國立臺灣大學管理學院財務金融學系

碩士論文

Department of Finance

College of Management

National Taiwan University

Master’s Thesis

GARCH-MIDAS 對波動性與涉險值預測的應用

An Empirical Application of GARCH-MIDAS for Volatility and Value-at-Risk Forecasts

蔡秉叡

PING-JUI Tsai

指導教授: 陳宜廷 博士

Advisor: YI-TING Chen Ph.D.

中華民國 114 年 7 月

July, 2025

doi:10.6342/NTU202502186

摘要

本文應用 Fang et al. (2020, Journal of Empirical Finance) 所提出結合懲罰項之自我迴歸條件異質變異數-混合頻率模型 (GARCH-MIDAS)，進行股票與債券市場報酬率之波動度與涉險值（VaR）的日頻率預測。在實證過程中考慮了多個月頻率的預測變數；此外，也加入傳染病風險變數做為新的解釋變數。實證結果表示所考慮的實證設定在股票市場波動度與涉險值預測均具有相對的優勢；另外，就債券市場的波動度與涉險值預測而言，雖整體的預測能力下降，但本文的實證設定尚能展現一定的預測能力。

關鍵字：結合懲罰項之混合頻率-自我迴歸條件異質變異數模型、波動度預測、涉險值預測

doi:10.6342/NTU202502186

Abstract

This thesis applies the generalized autoregressive conditional heteroskedasticity mixed data sampling (GARCH-MIDAS) with penalized framework proposed by Fang et al. (2020, Journal of Empirical Finance) to generate daily forecasts of volatility and Value at Risk (VaR) for stock and bond index returns. Beyond daily returns, we incorporate several monthly macroeconomic and financial predictors, and we further augment the model with indicators that capture infectious disease risk. Empirically, our specification outperforms both the traditional GJR-GARCH model and the quarterly-frequency GARCH-MIDAS framework. While its overall predictive power declines for bond index returns, the proposed specification still delivers a meaningful level of accuracy in that market.

Keywords: GARCH-MIDAS with penlized framework, Volatility forecast, Value at Risk forecast

doi:10.6342/NTU202502186

Contents

doi:10.6342/NTU202502186

doi:10.6342/NTU202502186

List of Figures

doi:10.6342/NTU202502186

List of Tables

doi:10.6342/NTU202502186

Chapter 1 Introduction

Effective risk management relies on accurate volatility forecasts, which are crucial for informed financial decisions and portfolio allocation. The generalized autoregressive conditional heteroskedasticity mixed data sampling (GARCH-MIDAS) model, proposed by Engle et al. (2013), stands out. The model is notable for decomposing volatility into short- and long-term components, where the persistent element is governed by low-frequency explanatory factors. This structure allows macroeconomic fundamentals to inform high-frequency volatility modeling. Empirical studies consistently show that GARCH-MIDAS models outperform alternatives across forecasting horizons due to their ability to incorporate macroeconomic drivers of long-term volatility (Engle et al., 2013; Asgharian et al., 2013; Conrad and Loch, 2015; Conrad and Kleen, 2020). The GARCH-MIDAS model has been used to examine risk factors influencing diverse markets, including energy markets (Salisu et al., 2024), European Union Allowance futures (Lu et al., 2024), and foreign exchange markets (Eniayewu et al., 2024). Additionally, the GARCH-MIDAS framework has been employed to investigate interdependencies among different markets (Jiang et al., 2021).

However, standard GARCH-MIDAS models often rely on the maximum-likelihood (ML) method, which limits the number of explanatory variables (Conrad and Kleen, 2020). To address this, Fang et al. (2020) extend the model using adaptive Lasso introduced by

doi:10.6342/NTU202502186

Zou (2006), which enables estimation with twenty variables in a ML estimator (MLE) framework.

Given the strong potential of combining the GARCH-MIDAS model with a variety of predictors, this empirical study adopts the GARCH-MIDAS framework and the penalized estimation approach introduced by Fang et al. (2020), but extends the empirical analysis in several key ways. Specifically, we use the model that incorporates multiple monthly economic and financial variables, in contrast to Fang et al. (2020), which employs quarterly variables. The use of monthly variables aligns with prior studies (Engle et al., 2013; Conrad and Loch, 2015; Conrad and Kleen, 2020), where penalized estimation is not employed, as quarterly variables may be too infrequent to effectively capture the dynamic nature of volatility.

Furthermore, we incorporate the infectious disease index (INFECT), developed by Baker et al. (2019), into the GARCH-MIDAS framework.[1] The relevance of INFECT is underscored in the post-COVID-19 environment, where pandemic-related disruptions have significantly influenced financial market volatility (Albulescu, 2021). Moreover, we revisit Segnon et al. (2024), where two geopolitical risk indices, the historical geopolitical threats index (GPRHT) and the historical geopolitical acts index (GPRHA), are introduced to the GARCH-MIDAS literature. These indices, which capture perceived geopolitical threats and actual geopolitical events, respectively, have demonstrated strong predictive power in recent applications and offer further insights into the role of geopolitical uncertainty in financial markets. We also incorporate the global economic policy uncertainty index (GEPU), proposed by Baker et al. (2016) and introduced to GARCH-MIDAS mod-

[1] Liu et al. (2022) and Raza et al. (2023) investigated how the parameters of the GARCH-MIDAS model changed before and after the COVID-19 pandemic, rather than explicitly incorporating pandemic-related indices into the model.

doi:10.6342/NTU202502186

els by Fang et al. (2018). Considering the increasing uncertainty of economic policy, we adopt the GEPU index as a relevant explanatory variable.

In total, we incorporate twenty macroeconomic and financial variables along with their respective volatility measures (excluding realized volatility). This broad set of predictors enables the model to capture a wide spectrum of macroeconomic influences on financial market uncertainty.

An additional contribution of this study is the extension of the model on bond market volatility, a domain largely overlooked in prior research, which has primarily focused on equity markets (Engle et al., 2013; Asgharian et al., 2013; Conrad and Kleen, 2020; Fang et al., 2020). Our empirical analysis considers two benchmark ETFs: the SPDR S&P 500 ETF Trust (SPY) representing the stock market, and the iShares Core U.S. Aggregate Bond ETF (AGG) representing the bond market. This focus allows us to examine the effectiveness of the model across different asset classes and to identify key macroeconomic drivers specific to bond market volatility.

Additionally, the thesis emphasizes the prediction of tail risk, which standard volatility models often fail to capture. By utilizing multiple explanatory variables within the GARCH-MIDAS framework, we aim to improve Value-at-Risk (VaR) forecasting performance, providing a more robust measure of downside risk.

Overall, while this thesis adopts the penalized framework proposed by Fang et al. (2020), it departs in several important empirical respects. First, we utilize monthly variables rather than quarterly variables. Second, we introduce INFECT to the GARCH-MIDAS framework. Third, the volatility of macroeconomic and financial variables is utilized into the model specification. Fourth, we extend the scope of the GARCH-MIDAS

doi:10.6342/NTU202502186

framework by applying it to the bond market. Fifth, beyond volatility forecasting, we apply the model to VaR prediction.

Our empirical analysis yields the following key findings. First, INFECT and its volatility exhibit a significant influence on stock market uncertainty. Second, the predictive power of GPRHT and GPRHA, which is emphasized in Segnon et al. (2024), diminishes within the penalized estimation framework, rendering them statistically insignificant. A similar lack of significance is also observed for GEPU. Third, the GARCH-MIDAS model with the penalized estimation using monthly variables achieves relatively strong predictive accuracy for volatility across competing models. However, the model tends to underpredict volatility in the bond market. Fourth, for VaR forecasting, the GARCH-MIDAS model with the penalized estimation with monthly variables outperforms alternative models for SPY. However, none of the models successfully capture the tail risk of AGG. Fifth, the GARCH-MIDAS model demonstrates superior performance across all evaluation metrics when using monthly variables, compared to the use of quarterly variables. Sixth, the results indicate that a penalized estimation framework is essential for the GARCH-MIDAS model when incorporating multiple explanatory variables.

Following are the remaining sections of the thesis. Chapter 2 introduces the modeling framework, estimation procedures, and forecasting methods. Chapter 3 outlines the empirical design and provides summary statistics. Chapter 4 presents the estimation and forecasting performance results. Finally, Chapter 5 concludes the thesis.

doi:10.6342/NTU202502186

Chapter 2 Econometric Method

We introduce this model and the estimation method for completeness, without claiming new methodological contributions. The model framework discussed here follows Engle et al. (2013) and Conrad and Loch (2015). The estimation procedure mainly follows Fang et al. (2020).

2.1 Model

The daily return $r_{i,t}$ at day i in month t is modeled as:[1]

$$
\begin{aligned}
r_{i,t} &= \mu + \sqrt{h_{i,t}}\,\varepsilon_{i,t},\ \forall i = 1, \ldots, N_t, \\
h_{i,t} &= \tau_t g_{i,t},
\end{aligned} \tag{2.1}
$$

where $\mu = \mathrm{E}(r_{i,t} \mid \mathcal{F}_{i-1,t})$ is a conditional expected value of daily return given the information set $\mathcal{F}_{i-1,t}$ up to day $i-1$ in month t, which is treated as a constant in our thesis, the conditional variance of daily return $h_{i,t} = \mathrm{var}(r_{i,t} \mid \mathcal{F}_{i-1,t})$ is decomposed to a long-term component τ_t in month t and a short-term component $g_{i,t}$ on day i of month t, the error term $\varepsilon_{i,t}$ is a random variable that follows a standardized skewed-t distribution, denoted as S, with zero mean and unit standard deviation proposed by Hansen (1994).

[1] See Equation (3) of Engle et al. (2013).

doi:10.6342/NTU202502186

More specifically, the long-term component τ_t with J explanatory variables is modeled as:[2]

$$\ln \tau_t = m + \sum_{j=1}^{J} \theta_j \sum_{k=1}^{K} \psi_k(\omega_{j,1}, \omega_{j,2}) X_{j,t-k}, \tag{2.2}$$

where θ_j is the impact of the jth variable on $\ln \tau_t$, $X_{j,t}$ is the jth explanatory in the month t, and $\psi_k(\omega_{j,1}, \omega_{j,2})$ is the beta weighting function of jth variable for past K months, which is specified as:

$$\psi_k(\omega_{j,1}, \omega_{j,2}) = \frac{\left(\frac{k}{K}\right)^{\omega_{j,1}-1}\left(1-\frac{k}{K}\right)^{\omega_{j,2}-1}}{\sum_{k=1}^{K}\left(\frac{k}{K}\right)^{\omega_{j,1}-1}\left(1-\frac{k}{K}\right)^{\omega_{j,2}-1}}, \tag{2.3}$$

where $\omega_{j,1} \geq 1$ and $\omega_{j,2} > 0$ determine the humped shape of the weighting scheme and the function sums up to 1.[3]

The short-term component $g_{i,t}$ captures daily fluctuations and follows a mean-reverting unit-variance GJR-GARCH process as:[4]

$$g_{i,t} = (1 - \alpha - \beta - \frac{\gamma}{2}) + (\alpha + \gamma \mathbf{1}_{r_{i-1,t}-\mu<0}) \frac{(r_{i-1,t} - \mu)^2}{\tau_t} + \beta g_{i-1,t}, \tag{2.4}$$

where $\alpha > 0$, $\beta > 0$, $\alpha + \gamma > 0$, $\alpha + \beta + \gamma/2 < 1$, and $\mathbf{1}$ is the indicator function that equals one if $r_{i-1,t} - \mu < 0$, and zero otherwise. According to Conrad and Kleen (2020), $\mathrm{E}(g_{i,t}) = 1$ holds under certain assumptions.[5]

The probability density function of S has the degree-of-freedom parameter η and the

[2]See Equation (14) of Engle et al. (2013).
[3]See Ghysels et al. (2007) for more details.
[4]See Equation (3) of Conrad and Loch (2015).
[5]See Conrad and Kleen (2020, p. 3).

doi:10.6342/NTU202502186

skewness parameter λ. This density function is of the form:[6]

$$s(\varepsilon \mid \eta, \lambda) = \begin{cases} bc\left(1 + \frac{1}{\eta-2}\left(\frac{b\varepsilon+d}{1-\lambda}\right)^2\right)^{-\frac{\eta+1}{2}}, & \varepsilon < -\frac{d}{b}, \\ bc\left(1 + \frac{1}{\eta-2}\left(\frac{b\varepsilon+d}{1+\lambda}\right)^2\right)^{-\frac{\eta+1}{2}}, & \varepsilon \geq -\frac{d}{b}, \end{cases} \tag{2.5}$$

where $a = 4\lambda c\left(\frac{\eta-2}{\eta-1}\right)$, $b^2 = 1 + 3\lambda^2 - a^2$, and $c = \frac{\Gamma\left(\frac{\eta+1}{2}\right)}{\sqrt{\pi(\eta-2)}\Gamma\left(\frac{\eta}{2}\right)}$ with the parameters $2 < \eta < \infty$ and $-1 < \lambda < 1$.

