV-NewAdd-C	1.008	1.001	1.006	1.013	1.002
SAV-NewAdd-U	1.009	1.002	1.007	1.013	1.017
SAV-NewAdd-AS-C	1.007*	1.001	1.001	1.010	1.001
SAV-NewAdd-AS-U	1.007*	1.001	1.004	1.011	1.015
ES-X-CAViaR-X	1.016	0.973*	1.002	0.996	0.951*
ES-X-CAViaR-X-AS	1.016	0.973*	0.998*	0.994*	0.952

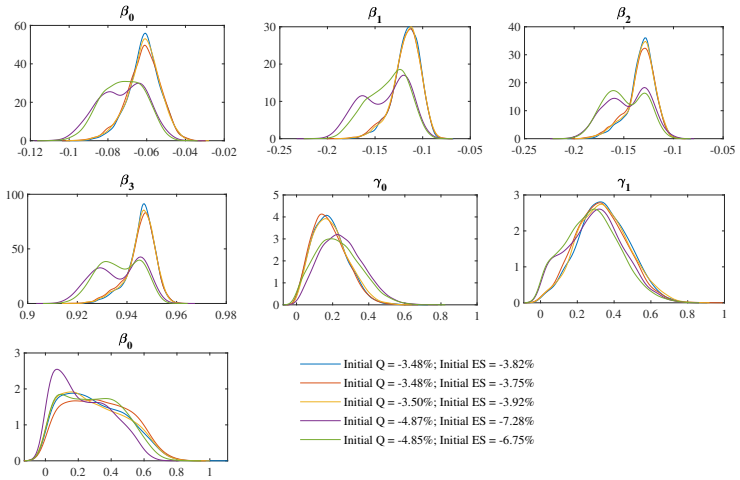
Notes:
Ratios of different methods' AL loss scores to that of the AS-Add model.
Bold indicates lowest values of loss score for a model with an Asymmetric Slope-type VaR component.
* indicates lowest values of loss score for a model with a Symmetric Absolute Value-type VaR component.

of the financial time series. In addition, Table 7 presents the counts of rejections of each model for the VaR and ES forecasts backtests over five datasets. The AS-ES-X-CAViaR-X-AS model is less likely to be rejected under all the tests relative to the other models. Overall, the backtest results are consistent with that of the AL joint loss scores, where the AS-ES-X-CAViaR-X-AS model outperforms other candidate models.

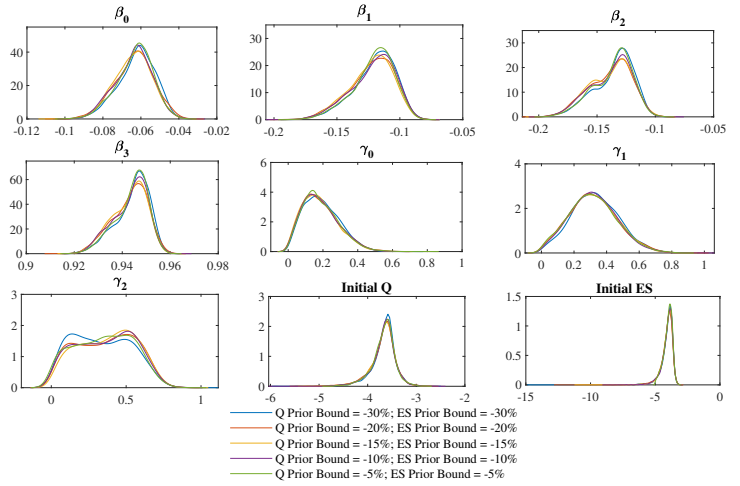
Table 7: 1% VaR and 1% ES Backtest Rejection Count

Model	DQ	ES Residual	ES Regression
AS-Add	1	3	**1**
AS-Mult	2	3	2
AS-NewAdd-C	1	3	2
AS-NewAdd-U	1	3	2
AS-NewAdd-AS-C	1	2	**1**
AS-NewAdd-AS-U	1	2	2
AS-ES-X-CAViaR-X-AS	**0**	**1**	**1**
SAV-Add	2	3	1*
SAV-Mult	2	3	1*
SAV-NewAdd-C	2	3	1*
SAV-NewAdd-U	2	2	1*
SAV-NewAdd-AS-C	2	1*	1*
SAV-NewAdd-AS-U	2	1*	1*
ES-X-CAViaR-X	1*	2	1*
ES-X-CAViaR-X-AS	1*	2	1*

Notes:
Counts of 1% VaR rejections with the DQ test and 1% ES rejections with ES residual and regression tests on five time series at a 5% significance level.
Bold indicates the fewest rejections for a model with an Asymmetric Slope-type VaR component.
* indicates the fewest rejections for a model with a Symmetric Absolute Value-type VaR component.



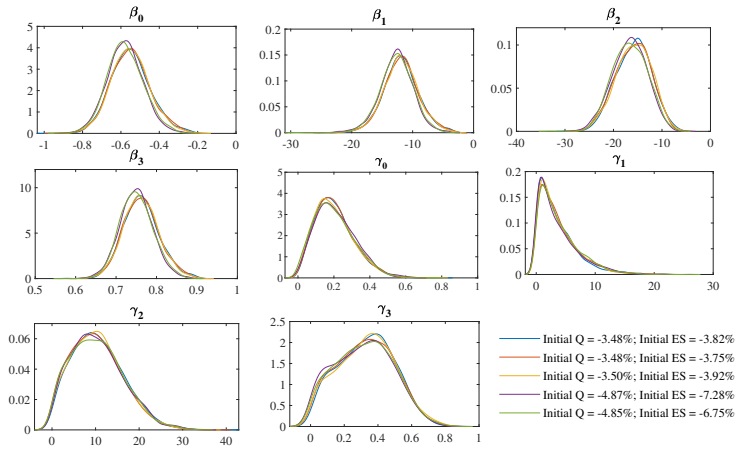
(a) AS-Add model with choices of Q_1 & ES_1



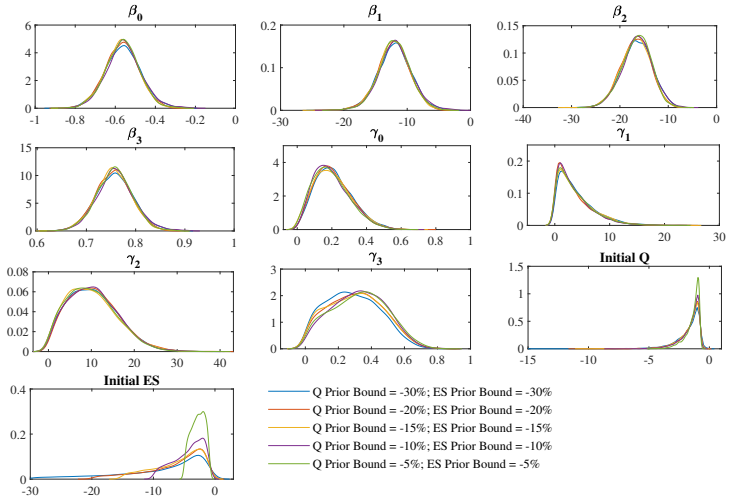
(b) AS-Add model with parameters Q_1 & ES_1

Figure 7: Marginal posterior distributions of the parameters of the AS-Add model based on 2000 USO daily returns.

The sensitivity issue is not unique to the S&P 500 but also appears under the other financial time series. Figure 7a shows inconsistent posterior distributions of the parameters of the AS-Add model based on a series of commodity returns. The issue can be addressed by treating initial conditions as unknown parameters, as demonstrated in Figure 7b. From Figure 8a, we can see that the AS-ES-X-CAViaR-X-AS model shows much less sensitivity to initial conditions than the AS-Add model. As shown in Figure 8b, with different priors, the marginal posterior distributions produced by treating the initial conditions as unknown parameters can become more consistent than the ones in Figure 8a.



