When MIDAS Meets LASSO: The Power of Low-frequency Variables in Forecasting Value-at-Risk and Expected Shortfall

Yi Luo

Address: Xi'an Jiaotong-Liverpool University, BS Building, 8 Chongwen Road, Suzhou Industrial Park, Suzhou, Jiangsu, 215123, China. Email: Yi.Luo@xjtlu.edu.cn. Tel: +86(0)512 88970289

Xiaohan Xue

Accounting, Finance and Law Division, School of Management, University of Bath

Marwan Izzeldin

Department of Economics, Lancaster University

Abstract: We propose a new framework for the joint estimation and forecasting of Value-at-Risk (VaR) and Expected Shortfall (ES) that integrates low-frequency variables. By maximizing the Asymmetric Laplace likelihood function with an Adaptive Lasso penalty, the most informative variables are selected on a rolling-window basis. In the empirical analysis, realized volatility, term spread, and housing starts serve as the strongest predictors of future tail risk. The out-of-sample backtesting results demonstrate that our method significantly outperforms other benchmarks, and achieves minimum loss in the joint forecasting of both the one-day-ahead and multi-day-ahead extreme S&P500 VaR and ES.

Keywords: Machine learning, Mixed frequency, Value-at-Risk, Expected Shortfall

JEL Classification: C14, C58, G17, G32

1 Introduction

In the context of recent credit and financial crises, appropriate risk measures have increasing importance for banks and other financial institutions. Value-at-Risk (VaR) and Expected Shortfall (ES) are two prevailing risk measures that currently dominate the financial regulatory framework. VaR measures the maximum loss that would occur within a certain period, for a predefined confidence level. Despite its conceptual simplicity and computational ease, VaR has inherent drawbacks. For example, VaR is not a "coherent" risk measure and therefore, it does not consider the benefits of diversification (Artzner et al. (1999)). After the global financial crisis of 2007-2008, Basel Committee on Banking Supervision (2019) proposed using the ES risk measure, which is defined as the conditional expectation of the returns that are below VaR. Unlike VaR, ES is a coherent risk measure, and it is more informative because it considers the tail shape of the loss distribution. Therefore, ES has been used as an alternative risk measure to complement the VaR measure. However, ES is not "elicitable", which means that no loss function exists for which ES is the solution that minimizes the loss.[1] Many studies have examined the problem of elicitability (Engle and Manganelli, 2004; Taylor, 2008; Zhu and Galbraith, 2011; Fissler and Ziegel, 2016; Du and Escanciano, 2017; Patton et al., 2019). Until the work of Fissler and Ziegel (2016), ES is found to be jointly elicitable with VaR and a set of suitable scoring functions are proposed.

With the development of methods, forecasting efficiency of VaR and ES have been significantly improved by incorporating the information from high-frequency data into the parametric models (Giot and Laurent, 2004; Hansen et al., 2012; Louzis et al., 2014) and the semiparametric models (Clements et al., 2008; Meng and Taylor, 2020; Lazar and Xue, 2020; Gerlach and Wang, 2020). Specifically, realized volatility, proposed by Andersen and Bollerslev (1998) and Alizadeh et al. (2002), is a near unbiased return variation measure and is one of the most widely exploited information variables in forecasting tail risk. The commonly used realized volatility is selected with 5-minute or 10-minute frequency. Existing empirical research documents that the models using realized volatility significantly outperform the earlier models in forecasting VaR and ES, which reveals that volatility strongly predicts tail risk.

Although the existing frameworks can accommodate additional high-frequency information in forecasting VaR and ES, they provide no clear way of utilizing the variables observed at a lower frequency. The benefits of heterogeneous sources of information in forecasting tasks are well documented in the literature. For example, many studies have demonstrated the superior predictability of economic variables on conditional return distribution features such as volatility (Engle et al., 2013) or return density (Cenesizoglu and Timmermann, 2008). However, the effect of low-frequency variables on forecasting tail dynamics is surprisingly much less exploited. The main difficulty of integrating low-frequency variables into forecasting tail risk lies in directly dealing with the frequency disalignment within the tail risk forecasting framework. Massacci (2017) conducts correlation analysis between tail risk and macroeconomic indicators using the returns on the United States (US) decile-sorted equity portfolios. The author finds that tail risk is countercyclical and the causal analysis reveals that the negative tail risk shocks to large firms Granger causes the tail shocks to the economy. However, this research, is limited in two aspects. First, it aims to approximate the conditional distribution of positive return exceedances implied by the Extreme Value Theory, which falls into the category of modeling the conditional distribution of extreme returns. Engle and Manganelli (2004) argue that the models in this category work only for very low probability quantiles, and its approximation is poor at less extreme probability levels. Second, low-frequency variables are still not formally entered into tail risk forecasting framework given that the correlation analysis is conducted by aggregating daily tail risk sequences into monthly values to align with the frequency of monthly macroeconomic indicators. More recent research has attempted to improve tail risk forecasting by formally integrating low-frequency variables (Candila et al., 2020; Le, 2020; Pan et al., 2021; Xu et al., 2021). However, the models proposed in such research are either limited to forecasting only VaR, leaving ES unexplored, or to accommodating a single predefined low-frequency variable. Thus, there are three challenges associated with incorporating low-frequency information into tail risk modeling that need to be solved: (1) the common data frequency disalignment between low-frequency variables and risk measures (Ghysels et al., 2004; Engle et al., 2013; Candila et al., 2020); (2) the need for a framework than can integrate a large amount of variables while utilizing only the most important variables among abundant candidates in forecasting both VaR and ES (Gu et al., 2020; Chen et al., 2020); (3) the need for a suitable loss function that can accommodate variable selection and parameter estimation in the proposed model (Patton et al., 2019; Taylor, 2019). In view of this, it is necessary to calibrate a model with a clear loss function that integrates different but effective information when modeling tail dynamics.

We address the first challenge through the channel of volatility, and we propose a new quantile-based model with an effective variable screening procedure to overcome the second and third challenges. Volatility is an important driving factor of tail risk and is found to be closely related to low-frequency macroeconomic and financial indicators through the Mixed Data Sampling (MIDAS) regression model proposed by Ghysels et al. (2004). Based on their work, many extended studies have demonstrated that the dynamic of return volatility is characterized by multiple components capturing information at different time horizons. For example, Engle et al.

(2013) propose the GARCH-MIDAS model to extract two components of volatility: the short-term component that follows a GARCH(1,1) process, and the long-term component that relates to low-frequency variables. Their study finds a significant macro-volatility relationship by directly incorporating low-frequency macroeconomic variables into the long-term volatility component. Their findings have been strengthened by many other studies, and the GARCH-MIDAS model has become the most popular model employed in investigating the relationships between aggregate financial volatility and macroeconomic and financial variables (Asgharian et al., 2013; Conrad et al., 2014; Conrad and Loch, 2015; Pan et al., 2017; Su et al., 2017; Conrad and Kleen, 2020). Following Engle et al. (2013), we decompose the return volatility into a short-term and a long-term component where an extensive set of macroeconomic and financial indicators are embedded. However, we extend the GARCH-MIDAS model to allow a direct forecast of tail risk, which complements the studies that examine the link between volatility and the macroeconomy.

Thus, we propose a new quantile-based model for jointly estimating VaR and ES that characterizes the relationships between risk measures and macroeconomic and financial indicators. Quantile-based modeling avoids the distributional assumptions on returns and allows the dynamics of the quantiles to vary for each probability level (Engle and Manganelli, 2004). This approach has been focused on VaR and produces superior VaR forecasts (e.g., Şener et al. (2012)). Taylor (2019) extends this approach and provides a way of producing ES forecasts in a quantile setting. By assuming the Asymmetric Laplace (AL) density on the pair of VaR and ES, Taylor (2019) demonstrates that the negative of AL likelihood function belongs to the function set proposed by Fissler and Ziegel (2016) and that it is therefore strictly consistent in the joint evaluation of VaR and ES. Following Taylor (2019), we assume the AL density on VaR and ES to obtain a strictly consistent scoring function. Our approach differs from that of Taylor (2019) by integrating a large set of low-frequency variables through the channel of volatility, allowing new information to support tail risk forecasting.

However, with a large number of variables being considered, the number of estimated parameters rises, which increases model complexity and reduces estimation efficiency. Fang et al. (2020) apply the Adaptive Lasso proposed by Zou (2006) to select the variables that display the strongest signal in predicting the long-term volatility component under the GARCH-MIDAS framework. They find that the GARCH-MIDAS model with variable selection outperforms all benchmark models except the model with realized volatility. The limitation of their approach is that the selected variables do not vary for every out-of-sample estimation, and it is unreasonable to assume that those variables have the same contribution in each rolling window of the testing sample. To address this problem, we propose an innovative dynamic variable selection process where the variables are selected on a rolling-window basis. This helps us to determine the most important variables, and allows us to visualize the evolution of supporting variables in predicting VaR and ES.

Our research makes two main contributions to the literature. First, we propose the semiparametric GARCH-MIDAS-X model to integrate low-frequency variables into high-frequency tail risk forecasting, allowing the use of richer information in forecasting tail risk. As discussed, the lack of a suitable model that deals with frequency disalignment makes it difficult to utilize low-frequency information such as macroeconomic indicators in forecasting tail risk. The model we propose addresses this problem through the channel of volatility and retains the semiparametric setting to avoid distributional assumptions on returns. The model efficiently produces forecasts for both VaR and ES. Simulation results demonstrate that the parameter estimates are nicely centered around their true values. The corresponding estimation framework avoids possible identification problems in estimating the model parameters.

Second, we embed a dynamic variable selection device to select and utilize the most important low-frequency variables in forecasting future tail risk. Instead of putting all candidate variables into the model or subjectively selecting the variables, we select the most informative variables that maximize the AL likelihood function with Adaptive Lasso penalty. Our proposed framework allows parameter estimation and variable selection simultaneously. No preprocessing is needed. Simulation results reveal that the Adaptive Lasso performs best under medium variable correlation across all the considered confidence levels. In addition, the selection process is conducted on a quarterly rolling basis, allowing possible changing effect of the low-frequency variables on tail risk to be visualized. The negative relationship between the selected macroeconomic variables and tail risk provides novel statistical evidence of countercyclical tail risk. The proposed model is then empirically evaluated using various backtesting approaches in forecasting one-day-ahead tail risk at four confidence levels. The results reveal that our proposed model with the selected variables significantly outperforms all benchmark models, particularly the model using only realized volatility in forecasting more extreme tail risk (i.e., 1%, 2.5%, and 5%). It also produces equivalently good forecasts as the model integrating realized volatility at less extreme tails (i.e., 10%). This reveals that the power of effective low-frequency information is stronger in forecasting extreme tail risk while the less extreme tail risk is more associated with the stock market volatility. When considering longer forecasting horizons, our proposed model also produces superior tail risk forecasts which overcomes the practical challenge in forecasting multi-day-ahead VaR and ES.

The remainder of this paper is structured as follows. Section 2 briefly introduces the proposed model and the procedure of variable selection. Section 3 presents the simulations of model parameters and the variable selection methods. Section 4 presents the data used in our empirical study. Section 5 discusses the dynamic variable selection results and the out-of-sample performance. Section 6 reports the long-term forecasting results. Section 7 provides the conclusions.

2 Models

2.1 Semiparametric GARCH-MIDAS-X model

Let $r_{i,t}$ be an asset's return on the ith day of the tth quarter, where $i = 1, \ldots, N_t$, $t = 1, \ldots, T$. This setting accommodates variations in the number of trading days per quarter. Following Engle et al. (2013), we specify the model of $r_{i,t}$ as follows:

$$r_{i,t} = \mu + \sqrt{\tau_t g_{i,t}} \epsilon_{i,t}, \tag{1}$$

where $\epsilon_{i,t}$ is identically and independently distributed (i.i.d.) innovations, and it follows an unknown distribution $F(\epsilon_{i,t})$ with zero mean and unit variance. μ denotes the mean of the asset returns. τ_t denotes the long-term volatility component in quarter t, and $g_{i,t}$ denotes the short-term volatility component on the ith day in the tth quarter.

Following Conrad and Kleen (2020), the short-term volatility component is assumed to follow a mean-reverting unit-variance GJR-GARCH(1,1) process: [2]

$$g_{i,t} = (1 - \beta_1 - \gamma/2 - \beta_2) + (\beta_1 + \gamma \cdot \mathbb{1}\{r_{i-1,t} < \mu\}) \frac{(r_{i-1,t} - \mu)^2}{\tau_t} + \beta_2 g_{i-1,t}, \tag{2}$$

where $\mathbb{1}$ is an indicator function which returns one when $r_{i-1,t} < \mu$ and returns zero otherwise. The parameter γ captures the leverage effect.

The long-term volatility component with a single low-frequency variable is given by:

$$\log(\tau_t) = m + \theta \sum_{k=1}^{K} \psi_k(w_1, w_2) X_{t-k}, \tag{3}$$

where X_t is the exogenous variable in quarter t, K is the lagged order, and $\log(\tau_t)$ is considered rather than τ_t in order to ensure a non-negative long-term volatility. Here, $\psi_k(w_1, w_2)$ is the Beta weighting scheme, which is defined as:

$$\psi_k(w_1, w_2) = \frac{(k/(K+1))^{w_1-1} \cdot (1 - k/(K+1))^{w_2-1}}{\sum_{s=1}^{K} (s/(K+1))^{w_1-1} \cdot (1 - s/(K+1))^{w_2-1}}, \tag{4}$$

where $\psi_k(w_1, w_2)$ measures the weight for the kth lag of the explanatory variable X_{t-k}, and it is determined by the two parameters w_1 and w_2. For $k = 1, \ldots, K$, we have $\psi_k \geq 0$ and $\sum_{k=1}^{K} \psi_k = 1$.

Given a probability level α for quantile, we characterize the quantile expression of Eq.(1) and obtain the tail risk process as:

$$\begin{aligned} v_{i,t}(r_{i,t}; \alpha|\Psi_{t-1}) - \mu &= a\sqrt{\tau_t g_{i,t}}, \text{ where } a = F_\alpha^{-1}(\epsilon_{i,t}), \\ e_{i,t}(r_{i,t}; \alpha|\Psi_{t-1}) - \mu &= b\sqrt{\tau_t g_{i,t}}, \text{ where } b = \mathbb{E}[\epsilon_{i,t}|\epsilon_{i,t} \leq a], \end{aligned} \tag{5}$$

where Ψ_{t-1} is the information set at time $t-1$, $F_\alpha^{-1}(\epsilon_{i,t})$ is the quantile of $\epsilon_{i,t}$ at level α. $v_{i,t}(r_{i,t}; \alpha|\Psi_{t-1})$ and $e_{i,t}(r_{i,t}; \alpha|\Psi_{t-1})$ denote VaR and ES based on the information up to the day $i-1$ of the quarter t at the α confidence level, respectively. We use $v_{i,t}$ and $e_{i,t}$ from now for notation convenience.

Recent studies that investigate the relationships between stock market tail risk and low-frequency variables (Candila et al., 2020; Le, 2020; Pan et al., 2021; Xu et al., 2021) focus on either the impact of a single predefined low-frequency variable or on the VaR forecasting only. Clearly, an alternative framework is necessary for multiple low-frequency variables to predict both VaR and ES. Therefore, we now propose a model that integrates multiple low-frequency variables for the joint forecasting of VaR and ES as follows:

$$\begin{cases} v_{i,t} - \mu = a\sqrt{\tau_t g_{i,t}} \\ e_{i,t} - \mu = b\sqrt{\tau_t g_{i,t}} \\ g_{i,t} = (1-\beta_1-\gamma/2-\beta_2) + (\beta_1 + \gamma \cdot \mathbb{1}\{r_{i-1,t} < \mu\})\dfrac{(r_{i-1,t}-\mu)^2}{\tau_t} + \beta_2 g_{i-1,t} \\ \tau_t^{\frac{1}{2}} = \exp\left(m + \sum_{j=1}^{J}\theta_j \sum_{k=1}^{K}\psi_k(w_{j,1}, w_{j,2})X_{j,t-k}\right) \\ \psi_k(w_1,w_2) = \dfrac{(k/(K+1))^{w_1-1}\cdot(1-k/(K+1))^{w_2-1}}{\sum_{s=1}^{K}(s/(K+1))^{w_1-1}\cdot(1-s/(K+1))^{w_2-1}}, \end{cases} \tag{6}$$

where J denotes the number of explanatory variables ($J = 20$ in this study). If $J = 1$, the model integrates any single low-frequency variable. θ_j measures the impact of the jth explanatory variable on the long-term volatility.

In the standard GARCH-family models, the asset returns is assumed to follow a specified conditional distribution. The maximum likelihood estimation is a more adaptable method for the parameter estimation. However, if the underlying conditional distribution is unclear, the maximum likelihood estimation method is no longer valid [3]. Patton et al. (2019) propose a semiparametric model, where the parameters are estimated by minimizing the FZ0 loss function (Fissler and Ziegel, 2016):

$$L_{FZ0}(r,v,e;\alpha) = -\frac{1}{\alpha e}\mathbb{1}\{r \leq v\}(v-r) + \frac{v}{e} + \log(-e) - 1. \tag{7}$$

Instead of adopting the FZ0 loss minimization, we follow Taylor (2019) to assume the AL density on the pair of VaR and ES. Therefore, our model falls into the semiparametric framework where the parameters are estimated by maximizing the AL likelihood function:[4]

$$LLF(\Phi) = \sum_{t=1}^{T}\sum_{i=1}^{N_t}\left[\log\left(\frac{\alpha-1}{e_{i,t}}\right) + \frac{(r_{i,t}-v_{i,t})(\alpha - \mathbb{1}\{r_{i,t} \leq v_{i,t}\})}{\alpha e_{i,t}}\right]. \tag{8}$$

In this study, we only focus on the MIDAS-extended semiparametric GARCH model and we call it the semiparametric GARCH-MIDAS-X model. However, this framework can be extended to other semiparametric models proposed by Patton et al. (2019), or to the alternative CAViaR frameworks introduced by Engle and Manganelli (2004) and Taylor (2019). This may include the Asymmetric Slope model, the Symmetric Absolute Value model, and the Adaptive and Indirect GARCH(1,1) models.

2.2 Semiparametric GARCH-MIDAS-X with variable selection

As the number of parameters rises with the number of variables included, difficulties arise in extracting those with the strongest signals from those who are less important. We therefore conduct variable screening by incorporating the Adaptive Lasso penalty proposed by Zou (2006) [5]. The penalized likelihood function is:

$$PLLF_\lambda(\Phi) = \sum_{t=1}^{T}\sum_{i=1}^{N_t}\left[\log\left(\frac{\alpha-1}{e_{i,t}}\right) + \frac{(r_{i,t}-v_{i,t})(\alpha - \mathbb{1}\{r_{i,t}\leq v_{i,t}\})}{\alpha e_{i,t}}\right] - \lambda\sum_{j=1}^{J}\hat{w}_j|\theta_j|, \tag{9}$$

where $\lambda > 0$ is the tuning parameter regulating the shrinkage power in the Adaptive Lasso penalty, $PLLF_\lambda(\Phi)$ denotes the penalized likelihood function for a given λ, and $\hat{w}_j$ is the adaptive weight for each explanatory variable X_j.

To obtain the adaptive weights calculated as $\hat{w}_j = \frac{1}{|\hat{\theta}_j|^\eta}$, we estimate the semiparametric GARCH-MIDAS-X model with all J variables to obtain the estimates $\hat{\theta_j}$. Following Zou (2006), we set $\eta = 2$ to achieve higher probability of obtaining the true model.

2.3 Tuning parameter selection

Choosing the value of the tuning parameter λ in the penalized likelihood function is vital for the success of the variable screening process. A small λ increases the number of selected variables while taking the risk of including noisy variables. In contrast, a sufficiently large λ rules out all variables, although some of them may display a strong signal in forecasting tail risk. Among the many techniques for tuning parameter selection, the cross-validation and the information criteria are familiar options. Considering the nature of time-series data in our empirical study, the large number of covariates considered, and the parameter estimation by the maximum likelihood, we apply the Generalized Information Criteria (GIC) to determine the tuning parameter.

For high-dimensional penalized likelihood, Fan and Tang (2013) recommend using the GIC to select the tuning parameter. The GIC draws from a trade-off between model fitting and model complexity by two components. The first component evaluates the goodness of fit, which increases with the number of explanatory variables. The second component penalizes model complexity as associated with the number of variables. The GIC for a given λ is calculated as follows:

$$GIC_\lambda = \frac{1}{N}\left\{2[LLF(\hat{\Phi}) - PLLF_\lambda(\hat{\Phi}_\lambda)] + a(N,p)|\hat{\theta}_\lambda|\right\} \tag{10}$$

where N is the number of observations. $LLF(\hat{\Phi})$ is the maximum value of the likelihood function including all candidate variables, and $PLLF_\lambda(\hat{\Phi}_\lambda)$ is the maximum value of the penalized likelihood function for a given λ with variable selection. $2[LLF(\hat{\Phi}) - PLLF_\lambda(\hat{\Phi}_\lambda)]$ is the scaled deviation between the original model including all variables and the model with the selected variables. $|\hat{\theta}_\lambda|$ is the $l_1 - norm$ of the parameter vector θ for a given λ. $a(N,p)$ is a positive value regulating the trade-off between model fitting and model complexity, which depends on the number of observations N and the number of estimated parameters p. Following Fang et al. (2020), we set $a(N,p) = log\{log(N)\}\cdot log(p)$. For the empirical analysis, the optimum tuning parameter λ^* is selected according to the minimum GIC_λ over the range $[0, \lambda_{max}]$.

