# Adaptively Weighted Combinations of Tail-Risk Forecasts

Alessandra Amendola[a], Vincenzo Candila[a], Antonio Naimoli[a], and Giuseppe Storti[a]

[a] Department of Economics and Statistics, University of Salerno, Italy,
alamendola@unisa.it, vcandila@unisa.it, anaimoli@unisa.it, storti@unisa.it

**Abstract**

Several methods have been proposed in the literature for forecasting daily Value-at-Risk (VaR) and Expected Shortfall (ES), with tail risk models generally classified into three main categories: parametric, semi-parametric, and non-parametric. However, given the various sources of uncertainty associated with the data, market conditions, estimation methods, and exogenous variables that can have a significant impact on the dynamics of tail risk, there is no single model that consistently outperforms all the others. To mitigate the impact of model misspecification and improve the predictive accuracy of individual models, we investigate the use of two forecast combination approaches. In the proposed approaches, the weight of the most accurate set of predictors is determined adaptively according to strictly consistent loss functions for VaR and ES employed in the Model Confidence Set procedures. Our findings reveal that combinations of VaR and ES forecasts result in higher predictive accuracy over a wide range of individual competitors.

***Keywords:*** Value-at-Risk, Expected Shortfall, Model Confidence Set, Forecast Combinations

## 1. Introduction

Value at Risk (VaR) and Expected Shortfall (ES) are widely accepted tail risk measures used by financial institutions to calculate the minimum capital requirements needed to protect their investments from adverse market conditions and to hedge against unexpected losses, as promoted by the Basel Capital Accord. Although a large number of models have been proposed for the generation of accurate VaR and ES forecasts, they can be mainly classified into parametric, non-parametric and semi-parametric approaches. Parametric models, such as GARCH-type models, rely heavily on assumptions about return distribution, which can significantly affect forecast accuracy. On the other hand, non-parametric techniques, like historical simulation, are widely used because of their ease of implementation and the absence of distributional assumptions. Finally, a framework for estimating VaR and ES that does not rely on the estimation of volatility is provided by the use of semi-parametric methods (Engle and Manganelli, 2004; Gerlach and Wang, 2020; Taylor, 2008a, 2019).

There is no evidence in the tail risk forecasting literature to support some specific models or approaches to forecast VaR and ES. This uncertainty arises from several sources, including data dynamics, parameter estimation, specification of measurement equations as well as the impact of exogenous factors. In this framework, combining forecasts allows controlling model uncertainty and exploiting the information from each individual model (Clemen, 1989; Timmermann, 2006; Wang et al., 2022).

Since the seminal research of Bates and Granger (1969), combining forecasts has received considerable attention in a variety of fields, with empirical evidence highlighting the benefits of reducing the risk of selecting a single predictor. For example, several strategies have been proposed to combine volatility forecasts in order to determine the optimal set of predictors and reduce model uncertainty in both univariate (Amendola and Storti, 2008; Becker and Clements, 2008; Ma et al., 2018) and multivariate (Amendola et al., 2020; Amendola and Storti, 2015; Caldeira et al., 2017) frameworks. Furthermore, several studies provide empirical evidence to support the use of forecast combination in risk assessment, as the comparison of forecasts over the forecast horizon might not always lead to a single optimal model or approach (Komunjer, 2013; Nieto and Ruiz, 2016). In this direction, strategies for the combination of VaR forecasts have been proposed, for instance, in Halbleib and Pohlmeier (2012), Jeon and Taylor (Jeon and Taylor) and Bayer (2018), among others. Finally, Taylor (2020a) and Storti and Wang (2023) developed approaches for for jointly combining VaR and ES forecasts.

The aim of this paper is to evaluate the merits of combining forecasts to improve the accuracy of VaR and ES obtained from individual models. The ability to mitigate model misspecification and synthesize multiple sources of information allows for different combination strategies, adaptively weighting the most accurate set of VaR and ES forecasts. For the selection of models to be included in the combination, we refer to the Model Confidence Set (MCS, Hansen et al., 2011). Specifically, forecast combinations are determined by minimizing strictly consistent joint VaR and ES loss functions of the class of Fissler and Ziegel (2016) according to different weighting schemes. It is also worth noting that the set of models included in the combination changes dynamically over time without having to estimate parameters.