2.2 Estimation Procedure

Let $\boldsymbol{\vartheta} = (\eta, \lambda, \mu, \alpha, \beta, \gamma, m, \theta_1, \cdots, \theta_J, \omega_{1,1}, \omega_{1,2}, \cdots, \omega_{J,1}, \omega_{J,2})$ be the parameter vector. ML method is used to estimate $\boldsymbol{\vartheta}$ with the log-likelihood function L based on the sample up to month T:

$$L(\boldsymbol{\vartheta}) = \sum_{t=1}^{T}\sum_{i=1}^{N_t} \ln\left(\frac{1}{\sqrt{\tau_t g_{i,t}}} s\left(\frac{r_{i,t}-\mu}{\sqrt{\tau_t g_{i,t}}} \mid \eta, \lambda\right)\right). \tag{2.6}$$

where N_t denotes the number of days in month t .

However, the parameter vector can be extremely large when J increases. A large parameter vector causes potential problem to identify the predictors which matter the most. Following Fang et al. (2020), we adopt the adaptive Lasso method introduced by Zou (2006) to address this issue. Combined with (2.6), the penalized log-likelihood function PL under the adaptive Lasso is:

$$PL_\kappa(\boldsymbol{\vartheta}) = L(\boldsymbol{\vartheta}) - \kappa \sum_{j=1}^{J} \hat{\mathrm{w}}_j \mid \theta_j \mid, \tag{2.7}$$

[6] See Equation (10) of Hansen (1994).

doi:10.6342/NTU202502186

where $\kappa > 0$ is the tuning parameter and $\hat{\mathrm{w}}$ is the adaptive weight. The adaptive weight $\hat{\mathrm{w}}_j$ is constructed based on initial estimates $\hat{\theta}_{o,j}$ obtained by maximizing the unpenalized log-likelihood function (2.6), with $\hat{\mathrm{w}}_j = 1/\mid \hat{\theta}_{o,j} \mid^2$ as recommended by Zou (2006).

For (2.7), selecting an appropriate tuning parameter κ for determining the suitable model is important. Following Fang et al. (2020), generalized information criteria (GIC) proposed by Fan and Tang (2012) is used to select the optimal κ, which is a BIC-type criterion. Fan and Tang (2012) showed that GIC outperforms AIC in selecting the optimal tuning parameter in simulations. The GIC of the tuning parameter κ, denoted as GIC_κ, is specified as:[7]

$$GIC_\kappa = \frac{1}{N_0}\left\{2\left[L(\hat{\boldsymbol{\vartheta}}_o) - PL_\kappa(\hat{\boldsymbol{\vartheta}}_\kappa)\right] + a(N_0, p)\left|\hat{\boldsymbol{\vartheta}}_\kappa\right|\right\}, \tag{2.8}$$

where $\hat{\boldsymbol{\vartheta}}_o$ is the MLE without the penalization, $\hat{\boldsymbol{\vartheta}}_\kappa$ is the penalized MLE with κ, N_0 is the number of total observations, $p = 3J + 1$ is the number of parameters in (2.2), and $\left|\hat{\boldsymbol{\vartheta}}_\kappa\right|$ shows the number of non-zero parameters. GIC_κ can be divided into two components: $L(\hat{\boldsymbol{\vartheta}}_o) - PL_\kappa(\hat{\boldsymbol{\vartheta}}_\kappa)$ which measures the goodness of fit, and $a(N_0, p)\left|\hat{\boldsymbol{\vartheta}}_\kappa\right|$ which represents the penalty on model complexity. Fan and Tang (2012) suggested the setting of $a(N_0, p) = \log\{\log(N_0)\}\log(p)$. We select κ in the range of $[0, 20]$.

In the estimation procedure, the Broyden – Fletcher – Goldfarb – Shanno algorithm[8] is utilized to maximize (2.6), while the proximal gradient descent method[9] is applied to maximize (2.7). To achieve convergence, we firstly maximize (2.6) with single explanatory variable in (2.2) for each X_j, then obtain its corresponding estimated θ_j, $\omega_{j,1}$ and $\omega_{j,2}$.

[7]See Equation (8) of Fang et al. (2020).

[8]The method is also used by Fang et al. (2020). The `constrOptim` function from the `stats` package in R is employed in this thesis.

[9]The backtracking line search algorithm is incorporated with proximal gradient descent to determine an appropriate step size.

doi:10.6342/NTU202502186

Denote the results as $\theta_j^\star$, $\omega_{j,1}^\star$ and $\omega_{j,2}^\star$, which represent the initial values for estimating the parameter vector $\boldsymbol{\vartheta}$ jointly with all J variables.

As noted by Fang et al. (2020, p. 4), some θ_j's are shrunk to zero as the tuning parameter κ increases. This potentially causes identification issue regarding $\omega_{j,1}$ and $\omega_{j,2}$ in (2.3). To address the identification issue and reduce computational complexity, we refer to the method suggested by Fang et al. (2020, p. 8). In their estimation method, the parameter vector $\boldsymbol{\vartheta}$ is divided into two subvectors, $\boldsymbol{\vartheta}_1$ and $\boldsymbol{\vartheta}_2$, where $\boldsymbol{\vartheta}_2 = (\omega_{1,1}, \omega_{1,2}, \cdots, \omega_{J,1}, \omega_{J,2})$ is the beta weighting function parameters, and $\boldsymbol{\vartheta}_1$ is the set of all remaining parameters in $\boldsymbol{\vartheta}$. We maximize (2.6) with all J variables and remain fixed in model selection process to estimate $\boldsymbol{\vartheta}_2$. See Fang et al. (2020, Table 1) for the computational procedure of their estimation method, and we let the parameter estimate vector be $\hat{\boldsymbol{\vartheta}} = \left(\hat{\eta}, \hat{\lambda}, \hat{\mu}, \hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{m}, \hat{\theta}_1, \cdots, \hat{\theta}_J, \hat{\omega}_{1,1}, \hat{\omega}_{1,2}, \cdots, \hat{\omega}_{J,1}, \hat{\omega}_{J,2}\right)$.

2.3 Forecasting

2.3.1 Volatility Forecast

Assume we aim to construct a k-day-ahead volatility forecast on the last day of month t, targeting a day i in month $t+f$ and denote the predicted volatility as $h_{i,t+f|\mathcal{F}_t}$, where $\mathcal{F}_t$ is the information set up to month t.

In the case where $f = 1$, since τ_{t+1} is contained in the information set $\mathcal{F}_t$, refer to (2.1), this forecast is defined as:

$$h_{i,t+1|\mathcal{F}_t} = \mathrm{var}(r_{i,t+1} \mid \mathcal{F}_t) = \mathrm{E}(\tau_{t+1} g_{i,t+1} \mid \mathcal{F}_t) = \tau_{t+1} \mathrm{E}\left(g_{i,t+1} \mid \mathcal{F}_t\right), \qquad (2.9)$$

doi:10.6342/NTU202502186

where $\mathrm{E}\left(g_{i,t+1} \mid \mathcal{F}_t\right) = 1 + \left(\alpha + \beta + \frac{\gamma}{2}\right)^{k-1} (g_{1,t+1|\mathcal{F}_t} - 1)$, in the population context.[10]

With the parameter estimate vector $\hat{\boldsymbol{\vartheta}}$, we replace unknown parameters with their estimates. Consequently, the aforementioned forecast is computed as:

$$\hat{h}_{i,t+1|\mathcal{F}_t} = \hat{\tau}_{t+1} \left(1 + \left(\hat{\alpha} + \hat{\beta} + \frac{\hat{\gamma}}{2}\right)^{k-1} (\hat{g}_{1,t+1} - 1)\right), \tag{2.10}$$

where the terms $\hat{\tau}_{t+1}$ and $\hat{g}_{1,t+1}$ are the estimated long-term and short-term volatility components, respectively, computed from (2.2) and (2.4) with estimated parameters.

In the case where $f > 1$, τ_{t+f} is not contained in the information set $\mathcal{F}_t$. Following Conrad and Kleen (2020, p. 10), we conduct the setting of forecast factorization: $\mathrm{E}(\tau_{t+f} \mid \mathcal{F}_t)\mathrm{E}(g_{i,t+f} \mid \mathcal{F}_t)$.[11] It is obvious that $\mathrm{E}(g_{i,t+f} \mid \mathcal{F}_t) = 1 + \left(\alpha + \beta + \frac{\gamma}{2}\right)^{k-1} (g_{1,t+1|\mathcal{F}_t} - 1)$. However, $\mathrm{E}(\tau_{t+f} \mid \mathcal{F}_t)$ is infeasible without modeling the dynamics of all explanatory variables. To solve this, we assume that the long-term component remains unchanged over the near forecast horizon, i.e., $\tau_{t+f|\mathcal{F}_t} = \tau_{t+1}$.[12]

With the parameter estimate vector $\hat{\boldsymbol{\vartheta}}$, we replace all unknown parameters with their estimates, then $h_{i,t+f|\mathcal{F}_t}$ can be computed similarly to (2.10).

2.3.2 Value-at-Risk Forecast

Assume we aim to construct a k-day-ahead VaR forecaset at probability level a, evaluated on the last day of month t and targeting day i in month $t+f$.[13] The corresponding

[10]See Conrad and Kleen (2020, p. 3) for the expression and $g_{1,t+1|\mathcal{F}_t}$ can be calculated from (2.4).
[11]This is based on the assumption that the short- and long-term components are independent.
[12]The assumption follows Conrad and Kleen (2020).
[13]The VaR at probability level a indicates that there is an a probability that the return will not fall below the VaR threshold.

doi:10.6342/NTU202502186

VaR is denoted by $VaR_{a,i,t+f|\mathcal{F}_t}$. This quantity is as:

$$VaR_{a,i,t+f|\mathcal{F}_t} = \mu + S^{-1}_{\eta,\lambda}(1-a)\sqrt{h_{i,t+f|\mathcal{F}_t}}, \tag{2.11}$$

where $S^{-1}_{\eta,\lambda}(\cdot)$ denotes the quantile function of S with parameters η and λ.

Given the parameter estimates $\hat{\boldsymbol{\vartheta}}$, and substituting in the estimated parameters, we compute (2.11) as:

$$\widehat{VaR}_{a,i,t+1|\mathcal{F}_t} = \hat{\mu} + S^{-1}_{\hat{\eta},\hat{\lambda}}(1-a)\sqrt{\hat{h}_{i,t+f|\mathcal{F}_t}}, \tag{2.12}$$

where $\widehat{VaR}_{a,i,t+1|\mathcal{F}_t}$ represents the predicted k-day-ahead VaR with estimated parameters.

doi:10.6342/NTU202502186

Chapter 3 Empirical Setting

Our dataset includes the stock price index SPY and the bond price index AGG, as well as monthly financial and macroeconomic variables from October 2003 to December 2024.[1] The dataset is partitioned into two distinct periods. The in-sample period, covering October 2003 to December 2021, is used for model estimation. The out-of-sample period, from January 2022 to December 2024, is reserved for evaluating the predictive performance of the models estimated during the in-sample period. We consider a two-year weighting scheme in (2.2), so $K = 24$ in our setting.

For the forecasting targets, SPY serves as a benchmark for overall stock market conditions. In comparison, AGG represents the broad bond market.

Regarding the predictors, we consider a total of seven financial variables and thirteen macroeconomic variables, including their respective volatility measures, except for the volatilities of realized volatilities. Compared to Fang et al. (2020), the set of financial variables almost remains the same, except for the realized volatility of AGG ($\mathrm{RV}_{\mathrm{AGG}}$) adding to our dataset. However, among macroeconomic variables, we exclude those provided exclusively at quarterly frequency, such as GDP, New Orders, and Corporate Profits.[2] This exclusion is also reasonable because these variables were found to be statistically

[1] The time span is 1969Q1 to 2018Q4 in Fang et al. (2020).

[2] Certain variables, such as Consumer Price Index (CPI) , are available at monthly frequency.

doi:10.6342/NTU202502186

insignificant in Fang et al. (2020).

Additionally, we attempt to enhance the empirical specification by including four interesting macroeconomic indicators: the INFECT, GEPU, GPRHT, and GPRHA. Also, we extend this approach by including the volatility measures for all macroeconomic and financial variables considered in the model, except for the realized volatilities. Tables 3.1 and 3.2 present the details of explanatory variables of financial and macroeconomic respectively. Table 3.3 presents the descriptive statistics and Table 3.4 shows the summary statistics for volatility data. In the following, we provide more details about the financial variables, the macroeconomic variables and their volatilities.

3.1 Financial Variables

Seven financial variables are utilized including: realized volatility of SPY ($\mathrm{RV}_{\mathrm{SPY}}$), realized volatility of AGG ($\mathrm{RV}_{\mathrm{AGG}}$), short-term reversal factor (STR), VIX of S&P 500 index (VIX), market return (MKT), default spread (DS), and term spread (TS). Among these variables, RV serves as a proxy for unobservable past volatility and is widely utilized in GARCH-MIDAS models (Engle et al., 2013; Conrad and Loch, 2015; Conrad and Kleen, 2020). The monthly realized volatilities for SPY and AGG are calculated as follows:[3]

$$RV_t = \sum_{i=1}^{N_t} r_{i,t}^2. \tag{3.1}$$

The MKT is the excess return of overall equity market, while the STR implies the uncertainty in the market. The VIX, derived from option prices, represents market expectations of future risk, making it a valuable explanatory variable in risk forecasting. The

[3] See Equation (9) in Fang et al. (2020).

doi:10.6342/NTU202502186

term spread reflects the shape of the yield curve. The default spread indicates the level of market risk aversion.[4] All financial variables are expressed in levels.