2.4 Estimation framework

Two problems arise with the estimation: dynamic selection and identification. With the dynamic selection problem, variable screening is usually static, offering limited support to the out-of-sample forecasting. It is also unlikely that the same variables would be selected, or that they would make the same contributions, at different times. To address this issue, we conduct a dynamic variable selection process where the optimum tuning parameter λ^* and the important variables are selected on a rolling-window basis.

The identification problem arises when variables are dropped from the model if their respective parameters are shrunk to zero by the Adaptive Lasso during the screening process. This leads to an identification problem as the Beta weighting parameters of the dropped variables will not enter into the penalized likelihood function and therefore will not be identified. To avoid this problem, we fix the Beta weighting parameters when estimating the other parameters. Specifically, suppose the penalized likelihood function for the semiparametric GARCH-MIDAS-X model depends on two sets of parameters $\Phi_1 = (\mu, m, \eta_0, \beta_1, \beta_2, \beta_3, \theta_1, \theta_2, ..., \theta_{20})$ and $\Phi_2 = (w_{1,1}, w_{1,2}, w_{2,1}, w_{2,2}, ..., w_{20,1}, w_{20,2})$. We fix the Beta weighting parameters $\Phi_2 = \hat{\Phi}_2$ before estimating Φ_1 where $\hat{\Phi}_2$ is obtained by estimating the semiparametric GARCH-MIDAS-X model with all candidate variables.

A two-stage procedure is proposed to address the above two problems in a single framework. The procedure is demonstrated below and additional details are provided in Figure 1.

Step 1: Obtain the Beta weighting estimates and calculate the adaptive weight

- Estimate the semiparametric GARCH-MIDAS-X model with all J variables by maximizing the AL likelihood function as in Eq.(8) under the linear constraints [6]. Obtain the Beta weighting estimates $\Phi_2 = \hat{\Phi}_2$ and the estimated effect of each low-frequency variable on future tail risk $\hat{\theta}_j$.
- Calculate the adaptive weight as $\hat{w}_j = \frac{1}{|\hat{\theta}_j|^\eta}$.

Step 2: Dynamically select the optimum tuning parameter and the variables

- Divide the entire out-of-sample into 66 rolling windows. For each rolling window, choose the optimum tuning parameter and select the variables.
- Set $\Phi_2 = \hat{\Phi}_2$ and $w_j = \hat{w}_j$. For each rolling window, estimate the semiparametric GARCH-MIDAS-X model with variable selection by maximizing the penalized AL likelihood function as in Eq.(9) under the linear constraints, with the tuning parameter λ on a grid of $[0, \lambda_{max}]$. Obtain the parameter estimates $\hat{\Phi}_1$ and calculate the GIC for each value of λ.
- Choose the optimum tuning parameter λ^* according to the minimum GIC and obtain the selected variables.
- Repeat the above process until the last window is reached.

[INSERT FIGURE 1 ABOUT HERE]

3 Simulation study

In this section, we conduct two simulations that serve different purposes. First, we evaluate the finite-sample performance of the Maximum Likelihood Estimator (MLE) of the proposed semiparametric GARCH-MIDAS-X model integrating a single low-frequency variable with the restricted Beta weighting scheme. The parameters are estimated by maximizing the AL likelihood function. Second, we examine the power of variable selection methods in selecting the most important variables under the semiparametric GARCH-MIDAS-X model with multiple low-frequency variables. In this case, we estimate the parameters by maximizing the penalized AL likelihood function. For each simulation, we perform 2,000 Monte Carlo replications.

3.1 Simulation 1: parameter estimates

Under the Data Generating Process (DGP) specified in Eq.(1), (VaR, ES) are proportional to $\sqrt{\tau_t g_{i,t}}$ with

$$(VaR^{\alpha}_{i,t}, ES^{\alpha}_{i,t}) = (a_\alpha, b_\alpha)\sqrt{\tau_t g_{i,t}} \tag{11}$$

where $a_\alpha = F^{-1}_\alpha(\epsilon_{i,t})$ and $b_\alpha = \mathbb{E}[\epsilon_{i,t}|\epsilon_{i,t} \leq a]$. Following Patton et al. (2019) and Conrad and Kleen (2020), we consider three choices for the distribution of $\epsilon_{i,t}$: the standard normal, the standardized Student's t distribution with five degrees of freedom, and the standardized skew t distribution with the degrees of freedom (ϑ) and the skewness (κ) being set to $\vartheta = 5$ and $\kappa = -0.5$, respectively. We select four confidence levels in the simulation study: $\alpha \in \{0.01, 0.025, 0.05, 0.10\}$. To match our empirical application, we replace b_α with $c_\alpha = b_\alpha/a_\alpha$, and our parameter vector now becomes $(a_\alpha, c_\alpha, \beta_1, \beta_2, \gamma, m, \theta, w_2)$.

Proportional parameters: The value of (a_α, c_α) depends on the distributions and the confidence levels. We report the chosen values of (a_α, c_α) for each distribution and confidence level in the first row of each results table.

Volatility parameters: Following Conrad and Kleen (2020), we set parameters for the short-term volatility component $g_{i,t}$ as $\beta_1 = 0.06$, $\beta_2 = 0.91$, and $\gamma = 0$. For the long-term volatility component τ_t, we choose its exponential specification as in Eq.(3). Following Conrad and Loch (2015) and Conrad and Kleen (2020), we specify the MIDAS parameters in the long-term volatility component as $m = 0.1$, $\theta = 0.3$, $w_2 = 4$, and $K = 12$ [7]. The chosen value of w_2 gives a monotonically decaying weighting scheme with weights close to zero for lags greater than two-thirds of the lag length. The exogenous variable X_t is set to follow an AR(1) process where $X_t = \phi X_{t-1} + e_t, e_t \overset{iid}{\sim} \mathcal{N}(0, \sigma_e^2)$ with $\phi = 0.9$ and $\sigma_e^2 = 0.3^2$.

We consider two sample sizes with $T \in \{2520, 5040\}$, representing 10 and 20 years of daily returns, respectively. To mitigate sensitivity to starting values, we set true parameter values as the starting values. We repeat all the simulations 2,000 times.

Table 1 presents the estimation results for the semiparametric GARCH-MIDAS-X model with a single low-frequency variable on the skew t innovations.[8] The first row of each panel lists the true parameter values with a_α

and c_α varying across the distributions and the confidence levels. The second row reports the median estimated parameters across the simulations and the third row presents the average bias of the estimated parameters. For all the considered innovation distributions, close median of the estimates and small bias are observed. This indicates that the estimated parameters are nicely centered around the true parameter values. In addition, the average bias decreases when the sample size and the confidence level increase (i.e., when the α increases), as expected. The fourth row displays the standard deviations of the estimated parameters across the simulations. We find that the standard deviations decrease with the sample size and increase when moving further into the tails (i.e., when the α decreases) for all the considered innovation distributions, both as expected.

The fifth row of each panel reports the coverage ratio of the 95% confidence intervals constructed by the estimated standard errors for each parameter. Patton et al. (2019) and Pan et al. (2021) demonstrate that the estimator of the GARCH-MIDAS model integrating a single low-frequency variable is asymptotically normal. Therefore, the computation of the standard errors relies on the estimation of the asymptotic covariance matrix [9]. For all the considered innovation distributions, we find that the coverage ratio is reasonably stable across the confidence levels. In particular, the coverage ratio of most parameters is still higher than 0.90 even in extreme tails (i.e., $\alpha = 0.01$ and 0.025). This demonstrates the superiority of the proposed semiparametric GARCH-MIDAS-X model with a single low-frequency variable.

3.2 Simulation 2: variable selection

In Simulation 2, we examine the performance of various variable selection methods under the semiparametric GARCH-MIDAS-X model with multiple low-frequency variables in Eq.(6). To demonstrate possible varying selection power under different conditions (i.e., high/low signal-to-noise ratio and high/low variable correlation), we consider four variable selection approaches: Lasso, Adaptive Lasso (ALasso), Elastic Net (ENet), and Adaptive Elastic Net (AENet). As in Simulation 1, we consider four confidence levels: $\alpha \in \{0.01, 0.025, 0.05, 0.10\}$, and repeat all simulations 2,000 times. However, because of space limitation, we report only the results for the skew t innovations [10].

Proportional and volatility parameters: In Simulation 2, we maintain the same setup as in Simulation 1 for all the proportional and volatility parameters except for the parameter θ and the low-frequency variables $\mathbf{X}_t$. We now include low-frequency variable vector $\mathbf{X}_t$, where $\mathbf{X}_t = [X_{t,1}, ..., X_{t,J}]'$, and $X_{t,j}, 1 \leqslant j \leqslant J$, represents the column vector storing values for the jth low-frequency variable at time t. We set $t = 160$ covering 40 years, and $J = 20$, which aligns with the number of low-frequency variables in our empirical study. As in Simulation 1, $\mathbf{X}_t$ is set to follow a multivariate VAR(1) process where $\mathbf{X}_t = \boldsymbol{\phi}\mathbf{X}_{t-1} + \mathbf{e}_t, \mathbf{e}_t \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{\Sigma})$, where $\boldsymbol{\phi}$ is a $J \times J$ diagonal matrix with all diagonal elements being 0.9, and $\mathbf{e}_t$ is a $J \times 1$ unobservable zero mean white noise vector process (time serially uncorrelated or independent) with a time-invariant covariance matrix. $\sigma^2 = 0.3^2$, and $\mathbf{0}$ is a $J \times 1$ vector of zeros. The $\boldsymbol{\Sigma}$ is a symmetric correlation matrix, and we set the pairwise correlation between the predictor X_{t,j_1} and X_{t,j_2} to be $corr(j_1, j_2) = \rho^{|j_1 - j_2|}$, $1 \leqslant j_1, j_2 \leqslant J$. To examine the performance of the variable selection methods under different pairwise correlations, we set $\rho \in \{0.5, 0.99\}$ allowing medium and high variable correlation. We now describe the setup for the signal parameter θ as follows.

Signal parameter: Under the semiparametric GARCH-MIDAS-X model in forecasting tail risk, the signal comes from the long-term volatility component while the noise comes from the short-term volatility component. Therefore, the θ measures the signal intensity of low-frequency variables in tail risk forecasting. We define $\theta = \zeta \mathbf{w}$, where ζ controls the signal-to-noise ratio and $\mathbf{w}$ is a vector storing weights assigned to each low-frequency variable. We assign 45% of weight to $X_{t,1}$, 25% to $X_{t,2}$, 30% to $X_{t,5}$, and all other low-frequency variables carry zero weight. That is, $X_{t,1}$, $X_{t,2}$, and $X_{t,5}$ are the correct variables to be chosen. To compare the performance of different variable selection methods under varying signal-to-noise ratios, we select $\zeta \in \{1.5, 1, 0.5\}$ representing high, medium, and low signal-to-noise ratios, respectively.

Table 2 reports the variable selection results at four confidence levels. The performance of four variable selection methods is measured in three dimensions: number of chosen variables, number of correctly chosen variables, and number of incorrectly chosen variables with their true values listed in the first row of each panel. For each considered variable selection method, we report the median and the mean performance of these three dimensions. The left and the right panels present the results for $\rho = 0.5$ and $\rho = 0.99$, respectively.

Given the medium variable correlation ($\rho = 0.5$), ALasso performs best for the number of chosen variables, the number of correctly chosen variables, and the number of incorrectly chosen variables for all the considered confidence levels and the signal-to-noise ratios. Its median in each of the three dimensions is the closest to their true values, and its mean subjects to the least downward or upward bias. AENet ranks second in performance. Although it achieves a median similar to ALasso in all dimensions, it selects more variables on average and therefore subjects to upward selection bias. Lasso ranks third in performance, followed by ENet. Both of ENet and Lasso have higher upward selection bias. However, when variable correlation increases ($\rho = 0.99$), AENet performs best while ALasso gains more upward selection bias. This is as expected given that the grouping effect is not allowed in ALasso and that highly correlated variables tend to be in or out of the model together under AENet (Zou and Hastie, 2005).

Although ALasso and AENet outperform the other methods under a medium and high variable correlation, respectively, the average performance of all methods in the three dimensions generally declines when the signal-to-noise ratio decreases. In particular, under a low signal-to-noise ratio ($\zeta = 0.5$), the least affected dimension is the number of correctly chosen variables, while all methods confront greater upward selection bias by incorrectly selecting more variables than they do under a high signal-to-noise ratio ($\zeta = 1.5$). This indicates that all the considered variable selection methods may suffice for selecting the correct variables but the risk of including the incorrect variables increases when the signal-to-noise ratio is reduced. This is quite logical, because, higher noise elevates the difficulty of extracting the effective signal. Despite the case, it is worth noting that the ALasso and AENet still possess the least upward selection bias under a medium and high variable correlation, respectively, when facing a low signal-to-noise ratio. The finding is consistent across all the considered confidence levels.

Comparing the adaptive methods with their counterparts (i.e., ALasso vs. Lasso, AENet vs. ENet), the adaptive method generally performs better in the number of chosen variables and the number of incorrectly chosen variables, as expected. A similar pattern is observed across all the considered confidence levels, the signal-to-noise ratios, and the variable correlations. Our finding is consistent with Zou (2006), where the author argues that the non-adaptive methods force coefficients to be penalized equally regardless of variable importance. However, the adaptive methods penalize coefficients differently through the working of the adaptive weights, and large coefficients of inherently important variables are penalized more lightly than the small coefficients of less important variables. Thus, the outperformance of the adaptive methods is more evident when there are strong effects (i.e., high signal-to-noise ratio). In contrast, under medium and low signal-to-noise ratios, the adaptive methods are subject to more downward selection bias on average across all the considered confidence levels.

All the considered variable selection methods perform consistently across four confidence levels, conditional on the variable correlation and the signal-to-noise ratio. Given the relatively weak variable correlation displayed in our empirical study and the superior performance of ALasso under $\rho = 0.5$, we use only ALasso as the variable selection method in the following empirical analysis.

4 Data

In this empirical analysis, we investigate the S&P500, the US macroeconomic and financial data from 1969Q1 to 2021Q2. The S&P500 index data are obtained from the Yahoo Finance, and we consider the daily stock market returns, which are calculated as the natural logarithm of the S&P500 index prices. With the macroeconomic and financial variables collected quarterly, it is necessary to align the frequency of the variables, which are recorded at daily/monthly frequencies, by taking the quarterly averages.

When considering macroeconomic and financial variables, substantial data revision can be frequent, and it has been found that using the revised instead of the real-time data may mislead the forecast evaluations (Conrad and Loch, 2015). The importance of using the real-time data has long been documented in the empirical studies. For example, Stark (2010) demonstrates that the forecast accuracy of the Survey of Professional Forecasters declines as the data are the revised over time, indicating the negative impact of using revised data. Croushore (2011) argues that forecasters generally produce forecasts based on the existing methodologies and cannot be expected to predict future changes in the methodology. Thus, forecast evaluations should focus on the early release of the data and ideally, the real-time data.

In this paper, we obtain the first-release (real-time) macroeconomic and financial data from the Real-time Data Research Center (RDRC) of the Federal Reserve Bank of Philadelphia. Other variables are collected from the FRED database at the Federal Reserve Bank of St Louis, Tradingeconomics.com, the Survey of Consumers from University of Michigan (SCUM), the Federal Reserve Bank of Chicago (FRBC), and the personal website of French and Manela. Descriptive statistics are reported in Table 3. [11] [12]

4.1 Macroeconomic variables

As discussed, low-frequency signal is integrated into tail risk forecasting in our proposed model through the channel of the long-term volatility component. We therefore include the variables that are considered important in the research on forecasting volatility (Paye, 2012; Engle et al., 2013; Asgharian et al., 2013; Conrad and Loch, 2015; Pan et al., 2017; Su et al., 2017; Conrad and Kleen, 2020; Fang et al., 2020). For example, stock market volatility forecasts are found to be improved by including industrial production growth rate and producer price index inflation rate (PPI) under the GARCH-MIDAS framework proposed by Engle et al. (2013). Unemployment rate, housing starts, and corporate profits are found to have the highest predictability for the long-term stock market volatility (Conrad and Loch (2015)).

Our proposed model enables investigating the predictability of the low-frequency variables for stock market tail risk. The macroeconomic variables used in our study are the real Gross Domestic Product growth rate (GDP), industrial production growth rate (IP), unemployment rate, housing starts, post-tax nominal corporate profits, real personal consumption, CPI, PPI, the Chicago Fed National Activity Index (CFNAI), the New Orders Index of the Institute of Supply Management, monetary base, consumer sentiment index of the University of Michigan, real GDP volatility, and CPI volatility.

The CFNAI is an index designed to gauge the overall economic activity and the related inflationary pressure. It can be seen as a proxy of business cycles and is found to complement high-frequency information in volatility forecasting (Conrad and Kleen (2020)). As another leading economic indicator, the New Orders Index provides an idea of future economic growth by measuring the changes in employment, production, inventories, supplier deliveries, and new orders. In addition, as indicated by the volatility of the first-release GDP growth rate and the CPI, macroeconomic volatility serves as an important determinant of stock market volatility (Schwert, 1989; Liljeblom and Stenius, 1997; Engle et al., 2013). Following Engle et al. (2013), we use the GARCH(1,1) model to estimate the quarterly macroeconomic volatility.

Following Conrad and Loch (2015) and Fang et al. (2020), we consider the CFNAI in levels. For the unemployment rate and the consumer sentiment index, we take the first difference of the level data. For other variables, we take the annualized quarterly percentage change as $100((X_t - X_{t-1})^4 - 1)$.

4.2 Financial variables

We include six financial variables: equity market returns (MKT), short-term reversal factor (STR), default spread, term spread, realized volatility (RV), and implied volatility (IV). As with the macroeconomic variables, these financial indicators are found to be important in forecasting volatility and are now examined under the tail risk prediction framework.

As taken from Fama and French (1992), the variable equity market returns (MKT) captures the leverage effects, that is, the negative relationship between the stock returns and the volatility (Black, 1976; Christiansen et al., 2012; Nonejad, 2017). In Nagel (2012), the expected return of the reversal strategy rises predictably and dramatically during the periods of market turmoil, thereby indicating that the short-term reversal factor (STR) is related to the stock market volatility. For the spread terms, the following proxies are used: the yield spread between the BAA and the AAA rated bonds for default spread; the yield spread between the 10-year treasury bond and the three-month treasury bill for term spread. We also consider realized volatility in forecasting tail risk. The quarterly volatility is calculated as:

$$RV_t = \sum_{i=1}^{N_t} r_{i,t}^2 \tag{12}$$

The Chicago Board Options Exchange's (CBOE) VIX and VXO are two implied volatility indices that measure the expected market volatility implied by the stock index option prices. Many studies find that the option-based volatility is more informative for forecasting purposes than the time-series volatility models based on the returns of the stock market index (Martens and Zein, 2004; Becker et al., 2009; Fernandes et al., 2014). Therefore, implied volatility is included to complement realized volatility. However, the data horizons of VIX and VXO do not match those of the other variables given their availability from 1990 and 1986, respectively. To address this problem, we use the news-based implied volatility index (NVIX) proposed by Manela and Moreira (2017) as a proxy of implied volatility before 1986. NVIX is an estimated VXO index based on the data from the front-page articles of the *Wall Street Journal*. Su et al. (2017) find that the NVIX is a source of financial aggregate volatility. We then switch to using the VXO between 1986 and 1990. CBOE VIX is used thereafter in our empirical analysis.

[INSERT TABLE 3 ABOUT HERE]

4.3 Principal Component Analysis for variables

Dimensionality reduction techniques such as the Principal Component Analysis (PCA) are common alternatives to the feature selection methods in big data analysis. As a feature selection method, Lasso focuses on dropping uninformative variables, while PCA creates a lower-dimensional representation of the original features (Jolliffe and Cadima, 2016). PCA is widely used to reduce the dimensionality of big datasets, increasing interpretability but minimizing the information loss at the same time. It does so by creating principal components that maximize data variation, therefore retaining as much the statistical information as possible. To examine the power of the lower-dimensional features represented by the principal components, we conduct the following PCA analysis.

After standardizing all variables, PCA is applied to 20 macroeconomic and financial variables. The PCA results indicate that the first, the second, and the third principal component accounts for 28.12%, 13.91%, and 11.22%, respectively, of the total variation in the 20 variables. With the first three components accounting for only 53.25% of the total variation in the data [13], PCA delivers inadequate representations of the original features. Therefore, instead of using PCA directly, we use the Adaptive Lasso to select informative variables for forecasting [14].

[INSERT TABLE 4 ABOUT HERE]

5 Empirical analysis

The out-of-sample is from 2005Q1 to 2021Q2 (66 quarters), and the estimation sample is the previous 144 quarters of each out-of-sample quarter. We evaluate the out-of-sample one-day-ahead VaR and ES forecasts for the daily log returns of the S&P 500 index at four confidence levels: 1%, 2.5%, 5%, and 10%. To generate one-day-ahead VaR and ES forecasts in one quarter, the parameter estimates and the informative low-frequency variables are obtained on a rolling-window basis. Each rolling window contains daily return observations in the previous 144 quarters, and we move the window one quarter forward after each estimation to produce the forecasts for the days in the next quarter. The entire out-of-sample produces 4158 (66 quarters) daily tail risk forecasts, and each daily prediction in each quarter is built on the selected low-frequency information over the past 12 quarters. These forecasts are then backtested using various approaches introduced in Section 5.2. Each backtest is conducted on all the out-of-sample forecasts, and we report the backtesing results in Section 5.3.