The performance of the proposed approach is assessed by a comprehensive backtesting analysis according to the Unconditional Coverage (UC) test (Kupiec, 1995), the Conditional Coverage (CC) test (Christoffersen, 1998) and the Dynamic Quantile (DQ) test (Engle and Manganelli, 2004). On the other hand, the Regression-Based Expected Shortfall Backtesting (BD) of Bayer and Dimitriadis (2022) is used to test the validity of ES forecasts. Finally, the MCS is employed to assess the statistical significance of differences among competing models.

The proposed approach is applied to combine VaR and ES forecasts of a wide range of tail risk models for the S&P500 Index. The empirical results show that the proposed combined specifications provide evidence of correct model specification under the backtesting procedures while entering the 75% MCS.

The rest of the paper is organized as follows. Section 2. reviews VaR and ES measures, Section 3. illustrates the implemented methodology and the proposed combination strategies. Section 4. is focuses on the empirical application. Finally, Section 5. provides concluding remarks.

## 2. Tail risk measures

VaR and ES are the most popular benchmarks from a risk management perspective (Christoffersen and Gonçalves, 2005; Sarykalin et al., 2008).

As a conditional quantile in the lower tail of the distribution of the return on a portfolio, VaR has become a popular measure of financial market risk used to meet regulatory and internal risk management requirements. VaR is a measure of the maximum potential loss on a portfolio of assets over a given time horizon and with a specified level of confidence. Formally,

$$Pr(r_t < VaR_t(\tau)|\mathcal{F}_{t-1}) = \tau \,, \tag{1}$$

or equivalently

$$VaR_t(\tau) = F^{-1}_{t,r}(\tau) \,, \tag{2}$$

where $r_t$ is the one-period log-return from time $t-1$ to $t$, $\tau \in (0,1)$ is the quantile level, $\mathcal{F}_{t-1}$ is the information set at time $t-1$ and $F_{t,r}^{-1}(\tau)$ is the Cumulative Distribution Function (CDF) of $r_t$ conditional on $\mathcal{F}_{t-1}$.

However, VaR has been criticized for failing to meet the criteria of a consistent risk metric (Artzner et al., 1999) and for its limitations in providing insight into potential outliers beyond the quantile.

The ES has been found to be a coherent measure of risk (Acerbi and Tasche, 2002) and is therefore suggested as an alternative for VaR in risk management applications. The ES is defined as the conditional expectation of returns exceeding the corresponding VaR threshold, namely

$$ES_t(\tau) = E\left[r_t | r_t < VaR_t(\tau), \mathcal{F}_{t-1}\right]. \tag{3}$$

Despite its appropriate theoretical properties, ES is not "elicitable" (Gneiting, 2011). This means that the ES cannot uniquely minimize the expectation of at least one loss or scoring function. However, Fissler and Ziegel (2016) found that VaR and ES are jointly elicitable, leading to a set of strictly consistent scoring functions that can be used for jointly evaluating the VaR and ES forecasts.

The general Fissler and Ziegel (2016) class of loss functions is defined as

$$\begin{aligned} FZ(VaR_t, ES_t | r_t; \tau) =& \{I(r_t \leq VaR_t) - \tau\} G_1(VaR_t) - I(r_t \leq VaR_t) G_1(r_t) + G_2(ES_t) \\ & \left\{ ES_t - VaR_t + I(r_t \leq VaR_t) \frac{VaR_t - r_t}{\tau} \right\} - \zeta_2(ES_t) + a(r_t), \end{aligned} \tag{4}$$

where $G_1$, $G_2$, $\zeta_2$ and $a$ are functions satisfying a number of conditions, including $G_2 = \zeta_2'$, $G_1$ is increasing and $\zeta_2$ is increasing and convex. Choosing $G_1 = 0$, $G_2(x) = -1/x$, $\zeta_2(x) = -\log(-x)$, and setting $a(r_t) = 0$, leads to the zero-degree homogeneous loss ($FZ0$) in Patton et al. (2019). Similarly, setting $G_1 = 0$, $G_2(x) = -1/x$, $\zeta_2(x) = -\log(-x)$, $a(r_t) = 1 - \log(1-\alpha)$, leads to $AL$ score in Taylor (2020b).