3.2 Macroeconomic Variables

We consider thirteen macroeconomic variables, including unemployment rate (UNRATE), consumer price index (CPI), product price index (PPI), consumer sentiment index (CS), monetary base (MB), personal consumption expenditure (PCE), industrial production index (IP), housing starts (HS), Chicago Fed national activity index (CFNAI), INFECT, GEPU, GPRHT, and GPRHA. Most of these variables have been considered in previous studies; see Engle et al. (2013), Conrad and Loch (2015), Conrad and Kleen (2020), Fang et al. (2020), and Segnon et al. (2024). The indices GPRHT and GPRHA represent geopolitical uncertainty risk related to threats and acts, respectively, as considered by Segnon et al. (2024), with further details provided by Caldara and Iacoviello (2022). The index GEPU captures global economic policy uncertainty, considered by Fang et al. (2018), which affects financial markets, corporate decisions, and risk management practices (Al-Thaqeb and Algharabali, 2019; Gulen and Ion, 2015). This thesis also introduces a new macroeconomic variable: INFECT. Post-COVID-19, infectious diseases have become a significant source of market uncertainty, prompting the inclusion of the INFECT index (Baker et al., 2020).

UNRATE, CS, GPRHT, GPRHA, GEPU, and INFECT are kept in levels. CPI, PPI, and PCE are expressed as year-over-year percentage changes.[5] All remaining variables are annualized month-over-month percentage changes.[6]

[4]See https://fred.stlouisfed.org/series/BAMLH0A0HYM2 for more details.
[5]Calculated as $100((X_t/X_{t-12}) - 1)$.
[6]Calculated as $100((X_t/X_{t-1})^{12} - 1)$.

doi:10.6342/NTU202502186

3.3 Volatility Variables

Engle et al. (2013) proposed that the volatility of explanatory variables can also impact overall market volatility. Building on the approach of Schwert (1989), we estimate monthly volatility for any given variable X (RV_{SPY} and RV_{AGG} are excluded) using an autoregressive model that includes twelve monthly dummy variables D_{jt}. Specifically, the squared residuals from the regression with OLS method are used to compute monthly volatility.[7] The regression is specified as:[8]

$$X_t = \sum_{j=1}^{12} \delta_j D_{jt} + \sum_{i=1}^{12} \zeta_i X_{t-i} + e_t. \tag{3.2}$$

Table 3.1: Financial variables.

Variable Name	Abbreviation	Description	Database
Financial Variables			
Realized Volatility of SPY Ticker	RV_{SPY}	The realized volatility of the return of SPY.	Yahoo
Realized Volatility of AGG Ticker	RV_{AGG}	The realized volatility of the return of AGG.	Yahoo
Market Return	MKT	The market return includes all NYSE, AMEX, and NASDAQ firms.	French
Short-Term Reversal Factor	STR	The index shows the degree of short-term reversal effects.	French
VIX of S&P 500 Index	VIX	The VIX of the S&P 500 index.	FRED
Default Spread	DS	The interest spreads between a computed OAS index of all bonds and a spot Treasury curve.	FRED
Term Spread	TS	The yield spreads between 3-month Treasury bills and 10-year Treasury bonds.	FRED

Note:
1. FRED: the database at the Federal Reserve Bank of St Louis (https://fred.stlouisfed.org).
2. Yahoo: Yahoo Finance (https://finance.yahoo.com).
3. French: The personal website of K. R. French (https://mba.tuck.dartmouth.edu/pages/faculty/ken.french).

[7] In the following context, the variable with the subscript ^{v} represents the volatility of the corresponding index.
[8] See Equation (21) of Engle et al. (2013).

doi:10.6342/NTU202502186

Table 3.2: Macroeconomic variables.

Variable Name	Abbreviation	Description	Database
Macroeconomic Variables			
Unemployment Rate	UNRATE	U.S. monthly unemployment rate.	FRED
Consumer Price Index	CPI	U.S. monthly consumer price index.	FRED
Producer Price Index	PPI	U.S. monthly producer price index.	FRED
Consumer Sentiment Index	CS	U.S. monthly consumer sentiment index.	SCUM
Monetary Base	MB	Currency in circulation plus reserve balances.	FRED
Personal Consumption Expenditure	PCE	U.S. monthly personal consumption index.	RDRC
Industrial Production Index	IP	U.S. real output of industrial sectors.	RDRC
Housing Starts	HS	Monthly count of new housing construction starts.	RDRC
Global Economic Policy Uncertainty Index	GEPU	News-based index on global economic uncertainty.	EPU
Historical Geopolitical Threats Index	GPRHT	Newspaper-based index on geopolitical threats.	Iacoviello
Historical Geopolitical Acts Index	GPRHA	Newspaper-based index on geopolitical acts.	Iacoviello
Chicago Fed National Activity Index	CFNAI	Index of U.S. economic activity relative to trend.	FRED
Infectious Disease Equity Market Volatility	INFECT	Volatility index based on pandemic-related news.	EPU

Note:
1. FRED: the database at the Federal Reserve Bank of St Louis (https://fred.stlouisfed.org).
2. SCUM: the Survey from University of Michigan (https://www.sca.isr.umich.edu/).
3. French: The personal website of K. R. French (https://mba.tuck.dartmouth.edu/pages/faculty/ken.french).
4. RDRC: Real-time Data Research Center of the Federal Reserve Bank of Philadelphia (https://www.philadelphiafed.org).
5. Iacoviello: the personal website of Matteo Iacoviello (https://www.matteoiacoviello.com/).
6. EPU: the Economic Policy Uncertainty team (https://www.policyuncertainty.com/).

doi:10.6342/NTU202502186

Table 3.3: Descriptive statistics.

	N	Min	Max	Mean	SD	Skewness	Kurtosis
Tickers							
SPY	5351	-3.984	2.456	0.032	0.768	-0.171	3.920
AGG	5351	-1.204	1.248	0.009	0.346	-0.094	2.762
Financial Variables							
RV_{SPY}	261	2.251	689.835	36.732	84.523	5.181	33.255
RV_{AGG}	261	0.231	97.752	2.828	6.964	9.108	97.516
MKT	261	-17.150	12.480	0.850	4.600	-0.162	0.885
STR	261	-11.630	15.770	0.008	2.849	0.655	2.750
VIX	261	10.125	62.669	20.087	9.277	1.943	4.123
DS	261	2.571	20.310	4.938	2.665	2.629	7.389
TS	261	-1.734	4.762	0.865	1.326	0.264	-0.387
Macroeconomic Variables							
UNRATE	261	-2.200	10.400	0.007	0.503	7.915	105.965
CPI	261	-1.958	8.999	2.672	1.963	0.614	0.409
PPI	261	-16.058	22.686	4.249	7.157	0.250	0.181
CS	261	50.000	103.800	81.234	12.944	-0.356	-0.235
MB	261	-66.801	1646.999	31.193	149.933	7.948	70.888
PCE	261	-14.385	30.084	4.420	5.357	1.297	5.769
IP	261	-76.016	88.085	3.304	12.985	0.368	7.155
HS	261	-97.605	1370.252	43.625	147.364	5.291	37.192
GPRHT	261	47.140	264.450	98.448	29.531	1.197	2.182
GPRHA	261	21.130	166.370	74.073	31.187	0.566	-0.253
GEPU	261	49.225	431.572	154.389	81.146	0.966	0.374
INFECT	261	0.096	50.215	5.452	8.876	2.619	7.107
CFNAI	261	-18.150	6.290	-0.186	1.606	-2.606	28.900

doi:10.6342/NTU202502186

Table 3.4: Descriptive statistics of volatility variables.

	N	Min	Max	Mean	SD	Skewness	Kurtosis
Financial Variables							
MKT^v	261	0.000	333.271	18.277	36.110	4.344	24.832
STR^v	261	0.000	178.089	9.885	21.517	3.990	20.610
VIX^v	261	0.000	1294.308	24.146	98.250	10.888	130.832
DS^v	261	0.000	26.092	0.531	1.866	10.897	135.486
TS^v	261	0.000	0.871	0.043	0.093	3.854	20.654
Economic Variables							
UNRATE^v	261	0.000	99.909	0.606	6.121	15.551	250.027
CPI^v	261	0.000	0.978	0.112	0.179	2.280	5.981
PPI^v	261	0.000	28.239	1.785	3.345	4.693	26.308
CS^v	261	0.000	278.945	20.693	34.290	3.968	20.915
MB^v	261	0.000	1429702.239	13581.614	93414.851	13.773	199.868
PCE^v	261	0.000	151.975	2.894	10.604	11.389	147.472
IP^v	261	0.000	4108.341	92.861	316.723	10.225	116.409
HS^v	261	0.000	1569663.955	30545.742	121227.487	11.439	141.105
GPRHT^v	261	0.000	9377.023	453.747	1052.701	5.410	34.618
GPRHA^v	261	0.000	5769.087	314.555	695.717	5.155	31.324
GEPU^v	261	0.000	16823.773	704.942	1507.048	5.991	43.814
INFECT^v	261	0.000	1655.790	6.965	45.693	31.687	1074.838
CFNAI^v	261	0.000	269.709	2.233	16.984	14.025	203.614

Note: ^{v} indicates that it’s volatility variables of the underlying variable.

doi:10.6342/NTU202502186

Chapter 4 Empirical Findings

4.1 In-sample Estimation

To facilitate the adaptive-Lasso penalty, all explanatory variables are standardized beforehand to control the scales. The estimation process is as outlined in Chapter 2.2. The tuning parameter κ is assessed over the range $[0, 20]$ with increments of 0.1 per iteration. The optimal κ is selected by the one has minimum GIC.

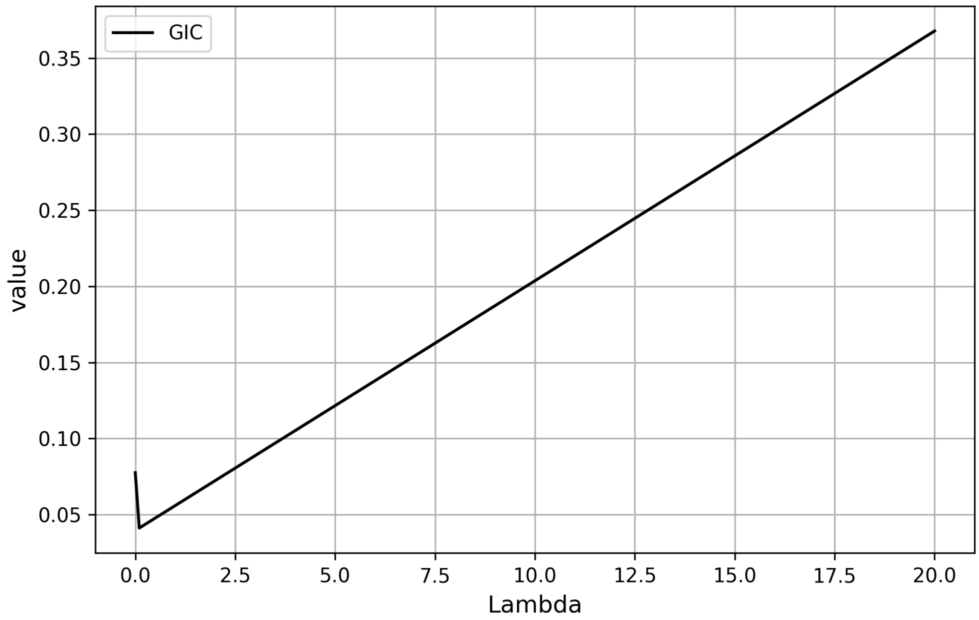
For the GARCH-MIDAS model applied to SPY, Figure 4.1a shows that the optimal value of κ is 0.1, resulting in the selection of 10 variables: RV_{AGG}, CSI, MKT, STR, INFECT, CFNAI, UNRATE^v, VIX^v, PCE^v, and INFECT^v. The corresponding estimation results are presented in Table 4.1. Although previous researches have emphasized the importance of RV_{SPY} in GARCH-MIDAS models (Engle et al., 2013; Asgharian et al., 2013; Fang et al., 2020), its significance disappears in our multivariate framework, a finding consistent with Fang et al. (2020). Similarly, the geopolitical risk indices GPRHT and GPRHA are excluded, which contrasts with the conclusions of Segnon et al. (2024). Moreover, GEPU is eliminated in the selection process, implying that economic policy uncertainty has no significant impact on stock market volatility. Notably, both the level and volatility components of INFECT are selected, aligning with the conclusion of Albulescu (2021) that pandemics contribute significantly to financial market uncertainty.