We backtest the VaR and ES forecasts of the proposed model with the selected variables and compare the performance of the proposed model with that of the following benchmark models: the semiparametric GARCH-MIDAS-X model that incorporates one variable at a time in the long-term volatility component (Conrad and Loch, 2015; Fang et al., 2020); the semiparametric GARCH-MIDAS-X model without any low-frequency variables; the semiparametric GARCH-MIDAS-X model with all 20 macroeconomic and financial variables; and the semiparametric GARCH-MIDAS-X model that incorporates the first three principal components (PC1, PC2, PC3) separately and jointly. In addition to the model integrating all the selected variables, we generate the combination forecasts based on the forecast results of the semiparametric GARCH-MIDAS-X model with each of the selected variables using the simple average combination and the loss-function-based combination (Conrad and Kleen, 2020; Happersberger et al., 2020). Descriptions of these models are presented in Table 5.

[INSERT TABLE 5 ABOUT HERE]

5.1 Dynamic variable selection using Adaptive Lasso

Using a dynamic variable selection process, the optimum tuning parameter and corresponding variables are selected on a quarterly rolling basis. This allows us to visualize the changes in the supporting power of low-frequency variables over time and provides a new perspective to add the most important low-frequency variables in predicting VaR and ES.

To select the tuning parameter λ, we search over the $[0, 30]$ range with increments of 0.1 and calculate the GIC for each value of λ. The λ with the minimum GIC in each window is considered as the optimum tuning parameter. The third panels in Figures 2 - 5 present the results of the dynamic variable selection process at four confidence levels: 1%, 2.5%, 5%, and 10%. The results are reported by time and then by confidence levels as follows.

Across the out-of-sample period, RV, term spread, and housing starts mostly serve to predict VaR and ES at all chosen confidence levels. Past market volatility, the yield spread between long-term and short-term government bonds and the volume of housing starts are predominant predictors of future tail risk. As indicated by the heat map, those three variables have a consistently strong but opposite effect on future VaR and ES. For RV, the green tabs in the heat map indicate its positive correlation with VaR and ES. Several studies have found that more accurate risk forecasts can be generated by incorporating RV into the dynamic volatility forecasting models (Giot and Laurent, 2004; Clements et al., 2008; Hansen et al., 2012; Louzis et al., 2014; Lazar and Xue, 2020). Unlike in the previous studies, we allow the data to speak for itself rather than assuming statistical models that include RV but may not fit the data well. Our study reveals that the RV is consistently selected as one of the strongest predictors of future tail risk by ALasso. However, term spread and housing starts, as indicated by their red tabs, are negatively associated with future tail risk. To the best of our knowledge, our study provides novel evidence for the negative relationship between housing starts and financial market tail risk, indicating the countercyclical pattern of tail risk. Inspired by Conrad et al. (2014) and Fang et al. (2020), the relevance of term spread and housing starts for tail risk prediction derives from the following. First, these two variables are powerful predictors of future changes in economic activity. Given that either a smaller term spread or yield curve inversion is generally followed by a recession, these are associated with increasing uncertainty and financial market risk. However, the presence of more housing starts implies future economic growth, which can be rationalized by the empirical observation of low mortgage interest rates that facilitate economic upturns and safer investment projects (Estrella and Trubin, 2006; Leamer, 2007; Kydland et al., 2012). Second, term spread is closely related to investors' expectations of investment risk. If short-term interest rates are expected to fall, this raises the price of long-term securities, thus lowering the long-term yields relative to short-term securities and indicating higher risk (Wheelock et al., 2009). However, more housing starts is associated with the expansion of the credit market, which drives economic growth and investment opportunities with lower risk. During the great recession period from 2007 to 2009, in addition to the three strongest predictors, corporate profits is also selected at all chosen confidence levels. The red tabs of corporate profits suggest its negative association with tail risk. Corporate profits has a strong impact on firm valuation and is a key determinant of financial investment decisions. Investors perceive that the firms with higher corporate profits, particularly during turbulent periods, are relatively safer investment target (Froot et al. (1993)). During the COVID-19 crisis period, RV, term spread, and housing starts dominate other variables in forecasting stock market tail risk. Although RV, term spread, and housing starts are the three strongest predictors of tail risk, the importance of other low-frequency variables should not be neglected given that many of them are also very frequently chosen by ALasso.

As shown in the second panels of Figures 2 - 5, across the confidence levels, the average number of selected variables is highest at the 1% level and lowest at the 10% level. This indicates that extreme VaR and ES rely on extra information in addition to what has been utilized in predicting less extreme tail risk. For example, at the 1% confidence level, other low-frequency variables such as unemployment rate, corporate profits, and MKT are also consistently selected for most of the rolling windows. This finding reveals the limitations of parametric methods in estimating tail risk. Under parametric methods where returns are usually assumed to follow a fixed elliptical distribution, VaR and ES can be expressed in a linear form of volatility which suggests that the selected variables at $\alpha = 10\%$ should be exactly the same as those selected at $\alpha = 1\%$. Our results show that more information is needed in predicting more extreme tail risk, which provides further evidence that the return distribution could be time-varying.

[INSERT FIGURES 2 - 5 ABOUT HERE]

5.2 Backtesting approaches

We implement six backtesting approaches to compare the performance of the out-of-sample forecasts generated by the proposed model with that generated by the benchmark models. We now introduce six prevailing backtesting methods for VaR and ES by assessing the quantile score, the unconditional coverage (UC) test, the Dynamic Quantile (DQ) regression, the bootstrap test for ES, and the Dynamic Expected Shortfall (DES) regression.

5.2.1 Unconditional coverage (UC) test

Several procedures for evaluating the performance of VaR forecasts are based on the VaR failures, that is:

$$I_t = \mathbb{1}\{r_t \leq v_t^{\alpha}\}.$$

A commonly used VaR backtesting method is the UC test proposed by Kupiec (1995). Focused on the proportion of failures, the hit proportion is defined as the percentage of the returns below the estimated VaR. The difference between the hit proportion and theoretical value α is then examined. The decision on the null hypothesis depends on the Likelihood Ratio test that is applied:

$$H_{UC}^{VaR} : \mathbb{E}_{t-1}[I_t] = \alpha.$$

5.2.2 Dynamic quantile (DQ) regression

However, applying the UC test to small samples is statistically invalid, and this test ignores the clustering of failures (Nieto and Ruiz, 2016). To address these drawbacks, the conditional coverage test is considered where the null hypothesis is:

$$H_{CC}^{VaR} : \mathbb{E}_{t-1}[I_t | I_{t-1}] = \alpha.$$

We employ the DQ test proposed by Engle and Manganelli (2004) to implement the conditional coverage test. The DQ test examines both whether the hit variable defined as $Hit_{v,t} = \mathbb{1}\{r_t \leq v_t\} - \alpha$, follows an i.i.d.

Bernoulli distribution with probability level α and whether it is independent of the VaR estimator. We include one lag of $Hit_{v,t}$ in the regression function of the test. Consider the following DQ regression:

$$Hit_{v,t} = a_0 + a_1 Hit_{v,t-1} + a_2 v_{t-1} + u_{v,t}, \tag{13}$$

where $\mathbf{a} = [a_0, a_1, a_2]$ is the set of parameters of the regression. We test whether all parameters in the set $\mathbf{a}$ are zero. The DQ test statistic follows an asymptotic $\mathcal{X}^2(3)$ distribution under the null hypothesis.

5.2.3 Quantile score function

An alternative way to assess the out-of-sample VaR forecasts is to use the quantile score function. Given its use in quantile regression, a reasonable choice for the score function is the linear piecewise loss function for VaR (Giacomini and Komunjer, 2005), which can be modified as the quantile score function expressed as:

$$S(v_t, r_t; \alpha) = (r_t - v_t)(\alpha - I_t). \tag{14}$$

Because VaR is an elicitable risk measure, the quantile score is strictly consistent for VaR. In this backtesting method, the best model is that which generates the lowest score from the above score function.

5.2.4 Bootstrap test for ES

To backtest ES, McNeil and Frey (2000) propose a bootstrap test that focuses on the discrepancies between the observed return and the ES forecast for the periods where the return exceeds the VaR forecast. Under the null hypothesis, the standardized discrepancies should have zero unconditional and conditional expectations. However, given typically small sample of the discrepancies, a test of zero conditional expectation is generally not performed. To address the problem of the sample size, McNeil and Frey (2000) employ a bootstrap test. This avoids the need for any distributional assumptions and tests for the zero unconditional mean of the VaR exceptions. Given that this test focuses on observations exceeding the VaR forecasts, the assessment of ES forecasts is not independent of the VaR forecasts.

5.2.5 Dynamic Expected Shortfall (DES) regression

We follow the backtesting method of Patton et al. (2019) to evaluate the ES estimates individually, using a DES regression test:

$$\lambda^s_{e,t} = b_0 + b_1 \lambda^s_{e,t-1} + b_2 e_{t-1} + u_{e,t}, \tag{15}$$

where $\lambda^s_{e,t}$ is the standardized version of $\lambda_{e,t}$, which is defined as:

$$\lambda^s_{e,t} = \frac{\lambda_{e,t}}{e_t} = \frac{1}{\alpha}\mathbb{1}\{r_t \leq v_t\}\frac{r_t}{e_t} - 1,$$

$\mathbf{b} = [b_0, b_1, b_2]$ is the set of parameters of the regression. Based on the null hypothesis, we test whether all the parameters in the set $\mathbf{b}$ are zero. The main intuition for this test is the same as that for the DQ regression discussed above.

5.2.6 FZ0 loss function for (VaR, ES)

To compare the VaR and ES forecasts jointly, a loss function proposed by Fissler and Ziegel (2016) is employed. The authors discuss how VaR and ES are jointly elicitable and present a group of loss functions for the estimation and backtesting of risk measures. We follow the choice of Patton et al. (2019) for the loss function FZ0, which is defined as:

$$L_{FZ0}(r, v, e; \alpha) = -\frac{1}{\alpha e}\mathbb{1}\{r \leq v\}(v - r) + \frac{v}{e} + \log(-e) - 1. \tag{16}$$

To compare the performance of each model using the FZ0 loss function, we calculate the average loss value $L_{FZ0} = \frac{1}{T}\sum_{t=1}^{T} L_{FZ0,t}$ for different α values.

5.3 Backtesting results

The out-of-sample forecasting period is from 2005Q1 to 2021Q2, which encompasses both the global financial crisis from 2007 to 2009 and the COVID-19 recession that began in early 2020. To evaluate the performance of forecasting one-day-ahead tail risk measures, we present the backtesting results at four confidence levels ($\alpha = 1\%, 2.5\%, 5\%$,and 10%) based on the above approaches in Table 6. We first report the overall findings below, followed by the detailed results of each backtesting approach.

At all the considered confidence levels except the 10% level, the loss-based combination using the forecasts generated by each selected variable and the proposed semiparametric GARCH-MIDAS-X model with all the selected variables achieve the minimum loss alternately in the joint forecasting of VaR and ES. The superiority of our proposed model is statistically significant based on the Diebold–Mariano (DM) test and the Model Confidence Set (MCS) test results. At the 10% level, the models that integrate RV and IV are the top two performers, while the proposed model incorporating all the selected variables and its loss-based combination forecasts rank third and fourth, respectively. However, the outperformance of the models integrate RV and IV are not statistically significant. Moreover, it is worth noting that the model with RV or IV performs much poorer at more extreme tail levels (e.g., at $\alpha = 1\%$) than the proposed model. This suggests that the effective low-frequency information is more valuable in forecasting more extreme tail risk (e.g., $\alpha = 1\%, 2.5\%$, and 5%) and it produces equivalently good forecasts as the model with RV or IV at less extreme tails (e.g., $\alpha = 10\%$). It also provides evidence that extreme tail risk is more associated with the macroeconomic fluctuations than with the stock market volatility. For most of the remaining backtests and most confidence levels, our proposed semiparametric GARCH-MIDAS-X model with all the selected variables and the loss-based combination forecast model outperform other benchmark models. The detailed backtesting results are now discussed.

Panel A in Table 6 presents the average loss of the VaR forecasts generated by each model. The loss-based combination forecasts model using the forecasts generated by each selected variable performs best across all the considered confidence levels. The model that integrates all the selected variables consistently ranks second. This indicates that the selected variables carry important low-frequency information that improves the VaR forecasts. However, the out-of-sample performance of the semiparametric GARCH-MIDAS-X that incorporates all 20 macroeconomic and financial variables has higher average loss. This indicates that a model without shrinkage and regularization suffers from the overfitting problem.

The hitting proportions and their corresponding p-values of the UC test are presented in Panel B of Table 6. We first count the number of VaR rejections for each model, then perform the UC backtest for all the forecasts at four confidence levels. For $\alpha = 5\%$ and $\alpha = 10\%$, the proposed semiparametric GARCH-MIDAS-X model with all the selected variables and two combination forecasts models achieve relatively fewer UC test rejections. In more extreme cases, (i.e., $\alpha = 1\%$ and $\alpha = 2.5\%$), almost all models fail to pass the UC test, having p-values less than 5%.

Panel C of Table 6 reports the p-values of the DQ test for the VaR forecasts. For $\alpha = 5\%$ and $\alpha = 10\%$, most of the models pass the DQ test. Both the proposed semiparametric GARCH-MIDAS-X model with all the selected variables and its loss-based combination forecasts based on each of the selected variables achieve a relatively large p-value at $\alpha = 10\%$. When considering the 1% and 2.5% VaR forecasting, only a few models pass the DQ test with the proposed model and its loss-based combination forecasts being the most successful. Overall, adding the selected variables generally improves the model performance in the DQ test, particularly at more extreme tail levels.

Panel D of Table 6 presents the average loss of the joint forecasting of VaR and ES based on the FZ0 loss function for all models. The model integrating all the selected variables and the loss-based combination forecast method achieve the minimum loss alternately at all considered confidence levels except at the 10% level. For $\alpha = 10\%$, the models that integrate RV and IV ranks first and second, respectively. The proposed semiparametric GARCH-MIDAS-X model with all the selected variables ranks third. This suggests that the effective low-frequency information is more valuable in forecasting more extreme tail risk (e.g., $\alpha = 1\%, 2.5\%$ and 5%) while the RV and IV provide better forecasts at less extreme tails (e.g., $\alpha = 10\%$). It also provides evidence that extreme tail risk is more associated with macroeconomic fluctuations than with the stock market volatility.

The bootstrap test results for the ES forecasting are presented in Panel E of Table 6. Overall, the semiparametric GARCH-MIDAS-X model with all the selected variables and the combination approaches pass the test with relatively larger p-values. We obtain similar results in the DES test on the ES forecasts reported in Panel F of Table 6. It is worth noting that in the most extreme case of $\alpha = 1\%$, the model with all the selected variables has the highest p-values in both tests. In general, the selected low-frequency variables improve ES forecasting, particularly at extreme tails.

[INSERT TABLE 6 ABOUT HERE]

To better visualize the model performance, the first panel of Table 7 presents the rankings of the out-of-sample forecasting performance based on the values of FZ0 loss function at four confidence levels. The best model ranks 1 while the worst ranks 29 given there are 29 competing models in total. In general, the top two models are the loss-based combination forecasts model using the forecasts generated by each selected variables and the semiparametric GARCH-MIDAS-X model with all the selected variables. The only exception is at the 10% level where these two models are outperformed by the model with RV and IV. At the 10% level, these two models rank fourth and third, respectively. Despite the models integrating RV and IV performing best at the 10% level, their ranking at the extreme tail is much lower (e.g., at $\alpha = 1\%$). Moreover, if we examine the ranking of the average loss achieved by each model across four confidence levels, we find that the proposed semiparametric GARCH-MIDAS-X model with all the selected variables and the loss-based combination forecast model still rank the top two. This indicates that the selected variables generally improve the joint forecasting of VaR and ES, and their power is stronger at more extreme tails than at less extreme ones.

[INSERT TABLE 7 ABOUT HERE]

Although the average losses of the joint forecasting of VaR and ES are a useful initial evaluation of the out-of-sample performance, it is not clear whether the forecasting gains are statistically significant. We apply the DM test to compare the out-of-sample performance of any two models. The null hypothesis of the DM test is that the column model and the row model achieve an equal average loss. Figure 6 presents the color map of the t-statistics of the DM test, which compares the average losses based on the FZ0 loss function at four confidence levels. A positive t-statistic (darker color on the map) indicates that the column model outperforms the row model. Absolute values greater than 1.96 (2.575 or 1.64) suggest that the difference of the average loss is statistically significant at the 95% (99% or 90%) confidence level. Overall, the loss-based combination forecasts model using the forecasts generated by each selected variable performs best, followed by the proposed semiparametric GARCH-MIDAS-X model with all the selected low-frequency variables [15] . The superiority of our model is statistically significant given the DM test results. It is worth noting that although the models integrating volatility indicators (RV and IV, respectively) achieves lower average loss than our proposed model at the 10% level, this difference is not statistically significant based on the DM test result.

[INSERT FIGURE 6 ABOUT HERE]

Following Taylor (2019) and Lazar and Xue (2020), we also apply the MCS test [16]. Table 8 presents the results of the 95% MCS test introduced by Hansen et al. (2011) [17]. This approach constructs the MCSs using one-sided elimination based on the DM test. The test compares models based on the FZ0 loss function using two methods: the R method, which uses the sum of absolute values in calculating the MCS test statistic, and the SQ method, which uses the summed squares. We find that the loss-based combination forecast model is the best performing model in general, and the proposed semiparametric GARCH-MIDAS-X model with all the selected low-frequency variables ranks second [18]. The outperformance achieved by our model is statistically significant based on the MCS test results. The MCS test results also suggest that the performance difference between the model integrating volatility indicators (RV and IV) and our proposed model is not statistically significant at the 10% level. Together with the DM test results, this result suggests that the power of informative low-frequency variables is stronger in predicting more extreme tail risk. At less extreme levels (e.g., $\alpha = 10\%$), the proposed model produces at least equivalently good tail risk forecasts as the model with RV and IV.

[INSERT TABLE 8 ABOUT HERE]

5.4 Robustness checks

5.4.1 Restricted Beta weighting schemes

Under the restricted Beta weighting scheme, w_1 is set to 1 to generate a decaying pattern of weights rather than the hump-shaped weights in the previous analysis. The use of restricted Beta weighting scheme may lead to new results. In this section, we fix $w_1 = 1$ to determine how this constraint changes the results with all other settings being the same as before. The Beta weighting scheme now becomes:

$$\psi_k(w_2) = \frac{(1 - k/(K+1))^{w_2 - 1}}{\sum_{s=1}^{K} (1 - s/(K+1))^{w_2 - 1}}, \tag{17}$$

In general, similar results are obtained under the restricted Beta weighting scheme. Industrial production, housing starts, term spread, and RV are predominant predictors of future tail risk. For housing starts, its red tabs in the heat map indicate that it is negatively correlated with future VaR and ES, providing evidence for the countercyclical pattern of tail risk. For term spread and RV, their positive relationship with future tail risk is shown by the green tabs in the heat map. During the great recession period from 2007 to 2009, in addition to those strongest predictors, corporate profits is also consistently selected and negatively correlated with tail risk. During the COVID-19 crisis period, housing starts, term spread, and RV dominate other low-frequency variables in forecasting tail risk. Moreover, similar to the results for the unrestricted Beta weighting schemes, the average number of selected variables is highest at the 1% confidence level and lowest at the 10% confidence level. For example, at the 1% level, in addition to the strongest predictors, other variables such as unemployment rate and MKT are also consistently selected for most of the rolling windows. However, fewer variables are selected when the confidence level increases. This indicates that more information is needed to predict more extreme VaR and ES. [19]

We conduct the same backtesting approaches as in the previous section under the restricted Beta weighting scheme. For most of the applied backtests and confidence levels, our proposed model with all the selected variables and the combination forecasts models consistently outperform other benchmark models. For the joint forecasting of VaR and ES, the loss-based combination forecasts model and the proposed semiparametric GARCH-MIDAS-X model with all the selected variables perform best alternately across all the considered confidence levels except at the 10% level. At the 10% level, the models integrating volatility indicators (e.g., RV and IV) achieve the minimum loss, and our proposed model with all the selected variables ranks third. However, the later DM and the MCS test results reveal that this outperformance is not statistically significant [20] [21]. For the individual forecasting of VaR and ES, the corresponding backtesting results also demonstrate the superiority of our proposed model by achieving either lower losses or higher p-values. [22]

The right panel of Table 7 presents the out-of-sample performance rankings for the one-day-ahead joint forecasting of VaR and ES under the restricted Beta weighting scheme based on the FZ0 loss function. The loss-based combination forecasts model using the forecasts generated by each selected variable ranks first overall, followed by the semiparametric GARCH-MIDAS-X model with all the selected variables. The model without variable selection ranks last, which indicates that simply adding the low-frequency information deteriorates the model performance while the effective variables selected by our proposed process improve the joint forecasting of VaR and ES.

6 Long-term forecasting

Engle (2011) argues that the financial crisis is predictable one day ahead, and therefore the key challenge lies in performing multi-day-ahead forecasts of tail risk. To evaluate the long-term forecasting performance of our proposed model, we consider the 1-/2-/3-/4-quarter-ahead joint forecasting of VaR and ES, following Fang et al. (2020). The estimation sample and the forecasting strategy remain the same as that in the one-day-ahead forecasts. The out-of-sample for the 2-quarter-ahead forecasts is from 2005Q2 to 2021Q2 ,and the out-of-sample for the 3-quarter-ahead forecasts is from 2005Q3 to 2021Q2 and so on.