# 3. Methodology

In this work, two different combined predictors for VaR and ES are proposed. All the combined predictors are based on the models which, dynamically, enter the Set of Superior Models (SSM) according to the MCS procedure. We defined as *training* MCS the procedure used to find the best models in the so-called training or in-sample period. The loss used in the *training* MCS belongs to the class of $FZ$. More in detail, the loss is that of Patton et al. (2019) labelled as $FZ0$. Formally:

$$\begin{aligned} FZ0(VaR_{i,t}(\tau), ES_{i,t}(\tau), r_{i,t}) =& \frac{1}{\tau ES_t(\tau)} \mathbb{1}_{(r_{i,t} \leq VaR_{i,t}(\tau))}(r_{i,t} - VaR_{i,t}(\tau)) + \\ & \frac{VaR_{i,t}(\tau)}{ES_{i,t}(\tau)} + \log(-ES_{i,t}(\tau)) - 1, \end{aligned} \tag{5}$$

where $\tau$ is the coverage level chosen, $r_{i,t}$ is the log-returns for day $i$ of the low-frequency period $t$ and $\mathbb{1}_{(\cdot)}$ is an indicator function.

The chosen set of candidate models covers a wide range of frequently used parametric, semi-parametric and non-parametric techniques, as well as methods based on intraday data and mixed frequency variables. Table 1 reports the set of candidate models.

In this work, we propose two different combined predictors: MCS-Comb and WL-MCS-Comb. In particular:

Table 1: Candidate models

| Model | Functional form | Err. Distr. |
|---|---|---|
| GARCH–N, GARCH–t (Bollerslev, 1986) | $r_{i,t}\|\mathcal{F}_{i-1,t} = \sqrt{h_{i,t}}\eta_{i,t}$ <br> $h_{i,t} = \omega + \alpha r^2_{i-1,t} + \beta h_{i-1,t}$ | $\eta_{i,t} \overset{i.i.d}{\sim} \mathcal{N}(0,1)$, $\eta_{i,t} \overset{i.i.d}{\sim} t_\nu$ |
| GJR–N, GJR–t (Glosten et al., 1993) | $r_{i,t}\|\mathcal{F}_{i-1,t} = \sqrt{h_{i,t}}\eta_{i,t}$ <br> $h_{i,t} = \omega + \left(\alpha + \gamma \mathbb{1}_{(r_{i-1,t}<0)}\right) r^2_{i-1,t} + \beta h_{i-1,t}$ | $\eta_{i,t} \overset{i.i.d}{\sim} \mathcal{N}(0,1)$, $\eta_{i,t} \overset{i.i.d}{\sim} t_\nu$ |
| GM–N, GM–t (Engle et al., 2013) | $r_{i,t}\|\mathcal{F}_{i-1,t} = \sqrt{\pi_t \times \xi_{i,t}}\eta_{i,t}$ <br> $\xi_{i,t} = (1-\alpha-\beta-\gamma/2) + \left(\alpha + \gamma \cdot \mathbb{1}_{(r_{i-1,t}<0)}\right) \frac{r^2_{i-1,t}}{\pi_t} + \beta \xi_{i-1,t}$ <br> $\pi_t = \exp\left\{m + \zeta \sum_{k=1}^{K} \delta_k(\omega) MV_{t-k}\right\}$ | $\eta_{i,t} \overset{i.i.d}{\sim} \mathcal{N}(0,1)$, $\eta_{i,t} \overset{i.i.d}{\sim} t_\nu$ |
| R–GARCH–N, R–GARCH–t (Hansen et al., 2012) | $r_{i,t}\|\mathcal{F}_{i-1,t} = \sqrt{h_{i,t}}\eta_{i,t}$ <br> $h_{i,t} = const + \beta h_{i-1,t} + \alpha x_{i-1,t}$ <br> $x_{i,t} = const_x + \delta h_{i,t} + \tau_1 \eta_{i,t} + \tau_2 \left(\eta^2_{i,t} - 1\right) + \sigma_u u_{i,t}$ | $\eta_{i,t} \overset{i.i.d}{\sim} \mathcal{N}(0,1)$, $\eta_{i,t} \overset{i.i.d}{\sim} t_\nu$ |
| HS (Hendricks, 1996) | $VaR_{i,t}(\tau) = Q_{\boldsymbol{r}^w_{i,t}}(\tau)$ <br> $\boldsymbol{r}^w_{i,t} = (r_{i-w,t}, r_{i-w+1,t}, \ldots, r_{i-1,t})$ | |
| CAViaR-SAV (Engle and Manganelli, 2004) | $VaR_{i,t}(\tau) = \beta_0 + \beta_1 VaR_{i-1,t}(\tau) + \beta_2 \|r_{i-1,t}\|$ | |
| CAViaR-AS (Engle and Manganelli, 2004) | $VaR_{i,t}(\tau) = \beta_0 + \beta_1 VaR_{i-1,t}(\tau) + (\beta_2 \mathbb{1}_{(r_{i-1,t}>0)} + \beta_3 \mathbb{1}_{(r_{i-1,t}<0)}) \|r_{i-1,t}\|$ | |
| CAViaR-IG (Engle and Manganelli, 2004) | $VaR_{i,t}(\tau) = -\sqrt{\beta_0 + \beta_1 VaR^2_{i-1,t}(\tau) + \beta_2 r^2_{i-1,t}}$ | |
| CAViaR-X (Gerlach and Wang, 2020) | $VaR_{i,t}(\tau) = \beta_0 + \beta_1 VaR_{i-1,t}(\tau) + \beta_2 x_{i-1,t}$ | |
| MF-X (Candila et al., 2023) | $r_{i,t}\|\mathcal{F}_{i-1,t} = \sqrt{h_{i,t}}\eta_{i,t}$ <br> $\sqrt{h_{i,t}} = (\beta_0 + \theta \|WS_{t-1}\| + \beta_1 \|r_{i-1,t}\| + \ldots + \beta_q \|r_{i-q,t}\| + \beta_X \|X_{i-1,t}\|)$ <br> $WS_{t-1} = \sum_{k=1}^{K} \delta_k(\omega) MV_{t-k}$ | $\eta_{i,t} \overset{i.i.d}{\sim} (0,1)$ |