doi:10.6342/NTU202502186

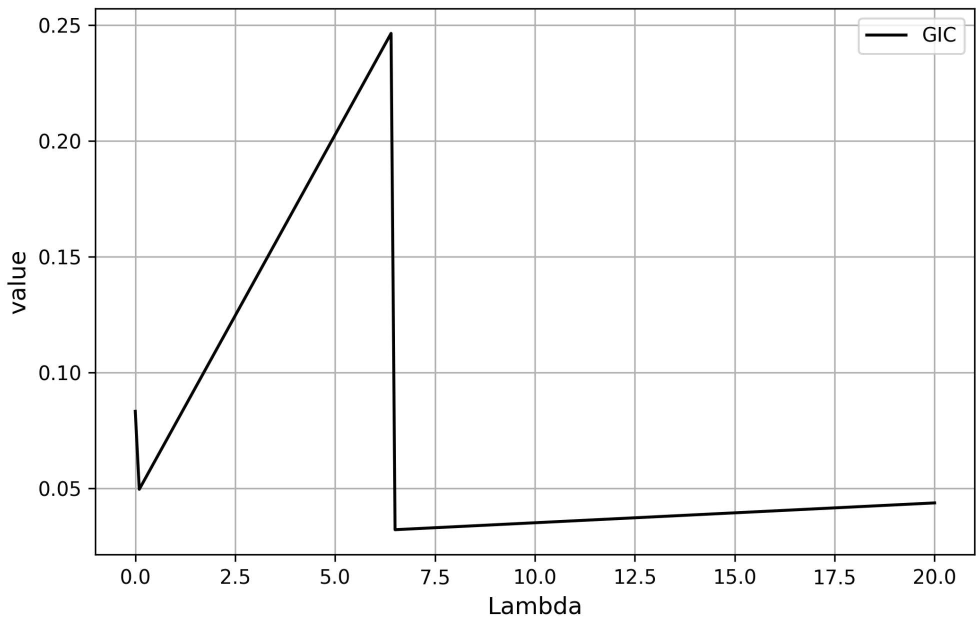
For the GARCH-MIDAS model of AGG, the tuning grid for λ_{Lasso} remains unchanged. As shown in Figure 4.1b, the optimal tuning parameter is 6.5, resulting in the selection of 2 variables: $\mathrm{RV}_{\mathrm{AGG}}$ and TS^{v}. Table 4.2 presents the corresponding estimation results. This implies that AGG volatility is influenced by its own past uncertainty and by the volatility of the term spread. It can be observed that $\mathrm{RV}_{\mathrm{AGG}}$ significantly affects both SPY and AGG returns, although the impacts exhibit opposite signs. This sign difference may reflect distinct market perceptions or risk transmission mechanisms across the stock and bond markets.

Figure 4.2 illustrates the weighting values assigned to several explanatory variables in both models. It can be observed that more recent observations receive more weights than the older ones, indicating that recent information has a stronger influence on future volatility. As shown in Figure 4.3, the standardized error distribution of SPY exhibits higher kurtosis and greater negative skewness than that of AGG, implying that the returns of SPY have heavier tails.

doi:10.6342/NTU202502186



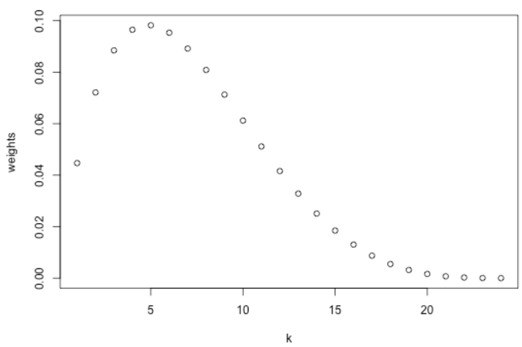
(a) The GIC corresponding to each λ on the grid [0,20] with 201 points on GARCH-MIDAS of SPY.



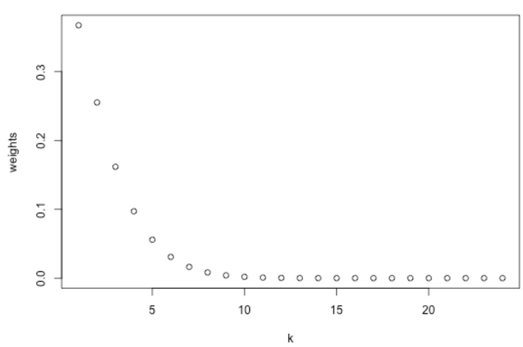
(b) The GIC corresponding to each λ on the grid [0,20] with 201 points on GARCH-MIDAS of AGG.

Figure 4.1: The GIC values for different λ on GARCH-MIDAS models of SPY and AGG.

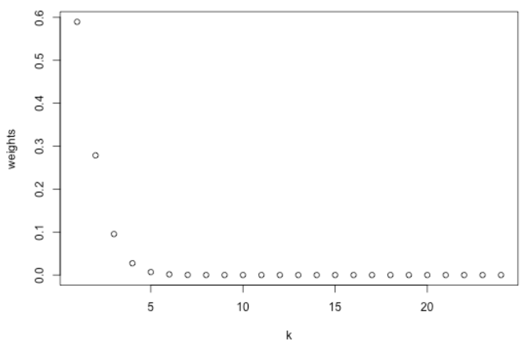
doi:10.6342/NTU202502186



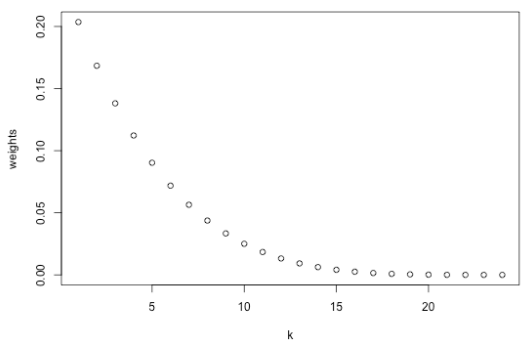
(a) Beta weighting of RV_{AGG} for GARCH-MIDAS model underlying SPY.



(b) Beta weighting of CSI for GARCH-MIDAS model underlying SPY.



(c) Beta weighting of RV_{AGG} for GARCH-MIDAS model underlying AGG.



(d) Beta weighting of TS^{v} for GARCH-MIDAS model underlying AGG.

Figure 4.2: Beta weightings for GARCH-MIDAS models.

doi:10.6342/NTU202502186

Table 4.1: Parameter estimates for GARCH-MIDAS of SPY.

μ and GARCH parameters				
μ	α	β	γ	
0.036	0.000	0.873***	0.243***	
(0.026)	(0.000)	(0.041)	(0.090)	
Parameters for the long-term component				
m	$\theta_{\text{RV}_{\text{AGG}}}$	θ_{CSI}	θ_{MKT}	θ_{STR}
1.388	−0.483**	0.166***	0.444***	−0.140***
(1.151)	(0.229)	(0.059)	(0.106)	(0.042)
θ_{INFECT}	θ_{CFNAI}	θ_{UNRATE^v}	θ_{VIX^v}	θ_{PCE^v}
0.343**	−0.550*	0.286*	−1.914*	−0.241***
(0.143)	(0.292)	(0.164)	(0.981)	(0.083)
θ_{INFECT^v}				
−0.499***				
(0.192)				
$\omega_{1,\text{RV}_{\text{AGG}}}$	$\omega_{1,\text{CSI}}$	$\omega_{1,\text{MKT}}$	$\omega_{1,\text{STR}}$	$\omega_{1,\text{INFECT}}$
1.929***	1.238***	10.660***	1.866***	1.116*
(0.493)	(0.348)	(1.384)	(0.539)	(0.621)
$\omega_{1,\text{CFNAI}}$	$\omega_{1,\text{UNRATE}^v}$	ω_{1,VIX^v}	ω_{1,PCE^v}	$\omega_{1,\text{INFECT}^v}$
1.836**	2.190	1.738***	8.117	1.316
(0.899)	(1.675)	(0.466)	(4.940)	(0.958)
$\omega_{2,\text{RV}_{\text{AGG}}}$	$\omega_{2,\text{CSI}}$	$\omega_{2,\text{MKT}}$	$\omega_{2,\text{STR}}$	$\omega_{2,\text{INFECT}}$
4.883***	13.432*	90.685***	2.662	12.748**
(1.050)	(7.028)	(2.660)	(1.952)	(5.850)
$\omega_{2,\text{CFNAI}}$	$\omega_{2,\text{UNRATE}^v}$	ω_{2,VIX^v}	ω_{2,PCE^v}	$\omega_{2,\text{INFECT}^v}$
4.138	5.698	2.975*	2.384	3.790*
(3.470)	(4.968)	(1.320)	(3.038)	(2.101)
Parameters for the distribution				
η_s	λ_s			
5.710***	−0.159			
(1.667)	(0.122)			

Note: The abbreviation indicates the variable to which the parameter corresponds. ^{v} means the volatility data. The numbers in parentheses are the Newey-West standard errors, and ***, **, and * indicate p-values of 1%, 5%, and 10%, respectively.

doi:10.6342/NTU202502186

Table 4.2: Parameter estimates for GARCH-MIDAS of AGG.

μ and GARCH parameters			
μ	α	β	γ
0.013**	0.039	0.894***	0.034
(0.007)	(0.056)	(0.060)	(0.060)
Parameters for the long-term component			
m	$\theta_{\text{RV}_{\text{AGG}}}$	θ_{TS^v}	
-2.928^{***}	0.465***	0.735***	
(0.085)	(0.184)	(0.310)	
$\omega_{1,\text{RV}_{\text{AGG}}}$	$\omega_{1,\text{TS}}$	$\omega_{2,\text{RV}_{\text{AGG}}}$	$\omega_{2,\text{TS}}$
1.906	1.000	33.374	5.459
(2.996)	(4.175)	(39.914)	(27.015)
Parameters for the distribution			
η	λ		
11.778***	-0.083		
(2.568)	(0.052)		

Note: The abbreviation indicates the variable to which the parameter corresponds. ^{v} means the volatility data. The numbers in parentheses are the Newey−West standard errors, and ***, **, and * indicate p-values of 1%, 5%, and 10%, respectively.

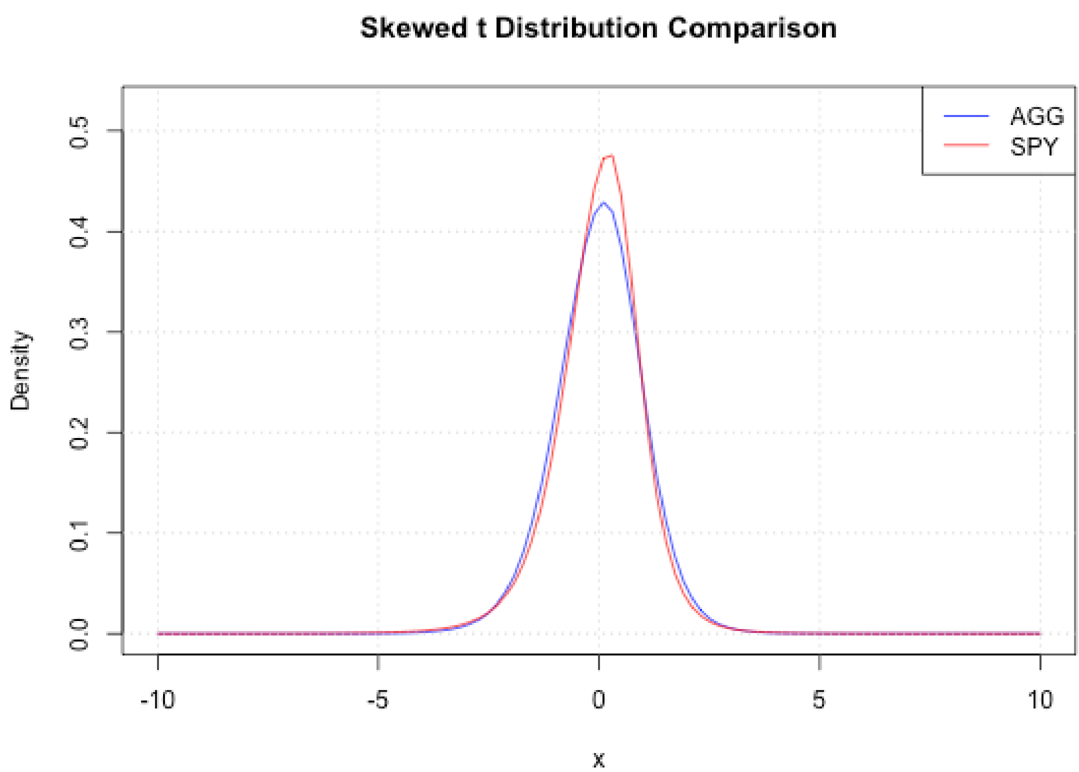


Figure 4.3: Standardized skewed-t distribution comparison between GARCH-MIDAS models underlying SPY and AGG respectively.

doi:10.6342/NTU202502186

4.2 Out-of-sample Forecast

4.2.1 Volatility Forecast

To assess the forecasting performance of volatility models during this period, it is necessary to compare the forecasts against the true underlying volatility. However, since true volatility is unobservable, we rely on an appropriate proxy. We adopt the Yang-Zhang volatility estimator proposed by Yang and Zhang (2000), an unbiased measure that effectively accounts for both overnight price jumps and intraday price movements.[1]

The Yang – Zhang volatility estimator, denoted as σ^2_{YZ}, combines three sources of return variation: the open-to-close variance (σ^2_o), the close-to-close variance (σ^2_c), and the Rogers – Satchell variance (σ^2_{RS}) proposed by Rogers and Satchell (1991).[2] It is defined as:[3]

$$\sigma^2_{\mathrm{YZ}} = \sigma^2_o + q\sigma^2_c + (1-q)\sigma^2_{\mathrm{RS}}, \tag{4.1}$$

where σ^2_o captures intraday volatility from open to close prices, σ^2_c reflects interday return variability based on consecutive closing prices, and the term σ^2_{RS} is the Rogers – Satchell variance, which adjusts for drift and is computed using the high, low, open, and close prices. The weighting parameter q determines the relative contributions of σ^2_c and σ^2_{rs}, and is specified as:

$$q = \frac{0.34}{1.34 + \frac{n+1}{n-1}},$$

[1] Most existing studies utilize high-frequency realized volatility, such as the 5-minute realized volatility, as a volatility proxy. However, due to limited access to high-frequency data, this thesis employs the Yang-Zhang volatility estimator as an alternative.