Table 9 reports the out-of-sample average losses of the joint forecasting of VaR and ES based on the FZ0 loss function at four confidence levels. The loss-based combination forecasts model using the forecasts generated by each selected variable and the semiparametric GARCH-MIDAS-X model with all the selected variables perform best alternately across all the considered long-term forecasting horizons at all the considered confidence levels [23]. The only exception is at the 10% level, where the models incorporating market volatility indicators (RV and IV) outperform.

[INSERT TABLE 9 ABOUT HERE]

To evaluate the statistical significance of the long-term forecasting performance, Figures 7 - 10 present the DM test results while Tables 10 and 11 present the MCS test results based on the R and the SQ method, respectively. We find that the superiority of our proposed model is statistically significant across all the considered long-term forecasting horizons at all the considered confidence levels except at the 10% level. At the 10% level, the forecasts are dominated by the model that integrates volatility indicators (i.e., RV or IV). This reveals that the power of informative low-frequency information is stronger in predicting more extreme tail risk. This also suggests that more extreme tail risk is associated with macroeconomic fluctuations while less extreme tail risk is more related to the stock market volatility.

[INSERT FIGURE 7 - 10 ABOUT HERE]

[INSERT TABLE 10 and TABLE 11 ABOUT HERE]

7 Conclusions

Accurate VaR and ES forecasts enhance statistical analysis that can inform decisions taken by financial and risk managers, regulators, and market participants. Although high-frequency information is widely exploited to improve tail risk forecasts, existing models provide no apparent way to utilize low-frequency signals. To address this gap, we propose the semiparametric GARCH-MIDAS-X model, which integrates effective low-frequency information, avoids distributional assumptions on financial returns, and produces efficient VaR and ES forecasts. To select the most informative low-frequency variables, we propose an innovative approach that maximizes the penalized Asymmetric Laplace likelihood function with an Adaptive Lasso penalty. By integrating the selected variables, our proposed model achieves the minimum loss in the joint forecasting of one-day-ahead S&P500 VaR and ES, particularly at more extreme quantile levels. Our proposed model also outperforms other benchmark models when considering longer forecasting horizons. In addition, our model can be extended to include more low-frequency variables without identification problems.

Three variables, namely, realized volatility, term spread, and housing starts are consistently selected for most of the rolling windows and are the strongest predictors of future tail risk. To the best of our knowledge, uncovering the empirical relationships between housing starts and tail risk is a unique and innovative result. In examining the effect of the selected macroeconomic variables on VaR and ES, we identify new statistical evidence of the countercyclical risk measures. In addition to the strongest predictors, the value of other low-frequency variables should not be ignored because many of these variables are also very frequently selected by the Adaptive Lasso.

In conclusion, our study provides novel evidence for the value of low-frequency macroeconomic and financial variables in high-frequency tail risk forecasting. The proposed new model enables market participants and regulators to utilize broader information in practical applications.

8 Appendix

In this Appendix, we introduce the details of four variable selections methods considered in our simulation study (Simulation 2). Specifically, we consider Lasso, Adaptive Lasso (ALasso), Elastic Net (ENet), and Adaptive Elastic Net (AENet). These methods generally differ in the penalty function when selecting variables. Recall the objective function is:

$$PLLF(\Phi) = \underbrace{LLF(\Phi)}_{\text{Likelihood Function}} - \underbrace{\phi(\theta;\cdot)}_{\text{Penalty Function}} \tag{18}$$

where $PLLF(\Phi)$ is the penalized Asymmetric Laplace (AL) likelihood function, and $LLF(\Phi)$ is the original AL likelihood function without variable selection. The details of $LLF(\Phi)$ are given in Eq.(8).

For the penalty function, we have:

$$\phi(\theta;\cdot)=\begin{cases}\lambda\sum_{j=1}^{J}|\theta_j|, & \text{Lasso}\\ \lambda\sum_{j=1}^{J}\hat{w}_j|\theta_j|, & \text{ALasso}\\ \lambda_1(1-\lambda_2)\sum_{j=1}^{J}|\theta_j|+\lambda_1\lambda_2\sum_{j=1}^{J}\theta_j^2, & \text{ENet}\\ \lambda_1(1-\lambda_2)\sum_{j=1}^{J}\hat{w}_j|\theta_j|+\lambda_1\lambda_2\sum_{j=1}^{J}\theta_j^2, & \text{AENet}\end{cases} \tag{19}$$

Lasso and ALasso are l_1 penalty while ENet and AENet additionally involve the l_2 penalty. ENet overcomes Lasso's limitation in the following conditions: 1) high-dimensional case where the number of predictors is larger than the number of observations; 2) highly correlated variables where Lasso tends to select only one variable from the group and ignores the grouping effect (see Zou and Hastie (2005) for more details). The adaptive methods (ALasso and AENet) allow different shrinkage power in selecting variables through the functioning of the adaptive weights. Inherently important variables (i.e., variables with large θ) are penalized more lightly than other variables in the selection process (Zou, 2006; Zou and Zhang, 2009).

The selection of the tuning parameter for ALasso has been discussed in detail in Section 2.3. The tuning parameter for Lasso is selected in a similar way without initializing the adaptive weights. For ENet and AENet, we follow Zou and Hastie (2005) to select the tuning parameters with the objective function being the GIC in Section 2.3 [24].

Acknowledgments

Y.L. was partly funded by the Research Development Fund of Xi'an Jiaotong-Liverpool University (RDF-22-02-031). This work was also supported by the High End Computing facility at Lancaster University. We appreciate the feedback from the audience at the 14th Annual Meeting of the Society for Financial Econometrics (SoFiE) in Cambridge, UK. We are also grateful to the discussant of the 2022 Greater China Area Finance Conference in Xiamen, China. We benefit a lot from the SoFiE Summer School in 2020 and 2021. We thank Gerald Steele for offering helpful editorial suggestions on early draft of this article.

9 Reference

Alizadeh, S., Brandt, M. W., and Diebold, F. X. 2002. Range-based estimation of stochastic volatility models. *Journal of Finance*. 57(3):1047–1091.

Andersen, T. G. and Bollerslev, T. 1998. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*. 39(4):885–905.

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

Asgharian, H., Javed, F., and Hou, A. J. 2013. The importance of the macroeconomic variables in forecasting stock return variance: A GARCH-MIDAS approach. *Journal of Forecasting*. 32(7):600–612.

Basel Committee on Banking Supervision. *Minimum capital requirements for market risk*. 2019. URL `http://https://www.bis.org/bcbs/publ/d457.pdf`. Available at `https://www.bis.org/bcbs/publ/d457.pdf`.

Becker, R., Clements, A., and Mcclelland, A. 2009. The jump component of s&p 500 volatility and the vix index. *Journal of Banking and Finance*. 33(6):1033–1038.

Black, F. 1976. Studies in stock price volatility changes. *In: Proceedings of the American Statistical Association, Business and Economic Statistics Section*. pages 177–181.

Candila, V., Gallo, G. M., and Petrella, L. 2020. Using mixed-frequency and realized measures in quantile regression. *Available at SSRN 3722927*.

Cenesizoglu, T. and Timmermann, A. 2008. Is the distribution of stock returns predictable? *Available at SSRN 1107185*.

Chen, L., Pelger, M., and Zhu, J. 2020. Deep learning in asset pricing. *Available at SSRN 3350138*.

Christiansen, C., Schmeling, M., and Schrimpf, A. 2012. A comprehensive look at financial volatility prediction by economic variables. *Journal of Applied Econometrics*. 27(6):956–977.

Clements, M. P., Galvão, A. B., and Kim, J. H. 2008. Quantile forecasts of daily exchange rate returns from forecasts of realized volatility. *Journal of Empirical Finance*. 15(4):729–750.

Conrad, C. and Kleen, O. 2020. Two are better than one: Volatility forecasting using multiplicative component garch-midas models. *Journal of Applied Econometrics*. 35(1):19–45.

Conrad, C. and Loch, K. 2015. Anticipating long-term stock market volatility. *Journal of Applied Econometrics*. 30(7):1090–1114.

Conrad, C., Loch, K., and Daniel, R. 2014. On the macroeconomic determinants of long-term volatilities and correlations in u.s. stock and crude oil markets. *Journal of Empirical Finance*. 29:26–40.

Croushore, D. 2011. Frontiers of real-time data analysis. *Journal of Economic Literature*. 49(1):72–100.

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

Engle, R. F. 2011. Long-term skewness and systemic risk. *Journal of Financial Econometrics*. 9(3):437–468.

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.

Engle, R. F., Ghysels, E., and Sohn, B. 2013. Stock market volatility and macroeconomic fundamentals. *Review of Economics and Statistics*. 95(3):776–797.

Estrella, A. and Trubin, M. 2006. The yield curve as a leading indicator: Some practical issues. *Current issues in Economics and Finance*. 12(5).

Fama, E. F. and French, K. R. 1992. The cross-section of expected stock returns. *Journal of Finance*. 47(2): 427–465.

Fan, Y. and Tang, C. Y. 2013. Tuning parameter selection in high dimensional penalized likelihood. *Journal of the Royal Statistical Society: SERIES B: Statistical Methodology*. pages 531–552.

Fang, T., Lee, T.-H., and Su, Z. 2020. Predicting the long-term stock market volatility: A garch-midas model with variable selection. *Journal of Empirical Finance*. 58:36–49.

Fernandes, D., Lynch Jr, J. G., and Netemeyer, R. G. 2014. Financial literacy, financial education and downstream financial behaviors. *Management Science*. 60(8):1861–2109.

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

Francq, C. and Zakoian, J.-M. 2007. Quasi-maximum likelihood estimation in garch processes when some coefficients are equal to zero. *Stochastic Processes and their Applications*. 117(9):1265–1284.

Froot, K. A., Scharfstein, D. S., and Stein, J. C. 1993. Risk management: Coordinating corporate investment and financing policies. *the Journal of Finance*. 48(5):1629–1658.

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.

Ghysels, E., Santa-Clara, P., and Valkanov, R. The midas touch: Mixed data sampling regression models. preprint on webpage at https://rady.ucsd.edu/faculty/directory/valkanov/pub/docs/midas-touch.pdf. 2004.

Giacomini, R. and Komunjer, I. 2005. Evaluation and combination of conditional quantile forecasts. *Journal of Business & Economic Statistics*. 23(4):416–431.

Giot, P. and Laurent, S. 2004. Modelling daily value-at-risk using realized volatility and arch type models. *Journal of Empirical Finance*. 11(3):379–398.

Gu, S., Kelly, B., and Xiu, D. 2020. Empirical asset pricing via machine learning. *The Review of Financial Studies*. 33(5):2223–2273.

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

Hansen, P. R., Huang, Z., and Shek, H. H. 2012. Realized garch: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*. 27(6):877–906.

Happersberger, D., Lohre, H., and Nolte, I. 2020. Estimating portfolio risk for tail risk protection strategies. *European Financial Management*. 26(4):1107–1146.

Jolliffe, I. T. and Cadima, J. 2016. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Socirty A*. 374(2065). ISSN 1364-503X. doi: doi.org/10.1098/rsta.2015.0202.

Kupiec, P. H. 1995. Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*. 3(2):73–84. ISSN 1074-1240. doi: 10.3905/jod.1995.407942.

Kydland, F. E., Rupert, P., and Sustek, R. 2012. Housing dynamics over the business cycle. *Departmental working papers. Department of Economics, UCSB, UC Santa Barbara*.

Lazar, E. and Xue, X. 2020. Forecasting risk measures using intraday data in a generalized autoregressive score framework. *International Journal of Forecasting*. 36(3):1057–1072.

Le, T. H. 2020. Forecasting value at risk and expected shortfall with mixed data sampling. *International Journal of Forecasting*. 36(4):1362–1379.

Leamer, E. E. 2007. Housing is the business cycle. *National Bureau of Economic Research Cambridge, Mass., USA*.

Liljeblom, E. and Stenius, M. 1997. Macroeconomic volatility and stock market volatility: Empirical evidence on finnish data. *Applied Financial Economics*. 7(4):419–426.

Louzis, D. P., Xanthopoulos-Sisinis, S., and Refenes, A. P. 2014. Realized volatility models and alternative value-at-risk prediction strategies. *Economic Modelling*. 40:101–116.

Manela, A. and Moreira, A. 2017. News implied volatility and disaster concerns. *Journal of Financial Economics*. 123(1):137–162.

Martens, M. and Zein, J. 2004. Predicting financial volatility: High-frequency time-series forecasts vis-à-vis implied volatility. *Journal of Futures Markets*. 24(11):1005–1028.

Massacci, D. 2017. Tail risk dynamics in stock returns: Links to the macroeconomy and global markets connectedness. *Management Science*. 63(9):2773–3145.

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.

Meng, X. and Taylor, J. W. 2020. Estimating value-at-risk and expected shortfall using the intraday low and range data. *European Journal of Operational Research*. 280(1):191–202.

Nagel, S. 2012. Evaporating liquidity. *Review of Financial Studies*. 25(7):2005–2039.

Nieto, M. R. and Ruiz, E. 2016. Frontiers in var forecasting and backtesting. *International Journal of Forecasting*. 32(2):475–501.

Nonejad, N. 2017. Modeling and forecasting aggregate stock market volatility in unstable environments using mixture innovation regressions. *Journal of Forecasting*. 26(6):718–740.

Pan, Z., Wang, Y., Wu, C., and Yin, L. 2017. Oil price volatility and macroeconomic fundamentals: A regime switching garch-midas model. *Journal of Empirical Finance*. 43:130–142.

Pan, Z., Wang, Y., and Liu, L. 2021. Macroeconomic uncertainty and expected shortfall (and value at risk): a new dynamic semiparametric model. *Quantitative Finance*. 21(11):1791–1805.

Patton, A., Politis, D. N., and White, H. 2009. Correction to "automatic block-length selection for the dependent bootstrap" by d. politis and h. white. *Econometric Reviews*. 28(4):372–375.

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.

Paye, B. S. 2012. 'Déjà vol': Predictive regressions for aggregate stock market volatility using macroeconomic variables. *Journal of Financial Economics*. 106(3):527–546.

Schwert, G. W. 1989. Why does stock market volatility change over time? *The Journal of Finance*. 44(5): 1115–1153.

Stark, T. 2010. Realistic evaluation of real-time forecasts in the survey of professional forecasters. *Technical Report Rap Special Report, Federal Reserve Bank of Philadelphia*.

Su, Z., Fang, T., and Yin, L. 2017. The role of news-based implied volatility among us financial markets. *Economics Letters*. 157:24–27.

Taylor, J. W. 2008. Estimating value at risk and expected shortfall using expectiles. *Journal of Financial Econometrics*. 6(2):231–252.

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

Wheelock, D. C., Wohar, M. E., et al. 2009. Can the term spread predict output growth and recessions? a survey of the literature. *Federal Reserve Bank of St. Louis Review*. 91(5 Part 1):419–440.

Xu, Y., Wang, X., and Liu, H. 2021. Quantile-based garch-midas: Estimating value-at-risk using mixed-frequency information. *Finance Research Letters*. 43:101965.

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

Zou, H. 2006. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*. 101 (476):1418–1429.

Zou, H. and Hastie, T. 2005. Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*. 67(2):301–320.

Zou, H. and Zhang, H. H. 2009. On the adaptive elastic-net with a diverging number of parameters. *Annals of statistics*. 37(4):1733.

Şener, E., Baronyan, S., and Mengütürk, L. A. 2012. Ranking the predictive performances of value-at-risk estimation methods. *International Journal of Forecasting*. 28:849–873.

10 Footnotes

1. A risk measure is elicitable if the correct forecast of the measure is the unique minimizer of the expectation of at least one scoring function (Fissler and Ziegel, 2016). Such scoring functions are considered strictly consistent for the risk measure.

2. To ensure covariance stationarity, two assumptions need to be made according to Conrad and Kleen (2020): (1) $\epsilon_{i,t}$ is identically and independently distributed innovations, and it follows an unknown distribution with zero mean and unit variance, and $1 < \kappa = E[\epsilon_{i,t}^4] < \infty$; (2) $\beta_1 > 0$, $\beta_2 \geq 0$, $\beta_1 + \gamma > 0$, and $\beta_1 + \gamma/2 + \beta_2 < 1$. Moreover, the parameters satisfy the condition $(\beta_1 + \gamma/2)^2\kappa + 2(\beta_1 + \gamma/2)\beta_2 + \beta_2^2 < 1$. These two assumptions imply that the $g_{i,t}$ is a covariance stationary GJR-GARCH(1,1) process and $E[g_{i,t}] = 1$.

3. The Quasi-maximum likelihood estimation could be used, under mild conditions, when the conditional distribution of the errors is unknown. See Francq and Zakoian (2007) for details.

4. Taylor (2019) demonstrates that the result achieved by maximizing the AL density function is consistent with the one obtained by minimizing the FZ0 loss function with the assumption of zero-mean returns.

5. We consider and compare various variable selection methods in Simulation part. As is discussed in the simulation results, we only consider Adaptive Lasso in the empirical analysis.

6. The linear constraints are $\beta_1 > 0$, $\beta_2 \geq 0$, $\beta_1 + \gamma > 0$, and $\beta_1 + \gamma/2 + \beta_2 < 1$.

7. Following Engle et al. (2013), we estimate the Beta weights for each of the considered low-frequency variable up to 24 MIDAS lag quarters at four confidence levels. We find that the lag weights decay to zero at around 10 MIDAS lag quarters for all variables at all the considered confidence levels. Hence, any K that is greater than 10 is an acceptable choice of MIDAS lag to capture reasonable dynamics of low-frequency variables. To ensure comparable simulation and empirical results with other papers, we set $K = 12$, which is consistent with, for example, Conrad and Loch (2015), Conrad and Kleen (2020), and Fang et al. (2020). We report the estimated Beta weights for each of the considered low-frequency variable at the 5% level in Figure A.1. Other results are available upon request.

8. Supplementary Appendix Table A.1 and A.2 display the results for the standard Normal and the Student's t innovations, respectively.

9. The detailed method of estimating the asymptotic covariance matrix can be found in Patton et al. (2019) and Pan et al. (2021).

10. Results for the standard Normal and the Student's t innovations are available upon request.

11. Tradingeconomics.com is a database that provides economic and financial data.

12. Variable correlation matrix is provided in Supplementary Appendix Table A.3.

13. More detailed results can be found in Table 4.

14. However, we incorporate the first three principal components into the proposed model as benchmarks.

15. The proposed semiparametric GARCH-MIDAS-X model with all the selected low-frequency variables performs best if we do not consider any combination forecasting techniques. We exclude two combination forecasting methods and report the DM test results in the Appendix Figure A.2. We have 27 models in this case with the Model 27 being the proposed semiparametric GARCH-MIDAS-X model integrating all the selected low-frequency variables.

16. To consider possible persistence of the loss differences in the MCS test, we choose the optimal block length using the method of Patton et al. (2009) where the autocorrelation of loss differences is considered for each of the out-of-sample forecasting horizons.

17. Details can be found on page 465 in Hansen et al. (2011); the MATLAB code for the MCS test can be downloaded from `https://github.com/bashtage/mfe-toolbox/`.

18. Similar to the previous DM test, we report the MCS test results without two combination forecasting methods in Appendix Table A.4. Our proposed semiparametric GARCH-MIDAS-X model with all the selected low-frequency variables still performs best when not considering the forecast combination techniques.

19. Dynamic variable selection results for $w_1 = 1$ at four confidence levels are provided in Supplementary Appendix Figures A.3 - A.6.

20. The MCS test results with fixed w_1 are reported in Supplementary Appendix Table A.6. The corresponding DM test results are shown in Figure A.7.

21. Under fixed w_1, the proposed semiparametric GARCH-MIDAS-X model with all the selected low-frequency variables performs best when we do not consider any combination forecasting techniques. We exclude two combination forecasting methods and report the MCS and DM test results in Supplementary Appendix Table A.7 and Figure A.8, respectively.

22. Backtesting results for $w_1 = 1$ are provided in Supplementary Appendix Table A.5.

23. In long-term forecasting, the proposed semiparametric GARCH-MIDAS-X model with all the selected low-frequency variables performs best when we do not consider any combination forecasting techniques. We exclude two combination forecasting methods and report the MCS and DM test results in Supplementary Appendix Table A.8 - A.9 and Figure A.9 - A.12, respectively.

24. We select λ_1 over the [1,15] range with increments of 1 and λ_2 over the [0,1] range with increments of 0.1. Adaptive weights in AENet are calculated in the same way as in ALasso.