(a) AS-ES-X-CAViaR-X-AS model with choices of Q_1 & ES_1



(b) AS-ES-X-CAViaR-X-AS model with parameters Q_1 & ES_1

Figure 8: Marginal posterior distributions of the parameters of the AS-ES-X-CAViaR-X-AS model based on 2000 USO daily returns.

6.2 Adjustment Results

6.2.1 Adjusted Posteriors

As demonstrated in the simulation study, the accuracy of the uncertainty interval estimates can be improved after adjusting the posterior samples. Before considering the interval estimates of VaR and ES forecasts, the posterior distributions after adjustments are shown in this section. The adjusted posterior distributions of the parameters of the models investigated in the simulation study are presented. In the case of the 1% quantile level in the simulation study, the 95% interval estimates for the quantile and ES forecasts and the parameters of the SAV-Mult model have low coverage rates of the true values, which indicates that the intervals are underestimated due to the misspecified working likelihood. In Section 5, we have demonstrated that the adjustment

method improved the accuracy of interval estimates for the simulated data. The effectiveness of the adjustment for the real data is discussed below.

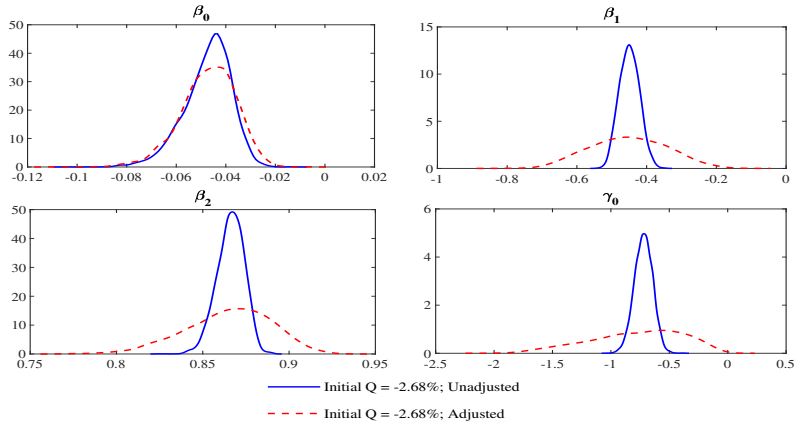


Figure 9: Adjusted marginal posterior distributions of the parameters of the SAV-Mult model based on 2250 S&P 500 daily returns. ($\alpha = 1\%$)

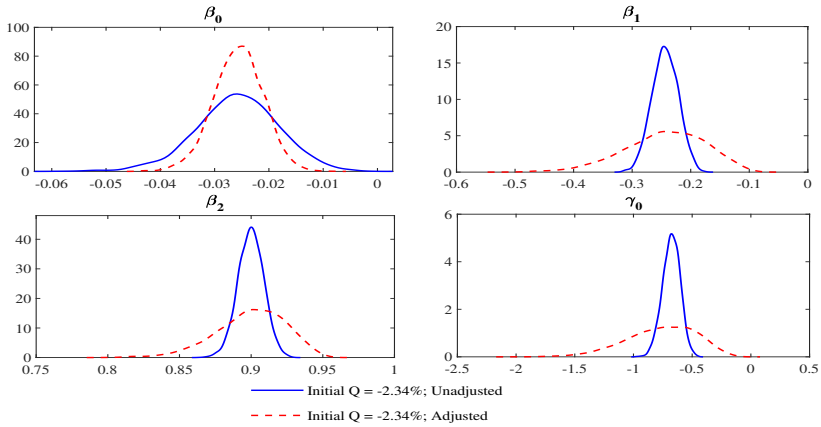


Figure 10: Adjusted marginal posterior distributions of the parameters of the SAV-Mult model based on 2250 S&P 500 daily returns. ($\alpha = 2.5\%$)

Given the S&P 500 data, Figures 9, 10, and 11 illustrate the marginal posterior distributions of the parameters of the SAV-Mult model before and after the adjustments. In the case of the 1% quantile level, most of the adjusted distributions become more dispersed compared to the unadjusted ones as desired, especially for the ES component related parameter. Since the simulation study shows that unadjusted uncertainty intervals are underestimated, the inflated variances after adjustments imply more accurate estimates of the uncertainty. The adjusted results for the SAV-NewAdd-C are given in Appendix C, where the adjusted uncertainty intervals of most parameters become wider. One parameter related to the ES component has a narrower uncertainty

estimate after the adjustment, which leads to relatively small uncertainty improvements in ES forecasts compared with the one from the SAV-Mult model, as discussed below.

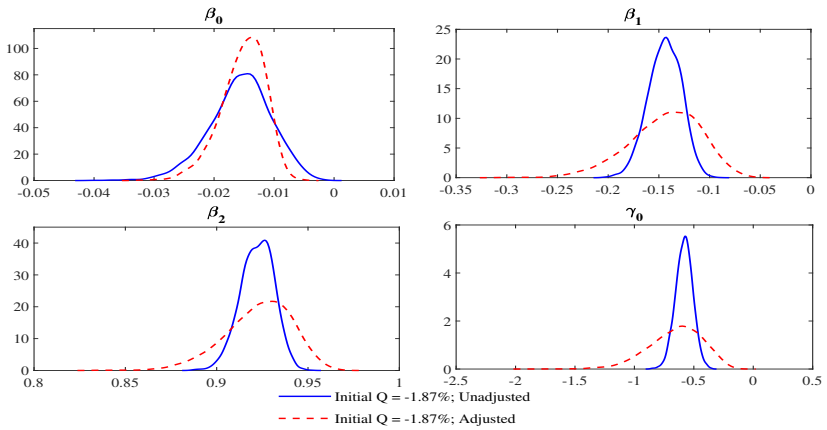
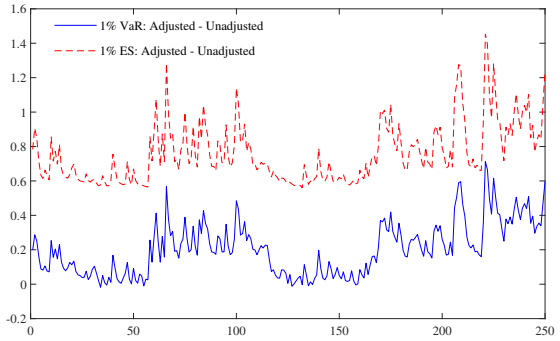


Figure 11: Adjusted marginal posterior distributions of the parameters of the SAV-Mult model based on 2250 S&P 500 daily returns. ($\alpha = 5\%$)

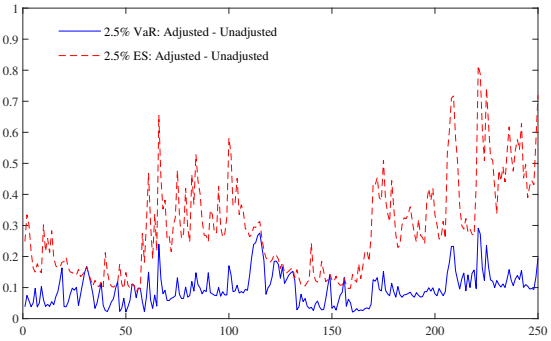
6.2.2 Adjusted Uncertainty

Based on the adjusted posterior distributions, we implement the Bayesian prediction approach to make forecasts for the real data, where the uncertainty around predicted tail risks can be obtained. The differences between adjusted and unadjusted widths of prediction interval (PI) estimates of the tail risk forecasts with 1%, 2.5% and 5% quantile levels for the SAV-Mult model are presented in Figures 12a, 12b and 12c. Positive differences indicate an increase in the uncertainty width after adjustment. In the case of the 1% quantile level of the simulation study, the unadjusted interval estimates from the SAV-Mult model are relatively narrow and overconfident on the parameters and tail risks estimates. Therefore, interval estimates with higher widths after the adjustment quantify the corresponding uncertainties more accurately. In the case of the 1% quantile level of the real data, wider prediction intervals for the tail risks are expected to be produced after the adjustment, as shown in Figure 12a. Meanwhile, as expected, the width increment of 1% ES is higher than the one of 1% VaR, which implies that the adjustment effect is larger on ES. This phenomenon is consistent with the simulation results, which demonstrate that, after the adjustment, more accurate uncertainty intervals can be gained for 1% ES forecasts than 1% VaR.