[2] See Equation (2) of Rogers and Satchell (1991).

[3] See Equation (7) of Yang and Zhang (2000).

doi:10.6342/NTU202502186

where n denotes the number of observations.

We assess the performance of volatility forecast using two loss measures: quasi-likelihood loss (QLIKE) and mean squared forecast error (MSFE). As shown in Patton (2011), QLIKE is especially robust to extreme observations by penalizing underprediction of volatility more heavily than overprediction, it aligns closely with risk management priorities. Lower QLIKE and MSFE values indicate better predictive performance.

For 1-day-ahead volatility forecast, let $\hat{h}_{i,t}$ be the forecasted volatility made 1 day prior and $\sigma^2_{\mathrm{YZ},i,t}$ be the σ^2_{YZ} on that day. The QLIKE is defined as:[4]

$$\mathrm{QLIKE} = \frac{1}{\sum_{t=T+1}^{T+M} N_t} \sum_{t=T+1}^{T+M} \sum_{i=1}^{N_t} \left\{ \frac{\sigma^2_{\mathrm{YZ},i,t}}{\hat{h}_{i,t}} - \ln\left(\frac{\sigma^2_{\mathrm{YZ},i,t}}{\hat{h}_{i,t}}\right) - 1 \right\}, \tag{4.2}$$

where M is the number of months in the out-of-sample period.

The corresponding MSFE is defined as:

$$\mathrm{MSFE} = \frac{1}{\sum_{t=T+1}^{T+M} N_t} \sum_{t=T+1}^{T+M} \sum_{i=1}^{N_t} \left(\sigma^2_{YZ,i,t} - \hat{h}_{i,t}\right)^2. \tag{4.3}$$

The forecasting capability of the proposed GARCH-MIDAS model is benchmarked against a range of competing volatility specifications across various forecasting horizons: 1-day, 1-week, 1-month, 1-quarter, 6-month, and 1-year ahead. The models considered are:

- **Model 1:** The GARCH-MIDAS model with the penalized estimation.

- **Model 2:** The GARCH-MIDAS model including all explanatory variables.

[4]For k-day-ahead volatility forecasts with $k > 1$, note that since predictions cannot be made earlier than the last day of month T, the forecasted time span will differ accordingly.

doi:10.6342/NTU202502186

- **Model 3:** The GARCH-MIDAS model with one explanatory variable.

- **Model 4:** The Standard GJR-GARCH model.

- **Model 5:** The GARCH-MIDAS model from Fang et al. (2020) which using quarterly data.

Model 1 is served as the benchmark and calculate the ratio against all the other models: $\text{QLIKE}/\text{QLIKE}_{\text{benchmark}}$ and $\text{MSFE}/\text{MSFE}_{\text{benchmark}}$. A ratio greater than 1 shows benchmark model is better than corresponding model. Additionally, to identify the superior models across different forecasting horizons, we apply the model confidence set (MCS) procedure developed by Hansen et al. (2011). Models retained in the MCS exhibit statistically indistinguishable predictive performance.[5]

In Tables 4.3 and 4.4, we report out-of-sample forecasting results for SPY using MSFE and QLIKE, respectively. Under MSFE, **Model 1** may not be the single best model at every horizon, but it consistently ranks among the top performers and outperforms most alternatives across all forecast periods. In fact, it is the sole model retained by the MCS procedure at the 1-day-ahead horizon. It also outperforms its unpenalized counterpart and the standard GJR-GARCH across most horizons, and surpasses the Fang et al. (2020) model at all except the 1-quarter-ahead horizon. The GARCH-MIDAS variants that include VIX exhibit particularly strong accuracy, while the version incorporating CFNAI shows superior long-term performance. In terms of QLIKE, **Model 1** once again delivers robust and stable forecasts relative to the other models. It achieves lower QLIKE than the Fang et al. (2020) model at every horizon, demonstrating its dominance under this loss cri-

[5] In MCS approach, all competing models are firstly put in a full set. At each step, the model demonstrating the poorest performance is eliminated if differences among models are statistically significant. This process continues until no significant differences remain. We set the confidence level to 90%.

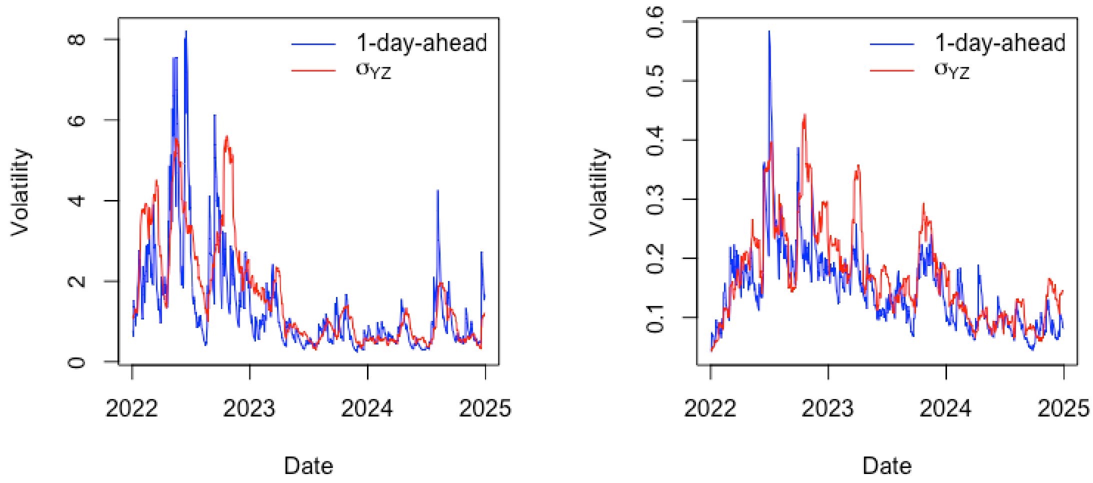
doi:10.6342/NTU202502186

terion. Figure 4.4a illustrates that **Model 1** forecasts one-day-ahead volatility reasonably, particularly during periods of relatively low market volatility.

In Tables 4.5 and 4.6, we present out-of-sample forecasting results under MSFE and QLIKE of AGG, respectively. Under the MSFE criterion, **Model 1** outperforms nearly all alternatives except at the 1-week-ahead and 1-month-ahead horizons. It is also the sole model retained by the MCS procedure at the 1-quarter-ahead horizon. Moreover, it consistently delivers lower MSFE than the unpenalized multivariate GARCH-MIDAS model, the standard GJR-GARCH, and the Fang et al. (2020) model across every horizon. By contrast, under QLIKE the performance of GARCH-MIDAS model with the penalized estimation using monthly variables deteriorates slightly, suggesting a tendency to under-predict volatility. This behavior aligns with the patterns observed in Figure 4.4b. This likely stems from the bond market becoming more volatile after 2022 driven by inflationary pressures and rising interest rates, which was not captured during the in-sample period. As a result, the model underperforms under these new conditions.

Overall, the results indicate that incorporating low-frequency components enhances the forecasting performance of the GARCH-MIDAS model. For predictive purposes, selecting an appropriate model within the multivariate GARCH-MIDAS framework yields further improvements. Additionally, the superior performance of our model relative to that of Fang et al. (2020) suggests that using monthly explanatory variables provides a forecasting advantage over quarterly ones. However, it should be noted that the GARCH-MIDAS model applied to AGG exhibits a tendency to underpredict volatility.

doi:10.6342/NTU202502186



(a) SPY (b) AGG

Figure 4.4: The Realized Volatility vs. 1-day-ahead Forecasting Volatility (**Model 1**).

doi:10.6342/NTU202502186

Table 4.3: Relative MSFE across models and horizons: SPY.

Model	1-day-ahead	1-week-ahead	1-month-ahead	1-quarter-ahead	6-month-ahead	1-year-ahead
Model 1	1.000	1.000	1.000	1.000	1.000	1.000
Model 2	1.894	2.794	5.431	11.329	21.850	31.124
Model 3						
RV_{SPY}	1.120	1.078	1.668	4.198	9.679	16.147
RV_{AGG}	1.220	1.577	5.249	15.778	32.633	43.065
UNRATE	1.045	1.093	3.072	7.188	16.788	24.456
VIX	1.054	$0.988^{\dagger}$	$0.942^{\dagger}$	$0.663^{\dagger}$	$0.569^{\dagger}$	$0.408^{\dagger}$
CPI	1.286	2.852	12.061	35.594	67.069	47.904
PPI	1.150	1.528	4.635	13.760	29.780	37.480
CSI	1.134	2.197	10.741	33.910	71.684	106.805
MB	1.231	1.172	1.483	2.789	5.365	7.391
DS	1.273	1.200	1.651	3.353	6.630	9.352
TS	1.105	1.106	2.020	5.010	10.657	17.606
PC	1.112	1.070	1.528	3.193	6.668	10.474
IP	1.148	1.134	1.731	4.084	7.224	10.708
MKT	1.150	1.475	4.462	13.684	27.296	38.733
STR	1.165	1.140	1.796	4.082	8.297	11.415
HS	1.129	1.115	1.844	4.346	9.321	14.851
GPRHT	1.127	1.050	1.307	2.467	5.462	10.305
GPRHA	1.069	1.028	2.021	6.085	15.260	30.496
GEPU	1.124	1.078	1.458	2.870	5.983	9.646
INFECT	1.121	1.088	1.523	3.062	6.415	10.375
CFNAI	1.390	1.295	1.152	$0.673^{\dagger}$	$0.399^{\dagger}$	$0.111^{\dagger}$
UNRATE^{v}	1.108	1.077	1.636	3.614	7.727	12.383
VIX^{v}	1.184	1.134	1.593	3.360	7.012	10.982
CPI^{v}	1.136	1.098	1.813	4.521	10.010	14.740
PPI^{v}	1.132	1.085	1.991	5.692	12.991	20.469
CSI^{v}	1.097	1.040	1.688	4.003	8.952	13.459
MB^{v}	1.039	$0.975^{\dagger}$	1.874	5.109	12.698	22.029
DS^{v}	1.239	1.190	1.733	3.791	7.764	11.376
TS^{v}	1.249	2.991	15.051	48.754	103.474	151.012
PC^{v}	1.059	1.006	1.787	4.396	9.496	14.384
IP^{v}	1.109	1.075	1.614	3.507	7.417	11.685
MKT^{v}	1.139	1.405	4.484	14.943	36.625	63.751
STR^{v}	1.101	1.036	1.683	4.097	9.664	16.310
HS^{v}	1.108	1.075	1.627	3.589	7.669	12.245
GPRHT^{v}	1.110	1.049	1.453	3.020	6.380	10.556
GPRHA^{v}	1.121	1.072	1.856	4.921	11.137	18.126
GEPU^{v}	1.132	1.107	1.576	3.081	6.047	8.913
EMV^{v}	1.108	1.077	1.648	3.626	7.753	12.473
CFNAI^{v}	1.108	1.076	1.635	3.629	7.789	12.540
Model 4	1.732	2.011	2.539	1.177	$0.647^{\dagger}$	$0.130^{\dagger}$
Model 5	2.881	2.945	1.765	1.191	$0.710^{\dagger}$	1.139

Note: All values represent the QLIKE relative to the baseline model (**Model 1**). A ratio $>$ 1 means that the corresponding model is worse than the baseline. In contrast, a ratio smaller than 1, marked with a dagger (†), indicates that the corresponding model shows better performance. Values in yellow cells indicate the selected models in the MCS approach in the corresponding forecasting window.

doi:10.6342/NTU202502186

Table 4.4: Relative QLIKE across models and horizons: SPY.