- MCS-Comb: equally weighting all the $VaR$'s and $ES$'s obtained from the models entering the SSM of the *training* MCS, using the (unweighted) $FZ0$;
- WL-MCS-Comb: equally weighting all the $VaR$'s and $ES$'s obtained from the models entering the SSM of the *training* MCS, using the *weighted* $FZ0$.

The *weighted* $FZ0$ adopted in the *training* MCS weights differently remote and recent observations. According to Taylor (2008b), the idea is that recent observations should have more weight with respect to remote observations. We adopt the same *exponentially weighted* approach of the RiskMetrics (Riskmetrics, 1996) model to weight differently the observations.

## 4. Empirical analysis

The empirical analysis uses daily data on S&P 500 index collected from the Oxford-Man Institute's Realized Library. The full sample covers the period from January 14, 2013 to May 31, 2022, for $T = 2350$ daily observations. High-frequency variables are the realized volatility (Andersen et al., 2001) at 5 minutes and the realized bipower variation (Barndorff-Nielsen and Shephard, 2004) with subsampling. The low-frequency variable is the Economic Policy Uncertainty (EPU, Baker et al., 2016), observed monthly.

Let $M$ be the set of candidate models, $T_{in}$ the length of the rolling period, $lstep$ the number of static one-step-ahead forecasts generated by each candidate model, $T$ the sample size, and $nstep = (T - T_{in})/lstep$ the number of steps employed in the algorithm. The algorithm for obtaining the combined predictors is as follows:

1. Estimate all the candidate models over the window including observations from the period $i = 1 + j$ to $i = T_{in} + j$. Conditionally on the estimated parameters, generate for the following $lstep$ days both $VaR$ and $ES$ one-step ahead forecasts, with $m = 1, \cdots, M$.
2. Compute the *training* MCS over the training period going from $i = 1 + j$ to $i = T_{in} + j$.
3. Obtain the proposed combined predictors $VaR^{Comb}_{(T_{in}+j+1):(T_{in}+lstep+j)}$ and $ES^{Comb}_{(T_{in}+j+1):(T_{in}+lstep+j)}$.
4. Iterate steps 1, 2, and 3, with $j = \{0, lstep, 2lstep, \cdots, (nstep-1)lstep\}$.

We use $M = 20$ models, $T_{in} = 1000$, and $lstep = 25$. Globally, the out-of-sample period consists of 1350 observations from 3 January 2017 to 31 May 2022. The coverage level chosen is $\alpha = 0.025$. Before evaluating the proposed combined predictors together with the two additional benchmarks, that is, the equally weighted combination (EW-Comb) and the median combination (Median-Comb), and all the set of competing models, we focus on the source of uncertainty associated with the selection of models. Figures 1 and 2 report the models entering the SSM in blue. Interestingly, no specification consistently enters the SSM for all the periods considered, independently of the use of the unweighted (Figure 1) or weighted (Figure 2) $FZ0$ loss. In other words, there is strong evidence of model uncertainty.