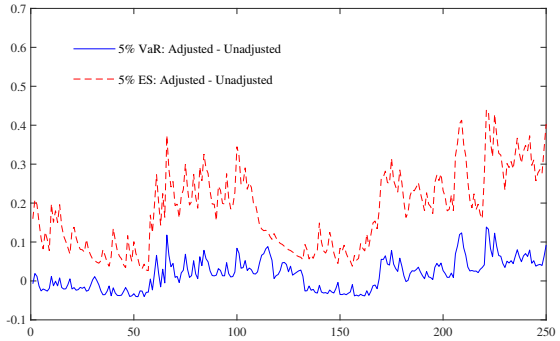
For the predicted tail risks with probability levels of 2.5% and 5%, their adjusted interval widths also become larger, which is consistent with their adjusted posterior distributions where



(a) $\alpha = 1\%$



(b) $\alpha = 2.5\%$



(c) $\alpha = 5\%$

Figure 12: Differences of 250 out-of-sample 95% prediction interval width, before and after the adjustment, of VaR and ES forecasts, for the SAV-Mult model with the probability levels of $\alpha = 1\%, 2.5\%, 5\%$ by using the Bayesian approach.

most of the distributions' variances increase. It is worth noting that for the case of the SAV-NewAdd-C model in Figure 17 given in Appendix C, the adjustment effect of the interval width of ES is relatively low from the SAV-Mult model since the adjusted marginal posterior of one ES parameter in Figure 16 has smaller variances.

6.3 Forecasts of Systemic Risk Measures with Joint VaR and ES Models

This section provides an empirical application of VaR and ES forecasts to estimate systemic risks. VaR and ES are not only critical risk measures for an individual institution, but they are also practically useful for systemic risk modelling. The systemic risk measures of CoVaR and CoES proposed by Adrian and Brunnermeier (2011) are considered. Following a similar approach taken earlier for jointly estimating VaR and ES, we forecast CoVaR and CoES jointly via quantile regression. To the best of our knowledge this approach has not been used in the literature. CoVaR is based on the concept of VaR and is defined as the VaR of a market index conditional on a particular institution being in distress. With a probability level of α, the CoVaR of a system m

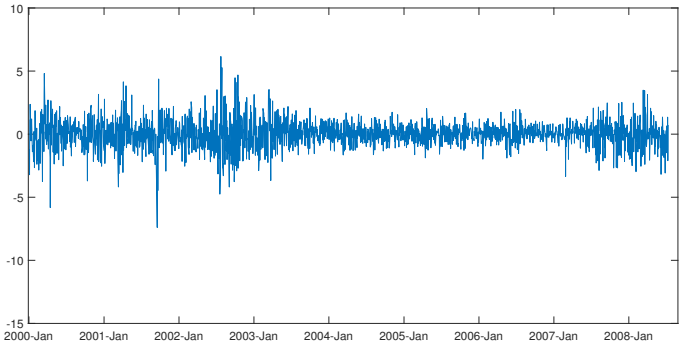
conditional on an institution i is given by

$$\Pr\left(r^m \leq \text{CoVaR}_\alpha^{m|i} | r^i = \text{VaR}_\alpha^i\right) = \alpha, \tag{23}$$

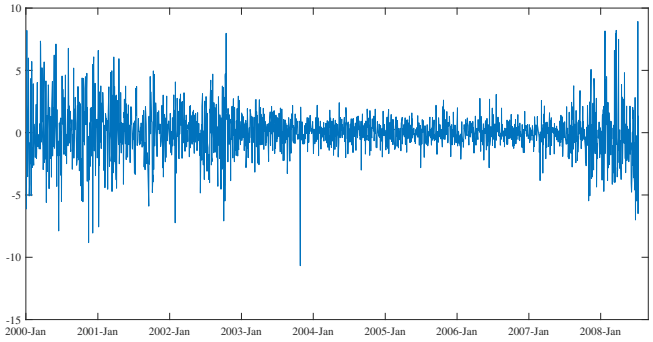
where r^m is return of a market index and r^i is financial return of an individual institution i. CoES is the ES of system m conditional on $r^i \leq \text{VaR}_\alpha^i$ of institution i:

$$\text{CoES}_\alpha^{m|i} = E\left[r^m | r^m \leq \text{CoVaR}_\alpha^{m|i}\right]. \tag{24}$$

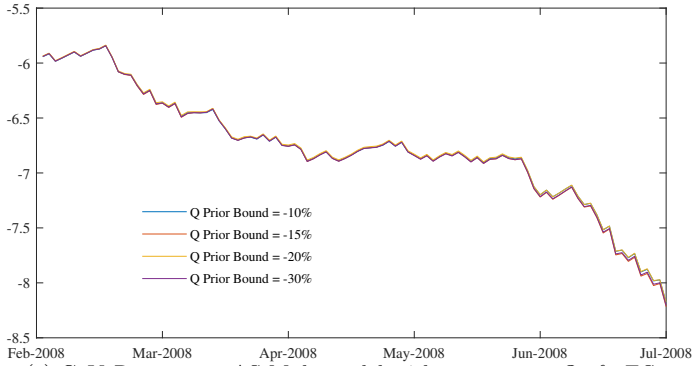
Given the dependence between CoVaR and VaR, we demonstrate that the sensitivity and uncertainty estimates issue discussed are translated to CoVaR and CoES forecasts.