Model	1-day-ahead	1-week-ahead	1-month-ahead	1-quarter-ahead	6-month-ahead	1-year-ahead
Model 1	1.000	1.000	1.000	1.000	1.000	1.000
Model 2	12.047	15.472	13.054	8.654	5.883	4.012
Model 3						
RV$_{\text{SPY}}$	1.079	0.872†	1.460	1.891	2.478	2.635
RV$_{\text{AGG}}$	0.833†	1.563	2.867	2.991	3.474	3.278
UNRATE	1.298	0.964†	1.536	1.676	2.114	2.074
VIX	1.180	0.962†	0.862†	0.862†	0.839†	0.737†
CPI	1.161	3.040	4.009	3.414	3.575	2.942
PPI	0.937†	1.752	2.834	2.831	3.261	2.995
CSI	0.624†	1.904	3.571	3.624	4.132	3.979
MB	1.293	0.934†	1.333	1.630	2.073	2.155
DS	1.288	0.876†	1.389	1.715	2.194	2.294
TS	0.963†	0.926†	1.649	2.015	2.568	2.719
PC	1.112	0.857†	1.357	1.705	2.224	2.379
IP	1.097	0.918†	1.436	1.783	2.200	2.296
MKT	0.965†	0.853†	1.798	2.273	2.852	3.012
STR	1.125	0.900†	1.429	1.761	2.253	2.329
HS	1.034	0.904†	1.584	1.936	2.476	2.607
GPRHT	1.364	0.816†	0.952†	1.309	1.877	2.271
GPRHA	0.985†	0.714†	1.449	2.035	2.745	3.046
GEPU	1.169	0.853†	1.276	1.616	2.138	2.324
INFECT	1.145	0.878†	1.351	1.676	2.198	2.375
CFNAI	1.741	1.423	1.338	0.942†	0.738†	0.392†
UNRATE^v	1.078	0.866†	1.435	1.794	2.334	2.491
VIX^v	1.184	0.872†	1.372	1.733	2.258	2.411
CPI^v	1.098	0.877†	1.525	1.945	2.519	2.595
PPI^v	1.039	0.854†	1.599	2.086	2.676	2.767
CSI^v	1.046	0.856†	1.486	1.864	2.417	2.512
MB^v	1.053	0.777†	1.390	1.848	2.438	2.618
DS^v	1.199	0.908†	1.479	1.829	2.338	2.434
TS^v	0.783†	1.768	3.242	3.219	3.707	3.518
PC^v	1.006	0.781†	1.477	1.876	2.407	2.518
IP^v	1.081	0.866†	1.424	1.775	2.305	2.453
MKT^v	0.901†	0.926†	2.048	2.572	3.267	3.377
STR^v	1.067	0.827†	1.427	1.839	2.423	2.597
HS^v	1.076	0.864†	1.431	1.790	2.330	2.484
GPRHT^v	1.163	0.797†	1.184	1.564	2.099	2.312
GPRHA^v	1.068	0.803†	1.480	1.962	2.575	2.736
GEPU^v	1.126	0.933†	1.404	1.667	2.138	2.258
EMV^v	1.081	0.869†	1.441	1.794	2.336	2.495
CFNAI^v	1.078	0.864†	1.433	1.796	2.340	2.500
Model 4	1.979	2.695	9.232	9.440	7.507	2.995
Model 5	2.536	3.350	2.865	2.149	1.511	1.162

Note: All values represent the QLIKE relative to the baseline model (**Model 1**). A ratio $>$ 1 means that the corresponding model is worse than the baseline. In contrast, a ratio smaller than 1, marked with a dagger (†), indicates that the corresponding model shows better performance. Values in yellow cells indicate the selected models in the MCS approach in the corresponding forecasting window.

doi:10.6342/NTU202502186

Table 4.5: Relative MSFE across models and horizons: AGG.

Model	1-day-ahead	1-week-ahead	1-month-ahead	1-quarter-ahead	6-month-ahead	1-year-ahead
Model 1	1.000	1.000	1.000	1.000	1.000	1.000
Model 2	16.565	13.723	8.245	5.893	6.506	16.263
Model 3						
RV_{SPY}	1.190	$0.872^\dagger$	$0.958^\dagger$	1.229	1.212	1.291
RV_{AGG}	1.188	$0.872^\dagger$	$0.961^\dagger$	1.236	1.220	1.309
UNRATE	1.184	$0.894^\dagger$	1.033	1.364	1.365	1.524
VIX	1.197	$0.868^\dagger$	$0.947^\dagger$	1.216	1.205	1.281
CPI	1.215	$0.904^\dagger$	1.012	1.337	1.344	1.486
PPI	1.208	$0.885^\dagger$	$0.979^\dagger$	1.280	1.280	1.393
CSI	1.158	$0.816^\dagger$	$0.875^\dagger$	1.098	1.075	1.086
MB	1.188	$0.864^\dagger$	$0.951^\dagger$	1.234	1.225	1.306
DS	1.195	$0.868^\dagger$	$0.956^\dagger$	1.250	1.250	1.348
TS	1.212	$0.880^\dagger$	$0.952^\dagger$	1.195	1.167	1.223
PC	1.229	$0.923^\dagger$	1.027	1.335	1.326	1.460
IP	1.199	$0.871^\dagger$	$0.961^\dagger$	1.254	1.253	1.358
MKT	1.414	1.045	$0.956^\dagger$	1.111	$0.996^\dagger$	$0.998^\dagger$
STR	1.177	$0.846^\dagger$	$0.945^\dagger$	1.239	1.234	1.321
HS	1.181	$0.862^\dagger$	$0.953^\dagger$	1.223	1.209	1.295
GPRHT	1.220	$0.915^\dagger$	1.017	1.373	1.363	1.511
GPRHA	1.203	$0.853^\dagger$	$0.926^\dagger$	1.191	1.178	1.263
GEPU	1.210	$0.942^\dagger$	1.080	1.425	1.430	1.608
INFECT	1.180	$0.818^\dagger$	$0.845^\dagger$	1.053	1.032	1.032
CFNAI	1.198	$0.892^\dagger$	$0.987^\dagger$	1.306	1.316	1.456
UNRATE^v	1.186	$0.864^\dagger$	$0.952^\dagger$	1.232	1.222	1.306
VIX^v	1.187	$0.876^\dagger$	$0.959^\dagger$	1.216	1.182	1.251
CPI^v	1.277	1.019	1.141	1.335	1.245	1.320
PPI^v	1.276	1.011	1.111	1.289	1.195	1.257
CSI^v	1.062	$0.724^\dagger$	$0.790^\dagger$	1.136	1.226	1.368
MB^v	1.216	$0.897^\dagger$	$0.970^\dagger$	1.224	1.203	1.249
DS^v	1.180	$0.869^\dagger$	$0.961^\dagger$	1.237	1.223	1.307
TS^v	1.200	$0.880^\dagger$	$0.969^\dagger$	1.266	1.265	1.377
PC^v	1.143	$0.771^\dagger$	$0.810^\dagger$	1.039	1.047	$0.985^\dagger$
IP^v	$0.991^\dagger$	$0.721^\dagger$	$0.801^\dagger$	1.115	1.195	1.317
MKT^v	1.170	$0.853^\dagger$	$0.953^\dagger$	1.241	1.226	1.332
STR^v	1.224	$0.905^\dagger$	$0.976^\dagger$	1.214	1.169	1.205
HS^v	1.185	$0.842^\dagger$	$0.959^\dagger$	1.305	1.321	1.497
GPRHT^v	1.184	$0.835^\dagger$	$0.908^\dagger$	1.207	1.222	1.302
GPRHA^v	1.244	$0.980^\dagger$	1.090	1.294	1.220	1.277
GEPU^v	1.214	$0.914^\dagger$	1.026	1.329	1.325	1.432
EMV^v	1.185	$0.863^\dagger$	$0.953^\dagger$	1.238	1.232	1.322
CFNAI^v	1.284	$0.932^\dagger$	$0.938^\dagger$	1.117	1.099	1.039
Model 4	4.576	2.646	2.527	3.279	3.285	3.797
Model 5	5.519	3.745	1.105	3.811	3.553	1.115

Note: All values represent the MSFE relative to the baseline model (**Model 1**). A ratio > 1 means that the corresponding model is worse than the baseline. In contrast, a ratio smaller than 1, marked with a dagger (†), indicates that the corresponding model shows better performance. Values in yellow cells indicate the selected models in the MCS approach in the corresponding forecasting window.

doi:10.6342/NTU202502186

Table 4.6: Relative QLIKE across models and horizons: AGG.

Model	1-day-ahead	1-week-ahead	1-month-ahead	1-quarter-ahead	6-month-ahead	1-year-ahead
Model 1	1.000	1.000	1.000	1.000	1.000	1.000
Model 2	18.885	14.645	5.859	3.765	1.732	2.476
Model 3						
RV_{SPY}	1.025	$0.808^\dagger$	$0.829^\dagger$	1.101	1.212	1.286
RV_{AGG}	1.014	$0.796^\dagger$	$0.822^\dagger$	1.114	1.236	1.334
UNRATE	1.031	$0.852^\dagger$	$0.968^\dagger$	1.477	1.768	2.048
VIX	1.020	$0.797^\dagger$	$0.813^\dagger$	1.071	1.174	1.239
CPI	1.057	$0.860^\dagger$	$0.942^\dagger$	1.395	1.655	1.853
PPI	1.030	$0.819^\dagger$	$0.873^\dagger$	1.232	1.412	1.545
CSI	$0.950^\dagger$	$0.713^\dagger$	$0.695^\dagger$	$0.837^\dagger$	$0.866^\dagger$	$0.845^\dagger$
MB	1.020	$0.802^\dagger$	$0.830^\dagger$	1.116	1.238	1.310
DS	1.019	$0.802^\dagger$	$0.839^\dagger$	1.153	1.306	1.408
TS	1.008	$0.776^\dagger$	$0.799^\dagger$	1.022	1.075	1.107
PC	1.075	$0.881^\dagger$	$0.958^\dagger$	1.392	1.617	1.828
IP	1.019	$0.801^\dagger$	$0.840^\dagger$	1.161	1.317	1.436
MKT	1.411	1.185	1.033	1.143	$0.979^\dagger$	$0.872^\dagger$
STR	$0.993^\dagger$	$0.779^\dagger$	$0.830^\dagger$	1.144	1.276	1.361
HS	1.009	$0.790^\dagger$	$0.819^\dagger$	1.100	1.208	1.296
GPRHT	1.223	1.006	1.081	1.757	2.119	2.543
GPRHA	$0.984^\dagger$	$0.747^\dagger$	$0.769^\dagger$	1.034	1.134	1.229
GEPU	1.104	$0.943^\dagger$	1.064	1.668	2.033	2.336
INFECT	$0.975^\dagger$	$0.713^\dagger$	$0.655^\dagger$	$0.755^\dagger$	$0.779^\dagger$	$0.752^\dagger$
CFNAI	1.030	$0.831^\dagger$	$0.897^\dagger$	1.334	1.582	1.790
$UNRATE^v$	1.016	$0.800^\dagger$	$0.828^\dagger$	1.110	1.231	1.311
VIX^v	1.049	$0.830^\dagger$	$0.836^\dagger$	1.105	1.207	1.282
CPI^v	1.189	1.037	1.128	1.367	1.293	1.338
PPI^v	1.179	1.009	1.062	1.248	1.164	1.207
CSI^v	$0.840^\dagger$	$0.618^\dagger$	$0.621^\dagger$	$0.937^\dagger$	1.255	1.497
MB^v	1.057	$0.842^\dagger$	$0.856^\dagger$	1.110	1.209	1.227
DS^v	1.024	$0.811^\dagger$	$0.839^\dagger$	1.126	1.243	1.325
TS^v	1.007	$0.805^\dagger$	$0.847^\dagger$	1.191	1.361	1.488
PC^v	$0.956^\dagger$	$0.705^\dagger$	$0.677^\dagger$	$0.809^\dagger$	$0.885^\dagger$	$0.784^\dagger$
IP^v	$0.815^\dagger$	$0.621^\dagger$	$0.624^\dagger$	$0.886^\dagger$	1.148	1.326
MKT^v	1.000	$0.790^\dagger$	$0.833^\dagger$	1.143	1.261	1.385
STR^v	1.079	$0.856^\dagger$	$0.852^\dagger$	1.100	1.168	1.179
HS^v	1.007	$0.784^\dagger$	$0.853^\dagger$	1.378	1.765	2.231
$GPRHT^v$	$0.977^\dagger$	$0.739^\dagger$	$0.754^\dagger$	1.058	1.238	1.336
$GPRHA^v$	1.161	$0.998^\dagger$	1.054	1.267	1.222	1.229
$GEPU^v$	1.054	$0.860^\dagger$	$0.950^\dagger$	1.373	1.587	1.675
EMV^v	1.011	$0.797^\dagger$	$0.831^\dagger$	1.123	1.255	1.344
$CFNAI^v$	1.114	$0.866^\dagger$	$0.804^\dagger$	$0.965^\dagger$	1.036	$0.986^\dagger$
Model 4	4.576	2.646	2.527	3.279	3.285	3.797
Model 5	5.519	3.745	1.105	3.811	3.553	1.115

Note: All values represent the QLIKE relative to the baseline model (**Model 1**). A ratio >1 means that the corresponding model is worse than the baseline. In contrast, a ratio smaller than 1, marked with a dagger (†), indicates that the corresponding model shows better performance. Values in yellow cells indicate the selected models in the MCS approach in the corresponding forecasting window.

doi:10.6342/NTU202502186

4.2.2 Value-at-Risk Forecast

VaR is a commonly adopted measure of tail risk, offering a single, interpretable metric that is widely used by financial institutions, regulators, and portfolio managers to assess risk.