Table 1: Simulation results for the skew t innovations

	T = 2520						T = 5040					
	β_1	β_2	a_α	c_α	θ	w_2	β_1	β_2	a_α	c_α	θ	w_2
	$\alpha = 0.01$											
TRUE	0.060	0.910	-3.289	0.730	0.300	4.000	0.060	0.910	-3.289	0.730	0.300	4.000
Median	0.063	0.897	-3.289	0.757	0.301	4.000	0.062	0.902	-3.290	0.742	0.301	4.000
Avg bias	0.021	-0.057	-0.019	0.025	0.036	0.002	0.011	-0.029	-0.013	0.011	0.014	0.003
St dev	0.070	0.142	0.243	0.058	1.087	0.992	0.043	0.082	0.166	0.037	0.687	0.960
Coverage	0.916	0.900	0.917	0.925	0.883	0.901	0.914	0.918	0.920	0.950	0.909	0.911
	$\alpha = 0.025$											
TRUE	0.060	0.910	-2.408	0.695	0.300	4.000	0.060	0.910	-2.408	0.695	0.300	4.000
Median	0.061	0.898	-2.409	0.708	0.301	4.000	0.061	0.906	-2.405	0.699	0.301	4.000
Avg bias	0.010	-0.036	-0.016	0.012	0.033	0.002	0.004	-0.013	0.005	0.004	0.014	0.001
St dev	0.046	0.099	0.212	0.039	0.879	2.112	0.025	0.043	0.120	0.025	0.454	2.522
Coverage	0.954	0.959	0.955	0.952	0.878	0.960	0.955	0.958	0.959	0.949	0.950	0.947
	$\alpha = 0.05$											
TRUE	0.060	0.910	-1.801	0.651	0.300	4.000	0.060	0.910	-1.801	0.651	0.300	4.000
Median	0.060	0.902	-1.802	0.658	0.300	4.000	0.060	0.906	-1.799	0.654	0.300	4.000
Avg bias	0.007	-0.030	-0.007	0.006	0.014	0.001	0.002	-0.010	0.001	0.002	0.008	0.000
St dev	0.042	0.091	0.192	0.031	0.917	3.258	0.021	0.036	0.111	0.019	0.351	5.432
Coverage	0.950	0.926	0.864	0.914	0.880	0.927	0.959	0.962	0.909	0.948	0.970	0.904
	$\alpha = 0.1$											
TRUE	0.060	0.910	-1.224	0.577	0.300	4.000	0.060	0.910	-1.224	0.577	0.300	4.000
Median	0.060	0.901	-1.224	0.580	0.302	4.000	0.060	0.906	-1.226	0.577	0.301	4.000
Avg bias	0.005	-0.022	-0.021	0.003	0.055	0.014	0.001	-0.007	-0.017	0.001	0.007	0.004
St dev	0.033	0.069	0.186	0.023	0.866	7.722	0.019	0.027	0.151	0.016	0.344	13.102
Coverage	0.916	0.928	0.896	0.948	0.886	0.840	0.930	0.930	0.909	0.951	0.977	0.780

Note: This table presents the results of 2,000 Monte Carlo repetitions of the parameter estimation of (VaR, ES) from the proposed semiparametric GARCH-MIDAS-X model with the skew t innovations and the restricted Beta weighting scheme. The first row of each panel reports the true values of the parameters. The median, the average bias, and the standard deviation (across simulations) of the estimated parameters are presented in the second, the third, and the fourth row of each panel, respectively. The fifth row of each panel reports the coverage ratio of the 95% confidence intervals constructed by the estimated standard errors.

Table 2: Simulation results for variable selection methods

$\alpha = 0.01$																		
	$\rho = 0.5$									$\rho = 0.99$								
	$\zeta = 1.5$			$\zeta = 1$			$\zeta = 0.5$			$\zeta = 1.5$			$\zeta = 1$			$\zeta = 0.5$		
	C	C1	I	C	C1	I	C	C1	I	C	C1	I	C	C1	I	C	C1	I
TRUE	3	3	0	3	3	0	3	3	0	3	3	0	3	3	0	3	3	0
Lasso	3.5 (4.33)	3 (2.95)	1 (1.38)	4 (4.82)	3 (2.95)	1 (1.87)	4 (5.20)	3 (2.75)	1 (2.45)	4 (4.82)	3 (2.95)	1 (1.87)	4 (5.03)	3 (2.98)	1 (2.05)	5 (5.68)	3 (2.89)	2 (2.79)
ALasso	3 (3.03)	3 (2.97)	0 (0.06)	3 (3.08)	3 (2.91)	0 (0.17)	3 (3.20)	3 (2.65)	0 (0.55)	3 (3.08)	3 (2.75)	0 (0.33)	4 (4.72)	3 (2.82)	1 (0.90)	4 (4.89)	3 (2.85)	1 (2.04)
ENet	4 (4.28)	3 (2.99)	1 (1.29)	4 (4.56)	3 (2.99)	1 (1.57)	4 (5.14)	3 (2.89)	1 (2.25)	4 (4.95)	3 (2.99)	1 (1.96)	4 (5.18)	3 (2.96)	1 (2.12)	5 (5.83)	3 (2.91)	2 (2.92)
AENet	3 (4.21)	3 (2.90)	0 (1.31)	3 (4.35)	3 (2.98)	0 (1.37)	3 (5.03)	3 (2.83)	0 (2.20)	3 (3.68)	3 (2.98)	0 (0.70)	3 (4.03)	3 (2.90)	0 (1.13)	3 (4.26)	3 (2.93)	1 (1.33)

$\alpha = 0.025$																		
	$\rho = 0.5$									$\rho = 0.99$								
	$\zeta = 1.5$			$\zeta = 1$			$\zeta = 0.5$			$\zeta = 1.5$			$\zeta = 1$			$\zeta = 0.5$		
	C	C1	I	C	C1	I	C	C1	I	C	C1	I	C	C1	I	C	C1	I
TRUE	3	3	0	3	3	0	3	3	0	3	3	0	3	3	0	3	3	0
Lasso	4 (4.27)	3 (2.95)	1 (1.32)	4 (4.81)	3 (2.96)	1 (1.85)	4 (4.84)	3 (2.88)	1 (1.96)	4 (4.96)	3 (2.90)	1 (2.06)	4 (5.08)	3 (2.91)	1 (2.17)	5 (5.56)	3 (2.93)	2 (2.63)
ALasso	3 (3.04)	3 (2.95)	0 (0.09)	3 (3.07)	3 (2.93)	0 (0.14)	3 (3.35)	3 (2.75)	0 (0.60)	3 (2.91)	3 (2.90)	0 (0.01)	4 (4.82)	3 (2.99)	1 (1.83)	4 (4.93)	3 (2.94)	1 (1.99)
ENet	4 (4.39)	3 (2.97)	1 (1.42)	4 (4.57)	3 (2.94)	1 (1.63)	4 (5.21)	3 (2.90)	1 (2.31)	5 (5.12)	3 (2.92)	2 (2.20)	5 (5.29)	3 (2.94)	2(2.35)	5 (5.68)	3 (2.86)	2 (2.82)
AENet	3 (4.23)	3 (2.95)	0 (1.38)	3 (4.39)	3 (2.97)	0 (1.42)	3 (5.12)	3 (2.90)	0 (2.22)	3 (3.89)	3 (2.95)	0 (0.94)	3 (4.01)	3 (2.94)	0 (1.07)	4 (4.76)	3 (3.00)	1 (1.76)

$\alpha = 0.05$																		
	$\rho = 0.5$									$\rho = 0.99$								
	$\zeta = 1.5$			$\zeta = 1$			$\zeta = 0.5$			$\zeta = 1.5$			$\zeta = 1$			$\zeta = 0.5$		
	C	C1	I	C	C1	I	C	C1	I	C	C1	I	C	C1	I	C	C1	I
TRUE	3	3	0	3	3	0	3	3	0	3	3	0	3	3	0	3	3	0
Lasso	4 (4.25)	3 (2.94)	1 (1.31)	4 (4.85)	3 (2.97)	1 (1.88)	4 (5.03)	3 (2.85)	1 (2.18)	4 (5.03)	3 (2.97)	1 (1.06)	4 (5.29)	3 (2.93)	1 (1.36)	5 (5.77)	3 (2.89)	2 (2.86)
ALasso	3 (3.15)	3 (2.93)	0 (0.22)	3 (3.20)	3 (2.93)	0 (0.27)	3 (3.54)	3 (2.71)	0 (0.83)	3 (3.07)	3 (2.99)	0 (0.08)	3.5 (4.62)	3 (2.93)	1 (1.69)	4 (4.69)	3 (2.84)	1 (1.85)
ENet	4 (4.52)	3 (2.93)	1 (1.69)	4 (4.86)	3 (2.95)	1 (1.91)	4 (5.28)	3 (2.92)	1 (2.36)	5 (5.16)	3 (2.99)	2 (2.17)	5 (5.58)	3 (2.90)	2 (2.68)	5 (5.87)	3 (3.82)	2 (2.05)
AENet	3 (4.35)	3 (2.86)	0 (1.49)	3 (4.86)	3 (2.94)	0 (1.92)	3 (5.18)	3 (2.80)	0 (2.38)	3 (3.95)	3 (2.96)	0 (0.99)	3 (4.15)	3 (2.96)	0 (1.19)	4 (4.85)	3 (2.98)	1 (0.87)

$\alpha = 0.1$																		
	$\rho = 0.5$									$\rho = 0.99$								
	$\zeta = 1.5$			$\zeta = 1$			$\zeta = 0.5$			$\zeta = 1.5$			$\zeta = 1$			$\zeta = 0.5$		
	C	C1	I	C	C1	I	C	C1	I	C	C1	I	C	C1	I	C	C1	I
TRUE	3	3	0	3	3	0	3	3	0	3	3	0	3	3	0	3	3	0
Lasso	4 (4.23)	3 (2.96)	1 (1.27)	4 (4.77)	3 (2.97)	1 (1.80)	4 (4.96)	3 (2.93)	1 (2.03)	4 (5.02)	3 (2.98)	1 (2.04)	5 (5.37)	3 (2.95)	2 (2.42)	5 (5.86)	3 (2.87)	2 (2.99)
ALasso	3 (3.28)	3 (2.92)	0 (0.36)	3 (3.23)	3 (2.97)	0 (0.26)	3 (3.65)	3 (2.68)	0 (0.97)	3 (3.08)	3 (2.75)	0 (0.33)	3 (3.39)	3 (2.94)	0 (0.45)	3 (3.44)	3 (2.99)	0 (0.45)
ENet	4 (4.85)	3 (2.92)	1 (1.93)	4 (5.03)	3 (2.96)	1 (1.07)	5 (5.39)	3 (2.91)	2 (2.48)	5 (5.27)	3 (2.92)	2 (2.35)	5 (5.53)	3 (2.91)	2 (2.62)	5 (5.96)	3 (2.94)	2 (3.02)
AENet	3 (4.52)	3 (2.93)	0 (1.59)	4 (4.96)	3 (2.93)	1 (2.03)	4 (5.28)	3 (2.85)	1 (2.43)	3 (4.02)	3 (2.95)	0 (1.07)	3 (4.25)	3 (2.98)	0 (1.27)	3.5 (4.17)	3 (2.95)	1 (1.22)

Note: This table reports the simulation results with the skew t innovations for four variable selection methods: Lasso (Lasso), Adaptive Lasso (ALasso), Elastic Net (ENet), and Adaptive Elastic Net (AENet). Performance is measured in three dimensions: number of chosen variables (C), number of correctly chosen variables (C1), and number of incorrectly chosen variables (I). Both the median and the mean performance based on 2,000 replications for each variable selection methods are reported (mean performance in the bracket). Four confidence levels are considered: $\alpha \in \{0.01, 0.025, 0.05, 0.10\}$. We also consider medium ($\rho = 0.5$) and high variable correlation ($\rho = 0.99$) as well as low ($\zeta = 0.5$), medium ($\zeta = 1$), and high ($\zeta = 1.5$) signal-to-noise ratios.

Table 3: Summary statistics

Variable	Obs.	Min.	Med.	Max.	Mean	Std.	Skew.	Kurt.	Database
Stock market data									
S&P 500 returns	13240	-22.90	0.05	10.96	0.03	1.08	-1.02	25.21	DataStream
Macroeconomic data									
Real GDP	210	-32.90	2.53	33.08	2.38	4.41	-1.15	30.07	RDRC
Industrial production	210	-25.71	3.03	18.81	2.09	6.23	-1.22	4.19	RDRC
Unemployment rate	210	-4.20	-0.07	9.20	0.01	0.80	6.67	87.17	RDRC
Housing starts	210	-76.50	4.20	304.56	8.48	46.89	2.28	10.72	RDRC
Corporate profits	210	-88.01	8.59	407.35	12.20	40.24	5.53	48.98	FRED
Personal consumption	210	-34.61	3.07	40.70	2.92	4.70	-0.20	38.60	RDRC
CPI	210	-10.84	3.42	16.76	4.00	3.53	0.61	2.57	FRED
PPI	210	-37.79	3.41	31.60	3.94	7.98	-0.10	4.81	FRED
CFNAI	210	-3.62	0.07	2.01	-0.03	0.89	-1.59	4.44	FRBC
New orders	210	27.27	56.08	71.90	55.10	7.52	-0.81	1.27	Trading
Moneytary base	210	-19.31	6.48	605.97	11.97	46.93	10.68	128.23	FRED
Consumer sentiment	210	-22.57	-0.10	16.27	-0.03	5.31	-0.27	1.83	SCUM
Real GDP volatility	210	2.78	3.99	31.16	4.64	2.94	6.55	53.73	RDRC [a]
CPI volatility	210	2.56	3.96	13.79	4.72	2.22	1.87	3.40	RDRC [a]
Financial data									
MKT	210	-9.72	1.00	7.29	0.57	2.94	-0.68	0.94	French
STR	210	-8.66	0.49	7.66	0.47	1.89	-0.15	3.98	French
Default spread	210	0.56	0.96	3.02	1.08	0.43	1.81	4.24	FRED [a]
Term spread	210	-1.43	1.69	3.80	1.62	1.21	-0.38	-0.57	RDRC
RV	210	8.14	42.88	1143.58	73.48	123.40	6.14	44.16	DataStream [a]
IV	210	10.31	20.43	58.58	20.33	6.11	2.10	9.80	FRED & Manela

Note: This table reports descriptive statistics for daily S&P 500 returns and quarterly macroeconomic and financial variables. Descriptive statistics includes number of observations (Obs.), minimum (Min.), maximum (Max.), mean (Mean.), standard deviation (Std.), Skewness (Skew.), and Kurtosis (Kurt.).

[a] Indicates that the variable is computed by authors using the data from the stated database.

Table 4: Principal Component Analysis for 20 macroeconomic and financial variables

Component	Explanatory power	Cumulative sum
1	28.12%	28.12%
2	13.91%	42.03%
3	11.22%	53.25%
4	9.34%	62.59%
5	6.06%	68.65%
6	4.72%	73.37%
7	4.58%	77.96%
8	3.73%	81.69%
9	3.67%	85.36%
10	2.99%	88.35%
11	2.45%	90.79%
12	2.04%	92.84%
13	1.88%	94.72%
14	1.44%	96.16%
15	1.11%	97.27%
16	0.83%	98.10%
17	0.63%	98.74%
18	0.61%	99.34%
19	0.38%	99.72%
20	0.28%	100.00%

Note: This table reports the results of Principal Component Analysis for the considered low-frequency variables. The second column shows the explanatory power of each principal component, and the third column presents the cumulative sum of the explanatory power.

Table 5: Description of competing models

Model name	Description
Variable name	semiparametric GARCH-MIDAS-X model that incorporates one variable at a time
None	semiparametric GARCH-MIDAS-X model without any low-frequency variables
All	semiparametric GARCH-MIDAS-X model with all 20 macroeconomic & financial variables
PC1	semiparametric GARCH-MIDAS-X model with the first principal component PC1 only
PC2	semiparametric GARCH-MIDAS-X model with the second principal component PC2 only
PC3	semiparametric GARCH-MIDAS-X model with the third principal component PC3 only
PC1-3	semiparametric GARCH-MIDAS-X model with the first three principal components
Lasso	semiparametric GARCH-MIDAS-X model with all the selected variables
Combine1	Simple average combination forecasts using the forecasts generated by each selected variables
Combine2	Loss-based combination forecasts using the forecasts generated by each selected variables

Note: This table shows the detailed descriptions of all competing models. The first column reports the model name of all models considered in the backtests, and the second column presents the model description.

Table 6: Out-of-sample backtesting results for estimated risk measures (VaR, ES) of the S&P 500 index

	Panel A: Avergae loss (VaR)				Panel B: Hit proportion (VaR)				Panel C: DQ p-values (VaR)			
	1%	2.50%	5%	10%	1%	2.50%	5%	10%	1%	2.50%	5%	10%
Real GDP	0.036	0.075	0.125	0.201	1.61%	3.29%	**5.48%**	**9.96%**	0.025	*0.062*	**0.449**	**0.272**
IP	0.036	0.075	0.125	0.201	1.61%	3.39%	**5.51%**	**10.08%**	0.029	0.032	**0.459**	**0.265**
dunemp	0.036	0.076	0.126	0.204	1.64%	3.29%	**5.56%**	**10.10%**	0.033	*0.081*	**0.506**	**0.846**
dhous	0.036	0.074	**0.124**	**0.200**	1.52%	3.17%	**5.48%**	**10.22%**	0.066	**0.118**	**0.451**	**0.136**
dcprof	0.036	0.076	0.126	0.200	1.73%	3.58%	5.80%	**10.39%**	0.010	0.010	**0.153**	**0.295**
drcons	0.036	0.075	0.125	0.201	1.49%	3.20%	**5.36%**	**9.84%**	**0.104**	**0.147**	**0.652**	**0.408**
CPI	0.036	0.075	0.125	0.200	1.37%	3.13%	**5.27%**	**9.69%**	**0.180**	**0.118**	**0.577**	**0.186**
PPI	0.036	0.076	0.125	0.200	1.52%	3.44%	*5.58%*	**10.08%**	*0.077*	0.025	**0.412**	**0.269**
CFNAI	0.036	0.074	0.125	0.201	1.49%	3.15%	**5.32%**	**9.96%**	*0.081*	**0.122**	**0.499**	**0.226**
new orders	0.036	0.075	0.125	0.200	1.68%	3.39%	5.77%	**10.22%**	0.008	0.037	**0.118**	**0.204**
monetary base	**0.035**	0.075	0.125	0.200	1.39%	3.10%	**5.32%**	**9.86%**	**0.213**	**0.176**	**0.817**	**0.296**
consumer sentiment	0.036	0.074	0.125	0.200	1.56%	3.34%	**5.56%**	**10.08%**	0.036	0.040	**0.485**	**0.172**
GDP_vol	0.036	0.075	0.125	0.201	1.54%	3.20%	**5.27%**	**10.03%**	0.052	**0.103**	**0.582**	**0.326**
CPI_vol	0.036	0.075	0.125	0.200	1.56%	3.34%	**5.34%**	**9.96%**	0.022	0.043	**0.575**	**0.139**
MKT	0.037	0.076	0.125	0.200	1.76%	3.39%	**5.48%**	**10.05%**	0.007	0.040	**0.447**	**0.231**
STR	0.036	0.075	0.125	0.200	1.47%	3.32%	**5.29%**	**9.88%**	**0.107**	0.048	**0.584**	**0.337**
Dspread	0.036	0.075	0.125	0.200	1.52%	3.34%	**5.53%**	**9.93%**	0.037	0.037	**0.184**	**0.187**
Tspread	0.037	0.076	0.126	0.201	1.59%	3.32%	**5.48%**	**9.88%**	0.038	0.051	0.034	**0.168**
RV	0.038	0.075	0.126	0.203	1.68%	3.01%	**5.53%**	**9.88%**	0.020	**0.406**	**0.501**	**0.819**
IV	0.037	0.075	0.126	0.202	1.66%	3.29%	5.94%	**10.46%**	0.020	*0.069*	0.043	**0.294**
None	0.036	0.076	0.126	0.201	1.71%	3.56%	5.77%	**10.00%**	0.014	0.012	0.049	**0.274**
All	0.052	0.164	0.141	0.217	**1.23%**	**2.77%**	**5.00%**	**9.76%**	0.021	0.000	0.010	*0.061*
PC1	0.036	0.076	0.125	0.201	1.56%	3.46%	**5.51%**	**10.15%**	0.058	0.022	**0.393**	**0.425**
PC2	0.036	0.076	0.125	0.200	1.49%	3.49%	**5.53%**	**9.98%**	**0.109**	0.021	**0.495**	**0.209**
PC3	0.036	0.075	0.125	0.200	1.59%	3.46%	*5.60%*	**10.03%**	0.048	0.022	**0.351**	**0.257**
PC1-3	0.036	0.076	0.125	0.200	1.52%	3.44%	*5.58%*	**9.98%**	*0.097*	0.024	**0.379**	**0.282**
Lasso	**0.034**	**0.072**	**0.123**	**0.200**	**1.18%**	*2.98%*	*5.68%*	**10.20%**	**0.309**	**0.247**	0.037	**0.330**
Combine1	0.036	**0.074**	0.124	0.201	1.56%	3.34%	*5.65%*	**10.10%**	0.038	0.040	*0.057*	**0.326**
Combine2	**0.033**	**0.072**	**0.123**	**0.200**	1.49%	3.05%	*5.63%*	**10.20%**	0.024	**0.109**	**0.237**	**0.390**

(Continued on the next page)

(Continued) Out-of-sample backtesting results for estimated risk measures (VaR, ES) of the S&P 500 index