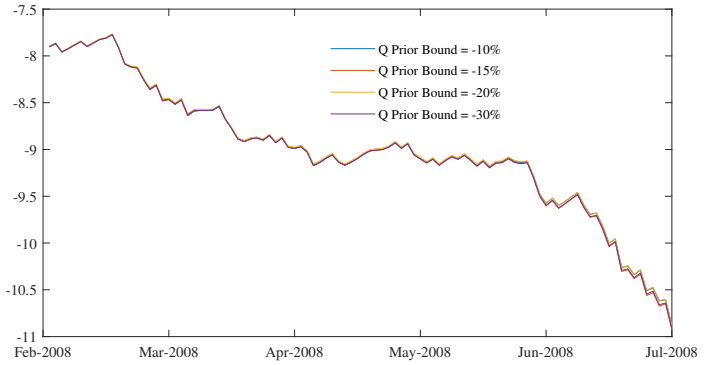
(a) Dow Jones (DJI) Return



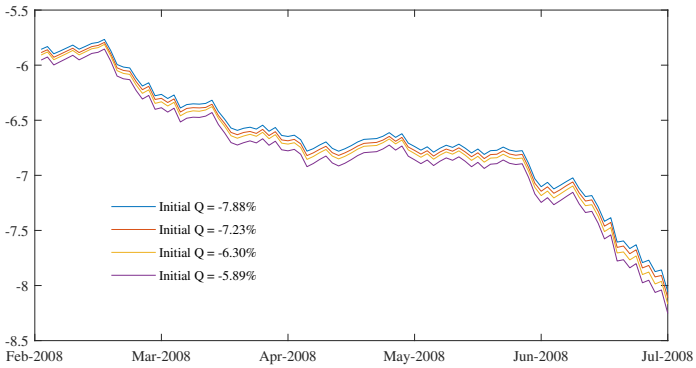
(b) Bank of America (BAC) Return



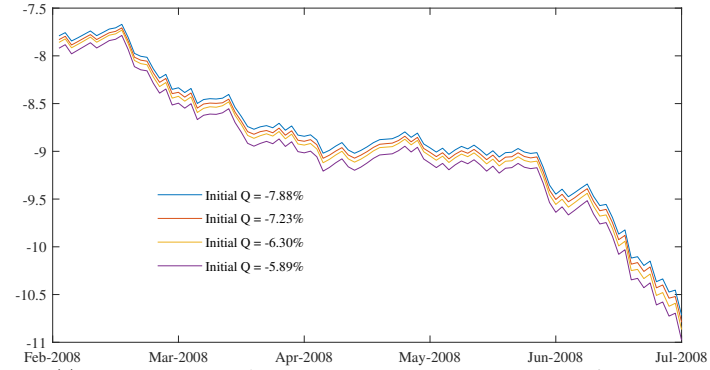
(c) $\text{CoVaR}_{1951:2050}$: AS-Mult model with parameters Q_1 & ES_1



(d) $\text{CoES}_{1951:2050}$: AS-Mult model with parameters Q_1 & ES_1



(e) $\text{CoVaR}_{1951:2050}$: AS-Mult model with choices of Q_1 & ES_1



(f) $\text{CoES}_{1951:2050}$: AS-Mult model with choices of Q_1 & ES_1

Figure 13: Sample data and posterior median of the last 100 out-of-sample CoVaR and CoES forecasts

To estimate CoVaR, we first consider a quantile regression of the system m on institution i for the quantile level α:

$$r_t^m = \beta_0^{m|i} + \beta_1^{m|i} r_t^i + \epsilon_t^{m|i}. \tag{25}$$

After obtaining the estimated parameters from the quantile regression model of Eq (25), according to Adrian and Brunnermeier (2011), the $\text{CoVaR}_t^{m|i}$ is given by:

$$\text{CoVaR}_t^{m|i} = \hat{\beta}_0^{m|i} + \hat{\beta}_1^{m|i} \text{VaR}_t^i. \tag{26}$$

The CoVaR models in (25) and (26) of Adrian and Brunnermeier (2011) contain a set of lagged state variables as covariates in order to generate time-varying risk estimates. Given the time-varying nature of the VaR forecasts from the joint VaR and ES models, such lagged variables are not included here. As we aim to explore how the robust approaches outlined earlier can be utilized to improve point and uncertainty estimates for CoVaR and CoES, the existence of lagged state variables in the models of Adrian and Brunnermeier (2011) do not affect the general results here. In this study, we extend the model of Eq 25 by incorporating an asymmetric property as follows:

$$r_t^m = \beta_0^{m|i} + \beta_1^{m|i} I(r_t > 0)|r_t^i| + \beta_2^{m|i} I(r_t \leq 0)|r_t^i| + \epsilon_t^{m|i}. \tag{27}$$

The CoES is in a similar manner to the multiplicative formulation for ES:

$$\text{ES}_t^m = 1 + \exp(\gamma_0) r_t^m, \tag{28}$$

where r_t^m is the quantile estimation in Eq (27). Then the $\text{CoES}_t^{m|i}$ is given by:

$$\text{CoES}_t^{m|i} = 1 + \exp(\hat{\gamma}_0) \text{CoVaR}_t^{m|i}. \tag{29}$$

The models of Eq (27) and (28) can be jointly estimated by the proposed Bayesian quantile regression approach described in Section 4. Based on the definition of CoVaR in Eq (26), we can see that VaR is an important input. Therefore, the reliability of VaR forecasts become critical.

In the application, we use the Dow Jones index (DJI) as a market proxy, and Bank of America (BAC) as an individual institution. The sample period is from 3 January 2000 to 8 June 2022. Figures 13a and 13b present the daily log returns of the two time series. The first 1800 observations are used for estimation, and 250 out-of-sample forecasts of VaR and ES are generated based on the in-sample estimation. Figures 13c and 13d provide the point estimate of CoVaR and CoES based on VaR and ES forecasts by treating initial conditions as parameters for the period of 13 Feb 2008

to 10 Jul 2008. With different choices of priors, consistent forecasting results are generated. In contrast, VaR and ES forecasts based on different predetermined values of initial conditions can lead to more volatile CoVaR and CoES forecasts, as shown in Figures 13e and 13f. The inconsistent outcomes indicate the lower reliability of the original additive-type models. Since the forecasts are used to determine a financial institution's regulatory capital, and thus can have real economic impacts, it is important for decision-makers and regulators to base their decisions on reliable risk estimates.

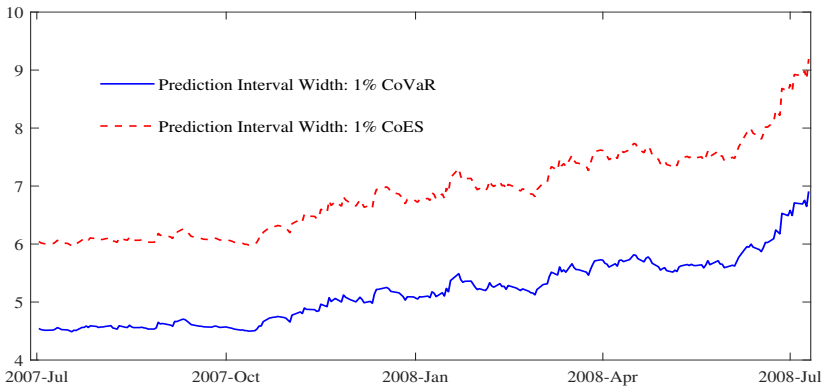


Figure 14: Adjusted 95% prediction interval width of 250 out-of-sample CoVaR and CoES forecasts, with the probability level of $\alpha = 1\%$ by using the Bayesian approach.

Uncertainties around the forecasts can be provided by the Bayesian quantile regression approach. However, the uncertainty interval estimates are inaccurate due to the adoption of a misspecified likelihood. The accuracy of uncertainty estimates for CoVaR and CoES can be improved by the proposed adjustment method discussed in detail in Section 4.3. Figure 14 shows the adjusted 95% prediction interval widths of the 250 out-of-sample forecasts of CoVaR and CoES. With the increasing volatility from Jul 2007 to Jul 2008, the uncertainties around the systemic risk measures become higher. Figure 15 presents the differences between adjusted and unadjusted widths of the 250 out-of-sample prediction interval estimates of CoVaR and CoES. We can see that their 95% prediction intervals are underestimated before making any adjustments. As the crisis develops, the differences between the adjusted and unadjusted interval widths increase, indicating a growing underestimation of the systemic risk measures' uncertainties.

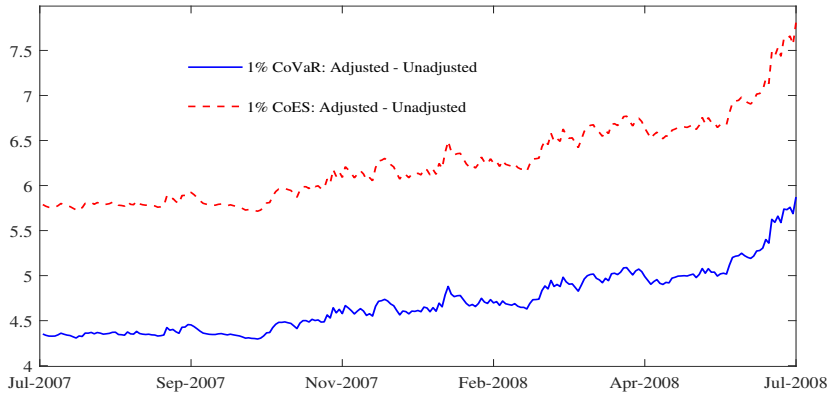


Figure 15: Differences of 250 out-of-sample 95% prediction interval width, before and after the adjustment, of CoVaR and CoES forecasts, with the probability level of $\alpha = 1\%$ by using the Bayesian approach.

