Master's Degree in Economics and Finance:
Financial Economics

Modelling Financial Tail Risk: Multi-Step Forecasting of Value-at-Risk and Expected Shortfall

Authors: Kai Penfold and Raphael Ravinet

Supervisor: Prof. Christian Brownlees

June 2024

Abstract in English:

It has been well documented in the literature that the inclusion of realized measures into GARCH models can lead to both statistical and economic gains through improved forecasting performance. In this paper, we employ the Realized HAR-GARCH model of Huang et al. (2016) to forecast multi-step-ahead financial tail risk. Specifically, we assess the model's performance and whether the resulting Value-at-Risk (VaR) and Expected Shortfall (ES) forecasts are well-specified. We also consider the economic significance of the model's accuracy through a hedging strategy exercise. Our analysis covers the period from October 10, 2003, to March 28, 2024, including the COVID-19 pandemic. The Realized HAR-GARCH model is found to outperform the Realized GARCH model in volatility forecasting, especially over longer horizons. However, we find this does not directly translate into superior VaR and ES forecasting performance.

Keywords in English:

Volatility Forecasting; Value-at-Risk; Realized Measures

Abstract in Spanish:

La inclusión de medidas ya realizadas a modelos GARCH y su impacto favorable sobre su capacidad de pronóstico es un hecho bien documentado en la literatura existente. En este artículo, aplicamos el modelo de Realized HAR-GARCH de Huang et al. (2016) para pronosticar riesgos financieros a la baja. En este ejercicio medimos la precisión del modelo y si las consiguientes medidas de Valor en Riesgo (VeR) y los pronósticos de la caída esperada están bien especificados. Además, medimos la significancia económica de la precisión del modelo con un ejercicio estratégico de cobertura de riesgo. Nuestro análisis cubre el periodo de octubre de 2003 a marzo de 2024, incluyendo la pandemia del COVID-19. Los resultados sugieren que el modelo de Realized HAR-GARCH supera los resultados del Realized GARCH en su habilidad de pronosticar volatilidad, sobre todo cuando el periodo de pronóstico es mayor. Sin embargo, esto no resulta en pronósticos superiores de VeR o caída esperada.

Keywords in Spanish:

Pronóstico de volatilidad; Valor en riesgo; Realized Measures

Abstract

It has been well documented in the literature that the inclusion of realized measures into GARCH models can lead to both statistical and economic gains through improved forecasting performance. In this paper, we employ the Realized HAR-GARCH model of Huang et al. (2016) to forecast multi-step-ahead financial tail risk. Specifically, we assess the model's performance and whether the resulting Value-at-Risk (VaR) and Expected Shortfall (ES) forecasts are well-specified. We also consider the economic significance of the model's accuracy through a hedging strategy exercise. Our analysis covers the period from October 10, 2003, to March 28, 2024, including the COVID-19 pandemic. The Realized HAR-GARCH model is found to outperform the Realized GARCH model in volatility forecasting, especially over longer horizons. However, we find this does not directly translate into superior VaR and ES forecasting performance.

Keywords: Volatility Models; Forecasting; Realized Measures; Value-at-Risk; Expected Shortfall

Contents

1. Introduction

Volatility refers to the variation in the returns of a financial asset over time. The accurate modelling and forecasting of volatility is an extremely important topic in finance that spans all corners of the industry. Some of these applications include asset pricing, portfolio management, pricing of derivatives, and risk management. For instance, both the buyers and sellers of fixed income derivatives rely on reliable estimates of volatility to ensure the correct pricing of contracts and the implementation of optimal hedging and risk mitigation strategies. Central banks, commercial banks and other financial institutions depend on accurate volatility estimates to manage risk exposure and ensure compliance with regulatory requirements. An example being the widely adopted value-at-risk (VaR) and Expected Shortfall (ES) metrics used to quantify financial downside market risk in a given time period.

Value at Risk (VaR) is a metric used for assessing market risk, offering a probabilistic estimation of the maximum expected loss under typical market conditions. Despite its broad adoption and straightforward calculation, VaR has notable limitations. Under extreme asset price fluctuations, as highlighted by Yamai and Yoshiba (2004)[1], VaR may significantly underestimate risk. Zaichao Du and Juan Carlos Escanciano (2017)[2] found that VaR is generally unresponsive to extreme events, whereas Expected Shortfall (ES) provides a more accurate description of the risk involved. This flaw became starkly evident during the 2008 global financial crisis.

In response to these limitations, the past decade has seen an increased emphasis placed on forecasting Expected Shortfall. The Basel III accord replaced VaR and stressed VaR as the market risk measure with a single ES metric. Expected Shortfall calculates the average of the losses that exceed the VaR threshold, providing a more comprehensive risk assessment. By accounting for the magnitude of losses beyond the VaR limit, ES addresses a critical question: if adverse conditions occur, how severe could the losses be? This approach offers a more reliable measure of tail risk, enhancing the risk management strategies of financial institutions and contributing to greater financial stability.

Since volatility is not directly observable, preliminary work in modelling volatility was focused on the use of latent variable models, most notably the autoregressive conditional heteroskedasticity (ARCH) model of Engle (1982)[3] and the generalized ARCH (GARCH) model of Bollerslev (1986)[4]. The GARCH model defines the conditional variance of a process as a deterministic function of the history of the process and the past conditional variances. Since then, the basic model has been extended to account for many common features in financial returns time series. Among these extensions include the Exponential Garch (EGARCH) model developed by Nelson (1991)[5], which uses a log specification to capture the asymmetry in volatility dynamics, whereby negative shocks to returns tend to increase volatility more than positive shocks of the same magnitude – commonly known as the "leverage effect" (Engle & Ng, 1993)[6]. Glosten, Jagannathan, and Runkle's (1993)[7] GJR-GARCH model features a dummy to allow for differing impacts of positive and negative returns on volatility, and Baillie, Bollerslev, and Mikkelsen (1996)[8] introduced the FIGARCH to better capture the long memory characteristics of financial time series data by including a fractional difference component in the variance equation.

Traditional GARCH models use squared daily returns to uncover information about underlying volatility and use this to make predictions about volatility in future periods. This reliance on a single value (squared daily returns) as a signal for underlying movements cause these models to react sluggishly during periods of high volatility. This is a result of the conditional variance taking a long time to adjust and reach a new given level over many periods (Andersen et al. (2003)[9]). Following the rise of high-frequency trading and thus availability of intra-daily financial data, a more sophisticated nonparametric estimator of daily volatility known as Realized Variance (RV) emerged. Unlike squared returns, realized variance is calculated by summing the squared *intra-daily* returns, which has been shown to be a far more informative and efficient proxy for ex-post volatility (Andersen and Bollerslev (1998)[10], Barndorff-Nielsen and Shephard (2002)[11], Andersen, Bollerslev, Diebold and Labys (2003))[9]. They also have the suitable property of being consistent estimators of underlying volatility when intra-daily data is high (Andersen and Bollerslev (1998))[10].

A popular realized variance model that stemmed from the availability of high-frequency data is that of the heterogeneous autoregressive (HAR) model pioneered by Corsi (2009)[12]. The HAR-RV model takes a non-parametric approach to forecasting volatility directly by using an additive cascade of realized volatility components over different time periods. The model's structure is based on the heterogeneity across participants in financial markets making decisions on daily, weekly, and monthly horizons. Despite the model's lack of true long-memory properties, the HAR-RV model has the characteristics of a 'pseudo long memory' model and has grown in popularity due to its parsimonious structure, ease of estimation, and strong performance in capturing key empirical features of financial returns, such as long memory and fat tails. Extensions to the standard HAR model have seen the use of bipower variation measures to improve forecasts by disentangling volatility into its jump and continuous components (Andersen et al., (2007)[13], Hua and Manzan, (2013))[14]. Wilms, Rombouts, and Croux (2021)[15] extend the model to incorporate market sentiment and expectations including the forward-looking measure of option implied variances (IV), finding it to substantially improve forecast accuracy.