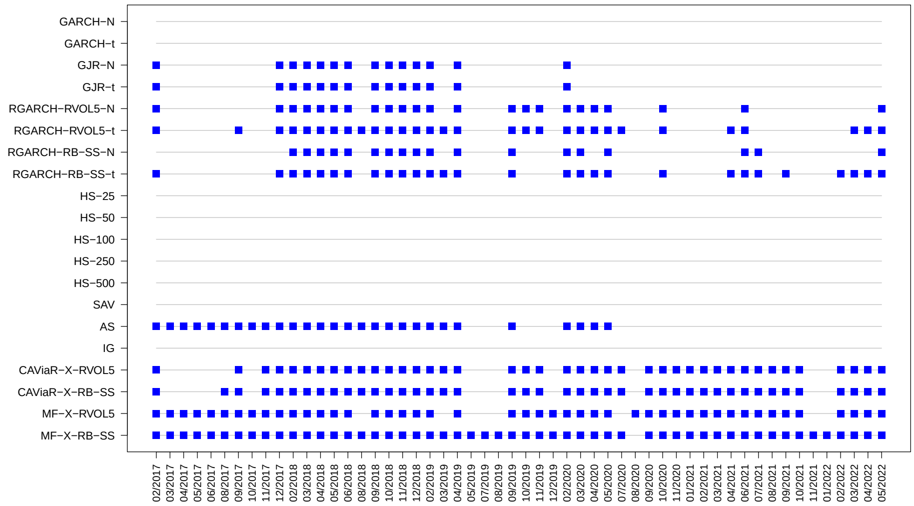


Figure 1: Training MCS, Unweighted $FZ0$ loss

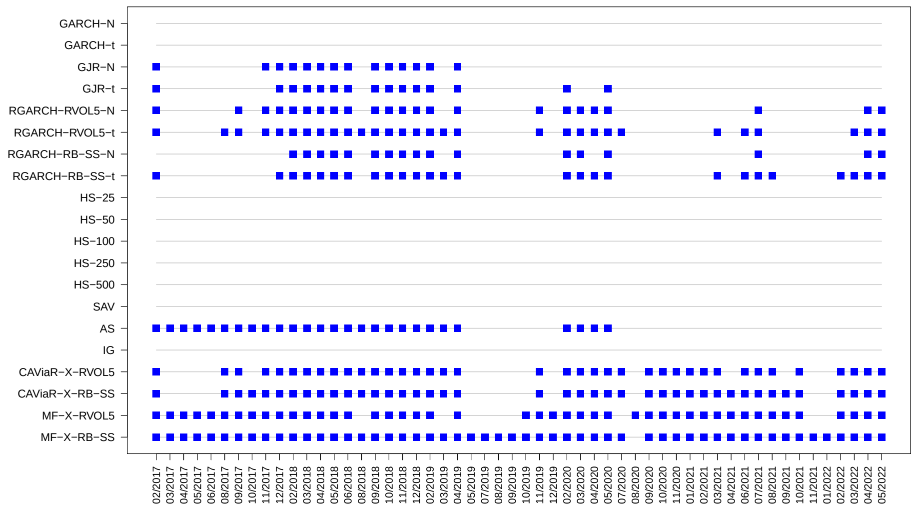


Figure 2: Training MCS, Weighted $FZ0$ loss

The combined predictors MCS-Comb and WL-MCS-Comb are then obtained equally weighting all the VaR and ES forecasts of the models enters the SSM as illustrated in Figures 1 and 2. Table 2 reports the violation rate (VR, in percentage) in the first column, the p-values of the UC, CC, and DQ tests for VaR in the columns from two to four, the p-values of the BD tests for

VaR and ES in columns from five to seven, the average of the $FZ0$ loss in the column eight, and the standard deviation (SD) of the VaR forecasts for all the candidate models as well as the two combined benchmarks and the two proposed combined predictors, for the whole out-of-sample period. Dark shades of gray in the penultimate column indicate that the model in the row enters the SSM of the MCS procedure at the significance level $\alpha = 0.25$. Light shades of gray indicate that the model in the row passes all the backtesting procedures at the significance level $\alpha = 0.05$. While many models (mainly some parametric and all the non-parametric specifications) do not perform well, the two proposed combined predictors pass the usual backtesting procedures at 5% significance level and, moreover, enter the SSM of the final MCS test. Notably, the proposed combined predictors have a smaller VaR standard deviations compared to the other models entering the SSM.