	Panel D: Avergae loss (VaR, ES)				Panel E: Bootstrap p-values (ES)				Panel F: DES p-values (ES)			
	1%	2.50%	5%	10%	1%	2.50%	5%	10%	1%	2.50%	5%	10%
Real GDP	1.188	0.982	0.787	0.551	**0.999**	**0.127**	**0.192**	**0.147**	*0.084*	**0.112**	**0.440**	**0.610**
IP	1.188	0.986	0.785	0.552	**1.000**	**0.165**	**0.147**	**0.293**	*0.086*	*0.059*	**0.405**	**0.575**
dunemp	1.184	0.980	0.782	0.552	**1.000**	**0.166**	**0.197**	**0.233**	*0.098*	**0.127**	**0.440**	**0.853**
dhous	1.188	0.978	0.777	0.547	**0.999**	**0.415**	**0.124**	**0.169**	**0.131**	**0.162**	**0.389**	**0.415**
dcprof	1.204	0.999	0.804	0.554	**0.997**	**0.302**	**0.224**	**0.179**	0.027	0.016	**0.129**	**0.348**
drcons	**1.177**	0.980	0.786	0.548	**0.998**	**0.188**	**0.166**	**0.306**	**0.230**	**0.216**	**0.602**	**0.754**
CPI	1.189	0.989	0.789	0.553	**1.000**	**0.165**	**0.166**	**0.218**	**0.348**	**0.196**	**0.645**	**0.637**
PPI	1.204	0.998	0.789	0.551	**0.994**	**0.215**	*0.075*	**0.162**	**0.154**	0.044	**0.322**	**0.542**
CFNAI	1.178	0.970	0.779	0.548	**1.000**	*0.094*	**0.205**	**0.343**	**0.201**	**0.207**	**0.539**	**0.540**
new orders	1.194	0.982	0.787	0.550	**1.000**	**0.269**	**0.369**	**0.178**	0.030	*0.058*	**0.147**	**0.439**
monetary base	1.181	0.978	0.792	0.550	**0.979**	**0.300**	**0.110**	**0.206**	**0.320**	**0.244**	**0.677**	**0.716**
consumer sentiment	1.182	0.974	0.781	0.549	**1.000**	*0.099*	**0.156**	**0.259**	*0.096*	*0.082*	**0.416**	**0.442**
GDP_vol	1.207	0.990	0.793	0.553	**1.000**	**0.190**	**0.124**	**0.206**	**0.124**	**0.158**	**0.612**	**0.622**
CPI_vol	1.210	0.995	0.791	0.551	**1.000**	**0.224**	*0.082*	**0.290**	*0.070*	*0.074*	**0.549**	**0.545**
MKT	1.230	1.006	0.793	0.552	**0.999**	**0.395**	*0.080*	**0.243**	0.023	*0.057*	**0.408**	**0.545**
STR	1.196	0.983	0.784	0.548	**1.000**	*0.091*	**0.170**	**0.291**	**0.210**	**0.103**	**0.604**	**0.734**
Dspread	1.201	0.983	0.782	0.548	**0.999**	**0.183**	**0.333**	**0.113**	**0.102**	*0.070*	**0.261**	**0.476**
Tspread	1.205	0.989	0.789	0.555	**1.000**	**0.414**	**0.366**	0.047	*0.093*	*0.086*	*0.095*	**0.362**
RV	1.194	**0.957**	**0.772**	**0.543**	**1.000**	**0.320**	**0.487**	**0.122**	*0.057*	**0.479**	**0.501**	**0.963**
IV	1.194	0.969	0.777	**0.545**	**0.967**	**0.545**	**0.333**	0.015	0.048	*0.095*	*0.058*	**0.281**
None	1.217	0.998	0.794	0.553	**0.997**	**0.324**	**0.333**	0.043	0.037	0.022	*0.082*	**0.542**
All	1.287	1.103	0.852	0.588	**0.998**	**0.727**	**0.237**	**0.635**	0.030	0.000	0.011	*0.054*
PC1	1.201	0.992	0.787	0.550	**0.997**	**0.133**	**0.124**	**0.201**	**0.119**	0.046	**0.357**	**0.587**
PC2	1.203	0.994	0.788	0.550	**0.985**	**0.125**	**0.112**	**0.221**	**0.168**	0.043	**0.408**	**0.554**
PC3	1.197	0.992	0.790	0.551	**0.999**	**0.141**	**0.221**	**0.248**	**0.107**	0.046	**0.344**	**0.590**
PC1-3	1.205	1.009	0.787	0.552	**0.991**	**0.167**	**0.227**	**0.125**	**0.157**	0.043	**0.348**	**0.599**
Lasso	**1.083**	**0.917**	**0.756**	**0.546**	**1.000**	**0.228**	**0.505**	**0.194**	**0.424**	**0.336**	*0.072*	**0.499**
Combine1	1.177	0.965	0.776	0.551	**1.000**	**0.114**	**0.488**	**0.172**	**0.104**	*0.083*	**0.122**	**0.558**
Combine2	**1.066**	**0.912**	**0.763**	0.547	**1.000**	0.025	**0.311**	**0.196**	**0.139**	**0.243**	**0.278**	**0.553**

Note: This table presents the out-of-sample backtesting results of forecasting one-day-ahead tail risk measures. Panel A and D respectively present the average loss using the quantile score function for VaR and the FZ0 loss function for (VaR, ES) with confidence levels: 1%, 2.5%, 5%, and 10%. The three lowest average loss in each column are highlighted in **bold**. Panel B shows the hit proportion results for the VaR forecasts. We **bold** the results if the p-values of the subsequent UC test are greater than 0.1 and *italicize* those between 0.05 and 0.1. Panel C, E, and F present the p-values for the DQ regression test for VaR, bootstrap test for ES, and DES regression test for ES, respectively. The p-values that are greater than 0.1 are in **bold**, and values between 0.05 and 0.1 are in *italics*.

Table 7: Out-of-sample performance ranking for various levels of α

	Average loss: (VaR, ES)						Average loss: (VaR, ES) with fixed w_1					
	1%	2.50%	5%	10%	Average	Rank	1%	2.50%	5%	10%	Average	Rank
Real GDP	9	13	16	19	14.3	12	18	10	16	13	14.3	13
IP	11	16	12	22	15.3	13	21	11	14	14	15.0	14
dunemp	8	11	9	23	12.8	10	8	8	8	16	10.0	8
dhous	10	9	5	5	7.3	8	5	6	7	8	6.5	7
dcprof	21	26	28	27	25.5	26	28	28	25	20	25.3	28
drcons	**3**	10	13	7	8.3	9	**3**	7	9	6	6.3	6
CPI	12	17	21	26	19.0	17	15	22	21	23	20.3	21
PPI	22	24	19	17	20.5	22	19	23	23	22	21.8	22
CFNAI	5	6	7	9	6.8	5	6	4	5	4	4.8	5
new orders	13	12	15	12	13.0	15	24	14	15	7	15.0	18
monetary base	6	8	24	11	12.3	11	26	17	18	26	21.8	24
consumer sentiment	7	7	8	10	8.0	6	16	9	6	5	9.0	9
GDP_vol	25	19	25	24	23.3	23	20	24	27	27	24.5	25
CPI_vol	26	23	23	18	22.5	24	10	25	20	19	18.5	20
MKT	28	27	26	20	25.3	28	27	27	24	24	25.5	27
STR	16	15	11	8	12.5	14	9	15	12	12	12.0	10
Dspread	19	14	10	6	12.3	16	22	12	13	11	14.5	16
Tspread	24	18	20	28	22.5	21	14	16	19	28	19.3	17
RV	14	**3**	**3**	**1**	**5.3**	**3**	7	5	**2**	**1**	3.8	**3**
IV	15	5	6	**2**	7.0	7	23	13	4	**2**	10.5	11
None	27	25	27	25	26.0	27	25	26	26	25	25.5	26
All	29	29	29	29	29.0	29	29	29	29	29	29.0	29
PC1	18	21	17	14	17.5	19	13	18	10	9	12.5	12
PC2	20	22	18	13	18.3	20	12	21	17	17	16.8	15
PC3	17	20	22	16	18.8	18	11	20	28	15	18.5	23
PC1-3	23	28	14	21	21.5	25	17	19	22	21	19.8	19
Lasso	**2**	**2**	**1**	**3**	**2.0**	**2**	**2**	**1**	**3**	**3**	**2.3**	**2**
Combine1	4	4	4	15	6.8	4	4	**3**	11	18	9.0	4
Combine2	**1**	**1**	**2**	4	**2.0**	**1**	**1**	**2**	**1**	10	**3.5**	**1**

Note: This table presents the ranking of the out-of-sample performance for forecasting (VaR,ES) based on the values of FZ0 loss function at four confidence levels. The left panel lists the results of the unrestricted Beta weighting scheme, and the right panel gives the results of the restricted ones. The best model ranks 1 and the worst ranks 29 given there are 29 competing models in total. The top three rankings are highlighted in **bold**. Columns 1-4 report the ranking at each confidence level. Column 5 presents the average ranking of each model across four confidence levels. Column 6 shows the ranking of the average loss achieved by each model across four confidence levels.

Table 8: The 95% Model Confidence Set based on the R and the SQ methods

	95% R					95% SQ				
	1%	2.50%	5%	10%	Total	1%	2.50%	5%	10%	Total
Real GDP	0	0	0	0	0	0	0	0	0	0
IP	0	0	0	0	0	0	0	0	0	0
dunemp	0	0	0	0	0	0	0	0	0	0
dhous	0	0	0	**1**	1	0	0	0	**1**	1
dcprof	0	0	0	0	0	0	0	0	0	0
drcons	0	0	0	0	0	0	0	0	0	0
CPI	0	0	0	0	0	0	0	0	0	0
PPI	0	0	0	0	0	0	0	0	0	0
CFNAI	0	0	0	**1**	1	0	0	0	**1**	1
new orders	0	0	0	0	0	0	0	0	0	0
monetary base	0	0	0	0	0	0	0	0	0	0
consumer sentiment	0	0	0	**1**	1	0	0	0	**1**	1
GDP_vol	0	0	0	0	0	0	0	0	0	0
CPI_vol	0	0	0	0	0	0	0	0	0	0
MKT	0	0	0	0	0	0	0	0	0	0
STR	0	0	0	0	0	0	0	0	0	0
Dspread	0	0	0	0	0	0	0	0	0	0
Tspread	0	0	0	0	0	0	0	0	0	0
RV	0	0	0	**1**	1	0	0	0	**1**	1
IV	0	0	0	**1**	1	0	0	0	**1**	1
None	0	0	0	0	0	0	0	0	0	0
All	0	0	0	0	0	0	0	0	0	0
PC1	0	0	0	**1**	1	0	0	0	1	1
PC2	0	0	0	0	0	0	0	0	0	0
PC3	0	0	0	0	0	0	0	0	0	0
PC1-3	0	0	0	0	0	0	0	0	0	0
Lasso	0	0	**1**	**1**	2	0	0	**1**	**1**	2
Combine1	0	0	0	0	0	0	0	0	0	0
Combine2	**1**	**1**	0	**1**	**3**	**1**	**1**	0	**1**	**3**

Note: This table presents the out-of-sample results of the Model Confidence Set (MCS) test based on the FZ0 loss function for the one-day-ahead joint forecasting of VaR and ES using the R and SQ methods. The number **1** indicates that the corresponding model (highlighted in **bold**) is within the 95% MCS, and the number 0 suggests the opposite. The Column Total shows the total number of times the corresponding model is within the 95% MCS. The highest value (in **bold**) in the Column Total means that the model is the most favored one across the considered confidence levels.

Table 9: Long-term out-of-sample forecasting results of (VaR, ES)

	1-quarter-ahead				2-quarter-ahead				3-quarter-ahead				4-quarter-ahead			
	1%	2.50%	5%	10%	1%	2.50%	5%	10%	1%	2.50%	5%	10%	1%	2.50%	5%	10%
Real GDP	1.219	0.997	0.796	0.557	1.217	1.002	0.806	0.562	1.242	1.015	0.819	0.573	1.248	1.025	0.826	0.580
IP	1.214	0.996	0.795	0.556	1.228	1.005	0.802	0.560	1.242	1.017	0.814	0.572	1.252	1.030	0.823	0.581
dunemp	1.199	0.992	0.792	0.567	1.222	1.003	0.799	0.566	1.230	1.017	0.816	0.590	1.246	1.039	0.821	0.599
dhous	1.198	1.002	0.787	0.554	1.232	1.000	0.796	0.560	1.227	1.013	0.808	0.573	1.243	1.019	0.814	0.577
dcprof	1.251	1.019	0.807	0.558	1.279	1.032	0.815	0.567	1.311	1.046	0.832	0.578	1.335	1.059	0.842	0.584
drcons	1.197	0.990	0.791	0.555	1.209	0.995	0.798	0.563	**1.218**	1.010	0.812	0.570	1.239	1.022	0.821	0.580
CPI	1.208	1.000	0.800	0.560	1.232	1.013	0.809	0.567	1.238	1.023	0.817	0.580	1.252	1.034	0.830	0.588
PPI	1.226	1.011	0.802	0.558	1.234	1.018	0.807	0.563	1.239	1.032	0.819	0.575	1.258	1.043	0.828	0.582
CFNAI	1.202	0.987	0.787	**0.553**	1.213	0.992	0.795	**0.558**	1.223	1.006	0.807	**0.569**	1.238	**1.017**	0.816	0.577
new orders	1.226	0.993	0.795	0.556	1.228	1.005	0.802	0.561	1.241	1.015	0.814	0.572	1.248	1.029	0.823	0.578
monetary base	1.225	0.991	0.788	0.558	1.230	1.013	0.797	0.563	1.251	1.013	0.816	0.582	1.253	1.031	0.818	0.591
consumer sentiment	1.219	**0.985**	0.787	0.554	1.225	1.001	0.795	0.559	1.241	1.017	0.806	0.571	1.251	1.024	0.816	0.578
GDP_vol	1.217	1.003	0.798	0.560	1.216	1.007	0.813	0.575	1.235	1.021	0.829	0.599	1.246	1.036	0.841	0.597
CPI_vol	1.215	1.011	0.798	0.558	1.215	1.017	0.804	0.563	1.247	1.028	0.817	0.573	1.257	1.035	0.826	0.580
MKT	1.225	1.017	0.802	0.560	1.235	1.018	0.808	0.566	1.258	1.030	0.821	0.576	1.279	1.048	0.831	0.584
STR	1.222	0.994	0.794	0.556	1.205	1.005	0.802	0.562	1.228	1.015	0.813	0.573	1.240	1.026	0.821	0.579
Dspread	1.216	0.993	0.793	0.555	1.240	1.007	0.801	0.561	1.246	1.021	0.813	0.572	1.266	1.031	0.822	0.579
Tspread	1.222	1.001	0.798	0.559	1.243	1.011	0.805	0.566	1.238	1.024	0.818	0.576	1.274	1.035	0.826	0.585
RV	1.201	0.990	**0.783**	**0.551**	1.219	0.991	**0.790**	**0.556**	1.250	**1.003**	**0.800**	**0.569**	1.248	1.017	**0.808**	**0.577**
IV	1.224	1.001	0.789	**0.550**	1.232	1.012	0.796	**0.557**	1.254	1.018	0.809	**0.569**	1.259	1.027	0.817	**0.576**
None	1.229	1.009	0.803	0.559	1.237	1.019	0.810	0.565	1.261	1.031	0.823	0.576	1.274	1.043	0.832	0.584
All	1.321	1.093	0.849	0.667	1.309	1.230	0.836	0.700	1.393	1.129	0.869	0.650	1.426	1.158	0.918	0.663
PC1	1.214	1.000	0.794	0.561	1.222	1.019	0.800	0.563	1.245	1.030	0.811	0.575	1.250	1.031	0.821	0.586
PC2	1.218	1.003	0.796	0.557	1.228	1.011	0.803	0.563	1.240	1.024	0.815	0.574	1.259	1.035	0.825	0.581
PC3	1.215	1.003	0.817	0.557	1.230	1.012	0.835	0.563	1.234	1.024	0.845	0.576	1.258	1.034	0.842	0.584
PC1-3	1.216	0.999	0.799	0.557	1.223	1.014	0.808	0.564	1.242	1.015	0.818	0.575	1.254	1.027	0.820	0.586
Lasso	**1.114**	**0.949**	**0.783**	0.555	**1.116**	**0.951**	**0.794**	0.560	**1.136**	**0.961**	**0.796**	0.573	**1.151**	**0.973**	**0.803**	**0.577**
Combine1	**1.193**	0.994	0.794	0.557	**1.205**	**0.986**	0.805	0.564	1.225	1.005	0.809	0.576	**1.232**	1.024	0.821	0.580
Combine2	**1.101**	**0.944**	**0.780**	0.555	**1.105**	**0.940**	**0.789**	0.562	**1.122**	**0.951**	**0.793**	0.574	**1.123**	**0.966**	**0.801**	0.579

Note: This table presents the average loss of the out-of-sample forecasting of (VaR, ES) based on the values of FZ0 loss function at four confidence levels in longer forecasting horizons. The three lowest average loss in each column are highlighted in **bold**.

Table 10: The 95% Model Confidence Set of the long-term forecasting based on the R method

	1-quarter-ahead					2-quarter-ahead					3-quarter-ahead					4-quarter-ahead				
	1%	2.50%	5%	10%	Total	1%	2.50%	5%	10%	Total	1%	2.50%	5%	10%	Total	1%	2.50%	5%	10%	Total
Real GDP	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
IP	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
dunemp	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
dhous	0	0	0	0	0	0	0	0	**1**	1	0	0	0	0	0	0	0	0	**1**	1
dcprof	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
drcons	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	1	0	0	0	**1**	1
CPI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PPI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CFNAI	0	0	0	0	0	0	0	0	**1**	1	0	0	0	**1**	1	0	0	0	**1**	1
new orders	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	1
monetary base	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
consumer sentiment	0	0	0	0	0	0	0	0	**1**	1	0	0	0	0	0	0	0	0	**1**	1
GDP_vol	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CPI_vol	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MKT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
STR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	1
Dspread	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Tspread	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RV	0	0	0	0	0	0	0	0	**1**	1	0	0	0	**1**	1	0	0	0	**1**	1
IV	0	0	0	**1**	1	0	0	0	**1**	1	0	0	0	**1**	1	0	0	0	**1**	1
None	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
All	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PC1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PC2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PC3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PC1-3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Lasso	**1**	**1**	**1**	0	**3**	**1**	0	0	**1**	2	0	**1**	**1**	0	2	0	**1**	**1**	**1**	**3**
Combine1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Combine2	**1**	**1**	**1**	0	**3**	**1**	**1**	**1**	0	**3**	**1**	**1**	**1**	0	**3**	**1**	**1**	**1**	0	**3**

Note: This table presents the out-of-sample results of the Model Confidence Set (MCS) test based on the FZ0 loss function for the long-term joint forecasting of VaR and ES using the R method. The number **1** indicates that the corresponding model (highlighted in **bold**) is within the 95% MCS, and the number 0 suggests the opposite. The Column Total shows the total number of times the corresponding model is within the 95% MCS. The highest value (in **bold**) in the Column Total means that the model is the most favored one across the considered confidence levels.

Table 11: The 95% Model Confidence Set of the long-term forecasting based on the SQ method

	1-quarter-ahead					2-quarter-ahead					3-quarter-ahead					4-quarter-ahead				
	1%	2.50%	5%	10%	Total	1%	2.50%	5%	10%	Total	1%	2.50%	5%	10%	Total	1%	2.50%	5%	10%	Total
Real GDP	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
IP	0	0	0	0	0	0	0	0	**1**	1	0	0	0	0	0	0	0	0	0	0
dunemp	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
dhous	0	0	0	0	0	0	0	0	**1**	1	0	0	0	0	0	0	0	0	**1**	1
dcprof	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
drcons	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	1	0	0	0	**1**	1
CPI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PPI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CFNAI	0	0	0	0	0	0	0	0	**1**	1	0	0	0	**1**	1	0	0	0	**1**	1
new orders	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	1
monetary base	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
consumer sentiment	0	0	0	0	0	0	0	0	**1**	1	0	0	0	**1**	1	0	0	0	**1**	1
GDP_vol	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CPI_vol	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MKT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
STR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	1
Dspread	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	1
Tspread	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RV	0	0	0	0	0	0	0	0	**1**	1	0	0	0	**1**	1	0	0	0	**1**	1
IV	0	0	0	**1**	1	0	0	0	**1**	1	0	0	0	**1**	1	0	0	0	**1**	1
None	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
All	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PC1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PC2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PC3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PC1-3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Lasso	**1**	**1**	**1**	0	**3**	**1**	0	0	**1**	2	0	**1**	**1**	0	2	0	**1**	**1**	**1**	**3**
Combine1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Combine2	**1**	**1**	**1**	0	**3**	**1**	**1**	**1**	0	**3**	**1**	**1**	**1**	0	**3**	**1**	**1**	**1**	**1**	**4**

Note: This table presents the out-of-sample results of the Model Confidence Set (MCS) test based on the FZ0 loss function for the long-term joint forecasting of VaR and ES using the SQ method. The number **1** indicates that the corresponding model (highlighted in **bold**) is within the 95% MCS, and the number 0 suggests the opposite. The Column Total shows the total number of times the corresponding model is within the 95% MCS. The highest value (in **bold**) in the Column Total means that the model is the most favored one across the considered confidence levels.