Early adoptions of RV in an ARCH/GARCH framework were known as GARCH-X type models, pioneered by Engle (2002)[16] by extending the volatility dynamics to include realized variance as an exogenous regressor in a multiplicative error model (MEM) context. This model type pays little attention to specifying the dynamics of the realized variance component and so the GARCH-X is restricted to forecasting returns and volatility to a single period ahead. However, this was later extended to allow for multi-step prediction by Engle & Gallo (2006)[17]. Shephard & Sheppard (2010)[18] and their high-frequency volatility (HEAVY) model provide an alternative extension to the GARCH-X by specifying a GARCH-like structure to each realized measure of volatility, forming a basis for a multi-step-ahead forecasting system. The HEAVY model typically excludes the squared returns component (setting $\alpha = 0$) unless the realized measure is not enough to entirely crowd out the lagged squared daily returns. There have been numerous studies that support the inclusion of realized measures into ARCH/GARCH type models both in terms of economic and statistical significance (Brownlees and Gallo, 2009[19], Christoffersen et al., 2010[20], Corsi et al., 2013[21], Christoffersen, P. et al., 2014[11], Huang et al., 2016[22]).

A perhaps more natural approach to including realized measures into a GARCH framework is that of Hansen et al. (2010)[23], who explicitly link realized volatility measures to the latent conditional variance through a measurement equation, implementing a joint estimation process to form the Realized-GARCH model. This approach encapsulates both the benefits associated with using high-frequency data as a proxy for volatility, whilst retaining convenient properties of a standard GARCH framework, such as being able to model and forecast both returns and volatility simultaneously. The Realized-GARCH model yields attractive forecasting results both in terms of economic and statistical gains compared to other GARCH family models (Christoffersen et al., 2014[15], Tian and Hamori, 2015[24], Gerlach & Wang, 2016[25], Huang, Wang & Hansen 2016[26]), and thus numerous extensions of the model have emerged throughout the literature. Hansen et al (2014)[27] develop a multivariate Realized GARCH framework to model the conditional Beta of asset returns. Hansen and Huang (2016)[28] introduce the Realized Exponential GARCH (Realized EGARCH), which is well suited for the use of multiple realized measures and features an additional leverage function to capture the asymmetric dependence between returns and volatility. Huang et al., 2016[22] develop the realized HAR-GARCH to better capture the long-memory component of volatility, and Chen and Watanabe (2019)[29] extend to the Realized TGARCH (Threshold GARCH) to allow for different responses on positive and negative shocks.

In this paper, we extend the work of Huang et al. (2016) by applying the HAR-GARCH model to a Value at Risk (VaR) and Expected Shortfall (ES) framework. Whilst previous papers have applied this model to 1-day VaR and Expected Shortfall forecasting, we consider the performance of the Realized HAR-GARCH model for multi-step forecasting of VaR and Expected Shortfall. We also consider the economic significance of the model in evaluation through a hedging strategy exercise on 3 US stocks.

We conduct our empirical analysis on the returns series of 1 market index and 25 stocks from the Dow Jones 30 over a sample period from 10 October, 2003, to 28 March, 2024. We find that the Realized HAR-GARCH model outperforms the Realized GARCH model in multi-step volatility forecasting, particularly over longer horizons. However, both models show similar performance in multi-step forecasting of Value at Risk (VaR) and Expected Shortfall (ES). Additionally, the Realized HAR-GARCH demonstrates effective performance in a hedging strategy application based on Expected Shortfall forecasts.

The remainder of this paper is organized as follows: Section 2 discusses Realized Variance and reviews the HAR, GARCH, Realized GARCH, and Realized HAR-GARCH models. Section 3 outlines the main features of the data used in the analysis. Section 4 discusses the forecasting results, including volatility forecasts, and the Value at Risk (VaR) and Expected Shortfall (ES) forecasts. Section 5 summarizes our findings and concludes the paper.

2. Methodology

2.1 Econometric Models

2.1.1 GARCH

The Generalized AutoRegressive Conditional Heteroskedasticity (GARCH) model is a widely used model for analyzing and forecasting volatility in financial markets. The classical GARCH model proposed by Bollerslev (1986)[4] can be expressed in its simplest form as:

$$r_t = \mu + \sqrt{\sigma_t^2} z_t$$

$$\sigma_t^2 = \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2$$

where r_t is a series of returns, σ_t^2 is the conditional variance, and $z_t \overset{i.i.d.}{\sim} \mathcal{D}(0,1)$.

2.1.2 Realized Measures

Realized measures aim to estimate the quadratic variation of a price process over some interval of time (Lily Liu, Andrew J. Patton, and Kevin Sheppard[30]). Barndorff-Nielsen and Shephard (2002)[31] show that realized variance, which can be defined as the sum of squared intra-day returns over a day, is a consistent predictor for quadratic variation. Let us define the realized variance of a given stock or index on day t for $t = 1, \ldots T$ as:

$$RV_t = \sum_{i=1}^{T} r_{t,i}^2, \tag{2.1}$$

where $r_{t,i}$ denotes the ith 5-min period return within day t.

2.1.3 HAR Model

Realized Variance (RV) has been shown to exhibit characteristics of a long-memory process and so the dynamics can be arguably specified by ARMA-type models. The Heterogeneous Autoregressive (HAR) model proposed by Corsi (2009)[12] seeks to capture these long memory dynamics through its cascade structure of RV lags, and has become a popular choice for volatility modelling due to its simple structure and strong forecasting performance. In it's simplest form the log version of the HAR model can be expressed as:

$$LRV_t = \gamma_0 + \gamma_d LRV_{t-1}^d + \gamma_w LRV_{t-1}^w + \gamma_m LRV_{t-1}^m + u_t,$$

where LRV_t is the logged realized variance of day t, and $LRV_{t-1}^d = LRV_{t-1}$, $LRV_{t-1}^w = \frac{1}{4}\sum_{i=2}^{5} LRV_{t-i}$, and $LRV_{t-1}^m = \frac{1}{17}\sum_{i=6}^{22} LRV_{t-i}$ represent the lagged daily, weekly and monthly logged realized variances, respectively.

Whilst the HAR model is an AR(22) model, thanks to its cascade lag structure it enjoys capturing RV's high degree of persistence in a more parsimonious manner. The parameters of the model can be estimated through a simple ordinary least squares regression.

2.1.4 Realized GARCH

The Realized GARCH model proposed by Hansen (2010)[23] builds on the standard GARCH model by replacing the squared returns component in the GARCH variance equation with a realized measure, x_t, motivated by the fact that RV is a more realistic proxy for latent volatility. The model also specifies a measurement equation to model the RV component. Let us denote $\tilde{\sigma}_t^2$ as $\log(\sigma_t^2)$ and $\tilde{x}_t$ as $\log(x_t)$. Then, the log-transformed version of the Realized GARCH model can be expressed by:

$$r_t = \mu_t + \sqrt{\sigma_t^2} z_t, \tag{2.2}$$

$$\tilde{\sigma}_t^2 = \omega + \beta \tilde{\sigma}_{t-1}^2 + \gamma \tilde{x}_{t-1} \tag{2.3}$$

$$\tilde{x}_t = \xi + \phi \tilde{\sigma}_t^2 + \tau(z_t) + u_t, \tag{2.4}$$

where $z_t \sim \mathcal{N}(0,1)$ and $u_t \sim \mathcal{N}(0, \sigma_u^2)$ with z_t and u_t independent of one another. An important feature of the measurement equation is the fact that it captures the return-volatility dependence. $\tau(z_t)$ is the leverage function that captures the asymmetry between returns and volatility. Specifically, it allows negative shocks to returns to have a greater impact on volatility than positive shocks of the same magnitude. In this paper we follow Hansen et al. (2012)[23] and set $\tau(z_t) = \tau_1 z_t + \tau_2(z_t^2 - 1)$.

2.1.5 Realized HAR-GARCH

We can extend the GARCH variance equation above to incorporate a HAR structure for the realized component. Following the specification of Huang et al. (2016)[22], we obtain the following volatility dynamics equation for the Realized HAR-GARCH model:

$$\tilde{\sigma}_t^2 = \omega + \beta\tilde{\sigma}_{t-1}^2 + \gamma_d \tilde{x}_{t-1} + \frac{\gamma_w}{4}\sum_{i=2}^{5}\tilde{x}_{t-i} + \frac{\gamma_m}{17}\sum_{i=6}^{22}\tilde{x}_{t-i} \tag{2.5}$$

Just as in the standard HAR model, each realized measure here corresponds to information from the previous day, week, and monthly log realized variance. However, a key difference here is that the standalone HAR model focuses on daily open-to-close returns, whereas the Realized HAR-GARCH model captures the daily close-to-close returns through the GARCH equation - meaning that overnight returns are inherently accounted for. This is particularly useful when forecasting volatility in a risk management context, where overnight returns need to be considered to produce accurate estimates of daily Value at Risk (VaR) and Expected Shortfall (ES).