The plot of the log-returns with the predicted VaR (red line) and ES (blue lines) obtained from the proposed MCS-Comb predictor for the out-of-sample period is reported in Figure 3.

Table 2: Backtesting and MCS evaluation

| | VR(%) | UC | CC | DQ | BD-1 | BD-2 | BD-3 | MCS | SD |
|---|---|---|---|---|---|---|---|---|---|
| GARCH-N | 3.630 | 0.013 | 0.044 | 0.115 | 0.254 | 0.253 | 0.001 | -3.492 | 0.014 |
| GARCH-t | 3.556 | 0.019 | 0.063 | 0.168 | 0.634 | 0.627 | 0.345 | -3.545 | 0.015 |
| GJR-N | 3.852 | 0.003 | 0.010 | 0.021 | 0.237 | 0.237 | 0.017 | -3.521 | 0.016 |
| GJR-t | 3.333 | 0.062 | 0.161 | 0.172 | 0.191 | 0.187 | 0.518 | -3.577 | 0.017 |
| RGARCH-RVOL5-N | 4.741 | 0.000 | 0.000 | 0.000 | 0.147 | 1.000 | 0.739 | -3.495 | 0.010 |
| RGARCH-RVOL5-t | 5.037 | 0.000 | 0.000 | 0.000 | 1.000 | 1.000 | 1.000 | -3.533 | 0.010 |
| RGARCH-RB-SS-N | 5.259 | 0.000 | 0.000 | 0.000 | 0.088 | 0.087 | 0.722 | -3.452 | 0.010 |
| RGARCH-RB-SS-t | 6.000 | 0.000 | 0.000 | 0.000 | 0.157 | 0.160 | 0.036 | -3.470 | 0.027 |
| HS-25 | 6.519 | 0.000 | 0.000 | 0.000 | 0.031 | 0.012 | 0.000 | -2.797 | 0.016 |
| HS-50 | 4.593 | 0.000 | 0.000 | 0.000 | 0.046 | 0.031 | 0.000 | -3.106 | 0.017 |
| HS-100 | 4.148 | 0.000 | 0.001 | 0.000 | 0.117 | 0.177 | 0.095 | -3.127 | 0.015 |
| HS-250 | 4.074 | 0.001 | 0.002 | 0.000 | 0.100 | 0.100 | 0.426 | -3.143 | 0.012 |
| HS-500 | 3.704 | 0.008 | 0.001 | 0.000 | 0.073 | 0.073 | 0.481 | -3.046 | 0.008 |
| SAV | 3.185 | 0.122 | 0.285 | 0.135 | 0.965 | 1.000 | 0.987 | -3.572 | 0.014 |
| AS | 4.074 | 0.001 | 0.003 | 0.001 | 0.484 | 0.515 | 0.665 | -3.552 | 0.016 |
| IG | 3.407 | 0.043 | 0.113 | 0.066 | 0.965 | 1.000 | 0.996 | -3.588 | 0.014 |
| CAViaR-X-RVOL5 | 2.889 | 0.372 | 0.500 | 0.224 | 0.938 | 0.908 | 0.739 | -3.723 | 0.017 |
| CAViaR-X-RB-SS | 2.815 | 0.468 | 0.544 | 0.231 | 0.951 | 0.881 | 0.999 | -3.750 | 0.016 |
| MF-X-RVOL5 | 3.111 | 0.166 | 0.367 | 0.548 | 0.913 | 0.925 | 0.748 | -3.649 | 0.019 |
| MF-X-RB-SS | 3.333 | 0.062 | 0.161 | 0.327 | 0.952 | 0.952 | 0.825 | -3.672 | 0.018 |
| EW-Comb | 3.111 | 0.166 | 0.323 | 0.143 | 0.742 | 0.774 | 0.576 | -3.649 | 0.012 |
| Median-Comb | 3.556 | 0.019 | 0.063 | 0.032 | 0.567 | 0.517 | 0.188 | -3.648 | 0.013 |
| MCS-Comb | 3.259 | 0.088 | 0.209 | 0.102 | 0.855 | 0.827 | 0.233 | -3.704 | 0.016 |
| WL-MCS-Comb | 3.111 | 0.166 | 0.323 | 0.123 | 0.857 | 0.839 | 0.339 | -3.712 | 0.015 |