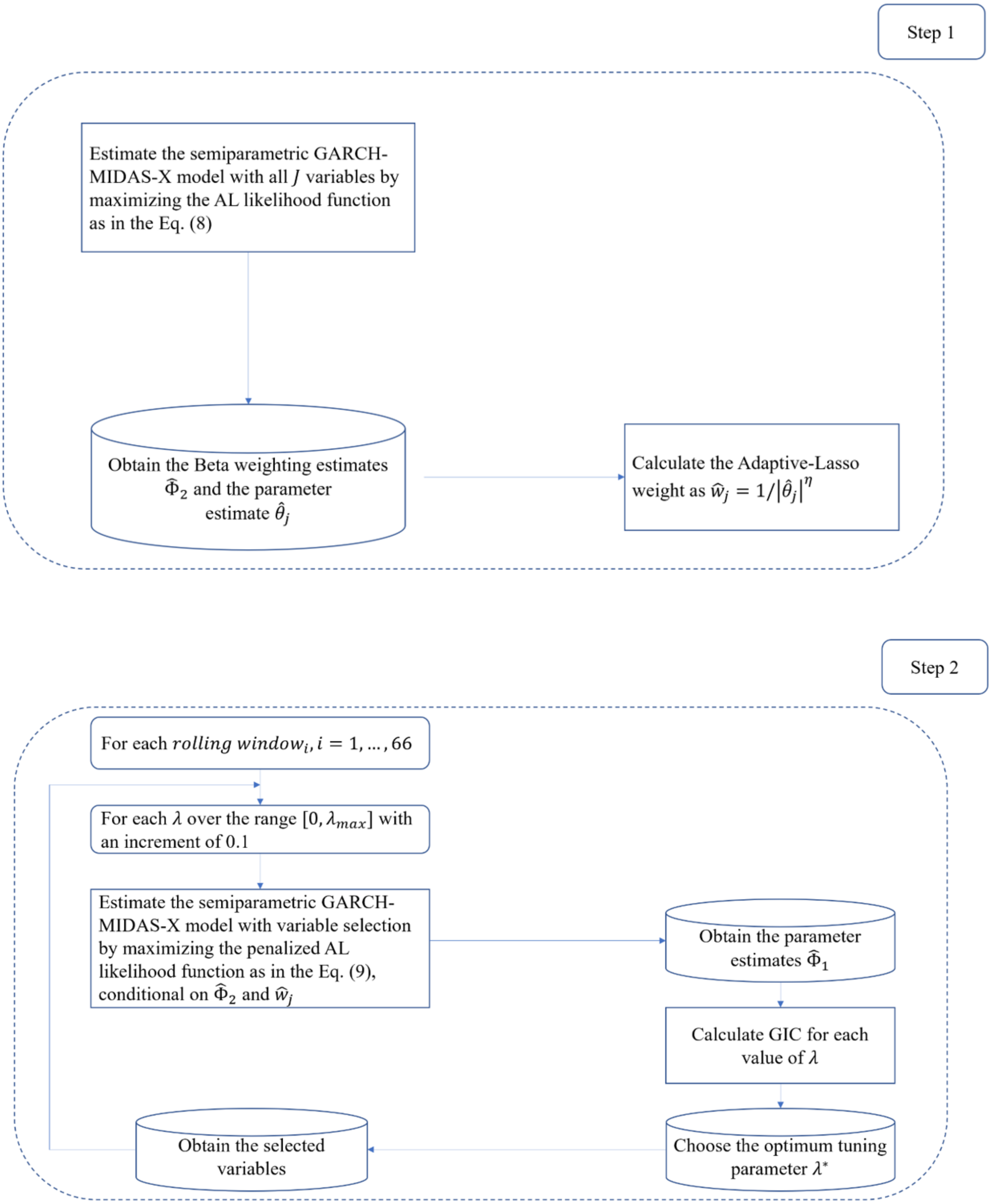


Figure 1: Estimation framework

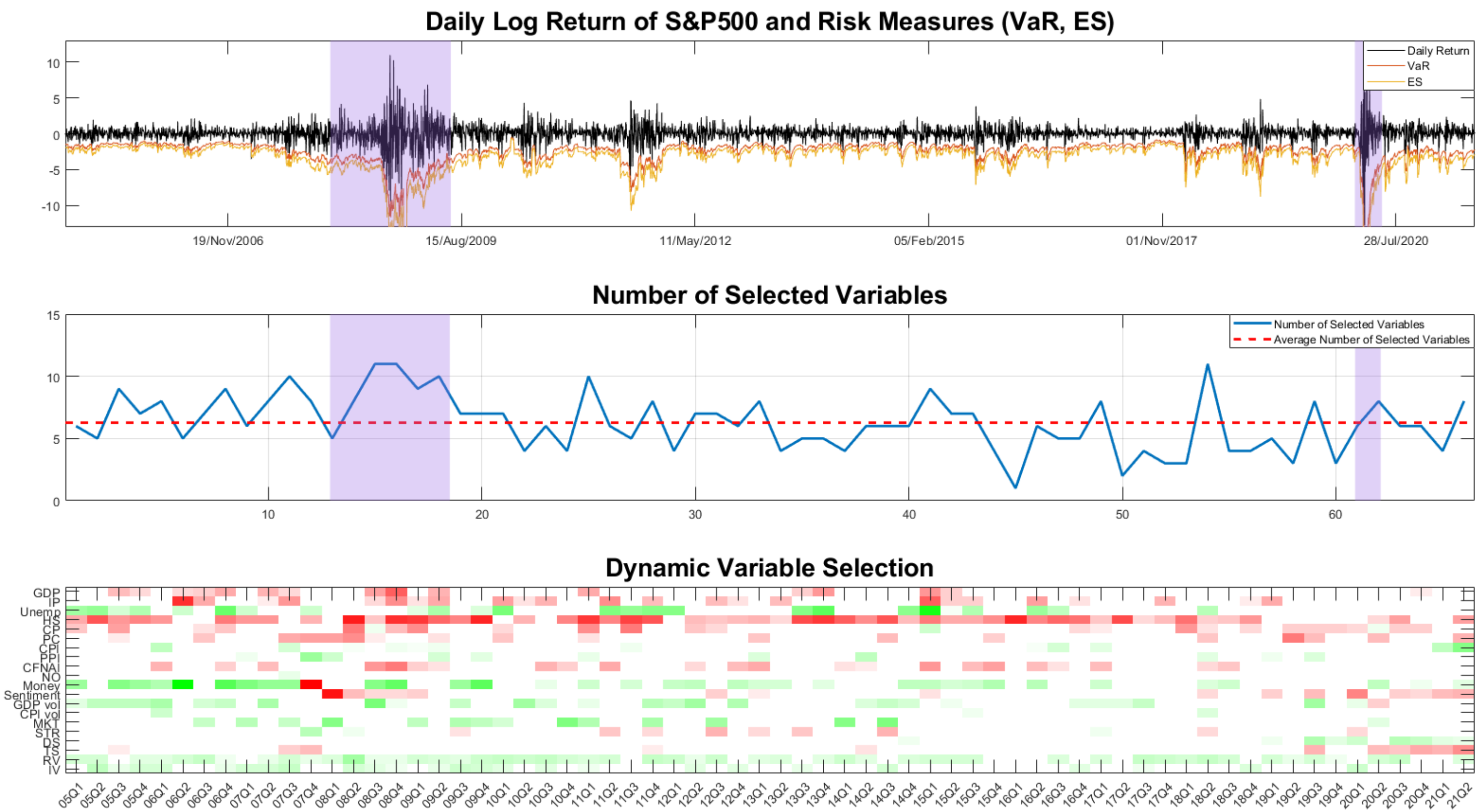


Figure 2: Dynamic variable selection at $\alpha = 1\%$

Note: This figure shows the dynamic variable selection results at the 1% confidence level. The color tab in this figure represents that the corresponding variable is selected by the Adaptive Lasso. Green tab indicates a positive relationship between the corresponding variable and future tail risk, and red tab suggests the opposite.

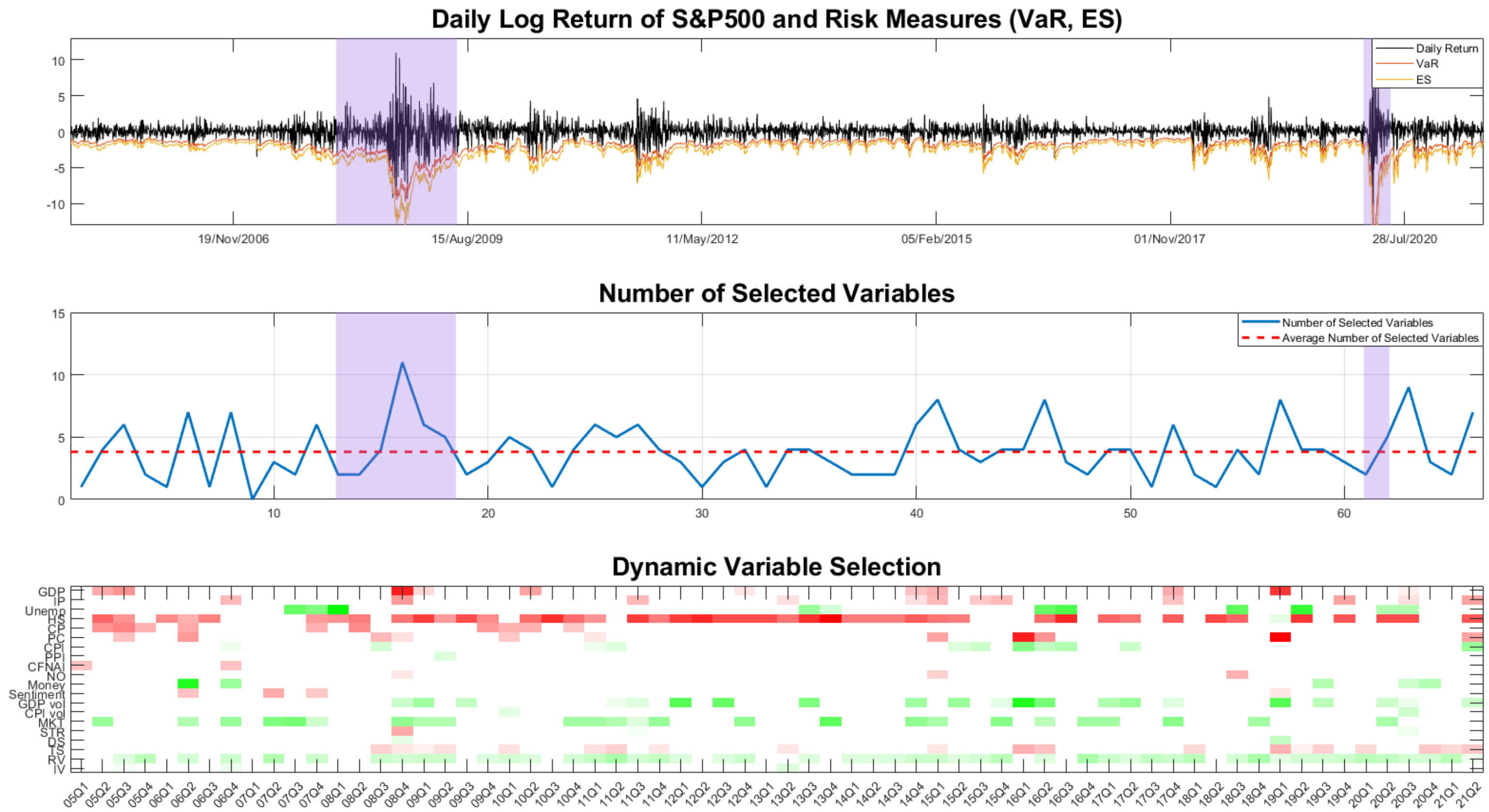


Figure 3: Dynamic variable selection at $\alpha = 2.5\%$

Note: This figure shows the dynamic variable selection results at the 2.5% confidence level. The color tab in this figure represents that the corresponding variable is selected by the Adaptive Lasso. Green tab indicates a positive relationship between the corresponding variable and future tail risk, and red tab suggests the opposite.

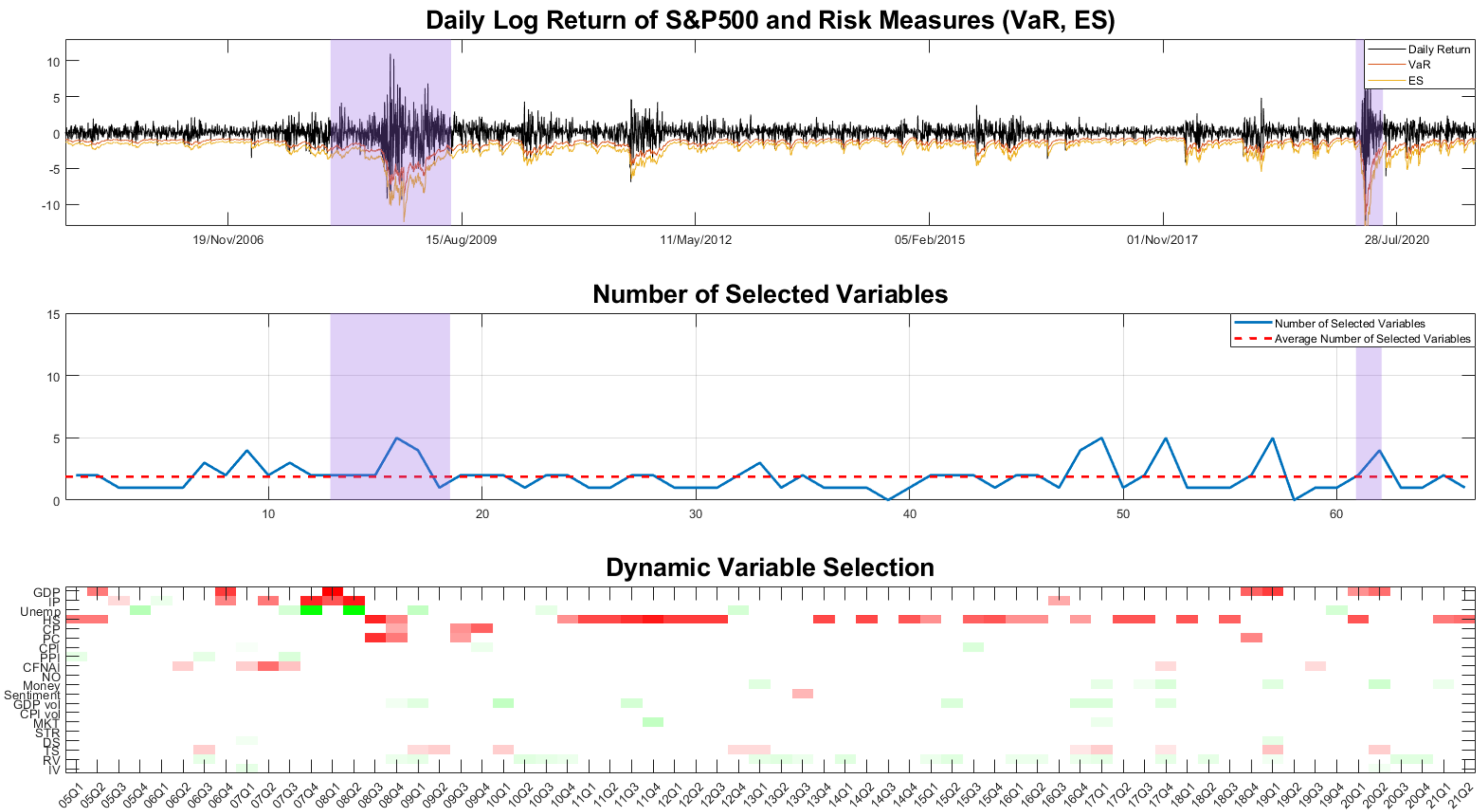


Figure 4: Dynamic variable selection at $\alpha = 5\%$

Note: This figure shows the dynamic variable selection results at the 5% confidence level. The color tab in this figure represents that the corresponding variable is selected by the Adaptive Lasso. Green tab indicates a positive relationship between the corresponding variable and future tail risk, and red tab suggests the opposite.

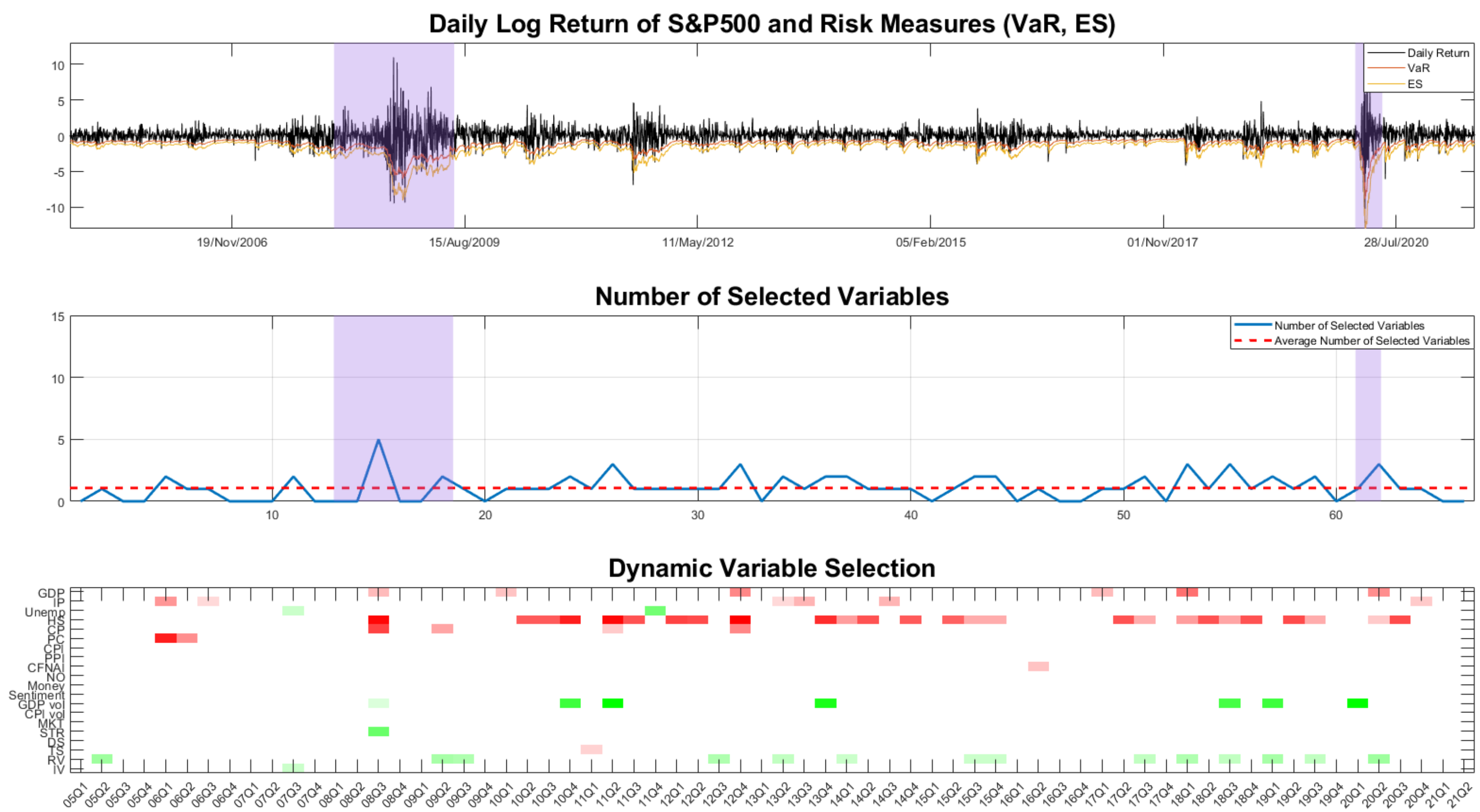
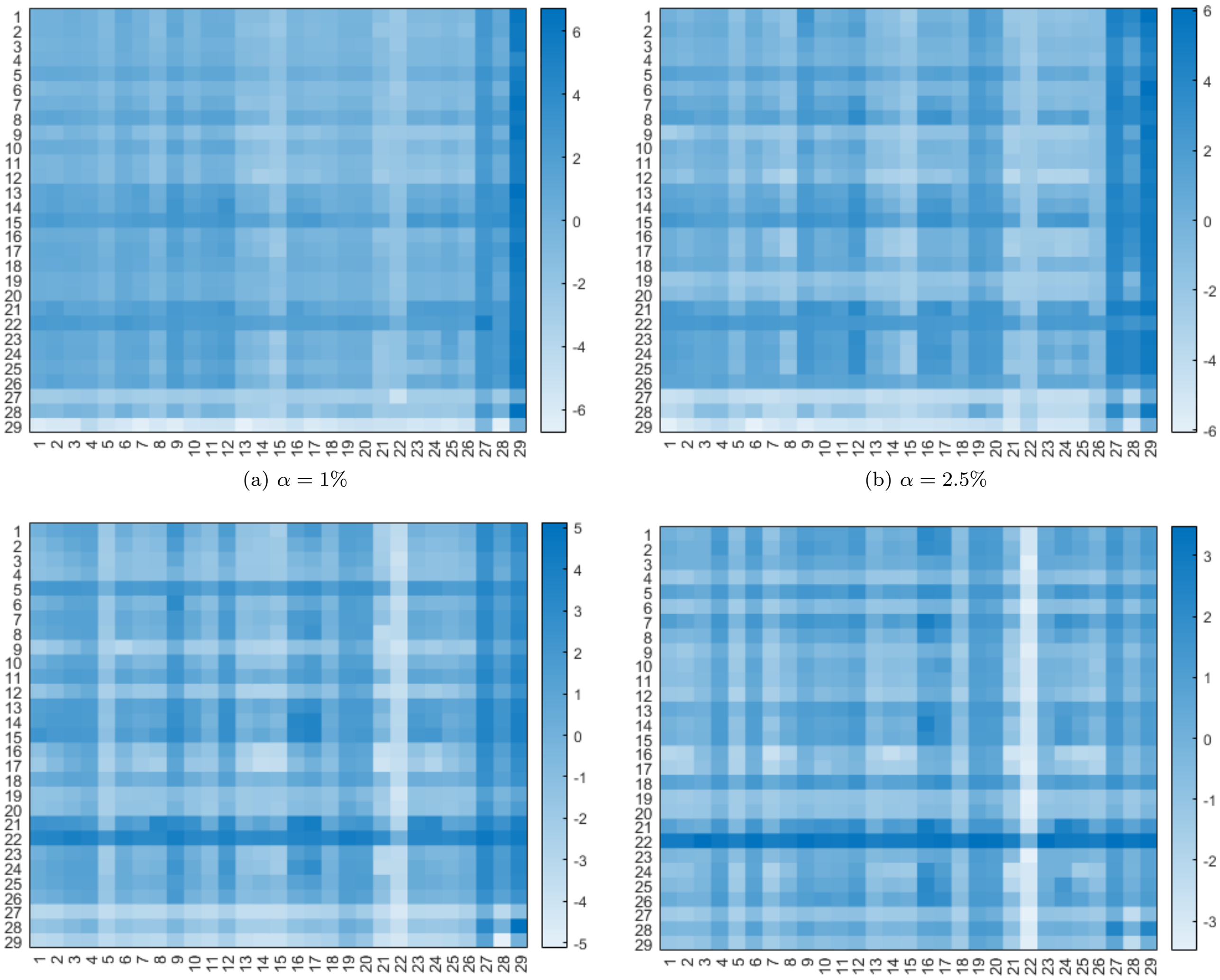