2.2 Likelihood Estimation

Given a set of observations, the likelihood function of a specific model can be expressed as: $\mathcal{L}(\theta \mid x_1, \ldots, x_T) = f(x_1, \ldots, x_T \mid \theta) = \prod_{t=1}^{T} f(x_t \mid \theta)$, where θ is a set of parameters and x_t is a series of random variables.

Maximizing this function with respect to the model parameters allows one to find the set of parameter estimates that corresponds to the desired probability distribution. The exact form of the likelihood function relies on assumptions regarding the distribution of the innovations.

We denote θ^{RG} and θ^{RHG} as the parameter sets for the Realized GARCH and Realized HAR-GARCH models, respectively, which are written as:

$$\theta^{RG} = (\omega, \beta, \gamma, \xi, \phi, \tau_1, \tau_2, \sigma_u^2)$$

$$\theta^{RHG} = (\omega, \beta, \gamma_d, \gamma_w, \gamma_m, \xi, \phi, \tau_1, \tau_2, \sigma_u^2)$$

These parameters can be estimated via the Quasi Log-Likelihood function:

$$\mathcal{L}(r, x; \theta) = \sum_{t=1}^{T} \log f(r_t, x_t \mid \mathcal{I}_{t-1})$$

Here, both the returns and realized measure are conditioned on $\mathcal{I}_{t-1}$, which is the set of information available at time t.

Adopting a Gaussian specification, the Log-Likelihood can be expressed as:

$$\begin{aligned}\ell(r, x; \theta) = &-\frac{1}{2}\sum_{t=1}^{n}\underbrace{\left[\log(2\pi) + \log(\sigma_t^2) + \frac{r_t^2}{\sigma_t^2}\right]}_{\ell(r;\theta)} \\ &-\frac{1}{2}\sum_{t=1}^{n}\underbrace{\left[\log(2\pi) + \log(\sigma_u^2) + \frac{u_t^2}{\sigma_u^2}\right]}_{\ell(r|x;\theta)}\end{aligned} \tag{2.6}$$

where $\ell(r;\theta)$ represents the partial log-likelihood focusing on the returns, while $\ell(r|x;\theta)$ includes the conditional part associated with the realized measures. This formulation allows for the separation of the likelihood contributions from the return series and the additional explanatory variables.

2.3 Estimation Results

We estimate our parameters via quasi maximum likelihood estimation (QMLE) using the joint log-likelihood function in Equation 2.6. For estimation of the Realized GARCH parameters we impose the following sta-

Table 1: Full sample parameter estimates for the Realized GARCH model

Ticker	ω	β	γ	ξ	ϕ	τ_1	τ_2	σ_{ut}
AAPL	0.066	0.789	0.207	-0.202	1.010	-0.390	0.207	0.857
AMGN	0.011	0.557	0.426	-0.026	1.041	-0.072	0.158	0.347
AMZN	0.012	0.584	0.392	-0.031	1.061	-0.115	0.176	0.383
AXP	0.011	0.516	0.460	-0.024	1.051	-0.171	0.119	0.370
BA	0.001	0.558	0.410	-0.002	1.070	-0.228	0.198	0.409
CAT	0.015	0.592	0.380	-0.039	1.072	-0.184	0.235	0.405
CSCO	0.001	0.618	0.364	-0.004	1.039	-0.110	0.163	0.377
CVX	0.0001	0.559	0.418	-0.0001	1.052	-0.125	0.248	0.372
DIS	0.011	0.511	0.471	-0.024	1.038	-0.127	0.123	0.351
GS	0.002	0.547	0.431	-0.006	1.049	-0.056	0.165	0.357
HD	0.001	0.555	0.424	-0.003	1.046	-0.146	0.183	0.363
HON	0.007	0.571	0.407	-0.0001	1.052	-0.082	0.218	0.364
IBM	0.00004	0.570	0.408	0.0005	1.051	-0.172	0.126	0.361
INTC	0.004	0.517	0.458	-0.009	1.052	-0.053	0.249	0.386
JNJ	0.002	0.551	0.433	-0.001	1.034	-0.093	0.176	0.346
JPM	0.00003	0.670	0.312	-0.0001	1.058	-0.119	0.200	0.370
KO	0.004	0.564	0.417	-0.011	1.045	-0.161	0.113	0.351
MCD	0.010	0.539	0.437	-0.023	1.052	-0.123	0.196	0.373
MMM	0.005	0.553	0.431	-0.013	1.032	-0.084	0.049	0.339
MRK	0.00006	0.581	0.401	-0.0006	1.039	-0.120	0.022	0.335
MSFT	0.013	0.693	0.286	-0.047	1.070	-0.099	0.143	0.391
NKE	0.0001	0.566	0.403	-0.0002	1.076	-0.155	0.158	0.381
PG	0.001	0.573	0.408	-0.003	1.047	-0.064	0.205	0.360
SPY	0.004	0.551	0.420	-0.009	1.068	-0.153	0.181	0.374
UNH	0.00008	0.490	0.484	0.00003	1.050	-0.066	0.185	0.366
WMT	0.0007	0.570	0.412	0.0001	1.042	-0.126	0.088	0.346

tionarity constraints:

$$\omega + \gamma\xi > 0,$$

$$0 < \beta + \gamma\phi < 1$$

As referenced in Gerlach and Wang (2016)[25], these restrictions ensure the positivity of each iteration of σ_t^2 and the existence and positivity of the long-run unconditional variance. We impose the same constraints during the parameter estimation of the Realized HAR-GARCH model, but now apply them to the sum of the daily, weekly, and monthly gamma terms:

$$\omega + (\gamma_d + \gamma_w + \gamma_m)\xi > 0,$$

$$0 < \beta + (\gamma_d + \gamma_w + \gamma_m)\phi < 1$$

Table 1 and Table 2 present the full sample parameter estimates for the Realized GARCH and Realized HAR-GARCH models, respectively. Our results are akin to those presented in Hansen et al. (2012)[23]. Considering Equation 2.4, one might expect an estimate of ϕ close to 1. If this is the case, then the bias in the realized measure is fully captured by the parameter ξ. In the case that $\xi = 0$, then the logged realized measure $\tilde{x}_t$ is an unbiased estimator for the true log of the latent volatility $\tilde{\sigma}_t^2$.

One should also consider the bias introduced by market micro-structure noise and the effects of out-of-hours trading activity. Since $\tilde{x}_t$ is constructed using intra-daily *open-close* returns, one would expect it to generally underestimate true log volatility, $\tilde{\sigma}_t^2$, which is based on daily *close-to-close* returns. This mismatch might lead us to expect a value of $\xi < 0$,

Table 2: Full sample estimates for the Realized HAR-GARCH model

Ticker	ω	β	γ_d	γ_w	γ_m	ξ	ϕ	τ_1	τ_2	σ_{ut}
AAPL	0.008	0.456	0.178	0.139	0.207	-0.015	1.033	-0.174	0.177	0.257
AMGN	0.119	0.286	0.295	0.398	0.009	-0.171	1.027	-0.090	0.160	0.493
AMZN	0.057	0.392	0.452	0.001	0.132	-0.097	1.039	-0.052	0.119	0.349
AXP	0.335	0.269	0.205	0.334	0.172	-0.472	1.028	0.003	0.194	0.450
BA	0.353	0.390	0.007	0.464	0.146	-0.580	0.992	-0.023	0.147	0.498
CAT	0.015	0.314	0.243	0.283	0.112	-0.023	1.076	-0.156	0.248	0.461
CSCO	0.206	0.347	0.443	0.116	0.071	-0.327	1.038	-0.165	0.223	0.247
CVX	0.022	0.012	0.331	0.303	0.330	-0.023	1.025	-0.121	0.132	0.530
DIS	0.089	0.202	0.458	0.105	0.192	-0.118	1.056	-0.213	0.188	0.444
GS	0.323	0.209	0.085	0.407	0.304	-0.406	0.992	-0.028	0.165	0.498
HD	0.0002	0.380	0.521	0.077	0.003	0.188	1.030	-0.157	0.171	0.794
HON	0.008	0.581	0.238	0.0003	0.159	-0.200	1.050	-0.222	0.173	0.495
IBM	0.021	0.246	0.135	0.400	0.183	-0.030	1.052	-0.038	0.085	0.412
INTC	0.028	0.397	0.198	0.238	0.139	-0.049	1.048	-0.072	0.135	0.362
JNJ	0.027	0.142	0.347	0.496	0.0001	-0.030	1.017	-0.293	0.136	0.582
JPM	0.094	0.120	0.449	0.0023	0.413	-0.109	1.020	-0.146	0.164	0.275
KO	0.119	0.280	0.365	0.283	0.047	-0.171	1.035	-0.108	0.121	0.365
MCD	0.110	0.303	0.395	0.231	0.046	-0.163	1.032	-0.056	0.218	0.363
MMM	0.037	0.162	0.544	0.002	0.248	-0.046	1.054	-0.107	0.055	0.324
MRK	0.350	0.228	0.375	0.155	0.224	-0.464	1.022	-0.302	0.0004	0.472
MSFT	0.008	0.189	0.209	0.011	0.601	-0.497	0.974	-0.238	0.084	0.484
NKE	0.064	0.273	0.374	0.001	0.311	-0.094	1.061	-0.115	0.113	0.336
PG	0.009	0.340	0.419	0.153	0.068	-0.014	1.031	-0.046	0.109	0.329
SPY	0.559	0.216	0.401	0.132	0.266	-0.696	0.978	-0.171	0.202	0.443
UNH	0.040	0.351	0.223	0.198	0.185	-0.0001	1.070	-0.137	0.201	0.404
WMT	0.040	0.440	0.210	0.133	0.190	-0.075	1.047	-0.118	0.127	0.309