7 Conclusion

This paper investigates problems with existing joint VaR and ES models. It demonstrates the sensitivity of the joint models of Taylor (2019), especially prominent for the additive-type VaR and ES models, and suggests approaches to address it. With different choices of starting values of the conditional quantile and ES, the resulting forecasts of tail risks can be very different. Employing a Bayesian quantile regression approach, the sensitivity issue can be largely addressed by treating the initial values of Q_1 and ES_1 as unknown parameters, where the corresponding prior distributions' ranges need to be wide enough. We argue that it is not appropriate to compare the performance of the sensitive joint models with others given the sensitivity issues outlined, whereas the proposed Bayesian approach allows such comparisons. In addition, new additive-type models are proposed, which mitigate the sensitivity issue to some degree, which can be further improved upon using our Bayesian approach. One of the proposed additive-type models can be linked to a parametric GARCH model to enable a simulation study which previously is restricted to the multiplicative-type models. Moreover, the paper extends the ES-X-CAViaR-X model of Gerlach and Wang (2020) by developing asymmetric properties. The models proposed by Gerlach and Wang (2020) incorporate different types of realized measures and demonstrate improved forecasting performance over only using returns. In this paper, RV is adopted for the proposed AS-ES-X-CAViaR-X-AS model. The performance of this proposed model can be further investigated by utilizing different realized measures.

We explore the risks associated with the forecasts of tail risk measures from the joint VaR and ES model by examining the corresponding posterior predictive distributions produced from

a Bayesian inference approach, issues that have not been investigated in the literature. We demonstrate that the interval estimates from the Bayesian quantile regression approach can be misleading due to the misspecified AL working likelihood for the joint VaR and ES models. We employ the OFS adjustment method to improve the estimation accuracy of risks of the tail risk measures. However, the adjustment established here has limited efficacy for improving the accuracy of interval estimates. Exploring more accurate adjustment methods is an interesting avenue for future work.

Finally, to highlight the practical impact of using inconsistent forecasts of tail risk measures for decision making, we conduct an empirical application to estimate systemic risk measures based on VaR. The results show that the sensitivity issue and inaccurate uncertainty estimates of VaR translate into systemic risk forecasts, highlighting the need for addressing these issues also in this context. A process for jointly estimating systemic risk measures, CoVaR and CoES, is also proposed. Comparing the forecasting performance of this approach with other systemic risk models could be a possible extension for future work. Moreover, given the risk measures considered in this work are important components for calculating regulatory capital requirements, the work can be extended by investigating the direct impact of the sensitivity and uncertainty underestimation issues within the regulatory environment.

Reference

Acerbi, C. and Szekely, B. (2014). Back-testing expected shortfall. *Risk*, 27(11):76–81.

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

Acharya, V. V., Pedersen, L. H., Philippon, T., and Richardson, M. (2017). Measuring systemic risk. *The Review of Financial Studies*, 30(1):2–47.

Adrian, T. and Brunnermeier, M. K. (2011). CoVaR. Technical report, National Bureau of Economic Research.

Alexander, C. and Sarabia, J. M. (2012). Quantile uncertainty and Value-at-Risk model risk. *Risk Analysis: An International Journal*, 32(8):1293–1308.

Anton, H. and Rorres, C. (2013). *Elementary linear algebra: applications version*. John Wiley & Sons.

Arismendi-Zambrano, J., Belitsky, V., Sobreiro, V. A., and Kimura, H. (2022). The implications of dependence, tail dependence, and bounds' measures for counterparty credit risk pricing. *Journal of Financial Stability*, 58:100969.

Bayer, S. and Dimitriadis, T. (2022). Regression-based expected shortfall backtesting. *Journal of Financial Econometrics*, 20(3):437–471.

Benoit, D. F. and Van den Poel, D. (2017). bayesQR: A Bayesian approach to quantile regression. *Journal of Statistical Software*, 76(7):1–32.

Bon, J. J., Lee, A., and Drovandi, C. (2021). Accelerating sequential Monte Carlo with surrogate likelihoods. *Statistics and Computing*, 31(5):1–26.

Chernozhukov, V. and Hong, H. (2003). An MCMC approach to classical estimation. *Journal of Econometrics*, 115(2):293–346.

Chopin, N. (2002). A sequential particle filter method for static models. *Biometrika*, 89(3):539–552.

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

Danielsson, J., James, K. R., Valenzuela, M., and Zer, I. (2016). Model risk of risk models. *Journal of Financial Stability*, 23:79–91.

Del Moral, P., Doucet, A., and Jasra, A. (2006). Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):411–436.

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

Farkas, W., Fringuellotti, F., and Tunaru, R. (2020). A cost-benefit analysis of capital requirements adjusted for model risk. *Journal of Corporate Finance*, 65:101753.

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

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

Gehrig, T. and Iannino, M. C. (2021). Did the Basel process of capital regulation enhance the resiliency of European banks? *Journal of Financial Stability*, 55:100904.

Gelman, A., Hwang, J., and Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, 24(6):997–1016.

Gerlach, R. and Wang, C. (2020). Semi-parametric dynamic asymmetric Laplace models for tail risk forecasting, incorporating realized measures. *International Journal of Forecasting*, 36(2):489–506.

Gerlach, R. H., Chen, C. W., and Chan, N. Y. (2011). Bayesian time-varying quantile forecasting for value-at-risk in financial markets. *Journal of Business & Economic Statistics*, 29(4):481–492.

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

Hastings, W. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109.

Kerkhof, J., Melenberg, B., and Schumacher, H. (2010). Model risk and capital reserves. *Journal of Banking & Finance*, 34(1):267–279.

Koenker, R. and Machado, J. A. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association*, 94(448):1296–1310.

Kupiec, P. H. (1995). Techniques for Verifying the Accuracy of Risk Measurement Models. *The Journal of Derivatives*, 3(2):73–84.

Lazar, E. and Zhang, N. (2019). Model risk of expected shortfall. *Journal of Banking & Finance*, 105:74–93.

Li, D., Clements, A., and Drovandi, C. (2021). Efficient bayesian estimation for garch-type models via sequential monte carlo. *Econometrics and Statistics*, 19:22–46.

Liu, Y., Li, M., Morris, J. S., et al. (2020). Function-on-scalar quantile regression with application to mass spectrometry proteomics data. *Annals of Applied Statistics*, 14(2):521–541.

McNeil, A. J. and 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.

Morgan, J. (1995). RiskMetrics technical manual. *New York: J.P. Morgan Bank*.

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

Salomone, R., South, L. F., Drovandi, C. C., and Kroese, D. P. (2018). Unbiased and consistent nested sampling via sequential Monte Carlo. *arXiv preprint arXiv:1805.03924*.

Shaby, B. A. (2014). The open-faced sandwich adjustment for MCMC using estimating functions. *Journal of Computational and Graphical Statistics*, 23(3):853–876.

Silva, W., Kimura, H., and Sobreiro, V. A. (2017). An analysis of the literature on systemic financial risk: A survey. *Journal of Financial Stability*, 28:91–114.