Figure 5: Dynamic variable selection at $\alpha = 10\%$

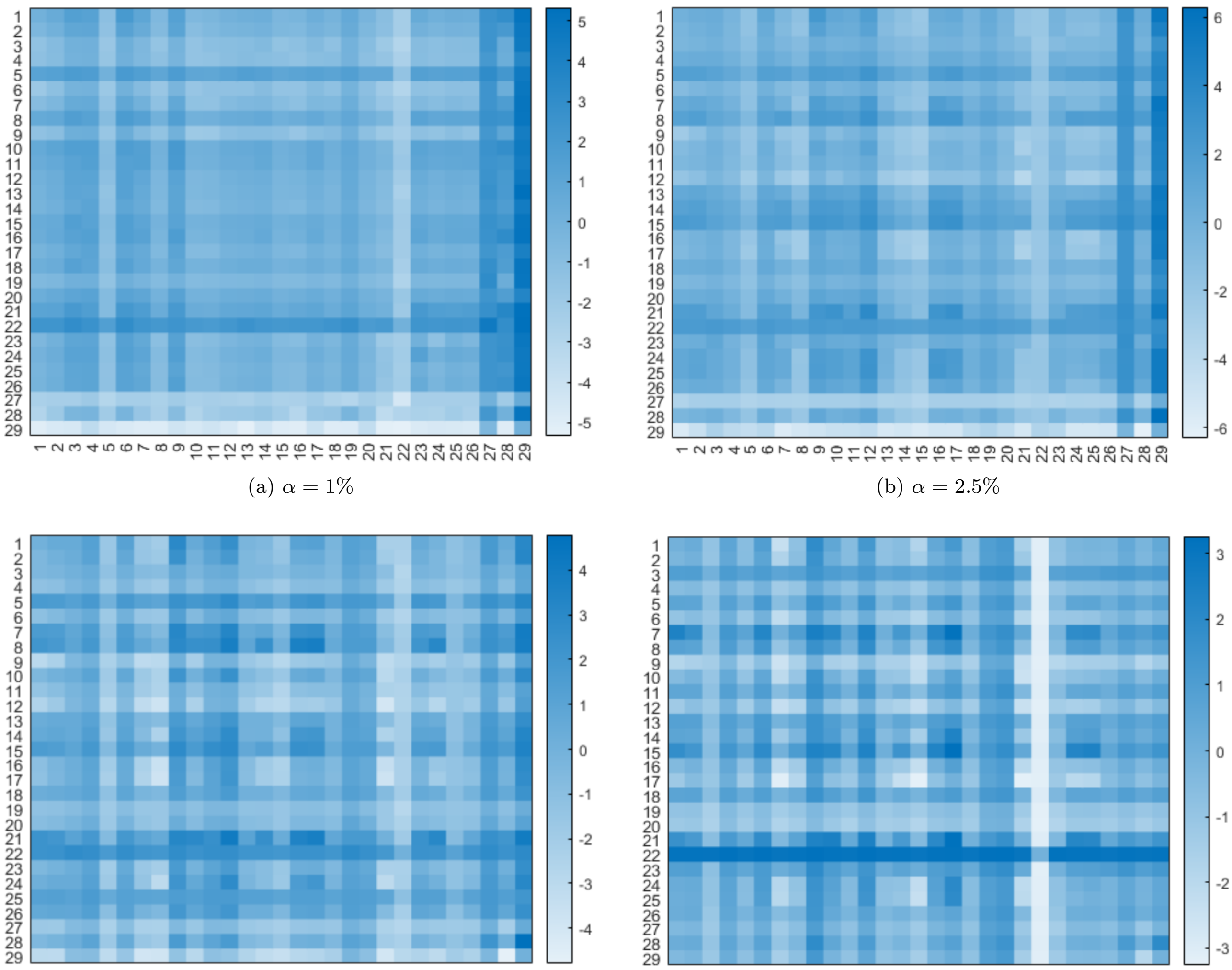
Note: This figure shows the dynamic variable selection results at the 10% confidence level. The color tab in this figure represents that the corresponding variable is selected by the Adaptive Lasso. Green tab indicates a positive relationship between the corresponding variable and future tail risk, and red tab suggests the opposite.



(a) $\alpha = 1\%$ (b) $\alpha = 2.5\%$ (c) $\alpha = 5\%$ (d) $\alpha = 10\%$

Figure 6: Color map of the DM test results based on the one-day-ahead forecasts

Note: This figure presents the color map of the Diebold-Mariano (DM) test results for 29 different models. The full name of these models can be obtained from the backtest results table. For example, model 1 is the semiparametric GARCH-MIDAS-X model that integrates the GDP only. The DM test compares the out-of-sample average losses for the one-day-ahead joint forecasting of VaR and ES based on the FZ0 loss function at: (a) 1% level; (b) 2.5% level; (c) 5% level; (d) 10% level. Each block reports the DM test t-statistic, which compares the performance of the column model with that of the row model. Darker blocks indicate the greater extent by which the column model outperforms the row model (achieves lower average losses).



(a) $\alpha = 1\%$

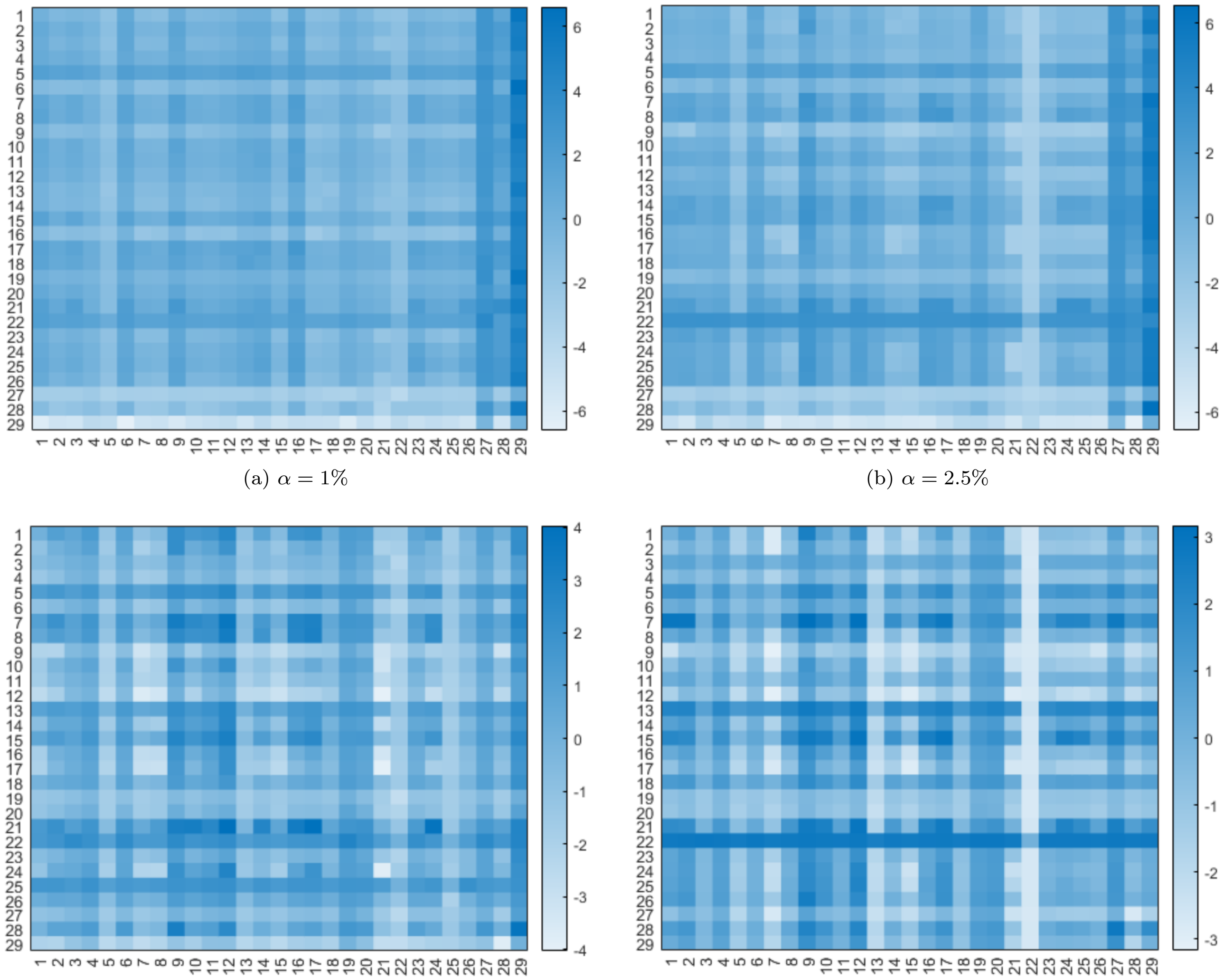
(b) $\alpha = 2.5\%$

(c) $\alpha = 5\%$

(d) $\alpha = 10\%$

Figure 7: Color map of the DM test results based on the 1-quarter-ahead forecasts

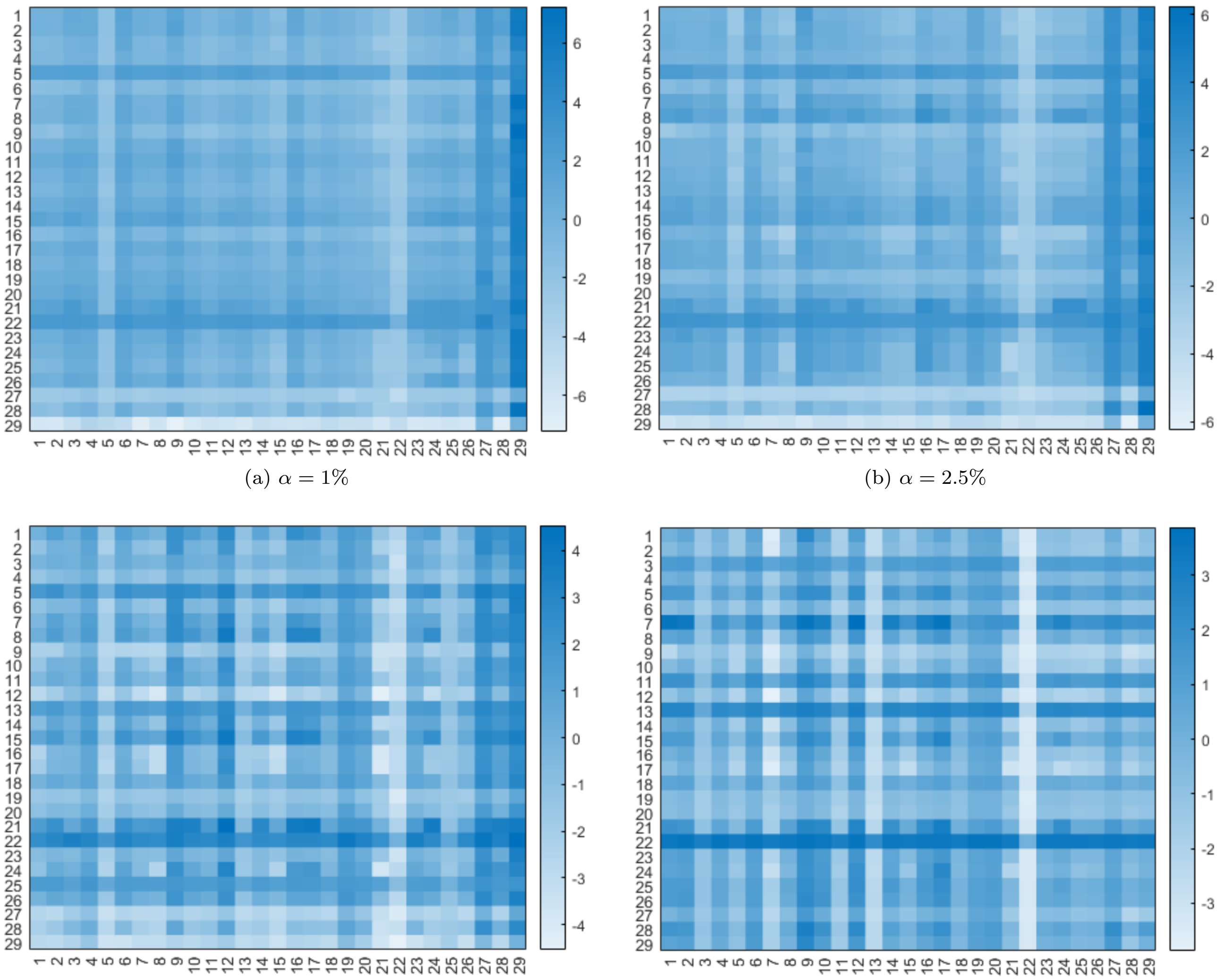
Note: This figure presents the color map of the Diebold-Mariano (DM) test results for 29 different models. The full name of these models can be obtained from the backtest results table. For example, model 1 is the semiparametric GARCH-MIDAS-X model that integrates the GDP only. The DM test compares the out-of-sample average losses for the 1-quarter-ahead joint forecasting of VaR and ES based on the FZ0 loss function at: (a) 1% level; (b) 2.5% level; (c) 5% level; (d) 10% level. Each block reports the DM test t-statistic, which compares the performance of the column model with that of the row model. Darker blocks indicate the greater extent by which the column model outperforms the row model (achieves lower average losses).



(a) $\alpha = 1\%$ (b) $\alpha = 2.5\%$ (c) $\alpha = 5\%$ (d) $\alpha = 10\%$

Figure 8: Color map of the DM test results based on the 2-quarter-ahead forecasts

Note: This figure presents the color map of the Diebold-Mariano (DM) test results for 29 different models. The full name of these models can be obtained from the backtest results table. For example, model 1 is the semiparametric GARCH-MIDAS-X model that integrates the GDP only. The DM test compares the out-of-sample average losses for the 2-quarter-ahead joint forecasting of VaR and ES based on the FZ0 loss function at: (a) 1% level; (b) 2.5% level; (c) 5% level; (d) 10% level. Each block reports the DM test t-statistic, which compares the performance of the column model with that of the row model. Darker blocks indicate the greater extent by which the column model outperforms the row model (achieves lower average losses).



(a) $\alpha = 1\%$

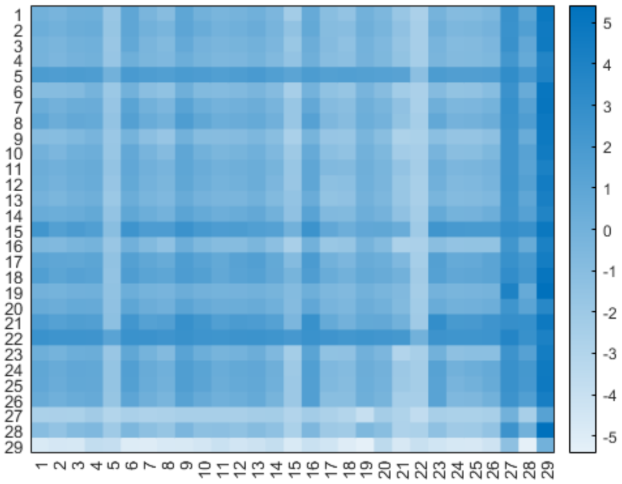
(b) $\alpha = 2.5\%$

(c) $\alpha = 5\%$

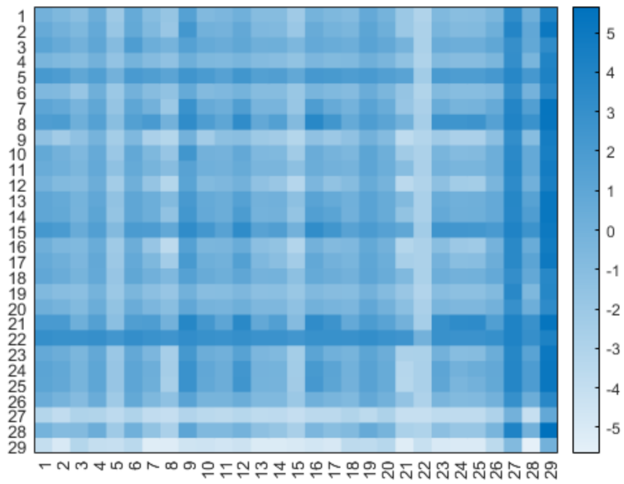
(d) $\alpha = 10\%$

Figure 9: Color map of the DM test results based on the 3-quarter-ahead forecasts

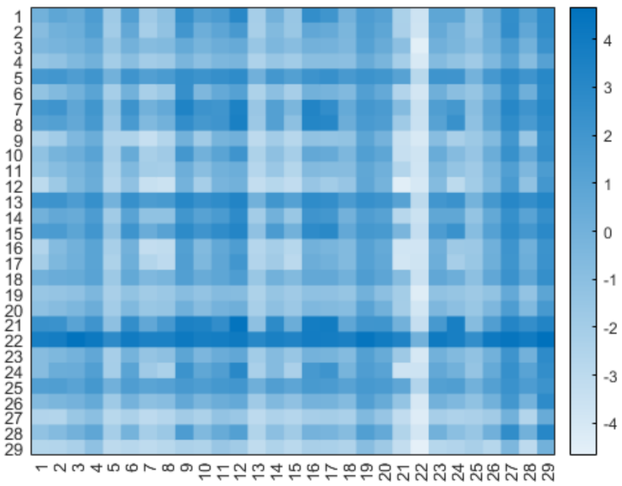
Note: This figure presents the color map of the Diebold-Mariano (DM) test results for 29 different models. The full name of these models can be obtained from the backtest results table. For example, model 1 is the semiparametric GARCH-MIDAS-X model that integrates the GDP only. The DM test compares the out-of-sample average losses for the 3-quarter-ahead joint forecasting of VaR and ES based on the FZ0 loss function at: (a) 1% level; (b) 2.5% level; (c) 5% level; (d) 10% level. Each block reports the DM test t-statistic, which compares the performance of the column model with that of the row model. Darker blocks indicate the greater extent by which the column model outperforms the row model (achieves lower average losses).



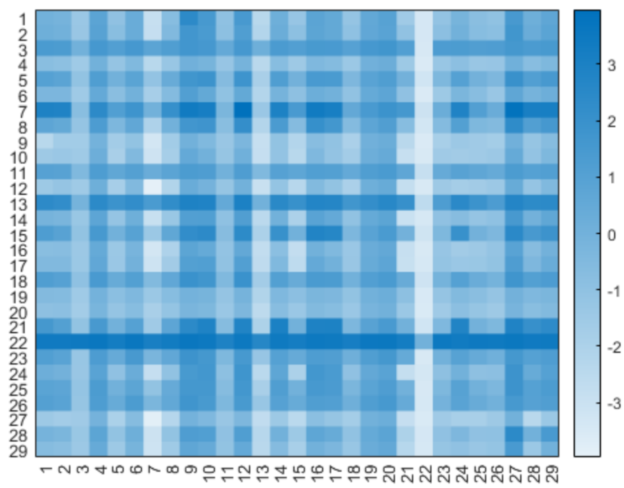
(a) $\alpha = 1\%$



(b) $\alpha = 2.5\%$



(c) $\alpha = 5\%$



(d) $\alpha = 10\%$

Figure 10: Color map of the DM test results based on the 4-quarter-ahead forecasts

Note: This figure presents the color map of the Diebold-Mariano (DM) test results for 29 different models. The full name of these models can be obtained from the backtest results table. For example, model 1 is the semiparametric GARCH-MIDAS-X model that integrates the GDP only. The DM test compares the out-of-sample average losses for the 4-quarter-ahead joint forecasting of VaR and ES based on the FZ0 loss function at: (a) 1% level; (b) 2.5% level; (c) 5% level; (d) 10% level. Each block reports the DM test t-statistic, which compares the performance of the column model with that of the row model. Darker blocks indicate the greater extent by which the column model outperforms the row model (achieves lower average losses).