which is seen across the vast majority of the returns series for both model estimates. We see negative estimates of τ_1 for both models throughout our datasets, which is consistent with the concept of asymmetry between returns and next-day volatility, known as the leverage effect. We calculate the average γ_d, γ_w, and γ_m coefficients across the series in Table 2 and find $\gamma_d = 0.312 > \gamma_w = 0.195 > \gamma_m = 0.183$, suggesting that there is a decreasing amount of information about tomorrow's volatility as we move from daily to weekly to monthly data.

2.4 Methods for Forecast Evaluation

Volatility Forecasts: True volatility is inherently unobserved, even ex-post, which poses a significant challenge in evaluating predictions across different models. Patton (2010)[30] derives two classes of loss functions that can *consistently* rank model forecasts, under the assumption that $\tilde{\sigma}_t$ is an unbiased proxy for volatility:

$$\mathrm{QL}: L\left(\tilde{\sigma}_t^2, \sigma_{t|t-1}^2\right) = \frac{\tilde{\sigma}_t^2}{\sigma_{t|t-1}^2} - \log\frac{\tilde{\sigma}_t^2}{\sigma_{t|t-1}^2} - 1$$

$$\mathrm{MSE}: L\left(\tilde{\sigma}_t^2, \sigma_{t|t-1}^2\right) = \left(\tilde{\sigma}_t^2 - \sigma_{t|t-1}^2\right)^2$$

where σ_t^2 represents a volatility proxy, and $\tilde{\sigma}_t^2$ represents the volatility forecast. We therefore employ both MSE and QL. Finally, we include Root Mean Squared Error (RMSE) in our metrics as it reduces the effect of outliers in the evaluation.

As our volatility proxy (Realized Variance) is constructed using intra-day returns whilst latent volatility is based on the close-to-close returns, we make the following adjustment during evaluation to account for this difference:

$$RV_t^{\mathrm{Adj}} = \frac{\sum_{s=1}^{T} r_s^2}{\sum_{s=1}^{T} RV_s} RV_t \tag{2.7}$$

where RV_s stands for the close-to-close Realized Variance and r_s is the close-to-close return.

VaR Forecasts: We can evaluate a model's forecast performance by comparing the predicted losses from a Value at Risk (VaR) model to the actual losses over a given period. Instances where actual losses surpass the VaR threshold at a given significance level p can be expressed as:

$$H_t = (r_t < \mathrm{VaR}_t^p). \tag{2.8}$$

where H_t, is the hit indicator.

The violation rate can then be calculated as follows:

$$\mathrm{VRate} = \frac{1}{m} \sum_{t=n+1}^{n+m} H_t \tag{2.9}$$

We also apply two common backtesting techniques to evaluate our predictions:

Unconditional Coverage (UC): Kupiec (1995)[32] proposed a log-likelihood ratio (LR) test to examine the frequency of VaR violations over a given time horizon. The LR test is given by:

$$LR_{uc} = -2 \log \left(\frac{p^{n_1}(1-p)^{n_0}}{\hat{\pi}^{n_1}(1-\hat{\pi})^{n_0}} \right) \sim \chi^2_{(1)}, \tag{2.10}$$

where n_0 and n_1 represent the number of 0's and 1's in the series, respectively, and $\hat{\pi} = n_1/(n_0+n_1)$. The null hypothesis of the UC test is that the failure probability is equal to p, and it is tested against the alternative of a failure rate different from p.

Dynamic Quantile (DQ): If H_t is an i.i.d process it should not be possible to predict future failure given past information. Engle and Manganelli (2004)[33] propose a regression-based test to examine the correct unconditional coverage of violations by testing whether the hit sequence H_t is i.i.d. The test statistic for the null hypothesis $H_0 : \beta = 0$, of H_t uncorrelated with past hits is given by:

$$DQ_{\mathrm{hit}} = \frac{\hat{\beta}'_{\mathrm{LS}} X' X \hat{\beta}_{\mathrm{LS}}}{p(1-p)} \sim \chi^2_q, \tag{2.11}$$

where q is the dimension of the β vector.

Expected Shortfall Forecasts: To evaluate our Expected Shortfall (ES) forecasts, we first employ the auxiliary ES regression-based backtesting methodology proposed by Bayer and Dimitriadis (2020)[34]. By regressing the returns on the ES forecasts, this method tests whether the ES predictions are well-specified, meaning they do not systematically underestimate or overestimate ES. In the Auxiliary ESR Backtest, the regression system is given by

$$\begin{aligned} Y_t &= \beta_1 + \beta_2 \hat{v}_t + u_t^q, \\ Y_t &= \gamma_1 + \gamma_2 \hat{e}_t + u_t^e, \end{aligned}$$

where $\hat{v}_t$ represents the VaR forecasts, $\hat{e}_t$ represents the ES forecasts, and Y_t is a series of realized returns.

The hypothesis tested is:

$$\begin{cases} H_0 : (\gamma_1, \gamma_2) = (0, 1) \\ H_A : (\gamma_1, \gamma_2) \neq (0, 1) \end{cases}$$

Rejection of H_0 indicates misspecification, while failure to reject suggests that our ES forecasts are properly specified.

Another way to evaluate forecasts is through a loss function. Although Expected Shortfall (ES) is not elicitable on its own—meaning there is no loss function that can be uniquely optimized by the "true" Expected Shortfall—Fissler and Ziegel (2016)[35] demonstrate that the pair (VaR, ES) is elicitable. Building on the (GAS) framework of Creal et al. (2013)[36], Patton (2019)[37] proposed a reparametrization of the FZ loss function that we will adopt for our ES evaluation:

$$FZ^0 = \frac{1}{\alpha ES_t^\alpha} I_t^\alpha (y_t - VaR_t^\alpha) + \frac{VaR_t^\alpha}{ES_t^\alpha} + \log(-ES_t^\alpha) - 1.$$

Here, y_t is the observed return at time t and

$$I_t^\alpha (y_t - VaR_t^\alpha) = \begin{cases} 1 & \text{if } y_t \leq VaR_t^\alpha, \\ 0 & \text{otherwise.} \end{cases}$$

2.4.1 Diebold-Mariano test

Diebold and Mariano (1995)[38] propose a test to compare the predictive accuracy of two competing models. First, we define the loss differential d_t as the difference between the loss functions of the two forecasts:

$$d_t = L\left(e_t^{RG}\right) - L\left(e_t^{RHG}\right)$$

where $L(\cdot)$ is the loss function and e_t^{RG} and e_t^{RHG} are the forecasting errors of the Realized GARCH and Realized HAR-GARCH models, respectively. The test statistic can be computed as:

$$\frac{\bar{d}}{\sqrt{\frac{2\pi \hat{f}_d(0)}{\hat{T}}}}$$

where $\hat{f}_d(0)$ is a consistent estimator of the true variance of d_t and $\bar{d} = \frac{1}{T}\sum_{t=1}^{T} d_t$.

We test the hypothesis of:

$$\begin{cases} H_0 : E(d_t) = 0 \\ H_A : E(d_t) \neq 0 \end{cases}$$

A large positive (negative) value would indicate that the Realized HAR-GARCH (Realized GARCH) provides superior forecast accuracy.