Sriram, K., Ramamoorthi, R., and Ghosh, P. (2013). Posterior consistency of Bayesian quantile regression based on the misspecified asymmetric Laplace density. *Bayesian Analysis*, 8(2):479–504.

Taylor, J. W. (2019). Forecasting value at risk and expected shortfall using a semiparametric approach based on the asymmetric laplace distribution. *Journal of Business & Economic Statistics*, 37(1):121–133.

Vehtari, A., Gelman, A., and Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 27(5):1413–1432.

Yang, Y., Wang, H. J., and He, X. (2016). Posterior inference in Bayesian quantile regression with asymmetric Laplace likelihood. *International Statistical Review*, 84(3):327–344.

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

Yu, K. and Stander, J. (2007). Bayesian analysis of a Tobit quantile regression model. *Journal of Econometrics*, 137(1):260–276.

Appendix A Gradients of log-likelihood functions of the SAV-Add and AS-Add models

As discussed in Section 4.3, the log-likelihood gradients of the additive-type models with treating Q_1 and ES_1 as unknown parameters are derived by a recursive approach. For the SAV-Add model, the conditional quantile at different time points $i = 1, 2, \ldots, t$ can be written as:

$$
\begin{aligned}
Q_{i=2} &= \beta_0 + \beta_1|r_1| + \beta_2 Q_1, \\
Q_{i=3} &= \beta_0 + \beta_1|r_2| + \beta_2(\beta_0 + \beta_1|r_1| + \beta_2 Q_1), \\
Q_{i=4} &= \beta_0 + \beta_1|r_3| + \beta_2\left[\beta_0 + \beta_1|r_2| + \beta_2(\beta_0 + \beta_1|r_1| + \beta_2 Q_1)\right] \\
&= \beta_0 + \beta_1|r_3| + \beta_2(\beta_0 + \beta_1|r_2|) + \beta_2^2(\beta_0 + \beta_1|r_1|) + \beta_2^3 Q_1, \\
&\vdots \\
Q_{i=t} &= \beta_0 + \beta_1|r_{t-1}| + \beta_2(\beta_0 + \beta_1|r_{n-2}|) + \beta_2^2(\beta_0 + \beta_1|r_{n-3}|) + \cdots + \beta_2^{n-2}(\beta_0 + \beta_1|r_1|) + \beta_2^{n-1} Q_1 \\
&= \beta_0(1 + \beta_2 + \beta_2^2 + \cdots + \beta_2^{n-2}) + \beta_1(|r_{n-1}| + \beta_2|r_{n-2}| + \beta_2^2|r_{n-3}| + \cdots + \beta_2^{n-2}|r_1|) + \beta_2^{n-1} Q_1 \\
&= \beta_0 \frac{1 - \beta_2^{n-1}}{1 - \beta_2} + \beta_1(|r_{n-1}| + \beta_2|r_{n-2}| + \beta_2^2|r_{n-3}| + \cdots + \beta_2^{n-2}|r_1|) + \beta_2^{n-1} Q_1, \qquad (30)
\end{aligned}
$$

where Q_t is expressed in terms of Q_1 which allows the computation of the derivative of a function of Q_t with respect to Q_1. For the ES component model in Eq (4), the additive difference x_t can be rewritten in terms of Q_1 by substituting Q_{t-1} that has an expression type of Eq (30). To express x_t in terms of ES_1, x_1 needs to be expressed as $Q_1 - ES_1$ first, and then similar to the recursive approach of Q_t in (30), x_t at other time points based upon x_1 can be obtained.

Similarly, for the AS-Add model the conditional quantiles can be written as:

$$
\begin{aligned}
Q_{i=2} &= \beta_0 + \beta_1 I(r_1 > 0)|r_1| + \beta_2 I(r_1 \leq 0)|r_1| + \beta_3 Q_1, \\
Q_{i=3} &= \beta_0 + \beta_1 I(r_2 > 0)|r_2| + \beta_2 I(r_2 \leq 0)|r_2| + \beta_3(\beta_0 + \beta_1 I(r_1 > 0)|r_1| + \beta_2 I(r_1 \leq 0)|r_1| + \beta_3 Q_1), \\
Q_{i=4} &= \beta_0 + \beta_1 I(r_3 > 0)|r_3| + \beta_2 I(r_3 \leq 0)|r_3| + \\
&\quad \beta_3 \left[\beta_0 + \beta_1 I(r_2 > 0)|r_2| + \beta_2 I(r_2 \leq 0)|r_2| + \beta_3(\beta_0 + \beta_1 I(r_1 > 0)|r_1| + \beta_2 I(r_1 \leq 0)|r_1| + \beta_3 Q_1)\right] \\
&\quad = \beta_0(1 + \beta_3 + \beta_3^2) + \left(\beta_1 I(r_3 > 0) + \beta_2 I(r_3 \leq 0)\right)|r_3| + \beta_3(\beta_1 I(r_2 > 0) + \beta_2 I(r_2 \leq 0))|r_2| + \\
&\quad \beta_3^2(\beta_1 I(r_1 > 0) + \beta_2 I(r_1 \leq 0))|r_1| + \beta_3^3 Q_1, \\
&\quad \vdots \\
Q_{i=t} &= \beta_0(1 + \beta_3 + \beta_3^2 + \cdots + \beta_3^{n-2}) + (\beta_1 I(r_{t-1} > 0) + \beta_2 I(r_{t-1} \leq 0))|r_{t-1}| + \\
&\quad \beta_3(\beta_1 I(r_{t-2} > 0) + \beta_2 I(r_{t-2} \leq 0))|r_{t-2}| + \cdots + \beta_3^{t-2}(\beta_1 I(r_1 > 0) + \beta_2 I(r_1 \leq 0))|r_1| + \beta_3^{n-1} Q_1 \\
&\quad = \beta_0 \frac{1 - \beta_3^{n-1}}{1 - \beta_3} + (\beta_1 I(r_{t-1} > 0) + \beta_2 I(r_{t-1} \leq 0))|r_{t-1}| + \beta_3(\beta_1 I(r_{t-2} > 0) + \beta_2 I(r_{t-2} \leq 0))|r_{t-2}| \\
&\quad + \cdots + \beta_3^{t-2}(\beta_1 I(r_1 > 0) + \beta_2 I(r_1 \leq 0))|r_1| + \beta_3^{n-1} Q_1. \qquad (31)
\end{aligned}
$$

Once obtaining the expressions of Q_t and ES_t in terms of all the unknown parameters, the partial derivatives of the log-likelihood function, which is a function of Q_t and ES_t, with respect to each of the parameters can be derived.

Appendix B Loss functions for jointly evaluating VaR and ES forecasts

The general expression of the loss function of Fissler and Ziegel (2016) is given by:

$$
\begin{aligned}
S_t(r_t, Q_t, ES_t) =& (I(r_t \leq Q_t) - \alpha)G_1(Q_t) - I(r_t \leq Q_t)G_1(r_t) \\
&+ G_2(ES_t)\left(ES_t - Q_t + \frac{I(r_t \leq Q_t)(Q_t - r_t)}{\alpha}\right) - \zeta_2(ES_t) + a(r_t), \qquad (32)
\end{aligned}
$$

where G_1 is an increasing function, ζ_2 is increasing and convex, $G_2 = \zeta_2^{'}$, and a is a real-valued integrable function for $t = 1, \ldots, T$. A variety of functions G_1, G_2 and a can be selected to meet these conditions. The loss score for a sample of size T is $S = \sum_{t=1}^{T} S_t$. Three specifications of 32 are employed in this work. The first loss function is the one used in Fissler et al. (2015), where the functions G_1 and G_2 are set as $G_1(x) = x$ and $G_2(x) = \exp(x)/(1 + \exp(x))$, and $a = ln(2)$. The expression of the first score function is given by

$$
\begin{aligned}
S_t(r_t, Q_t, ES_t) =& (I(r_t \leq Q_t) - \alpha)Q_t - I(r_t \leq Q_t)r_t + \frac{\exp(ES_t)}{1 + \exp(ES_t)} \times \\
&\left(ES_t - Q_t + \frac{I(r_t \leq Q_t) \times (Q_t - r_t)}{\alpha}\right) + \ln\left(\frac{2}{1 + \exp(ES_t)}\right). \qquad (33)
\end{aligned}
$$