The Kupiec test was proposed by Kupiec (1995) to assess the accuracy of VaR prediction models. It evaluates whether the observed frequency of VaR violations[6] is consistent with the expected frequency implied by the probability. For a VaR model at probability $1-\alpha$ (e.g., 95% VaR with $\alpha = 0.05$), the test examines whether the empirical violation rate significantly deviates from α. The test uses a likelihood ratio statistic, defined as:

$$LR_{uc} = -2\ln\left(\frac{(1-\alpha)^{N_o-V}\alpha^{V}}{(1-\hat{p})^{N_o-V}\hat{p}^{V}}\right), \tag{4.4}$$

where N_o is the number of observations, V is the number of violations, and $\hat{p} = V/N_o$ is the observed violation rate. Under the null hypothesis that the corresponding model can predict VaR correctly, the statistic follows $\chi^2(1)$, which allows for a hypothesis test to assess the model's reliability in capturing risk.

Table 4.7 shows the test results for VaR forecasts of SPY. The results indicate that **Model 1** exhibits significant forecasting ability within a one-month horizon for both 99%-VaR and 95%-VaR. It outperforms most alternatives except the one based on CFNAI, which shows a powerful forecasting ability within a half year. Also, the Fang et al. (2020) model shows no prediction power in most time horizons, highlighting the importance of choosing the appropriate frequency for explanatory variables in predicting VaR.

[6] In the k-day-ahead horizon, given the predicted VaR is $\widehat{VaR}_{a,k,t+1|\mathcal{F}_t}$. If the realized return $r_{k,t+1} < \widehat{VaR}_{a,k,t+1|\mathcal{F}_t}$, then it is classified as a violation.

doi:10.6342/NTU202502186

Similarly, Table 4.8 reports the Kupiec test results for VaR forecasts of AGG, where all models fail to accurately forecast the 99%-VaR and 95%-VaR across all forecasting horizons except for unpenalized model, which shows a significant result in 6-month horizon. The reason behind this is possibly due to the interest rate that was raised rapidly in a short period in the last three years, which was a rare circumstance in the in-sample period, which caused the real daily return to be far from the estimated expected return parameter $\hat{\mu}$ in (2.12).[7] Figure 4.5 illustrates the comparison of the realized returns and the predicted 1-day-ahead VaR of SPY and AGG.

[7]The average return of AGG in the out-of sample period is -0.009% and $\hat{\mu}$ is 0.013%.

doi:10.6342/NTU202502186

Table 4.7: LR statistics across horizons and probability levels: SPY.

Model	1 Day		1 Week		1 Month		1 Quarter		6 Months		1 Year	
	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%
Model 1	2.016	0.050*	1.978	0.060*	0.836	5.171*	13.890***	11.452***	12.623***	44.042***	10.091***	51.498***
Model 2	0.974	0.384	0.321	4.108*	0.256	10.729***	13.890***	10.040***	6.930**	35.578***	10.091***	37.182***
Model 3												
RV_{SPY}	0.029*	7.414**	3.517	0.876	8.714***	17.104***	13.890***	55.291***	12.623***	64.424***	10.091***	51.498***
RV_{AGG}	2.016	0.767	3.517	10.234***	14.714***	21.079***	13.890***	45.251***	12.623***	55.430***	10.091***	51.498***
UNRATE	0.039*	7.414**	0.321	0.172	1.821	23.297***	13.890***	45.251***	12.623***	55.430***	10.091***	51.498***
VIX	0.270	12.940***	0.029*	0.560	0.256	9.424**	13.890***	12.990***	6.930**	28.770***	10.091***	37.182***
CPI	2.016	0.080*	3.517	7.834**	3.311	21.079***	13.890***	30.753***	12.623***	55.430***	10.091***	51.498***
PPI	0.340	1.830	3.311	5.790***	3.311	17.104***	13.890***	33.889***	12.623***	55.430***	10.091***	51.498***
CSI	3.570	0.940	9.014***	22.221***	14.714***	40.833***	13.890***	70.890***	12.623***	64.424***	10.091***	51.498***
MB	0.029*	8.244**	3.517	0.074*	8.714***	13.670***	13.890***	33.889***	12.623***	55.430***	10.091***	51.498***
DS	0.270	7.414**	1.978	0.010*	14.714***	12.140***	13.890***	27.870***	12.623***	55.430***	10.091***	51.498***
TS	0.039*	3.880*	3.517	1.240	14.714***	17.104***	13.890***	55.291***	12.623***	64.424***	10.091***	51.498***
PC	0.029*	7.414**	3.517	0.060*	8.714***	13.670***	13.890***	37.320***	12.623***	55.430***	10.091***	51.498***
IP	0.029*	6.630**	1.978	0.580	5.489**	12.140***	13.890***	33.889***	12.623***	49.210***	10.091***	51.498***
MKT	0.039*	3.300	5.739***	3.398	8.714***	19.020***	13.890***	37.320***	12.623***	49.210***	10.091***	51.498***
STR	0.039*	8.244**	1.978	0.060*	5.489**	15.326***	13.890***	30.753***	12.623***	55.430***	10.091***	51.498***
HS	0.039*	6.630**	3.517	0.876	14.714***	15.326***	13.890***	41.080***	12.623***	55.430***	10.091***	51.498***
GPRHT	0.270	10.950***	1.978	0.074*	8.714***	7.112**	13.890***	33.889***	12.623***	44.042***	10.091***	51.498***
GPRHA	0.029*	8.244**	3.517	1.240	14.714***	17.104***	13.890***	61.700***	12.623***	64.424***	10.091***	51.498***
GEPU	0.029*	10.010**	1.978	0.010*	8.714***	12.140***	13.890***	33.889***	12.623***	55.430***	10.091***	51.498***
INFECT	0.029*	9.106**	1.978	0.060*	5.489**	12.145***	13.890***	33.889***	12.623***	55.430***	10.091***	51.498***
CFNAI	1.410	10.950***	0.029*	2.380	0.060*	2.310	0.580	3.060	2.140	11.080***	10.091***	19.000***
$UNRATE^v$	0.039*	6.630**	3.517	0.340	8.714***	13.670***	13.890***	37.320***	12.623***	55.430***	10.091***	51.498***
VIX^v	0.029*	9.106**	3.517	0.060*	8.714***	13.670***	13.890***	37.320***	12.623***	55.430***	10.091***	51.498***
CPI^v	0.039*	6.630**	3.517	0.876	14.714***	13.670***	13.890***	37.320***	12.623***	55.430***	10.091***	51.498***
PPI^v	0.340	6.630**	3.517	1.670	14.714***	19.020***	13.890***	61.700***	12.623***	64.424***	10.091***	51.498***
CSI^v	0.029*	5.880***	3.517	1.670	8.714***	17.104***	13.890***	55.291***	12.623***	64.424***	10.091***	51.498***
MB^v	0.039*	5.880***	3.517	0.876	8.714***	19.020***	13.890***	55.291***	12.623***	64.424***	10.091***	51.498***
DS^v	0.039*	6.630**	3.517	0.060*	14.714***	17.104***	13.890***	37.320***	12.623***	55.430***	10.091***	51.498***
TS^v	2.016	0.010*	5.739***	4.910*	5.489**	19.020***	13.890***	30.753***	12.623***	55.430***	10.091***	51.498***
PC^v	0.039*	6.630**	3.517	0.876	8.714***	15.326***	13.890***	45.251***	12.623***	64.424***	10.091***	51.498***
IP^v	0.039*	7.414**	3.517	0.340	8.714***	13.670***	13.890***	37.320***	12.623***	55.430***	10.091***	51.498***
MKT^v	0.974	5.171*	9.014***	4.108*	14.714***	21.079***	13.890***	41.080***	12.623***	64.424***	10.091***	51.498***
STR^v	0.039*	5.880***	3.517	0.580	8.714***	19.020***	13.890***	55.291***	12.623***	64.424***	10.091***	51.498***
HS^v	0.039*	6.630**	3.517	0.340	8.714***	15.326***	13.890***	37.320***	12.623***	55.430***	10.091***	51.498***
$GPRHT^v$	0.029*	7.414**	3.517	0.074*	14.714***	13.670***	13.890***	37.320***	12.623***	55.430***	10.091***	51.498***
$GPRHA^v$	0.039*	5.880***	1.978	1.670	14.714***	15.326***	13.890***	49.930***	12.623***	64.424***	10.091***	51.498***
$GEPU^v$	0.039*	8.244**	3.517	0.010*	8.714***	15.326***	13.890***	37.316***	12.623***	55.430***	10.091***	51.498***
EMV^v	0.029*	6.630**	3.517	0.345	8.714***	13.675***	13.890***	37.316***	12.623***	55.430***	10.091***	51.498***
$CFNAI^v$	0.029*	6.630**	3.517	0.345	8.714***	13.675***	13.890***	37.316***	12.623***	55.430***	10.091***	51.498***
Model 4	8.868***	4.507*	5.739***	22.221***	383.372***	222.352***	784.404***	413.103***	728.296***	376.024***	453.650***	235.263***
Model 5	8.868***	3.883***	4.551*	2.380*	11.190***	2.358*	21.111***	2.025*	16.891***	2.290*	8.908***	0.341

Note: ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively; no asterisk indicates non-rejection of the null hypothesis.

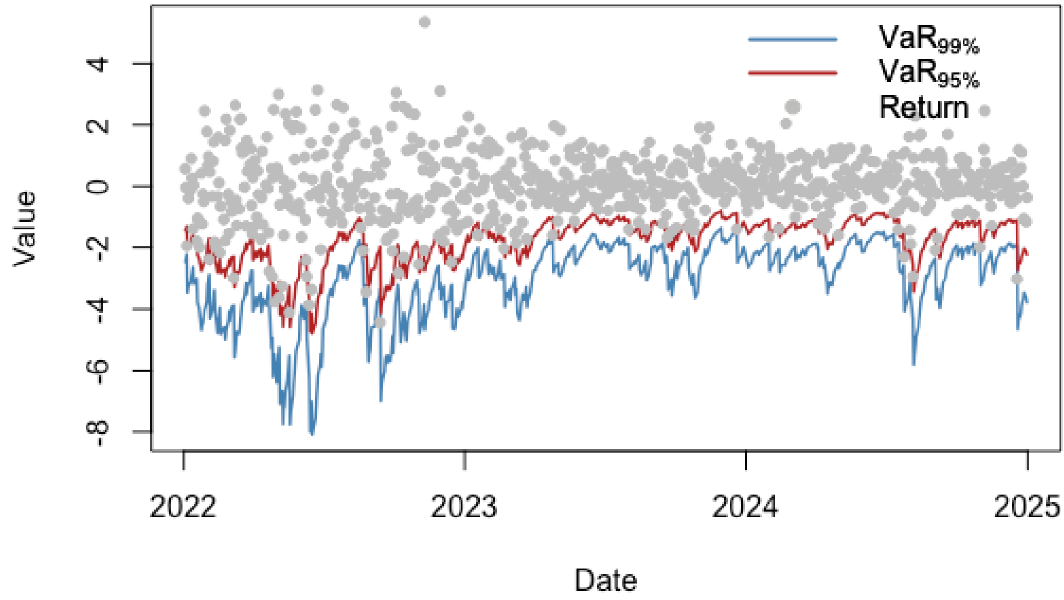
doi:10.6342/NTU202502186

Table 4.8: LR statistics across horizons and probability levels: AGG.