2.5 Design of Forecast Study

To assess the robustness and accuracy of our two forecasting models, we evaluate their performance on out-of-sample data over several distinct time horizons. Specifically, we analyze volatility predictions at $h = 1$, 5, 10, and 20 day-ahead intervals. We also examine Value at Risk (VaR) and Expected Shortfall (ES) forecasts at $h = 1$ and $h = 10$ day-ahead horizons, focusing on two significance levels of $\alpha = 0.01$ and 0.05. Evaluating VaR and ES over these horizons provides insights into the potential risk exposure and the tail-end risk of extreme losses under different market conditions.

To further contextualize our findings, we integrate an economic application of our ES forecasts by exploring their use in portfolio hedging strategies. This involves an investor holding a long position in a single stock and using ES predictions to determine the optimal number of shares to hedge, given a certain significance level and a maximum accepted expected loss.

3. Data

Our full sample spans over twenty years from 10 October, 2003 to 28 March, 2024. We use data on close-to-close returns for the S&P 500 market index as well as 25 individual stocks from the Dow Jones Industrial Average Index (DJI). Our choice of equities was based on the following criteria: 1) Sufficient liquidity for 5-minute data sampling to be applied. 2) Availability of intra-daily data for the entire sample period. 3) Maintaining a position in the Dow Jones 30 Index for the majority of the sample period. These 25 stocks matched these criteria and thus were selected for this study.

We use 5-minute return data collected from Polygon API to calculate the daily Realized Variance of each returns series, which we use as our realized measure. Sampling at the 5-minute frequency has been empirically shown to strike a good balance between estimation accuracy of unobserved volatility and minimisation of market micro-structure noise, and is thus why we sample at this frequency. The official NYSE and NASDAQ market trading hours begin at 9:30 AM and end at 4:00 PM each day, however, it is not uncommon for trades to be processed after official market closure. These trades typically stem from orders placed in the final moments preceding market closure and those placed in out-of-hours trading sessions. This creates the problem of defining an appropriate cut-off to define the true closing price of each day. To overcome this, we follow Brownlees and Gallo (2006)[19] and define daily trading hours to span from 9:30 AM to 4:05 PM – allowing us to capture trades placed in the final moments before market closure but discard those placed in out-of-hours sessions.

We follow the approach of Hansen & Huang (2015)[28] when removing short trading days from our sample to avoid outliers. Trading days containing too few 5-minute data points results in the realized variance calculation considering a smaller percentage of daily volatility, and thus are likely to be significantly smaller. To overcome this, we remove days that contain less than 66 high frequency data points. A standard trading day contains 78 5-minute intervals, and so we drop those containing less than 80% of this, which is 66. This removed on average 43 daily observations from each sample across our returns series, or approximately 2 days per year.

Table 3 contains summary statistics for the daily close-to-close returns and realized variance (RV) for each series. Each returns series exhibits significant kurtosis, indicating that the returns distributions in our sample are leptokurtic so extreme events are more likely than those implied under normality. We confirmed the stationarity of each returns series with Augmented Dickey-Fuller tests.

Across many series, the distribution of RV appears generally right-skewed, indicated by median values that fall below the mean. This is further confirmed by the values for Q3 often being very similar to the mean, indicating presence of a few large values in the data that are skewing the mean of the realized variance upwards.

Figure 1 plots the returns, realized variance, and logarithm of realized variance for a number of series. To save space we only present the results for the following tickers for the remainder of the paper: Apple Inc. (AAPL), Boeing Co (BA), IBM (IBM), JPMorgan Chase & Co (JPM), 3M (MMM) and the S&P 500 market index (SPY). These stocks were selected to represent a diverse range of market sectors. Additionally, we include the S&P 500 as it serves as a broad benchmark for the overall market performance. Full results for the remainder of the stocks are presented in an unpublished appendix, which is available upon request.

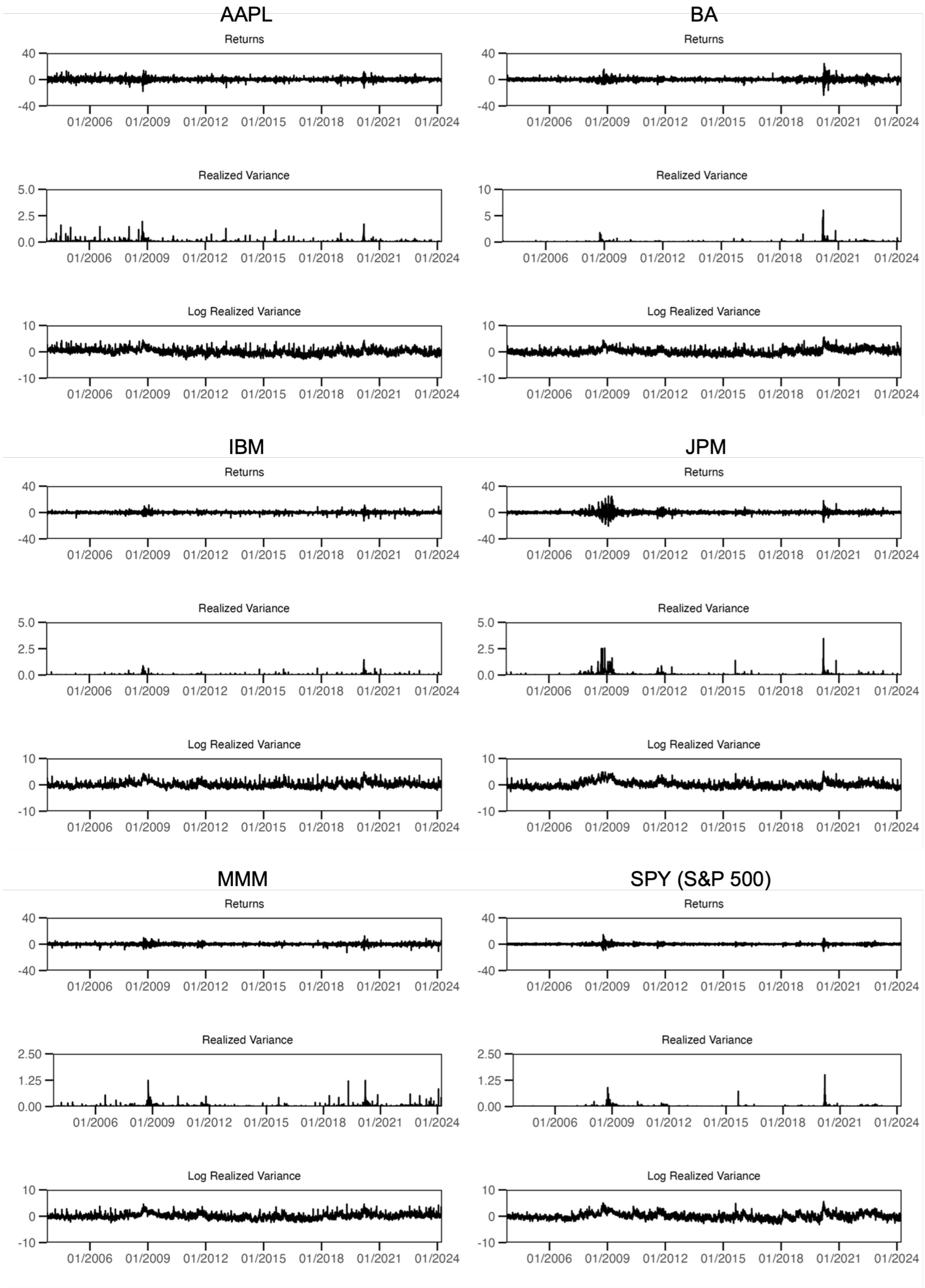


Figure 1: Returns, realized variance, and log-realized variance for various series