The second loss function is initially proposed by Acerbi and Szekely (2014), and we set $W = 4$ in the following expression to ensure $WQ_t < ES_t$:

$$\begin{aligned} S_t(r_t, Q_t, ES_t) =& \alpha \left(ES_t^2/2 + WQ_t^2/2 - Q_t ES_t \right) + I(r_t \leq Q_t) \times \\ & \left(-ES_t(r_t - Q_t) + W(r_t^2 - Q_t^2)/2 \right). \end{aligned} \tag{34}$$

The third loss function is the negative of the AL log-likelihood function proposed by Taylor (2019), and the expression is given by

$$S_t(r_t, Q_t, ES_t) = -\ln\left(\frac{\alpha - 1}{ES_t}\right) - \frac{(r_t - Q_t)(\alpha - I(r_t \leq Q_t))}{\alpha ES_t}. \tag{35}$$

Appendix C

Adjusted posteriors:

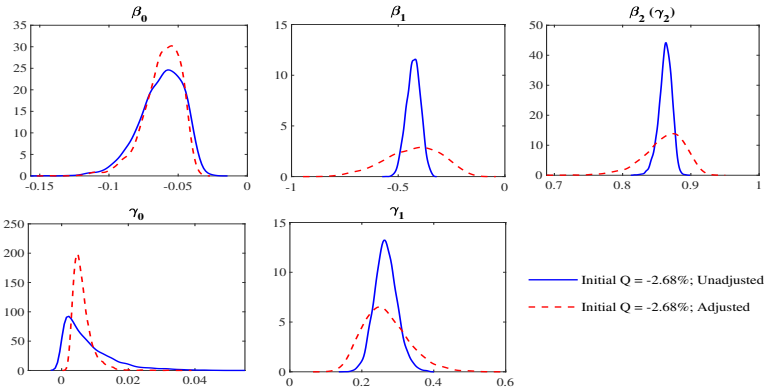


Figure 16: Adjusted marginal posterior distributions of the parameters of the SAV-NewAdd-C model based on 2250 S&P 500 daily returns. ($\alpha = 1\%$)

Adjusted uncertainty:

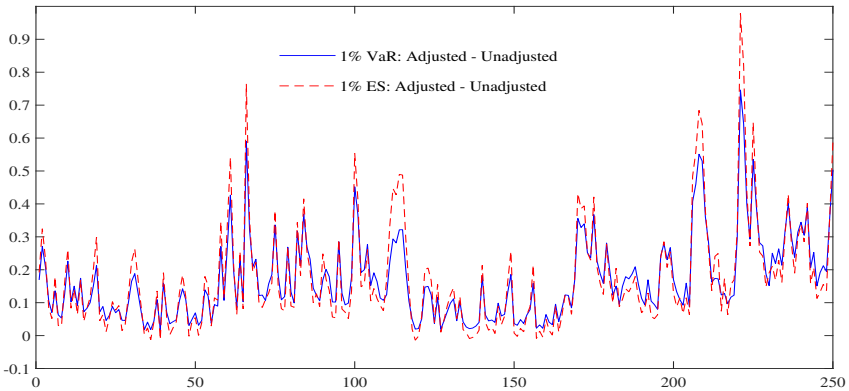
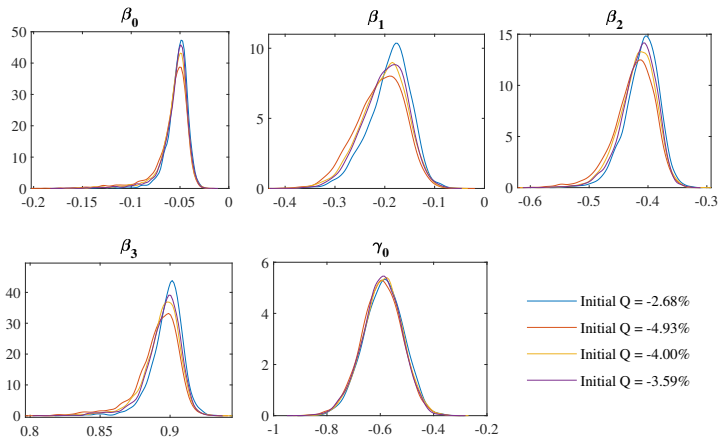
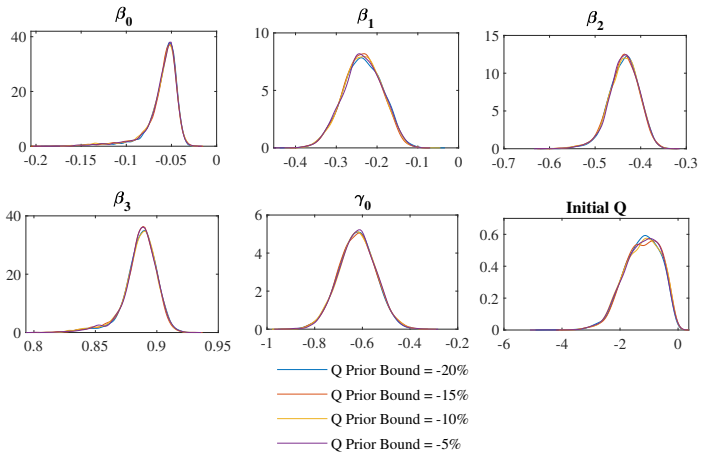


Figure 17: Differences of 250 out-of-sample 95% prediction interval width, before and after the adjustment, of VaR and ES forecasts, for the SAV-NewAdd-C model with the probability level of $\alpha = 1\%$ by using the Bayesian approach.

Marginal posterior distributions for the AS-Mult model and other variations of additive models:

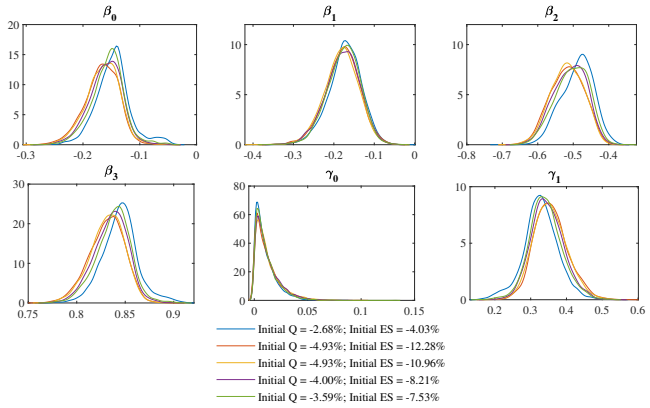


(a) AS-Mult model with choices of Q_1 & ES_1

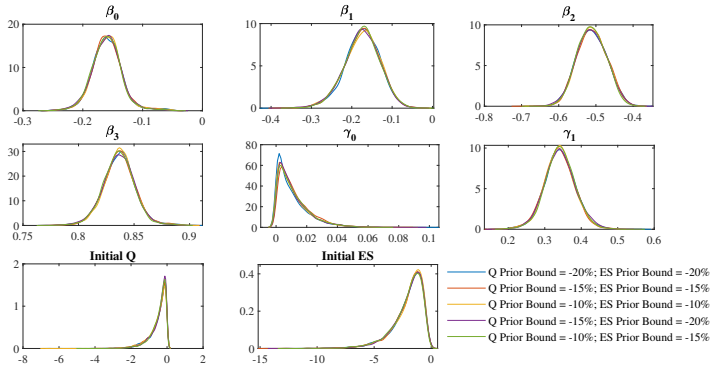


(b) AS-Mult model with parameters Q_1 & ES_1

Figure 18: Marginal posterior distributions of the parameters of the AS-Mult model by using SMC.