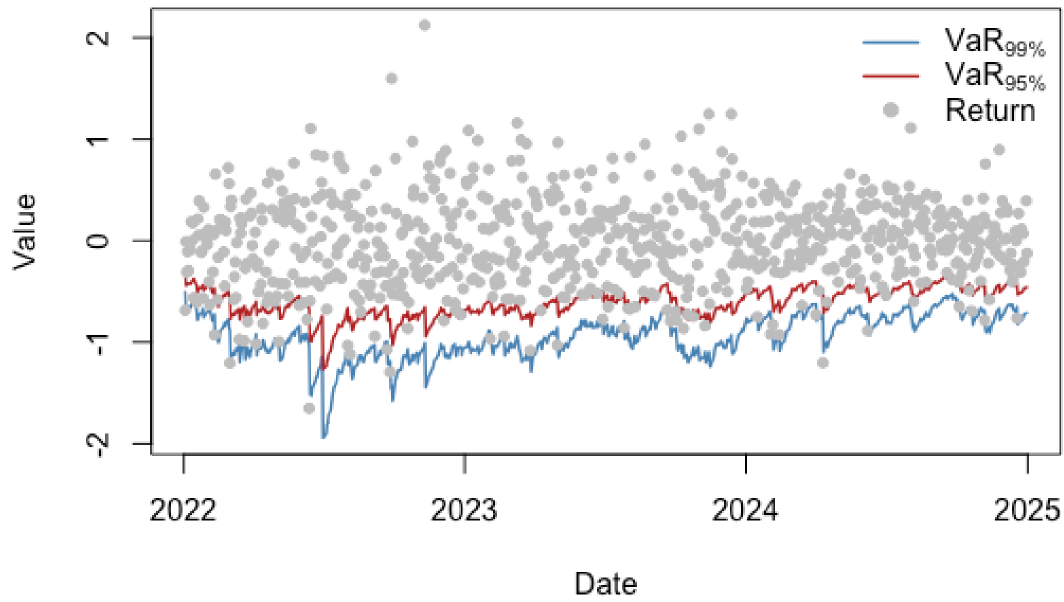
Model	1 Day		1 Week		1 Month		1 Quarter		6 Months		1 Year	
	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%
Model 1	33.170***	23.610***	25.660***	22.650***	65.130***	59.370***	102.430***	102.830***	94.220***	71.470***	31.640***	45.270***
Model 2	7.270**	10.950***	10.690***	9.330***	9.420***	10.310***	16.580***	8.090***	1.890	0.200	10.090***	8.850***
Model 3												
RV_{SPY}	16.380***	16.190***	16.530***	15.360***	45.900***	53.600***	86.990***	118.060***	106.350***	137.920***	73.840***	84.820***
RV_{AGG}	14.340***	16.190***	16.530***	15.360***	42.910***	51.730***	83.250***	123.310***	110.490***	135.070***	77.820***	84.820***
UNRATE	18.520***	17.350***	20.920***	15.360***	58.480***	63.340***	135.460***	153.650***	172.890***	173.700***	120.740***	117.970***
VIX	16.380***	16.190***	16.530***	15.360***	45.900***	55.500***	83.250***	123.310***	94.220***	126.670***	69.930***	84.820***
CPI	16.380***	16.190***	20.920***	15.360***	52.070***	65.360***	118.610***	159.430***	154.260***	167.530***	98.600***	112.170***
PPI	14.340***	16.190***	18.670***	15.360***	42.910***	57.420***	110.430***	131.340***	123.170***	149.510***	77.820***	98.150***
CSI	14.340***	16.190***	12.530***	14.270***	37.140***	41.080***	72.340***	100.370***	67.730***	97.610***	47.880***	55.920***
MB	16.380***	16.190***	16.530***	15.360***	45.900***	55.500***	90.780***	134.060***	118.900***	143.670***	77.820***	87.430***
DS	16.380***	16.190***	16.530***	15.360***	45.900***	57.420***	86.990***	136.800***	118.900***	146.580***	81.860***	95.420***
TS	14.340***	15.070***	16.530***	14.270***	42.910***	51.730***	83.250***	112.900***	102.260***	113.110***	73.840***	69.810***
PC	20.750***	16.190***	20.920***	16.500***	48.950***	65.360***	118.610***	150.800***	163.500***	167.530***	90.110***	103.670***
IP	16.380***	16.190***	16.530***	15.360***	48.950***	53.600***	94.610***	134.060***	123.170***	140.780***	77.820***	92.730***
MKT	23.070***	21.010***	16.530***	22.650***	65.130***	65.360***	110.430***	128.650***	67.730***	115.780***	23.020***	53.720***
STR	14.340***	16.190***	18.670***	15.360***	48.950***	51.730***	98.500***	134.060***	114.670***	143.670***	81.860***	87.430***
HS	16.380***	16.190***	14.480***	15.360***	52.070***	53.600***	90.780***	128.650***	102.260***	126.670***	81.860***	77.180***
GPRHT	23.070***	15.070***	20.920***	20.090***	79.080***	61.340***	135.460***	174.200***	182.430***	179.940***	134.620***	136.000***
GPRHA	14.340***	16.190***	16.530***	14.270***	45.900***	49.890***	90.780***	110.350***	94.220***	113.110***	77.820***	82.240***
GEPU	20.750***	17.350***	20.920***	16.500***	58.480***	69.480***	139.780***	174.200***	206.870***	192.680***	153.760***	126.870***
INFECT	14.340***	13.990***	12.530***	15.360***	37.140***	41.080***	61.910***	95.520***	78.760***	82.970***	44.460***	49.430***
CFNAI	16.380***	17.350***	18.670***	15.360***	52.070***	55.500***	114.500***	147.960***	136.250***	164.470***	98.600***	106.480***
$UNRATE^v$	16.380***	17.350***	16.530***	15.360***	45.900***	55.500***	86.990***	128.650***	114.670***	143.670***	73.840***	87.430***
VIX^v	16.380***	16.190***	16.530***	14.270***	39.990***	55.500***	94.610***	125.970***	114.670***	140.780***	77.820***	84.820***
CPI^v	20.750***	21.010***	23.250***	21.350***	58.480***	69.480***	122.760***	159.430***	118.900***	137.920***	73.840***	87.430***
PPI^v	20.750***	18.530***	20.920***	18.860***	55.240***	63.340***	102.430***	139.560***	106.350***	132.250***	73.840***	79.690***
CSI^v	10.580***	12.940***	12.530***	13.210***	29.030***	29.940***	61.910***	115.470***	123.170***	137.920***	94.330***	92.730***
MB^v	16.380***	16.190***	16.530***	16.500***	42.910***	53.600***	90.780***	125.970***	110.490***	135.070***	77.820***	79.690***
DS^v	16.380***	16.190***	16.530***	15.360***	45.900***	55.500***	90.780***	128.650***	123.170***	140.780***	77.820***	84.820***
TS^v	16.380***	15.070***	16.530***	16.500***	42.910***	49.890***	94.610***	128.650***	123.170***	137.920***	90.110***	98.150***
PC^v	14.340***	16.190***	12.530***	15.360***	37.140***	42.790***	72.340***	100.370***	82.540***	97.610***	34.710***	62.720***
IP^v	12.410***	12.940***	10.690***	14.270***	34.370***	32.980***	68.810***	110.350***	102.260***	123.920***	85.960***	82.240***
MKT^v	16.380***	17.350***	16.530***	14.270***	45.900***	57.420***	94.610***	136.800***	114.670***	140.780***	73.840***	90.060***
STR^v	14.340***	16.190***	16.530***	14.270***	45.900***	57.420***	90.780***	115.470***	106.350***	132.250***	73.840***	72.230***
HS^v	18.520***	16.190***	20.920***	15.360***	52.070***	53.600***	106.410***	147.960***	154.260***	152.460***	111.730***	106.480***
$GPRHT^v$	18.520***	12.940***	14.480***	15.360***	37.140***	53.600***	72.340***	123.310***	110.490***	115.780***	62.300***	77.180***
$GPRHA^v$	18.520***	19.750***	20.920***	16.500***	58.480***	65.360***	126.950***	147.960***	110.490***	129.450***	77.820***	87.430***
$GEPU^v$	14.340***	16.190***	18.670***	20.090***	55.240***	61.340***	114.500***	145.140***	136.250***	158.430***	98.600***	115.060***
EMV^v	16.380***	17.350***	16.530***	15.360***	45.900***	55.500***	86.990***	134.060***	114.670***	137.920***	77.820***	90.060***
$CFNAI^v$	16.380***	16.190***	16.530***	14.270***	37.140***	48.080***	79.560***	105.320***	90.280***	113.110***	73.840***	67.410***
Model 4	70.090***	58.000***	88.290***	94.910***	∞***	871.650***	∞***	1012.380***	∞***	905.280***	∞***	696.950***
Model 5	35.890***	19.750***	41.800***	42.280***	138.100***	136.590***	238.530***	183.300***	187.250***	132.250***	153.760***	87.430***

Note: ***, **, and * indicate p-values of 1%, 5%, and 10%, respectively. No asterisk indicates that the null hypothesis is not rejected.

doi:10.6342/NTU202502186



(a) SPY



(b) AGG

Figure 4.5: The 1-day-ahead forecasting VaR and return (**Model 1**).

doi:10.6342/NTU202502186

Chapter 5　Conclusion

We demonstrate that the GARCH-MIDAS model with the penalized estimation using monthly variables delivers superior out-of-sample volatility forecasts for both SPY and AGG, consistently outperforming its unpenalized counterpart, standard GJR-GARCH model, and the Fang et al. (2020) model across most horizons. The out-of-sample evaluations shows that proper model selection should be undertaken and it is better to adopt higher-frequency data to strengthen the model from all aspects. However, in contrast to the result from Fang et al. (2020) that GARCH-MIDAS model with the penalized estimation is superior to most competitive models in predicting volatility, GARCH-MIDAS model with the penalized estimation using monthly variables doesn't outperform all GARCH-MIDAS models with one explanatory variable when using monthly data. Also, the GARCH-MIDAS model with the penalized estimation using monthly variables proved to be a suitable method to predict the volatility in the bond market, but the tendency to underpredict is an issue that warrants attention.

In evaluating VaR forecasts, the GARCH-MIDAS model with the penalized estimation using monthly variables outperforms alternative specifications for SPY. However, all models struggle to accurately forecast the AGG VaR, likely reflecting the change in dynamics in the bond market over the out-of-sample period.

doi:10.6342/NTU202502186

The thesis considers an interesting explanatory variable INFECT, revisits the geopolitical risk indices GPRHT and GPRHA introduced by Segnon et al. (2024), and adopt GEPU following Fang et al. (2018). Our results show that INFECT as well as its volatility data have a significant impact on stock market uncertainty, whereas GPRHT, GPRHA, and GEPU exhibit only marginal influence within the multivariate GARCH-MIDAS framework.

Several avenues for future research emerge. First, incorporating ex-ante forecasts of the explanatory variables rather than treating them as predetermined may further enhance predictive accuracy. Second, allowing the expected return of the model to vary with time, instead of remaining fixed, might better capture evolving tradeoffs between risk and return. Third, adopting a rolling estimation framework, rather than fixing model parameters throughout the sample period, may offer greater practical value for forecasting. Finally, applying this framework to additional asset classes or emerging markets would test its robustness and broaden its practical applicability. Altogether, our thesis explore the applications of GARCH-MIDAS model for advanced volatility modeling.

doi:10.6342/NTU202502186

References

Al-Thaqeb, S. A. and Algharabali, B. G. (2019). Economic policy uncertainty: A literature review. The Journal of Economic Asymmetries, 20:e00133.

Albulescu, C. T. (2021). COVID-19 and the united states financial markets' volatility. Finance Research Letters, 38:101699.

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

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

Baker, S. R., Bloom, N., Davis, S. J., and Kost, K. J. (2019). Policy news and stock market volatility. Working Paper 25720, National Bureau of Economic Research.

Baker, S. R., Bloom, N., Davis, S. J., Kost, K. J., Sammon, M. C., and Viratyosin, T. (2020). The unprecedented stock market impact of COVID-19. Working Paper 26945, National Bureau of Economic Research.

Caldara, D. and Iacoviello, M. (2022). Measuring geopolitical risk. American Economic Review, 112(4):1194−1225.

doi:10.6342/NTU202502186

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.

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

Eniayewu, P. E., Samuel, G. T., Joshua, J. D., Samuel, B. T., Dogo, B. S., Yusuf, U., Ihekuna, R. O., and Mevweroso, C. R. (2024). Forecasting exchange rate volatility with monetary fundamentals: A GARCH-MIDAS approach. Scientific African, 23:e02101.

Fan, Y. and Tang, C. Y. (2012). Tuning parameter selection in high dimensional penalized likelihood. Journal of the Royal Statistical Society Series B: Statistical Methodology, 75(3):531–552.

Fang, L., Chen, B., Yu, H., and Qian, Y. (2018). The importance of global economic policy uncertainty in predicting gold futures market volatility: A GARCH-MIDAS approach. Journal of Futures Markets, 38(3):413–422.

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.

Ghysels, E., Sinko, A., and Valkanov, R. (2007). Midas regressions: Further results and new directions. Econometric Reviews, 26(1):53–90.

doi:10.6342/NTU202502186

Gulen, H. and Ion, M. (2015). Policy uncertainty and corporate investment. The Review of Financial Studies, 29(3):523–564.

Hansen, B. E. (1994). Autoregressive conditional density estimation. International Economic Review, 35(3):705–730.

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

Jiang, C., Li, Y., Xu, Q., and Liu, Y. (2021). Measuring risk spillovers from multiple developed stock markets to china: A vine-copula-GARCH-MIDAS model. International Review of Economics Finance, 75:386–398.

Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. The Journal of Derivatives, 2(4):73–84.

Liu, R., Zhou, Z., and Li, H. (2022). The influence of COVID-19 epidemic on bdi volatility: An evidence from GARCH-MIDAS model. Ocean & Coastal Management, 229:106330.

Lu, H., Gao, Q., Xiao, L., and Dhesi, G. (2024). Forecasting EUA futures volatility with geopolitical risk: Evidence from GARCH-MIDAS models. Review of Managerial Science, 18:1917–1943.

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. Journal of Econometrics, 160(1):246–256.

Raza, S. A., Masood, A., Benkraiem, R., and Urom, C. (2023). Forecasting the volatility of precious metals prices with global economic policy uncertainty in pre and during

doi:10.6342/NTU202502186

the COVID-19 period: Novel evidence from the GARCH-MIDAS approach. Energy Economics, 120:106591.

Rogers, L. C. G. and Satchell, S. E. (1991). Estimating variance from high, low and closing prices. The Annals of Applied Probability, 1(4):504–512.

Salisu, A. A., Ogbonna, A. E., Gupta, R., and Ji, Q. (2024). Energy market uncertainties and exchange rate volatility: A GARCH-MIDAS approach. Finance Research Letters, 67:105847.

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

Segnon, M., Gupta, R., and Wilfling, B. (2024). Forecasting stock market volatility with regime-switching GARCH-MIDAS: The role of geopolitical risks. International Journal of Forecasting, 40(1):29–43.

Yang, D. and Zhang, Q. (2000). Drift-independent volatility estimation based on high, low, open, and close prices. The Journal of Business, 73(3):477–492.

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

doi:10.6342/NTU202502186