Table 3: Summary statistics for returns and realized variance

		Return					RV				
Ticker	Obs	Mean	Std	Skew	Kurt	Median	Mean	Std	Q1	Median	Q3
AAPL	5140	0.04	0.09	0.10	8.32	0.02	0.04	0.09	0.01	0.02	0.04
AMGN	5106	0.03	0.08	0.66	11.07	0.02	0.03	0.08	0.01	0.02	0.03
AMZN	5109	0.06	0.16	0.78	17.01	0.03	0.06	0.16	0.02	0.03	0.05
AXP	5107	0.05	0.14	0.76	19.56	0.01	0.05	0.14	0.01	0.01	0.03
BA	5110	0.05	0.18	0.26	19.56	0.02	0.05	0.18	0.01	0.02	0.04
CAT	5107	0.04	0.09	-0.05	8.31	0.02	0.04	0.09	0.01	0.02	0.04
CSCO	5107	0.04	0.09	-0.14	14.03	0.02	0.04	0.09	0.01	0.02	0.03
CVX	5108	0.03	0.10	0.12	23.95	0.02	0.03	0.10	0.01	0.02	0.03
DIS	5110	0.03	0.09	0.54	13.66	0.01	0.03	0.09	0.01	0.01	0.03
GS	5107	0.05	0.21	0.80	21.60	0.02	0.05	0.21	0.01	0.02	0.04
HD	5107	0.03	0.08	-0.00	13.55	0.01	0.03	0.08	0.01	0.01	0.03
HON	5107	0.03	0.09	0.08	10.18	0.01	0.03	0.09	0.01	0.01	0.03
IBM	5107	0.02	0.05	-0.17	11.63	0.01	0.02	0.05	0.01	0.01	0.02
INTC	5111	0.04	0.08	-0.15	11.27	0.02	0.04	0.08	0.01	0.02	0.04
JNJ	5107	0.01	0.05	0.21	14.64	0.01	0.01	0.05	0.00	0.01	0.01
JPM	5109	0.05	0.15	0.97	23.25	0.02	0.05	0.15	0.01	0.02	0.03
KO	5109	0.02	0.05	0.01	15.77	0.01	0.02	0.05	0.00	0.01	0.01
MCD	5106	0.02	0.06	0.35	21.29	0.01	0.02	0.06	0.01	0.01	0.02
MMM	5107	0.02	0.05	-0.33	11.06	0.01	0.02	0.05	0.01	0.01	0.02
MRK	5108	0.03	0.13	-0.99	28.07	0.01	0.03	0.13	0.01	0.01	0.02
MSFT	5111	0.03	0.07	0.23	13.09	0.01	0.03	0.07	0.01	0.01	0.03
NKE	5108	0.03	0.09	0.47	12.74	0.01	0.03	0.09	0.01	0.01	0.03
PG	5107	0.02	0.05	0.11	13.47	0.01	0.02	0.05	0.00	0.01	0.01
SPY	5140	0.01	0.04	-0.07	18.10	0.00	0.01	0.04	0.00	0.00	0.01
UNH	5106	0.04	0.10	1.03	34.93	0.02	0.04	0.10	0.01	0.02	0.03
WMT	5107	0.02	0.05	0.18	15.61	0.01	0.02	0.05	0.01	0.01	0.02

4. Forecasting Results

This section presents our multi-period forecasting results. The results are presented as follows:

1. Explanation of the prediction mechanism used for horizons $h = 1$ and $h > 1$.
2. Out of sample volatility forecast results.
3. Value at Risk (VaR) and Expected Shortfall (ES) forecast results and evaluation.
4. Economic evaluation of forecast accuracy.

4.1 Multi-step Forecasting

Our out-of-sample period spans from 19 February, 2019, to 28 March, 2024, containing 1278 daily observations. For the one-step ahead forecast, $h = 1$, the procedure is straightforward as we can simply use our estimated model parameters, θ^{RG} and θ^{RHG}, to directly estimate Equations 2.4 and 2.5. However, when forecasting multiple steps ahead for $h = k$, where $k > 1$, the process becomes more challenging and so we resort to the use of simulation techniques.

To compute the multi-step forecasts for $\tilde{x}_t$ we substitute the GARCH variance Equation (2.5) into the measurement Equation (2.4) to achieve the result below. For conciseness, we only show this for the Realized HAR-GARCH model, but the same approach applies to that of the Realized GARCH model:

$$\tilde{x}_t = \mu_x + (\beta + \phi\gamma_d)\tilde{x}_{t-1} + \frac{\phi\gamma_w}{4}\sum_{i=2}^{5}\tilde{x}_{t-i} + \frac{\phi\gamma_m}{17}\sum_{i=6}^{22}\tilde{x}_{t-i} + \epsilon_t - \beta\epsilon_{t-1}, \quad (4.1)$$

where $\mu_x = \phi\omega + \xi(1-\beta)$ and $\epsilon_t = \tau(z_t) + u_t$.

This representation is that of an ARMA(22,1) model, and we use this to forecast $\tilde{x}_{t+h}$ in the following simulation exercise.

4.1.1 Simulation Procedure

To obtain a forecast for the n-period ahead volatility forecast σ^2_{t+n} at time t we perform the following steps:

1. We initialize the necessary values from the end of the training dataset and the estimated model parameters. This includes the most recent values of the variance $\tilde{\sigma}^2_t$, the realized variance $\tilde{x}_t$, and the error terms z_t and u_t.
2. We then calculate the $t \times 2$ matrix W_t containing the pairs of error terms z_t and u_t from our in-sample data, where t is the length of our in-sample data. We then draw $s = 5000$ random pairs from W_t and simulate $\tilde{x}_{t+n}$ via the measurement equation (2.4) for each draw. In the case of the HAR-GARCH model, we also update the corresponding weekly $\left(\sum_{i=2}^{5}\tilde{x}_{t+n-i}\right)$ and monthly $\left(\sum_{i=6}^{22}\tilde{x}_{t+n-i}\right)$ volatility terms.
3. At each draw, we use the simulated $\tilde{x}_{t+n}$ to calculate the logarithm of volatility, $\tilde{\sigma}^2_{t+n+1}$. We then obtain an estimate of the $n+1$ period ahead forecast $\hat{\sigma}_{t+n+1}$ by taking the average of the exponentiated simulated values:

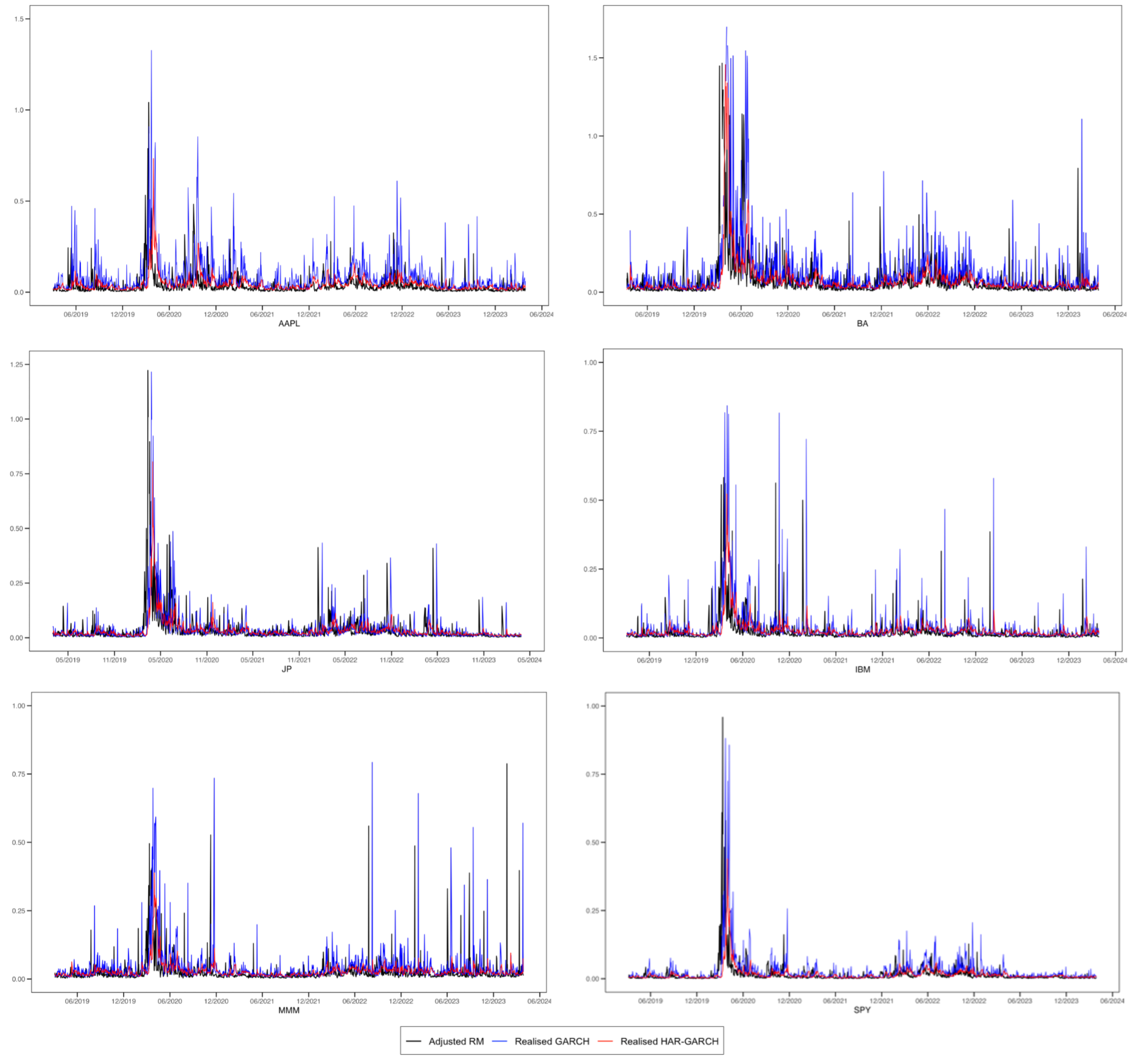
$$\hat{\sigma}^2_{t+n+1} = \frac{1}{5000}\sum_{s=1}^{5000}\exp\left(\tilde{\sigma}^2_{t+n+1}\right)$$