**Notes**: Sample period: 2017-01-03 to 2022-05-31 (1350 observations). Column MCS represents the averages of the $FZLoss$. Dark shades of gray denote the inclusion in the SSM, at significance level $\alpha = 0.25$. Light shades of gray denote the success in all the considered backtesting procedures, at significance level $\alpha = 0.05$.

## 5. Conclusion

In this paper, we propose two forecast combination strategies based on the use of strictly consistent joint VaR and ES loss functions and MCS in order to mitigate the impact of model

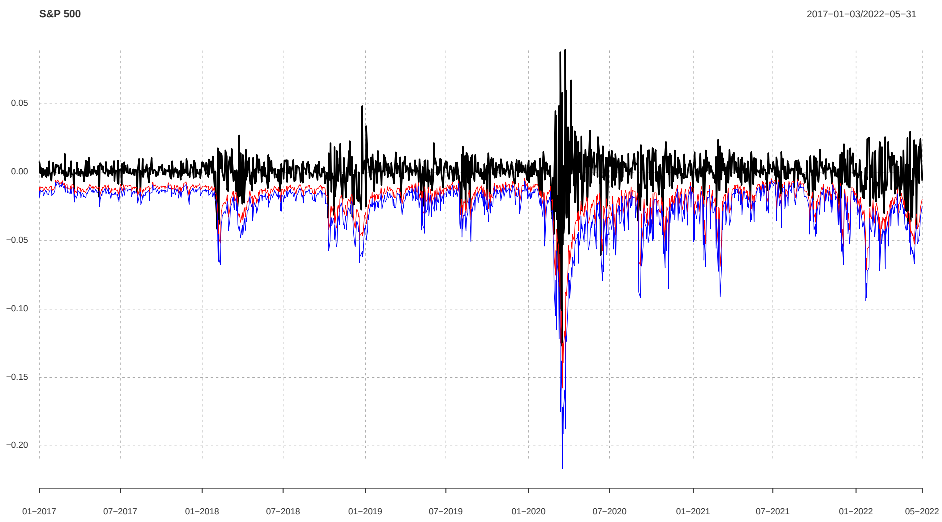


Figure 3: S&P 500 Log-returns (black line), VaR (red line) and ES (blue line) of the MCS-Comb predictor

uncertainty in tail risk forecasting. The proposed combined predictors are adaptive, as the composition and size of the set of best-performing models used in the combined forecast changes over time without requiring parameter estimation. The empirical study that combines predictions from 20 independent models on the S&P 500 index shows that the proposed combined predictors successfully pass the backtesting procedures and, compared with benchmarks and most individual models, enter the 75% evaluation MCS. Finally, the standard deviations of VaR forecasts for the combined predictors are generally lower than those of competing models. For future research, it would be interesting to investigate further combination strategies and expand the set of individual models by also considering the impact of a larger number of exogenous variables.

## References


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

Amendola, A., M. Braione, V. Candila, and G. Storti (2020). A model confidence set approach to the combination of multivariate volatility forecasts. *International Journal of Forecasting 36*(3), 873–891.

Amendola, A. and G. Storti (2008). A GMM procedure for combining volatility forecasts. *Computational Statistics & Data Analysis 52*(6), 3047–3060.

Amendola, A. and G. Storti (2015). Model uncertainty and forecast combination in high-dimensional multivariate volatility prediction. *Journal of Forecasting 34*(2), 83–91.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association 96*(453), 42–55.

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

Baker, S. R., N. Bloom, and S. J. Davis (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics 131*(4), 1593–1636.

Barndorff-Nielsen, O. and N. Shephard (2004). Power and bipower variation with stochastic volatility and jumps (with discussion). *Journal of Financial Econometrics 2*, 1–48.