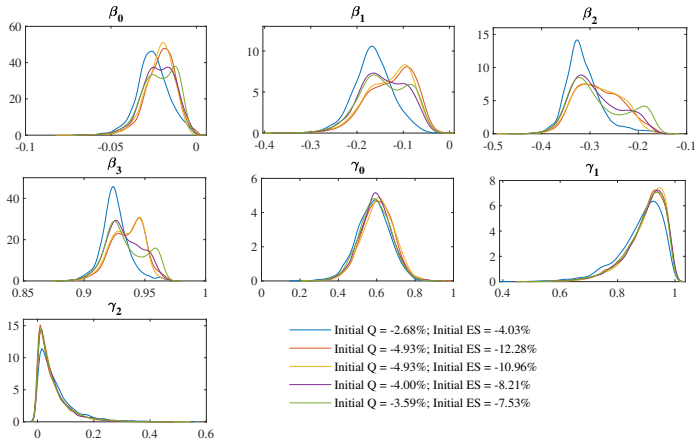


(a) AS-NewAdd-C with choices of Q_1 & ES_1

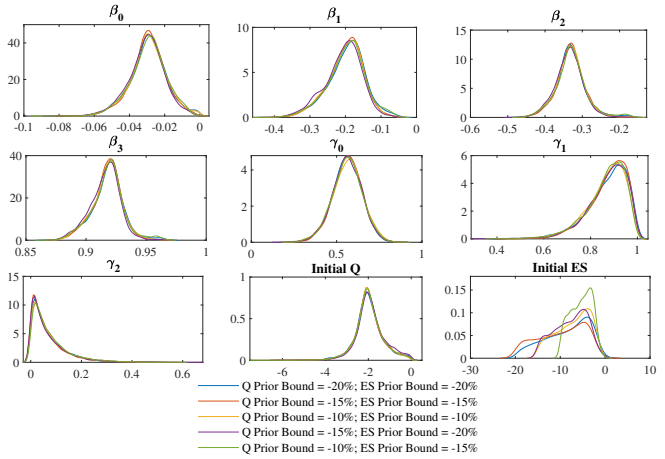


(b) AS-NewAdd-C with parameters Q_1 & ES_1

Figure 19: Marginal posterior distributions of the parameters of the AS-NewAdd-C model by using SMC.

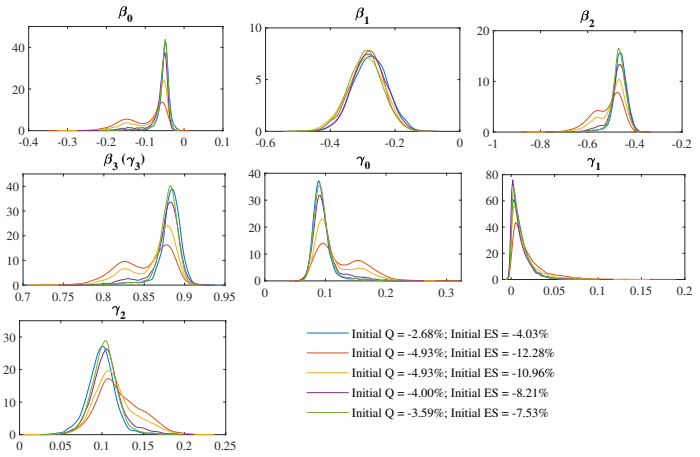


(a) AS-NewAdd-U with choices of Q_1 & ES_1

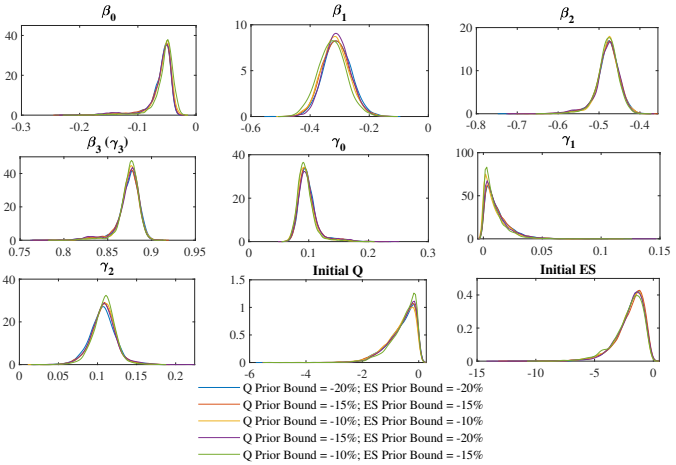


(b) AS-NewAdd-U with parameters Q_1 & ES_1

Figure 20: Marginal posterior distributions of the parameters of the AS-NewAdd-U model by using SMC.

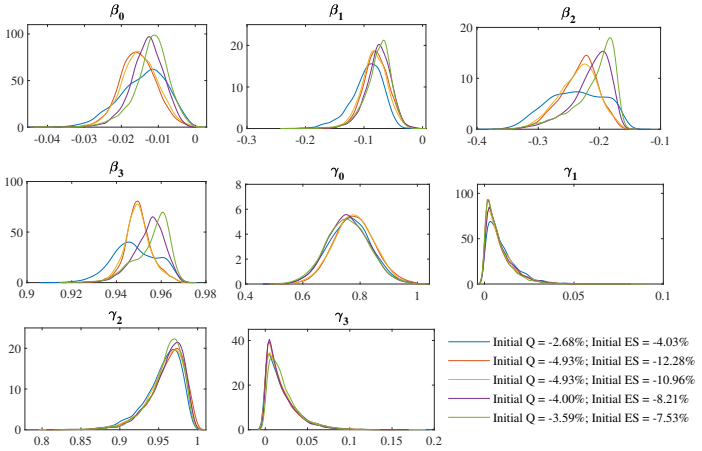


(a) AS-NewAdd-AS-C with choices of Q_1 & ES_1

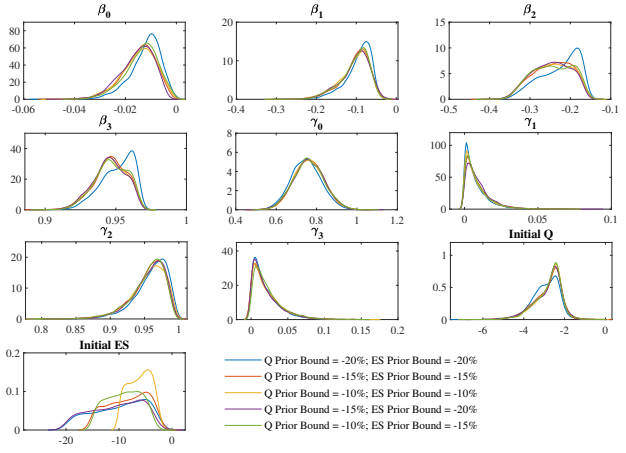


(b) AS-NewAdd-AS-C with parameters Q_1 & ES_1

Figure 21: Marginal posterior distributions of the parameters of the AS-NewAdd-AS-C model by using SMC.



(a) AS-NewAdd-AS-U with choices of Q_1 & ES_1



(b) AS-NewAdd-AS-U with parameters Q_1 & ES_1

Figure 22: Marginal posterior distributions of the parameters of the AS-NewAdd-AS-U model by using SMC.