4. We then use the average values as the new initial values and repeat the simulation process until reaching the end of the desired forecasting horizon such that $n = k$.

4.2 Volatility Forecasts

Figure 2 plots the 10-day rolling forecasts for the Realized GARCH and Realized HAR-GARCH models across six series. We notice that during periods of volatility spikes, the Realized GARCH model tends to predict higher volatility compared to its Realized

Figure 2: Realized GARCH & Realized HAR-GARCH: 10-days ahead volatility forecasts



HAR-GARCH counterpart. This difference is due to the Realized GARCH model's reliance on limited information about the dynamics of volatility, captured solely through its previous-day forecast of RV by γ_d. In contrast, the Realized HAR-GARCH leverages both the one-step forecast and information from weekly and monthly volatility through γ_w and γ_m, resulting in smoother predictions over longer forecast horizons.

These visual results are confirmed upon viewing Table 4, which presents the performance metrics for the competing models across a set of different stocks and forecast horizons $h = 1, 5, 10$, and 20. For the 1 day-ahead forecasts the Realized HAR-GARCH model performs best in MSE and RMSE for 4 out of the 6 series, yet no model appears dominant in terms of QL. Diebold Mariano (DM) tests confirm that overall no model is significantly more accurate than the other at this horizon. For longer forecast horizons we see a greater difference in performance, with the Realized HAR-GARCH model outperforming Realized GARCH in all series across horizons $h = 5, 10$, and 20. This performance gap widens as the forecast horizon increases, which is confirmed by DM tests. The strong forecasting ability of the Realized HAR-GARCH model across longer horizons points to its capacity to capture the long-memory characteristics of volatility through γ_w and γ_m.

Table 4: Mean squared error (MSE), RMSE, and quasi-likelihood (QLIKE) of volatility forecasts

Horizon	**Ticker**	**MSE**		**RMSE**		**QLIKE**	
		RG	RHG	RG	RHG	RG	RHG
$h = 1$	AAPL	0.855	0.622†	0.925	0.789†	0.380	0.407
	BA	0.708	0.700	0.842	0.837	0.623	0.674
	IBM	0.828	0.850	0.910	0.922	0.424	0.436
	JPM	0.634	0.660	0.796	0.812	0.458	0.395
	MMM	0.770	0.747	0.878	0.864	0.466	0.465
	SPY	0.857	0.721†	0.926	0.849†	0.470	0.429†
$h = 5$	AAPL	1.722	0.985†	1.312	0.992†	0.582	0.467†
	BA	1.258	0.865†	1.122	0.930†	0.564	0.534†
	IBM	1.274	1.056†	1.129	1.027†	0.583	0.516†
	JPM	0.961	0.765†	0.980	0.875†	0.615	0.456†
	MMM	1.183	0.842†	1.088	0.918†	0.572	0.572†
	SPY	1.442	1.070†	1.201	1.034†	0.568	0.531†
$h = 10$	AAPL	1.969	1.075†	1.403	1.037†	0.657	0.535†
	BA	1.442	0.920†	1.201	0.959†	0.656	0.740†
	IBM	1.406	1.022†	1.186	1.011†	0.621	0.542†
	JPM	1.154	0.853†	1.074	0.923†	0.797	0.660‡
	MMM	1.265	0.847†	1.125	0.920†	0.686	0.538†
	SPY	1.797	1.201†	1.341	1.096†	0.735	0.729†
$h = 20$	AAPL	2.073	1.103†	1.440	1.050†	0.783	0.624†
	BA	1.662	0.957†	1.289	0.978†	1.265	1.317‡
	IBM	1.611	1.069†	1.269	1.034†	0.936	0.740†
	JPM	1.219	0.948†	1.104	0.973†	1.737	1.030†
	MMM	1.408	0.835†	1.186	0.914†	0.734	0.621†
	SPY	2.139	1.344†	1.462	1.159†	1.546	1.297†

Notes: The table presents the Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Quasi-Likelihood (QLIKE) metrics at the following forecast horizons: 1 day, 5 days, 10 days, and 20 days. A single dagger (†) indicates that the model is significantly better at the 1% significance level, while a double dagger (‡) indicates significance at the 5% level according to the Diebold-Mariano (DM) test.

4.3 VaR and ES Forecasts

In this section, we evaluate the performance of our models in forecasting the 1-day and 10-day ahead Value at Risk (VaR) and Expected Shortfall (ES). Specifically, the 10-day-ahead VaR forecast at the 99% confidence level represents our 10-step ahead prediction for the 1-day VaR. In other words, we are estimating the value at risk for a single day, but 10 days into the future, with a 1% probability of being exceeded. The same applies to the 10-step-ahead ES estimates.

Figure 3 shows the 1-day ahead VaR forecasts for the Realized GARCH and Realized HAR-GARCH models and daily close-to-close returns at the $\alpha = 1\%$ and 5% level. To save space we only present the results for IBM. For plots of the other series please refer to the attached appendix.

Table 5 presents the violation rates for $\alpha = 1\%$ and 5% along with the p-values from two VaR backtests at horizons $h = 1$ and $h = 10$. We find the Realized HAR-GARCH model shows, on average, higher violation rates than the Realized GARCH model for $h = 1$.