Bates, J. M. and C. W. Granger (1969). The combination of forecasts. *Journal of the Operational*

*Research Society 20*(4), 451–468.
Bayer, S. (2018). Combining value-at-risk forecasts using penalized quantile regressions. *Econometrics and Statistics 8*, 56–77.
Bayer, S. and T. Dimitriadis (2022). Regression-based expected shortfall backtesting. *Journal of Financial Econometrics 20*(3), 437–471.
Becker, R. and A. E. Clements (2008). Are combination forecasts of S&P 500 volatility statistically superior? *International Journal of Forecasting 24*(1), 122–133.
Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*(3), 307–327.
Caldeira, J. F., G. V. Moura, F. J. Nogales, and A. A. Santos (2017). Combining multivariate volatility forecasts: an economic-based approach. *Journal of Financial Econometrics 15*(2), 247–285.
Candila, V., G. M. Gallo, and L. Petrella (2023). Mixed-frequency quantile regressions to forecast value-at-risk and expected shortfall. *Annals of Operations Research*, 1–34.
Christoffersen, P. and S. Gonçalves (2005). Estimation risk in financial risk management. *The Journal of Risk 7*(3), 1.
Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, 841–862.
Clemen, R. (1989). Combining forecast: a review and annotated bibliography. *International J. of Forecasting 5*(2), 559–581.
Engle, R. F., E. Ghysels, and B. Sohn (2013). Stock market volatility and macroeconomic fundamentals. *Review of Economics and Statistics 95*(3), 776–797.
Engle, R. F. and S. Manganelli (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics 22*(4), 367–381.
Fissler, T. and J. F. Ziegel (2016). Higher order elicitability and Osband's principle. *The Annals of Statistics 44*(4), 1680 – 1707.
Gerlach, R. and C. Wang (2020). Semi-parametric dynamic asymmetric Laplace models for tail risk forecasting, incorporating realized measures. *International Journal of Forecasting 36*(2), 489–506.
Glosten, L. R., R. Jagannanthan, and D. E. Runkle (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance 48*(5), 1779–1801.
Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association 106*(494), 746–762.
Halbleib, R. and W. Pohlmeier (2012). Improving the value at risk forecasts: Theory and evidence from financial crisis. *Journal of Economic Dynamics and Control 36*(8), 1212–1228.
Hansen, P. R., Z. Huang, and H. H. Shek (2012). Realized GARCH: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics 27*(6), 877–906.
Hansen, P. R., A. Lunde, and J. M. Nason (2011). The model confidence set. *Econometrica 79*(2), 453–497.
Hendricks, D. (1996). Evaluation of value-at-risk models using historical data. *Economic Policy Review 2*(1), 39–69.
Jeon, J. and J. W. Taylor. Using caviar models with implied volatility for value-at-risk estimation.
Komunjer, I. (2013). Quantile prediction. In G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting* , Volume 2, Chapter 17, pp. 961–994. Elsevier.
Kupiec, P. H. (1995). *Techniques for verifying the accuracy of risk measurement models*, Volume 95. Division of Research and Statistics, Division of Monetary Affairs, Federal.
Ma, F., Y. Li, L. Liu, and Y. Zhang (2018). Are low-frequency data really uninformative? A forecasting combination perspective. *The North American Journal of Economics and Finance 44*, 92–108.
Nieto, M. and E. Ruiz (2016). Frontiers in VaR forecasting and backtesting. *International Journal of Forecasting 32*, 475 – 501.

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

Riskmetrics (1996). Riskmetrics technical document.

Sarykalin, S., G. Serraino, and S. Uryasev (2008). *Value-at-Risk vs. Conditional Value-at-Risk in Risk Management and Optimization*, Chapter 13, pp. 270–294. Informs.

Storti, G. and C. Wang (2023). Modeling uncertainty in financial tail risk: A forecast combination and weighted quantile approach. *Journal of Forecasting*.

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

Taylor, J. W. (2008b). Exponentially weighted information criteria for selecting among forecasting models. *International Journal of Forecasting 24*(3), 513–524.

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.

Taylor, J. W. (2020a). Forecast combinations for value at risk and expected shortfall. *International Journal of Forecasting 36*(2), 428–441.

Taylor, J. W. (2020b). Model uncertainty and forecast combination in high-dimensional multivariate volatility prediction. *International Journal of Forecasting 36*, 428–441.

Timmermann, A. (2006). Forecast combination. In *Handbook of Economic Forecasting*, pp. 135–196. Elsevier.

Wang, X., R. Hyndman, and F. LI (2022). Forecast combinations: an over 50-year review. In *Workingpaper*. arXiv:2205.04216v2.