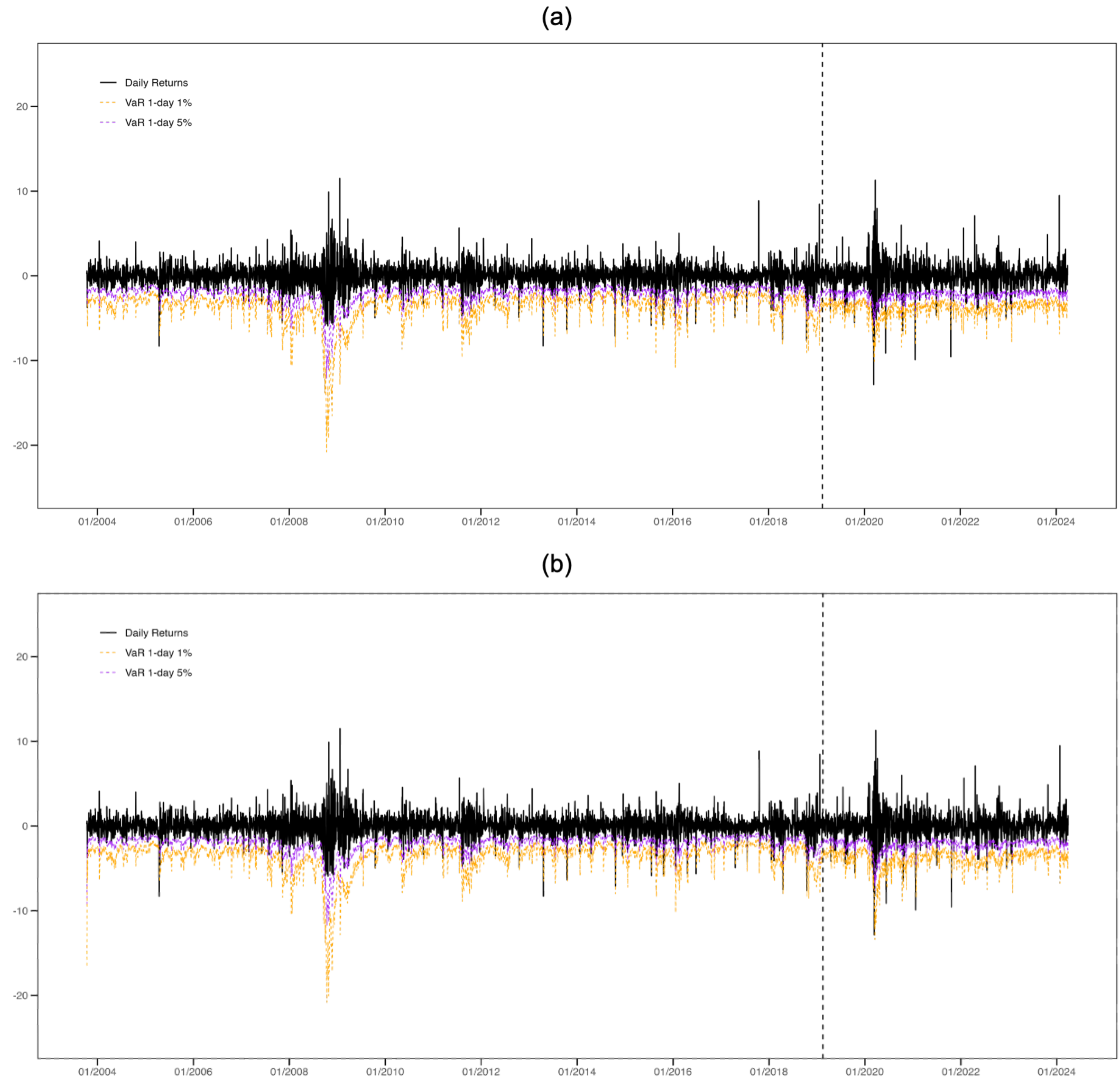


Figure 3: IBM One-day ahead VaR forecasts and daily returns for the Realized GARCH model (upper panel) and Realized HAR-GARCH model (lower panel) at the 5% and 1% significance levels. The vertical dashed line indicates the beginning of the out-of-sample period, beginning February 19, 2019, to 28 March, 2024.

In the case of IBM, the RHG model is remarkably accurate and passes both backtesting methods (UC and DQ) at both the 1% and 5% levels compared to RG passing just the UC backtest. Both models also pass the two backtests for MMM at both levels.

Violation rates for the 10-day ahead forecasts are in favor of the Realized GARCH model at the 5% level, and the Realized HAR-GARCH at the 1%. Both models however fail to pass the Unconditional Coverage test in over half of the cases. They also fail the Dynamic Quantile (DQ) test with low p-values more often than not, indicating a presence of clustering among violations that are neither uniformly distributed nor independent.

The furthest-right column lists p-values from the Auxiliary ESR backtest, which jointly tests the VaR and ES forecasts for misspecification. We find both models perform similarly across various horizons and significance levels, with RHG outperforming RG in some stocks, and RG outperforming RHG in others. Whilst both models generally fail to reject the null of correct specification at the 1% level, the Realized GARCH

Table 5: Evaluation of one and ten-day-ahead VaR performance

			No. of violations		Violation rate		UC p-value		DQ p-value		ES	
Period	**Confidence Level**	**Ticker**	**RG**	**RHG**	**RG**	**RHG**	**RG**	**RHG**	**RG**	**RHG**	**RG**	**RHG**
Horizon 1	**1%**											
		AAPL	16	29	1.2	2.3	**0.395**	0.000	0.000	0.000	0.008	**0.128**
		BA	36	46	2.8	3.6	0.000	0.000	0.000	0.000	0.006	0.000
		IBM	14	12	1.1	0.9	**0.733**	**0.827**	0.000	**0.307**	**0.148**	**0.170**
		JPM	19	49	1.5	3.8	**0.103**	0.000	0.000	0.000	**0.714**	**0.043**
		MMM	9	8	0.7	0.6	**0.263**	**0.150**	**0.930**	**0.906**	**0.362**	**0.257**
		SPY	47	34	3.7	2.6	0.000	0.000	0.000	0.000	0.007	0.004
	5%											
		AAPL	88	136	6.8	10.6	0.004	0.000	0.000	0.000	0.003	0.000
		BA	125	137	9.8	10.7	0.000	0.000	0.000	0.000	0.001	0.000
		IBM	69	64	5.4	5.0	**0.514**	**0.985**	0.001	**0.991**	0.037	**0.093**
		JPM	141	228	11.0	17.8	0.000	0.000	0.000	0.000	**0.912**	0.000
		MMM	65	69	5.1	5.4	**0.883**	**0.514**	**0.457**	**0.265**	**0.571**	**0.085**
		SPY	110	90	8.6	7.0	0.000	0.002	0.000	0.000	0.005	0.004
Horizon 10	**1%**											
		AAPL	4	21	0.3	1.6	0.004	**0.034**	**0.454**	0.000	0.011	**0.132**
		BA	26	35	2.0	2.8	0.001	0.000	0.000	0.000	**0.070**	**0.023**
		IBM	22	20	1.7	1.6	**0.017**	**0.057**	0.000	0.000	**0.086**	0.010
		JPM	38	52	3.0	4.1	0.000	0.000	0.000	0.000	**0.060**	**0.090**
		MMM	10	13	0.8	1.0	**0.432**	**0.928**	**0.782**	**0.054**	**0.036**	**0.213**
		SPY	38	35	3.0	2.7	0.000	0.000	0.000	0.000	0.002	0.006
	5%											
		AAPL	62	103	4.9	8.1	**0.816**	0.000	0.003	0.000	0.024	0.001
		BA	80	93	6.3	7.3	0.040	0.000	0.000	0.000	0.010	0.002
		IBM	74	67	5.8	5.3	**0.183**	**0.646**	0.000	0.000	0.034	0.035
		JPM	135	223	10.6	17.6	0.000	0.000	0.000	0.000	0.000	0.000
		MMM	69	81	5.4	6.4	**0.477**	0.029	0.007	0.000	0.012	**0.085**
		SPY	83	79	6.5	6.2	0.018	**0.059**	0.000	0.000	0.002	0.004

Notes: The table displays VaR performance over 1-day and 10-day-ahead forecast horizons. The "UC" and "DQ" columns represent the Unconditional Coverage and Dynamic Quantile tests, respectively. We set the number of lags = 4 in these tests. The "ES" column displays the p-values for the Auxiliary ESR Backtest, where a high p-value indicates a correct specification of the ES forecasts.

model often shows slightly better specification than the Realized HAR-GARCH model, particularly for $\alpha = 5\%$.

Figure 4 illustrates the 1-day ahead Expected Shortfall forecasts for the Realized GARCH and Realized HAR-GARCH model at the 1% and 5% levels for IBM. The plots for the remaining series can be found in the appendix. As previously mentioned, ES itself is not elicitable and thus we resort to the FZ^0 test, which evaluates the joint accuracy of VaR and ES forecasts using a loss function that accounts for both tail risk measures. The results can be seen in Table 6. Realized GARCH is the dominant model here, achieving a lower average loss than the Realized HAR-GARCH model across all tickers over both the 1-day and 10-day ahead forecast horizons.

To test whether this outperformance is statistically significant, we conduct the pairwise Diebold Mariano (DM) test for each model. Table 7 reports the test statistics and associated p-values, with statistically significant values highlighted in bold. We employ a two-sided test such that the alternative hypothesis states that the predictive accuracy of the two models is different. Interestingly, we find that the additional forecast accuracy of the Realized GARCH model is only statistically significant in 2 out of the 6 series for the 1-day ahead forecasts (AAPL & JPM) at the 1% level,

Table 6: FZ^0 Loss function results for out-of-sample forecasts of VaR and ES

Ticker	**Horizon 1**		**Horizon 10**	
	RG	RHG	RG	RHG
AAPL	-2.932	-2.692	-2.808	-2.654
BA	-1.988	-1.941	-1.938	-1.574
IBM	-2.920	-2.625	-2.751	-2.650
JPM	-2.970	-2.281	-2.641	-1.676
MMM	-2.604	-2.375	-2.593	-2.588
SPY	-2.828	-2.567	-3.100	-2.404

Notes: Values in the table correspond to the average FZ^0 loss for the Realized GARCH and Realized HAR-GARCH joint VaR & ES forecasts, tested at the 1% level. A lower value represents a lower overall loss.

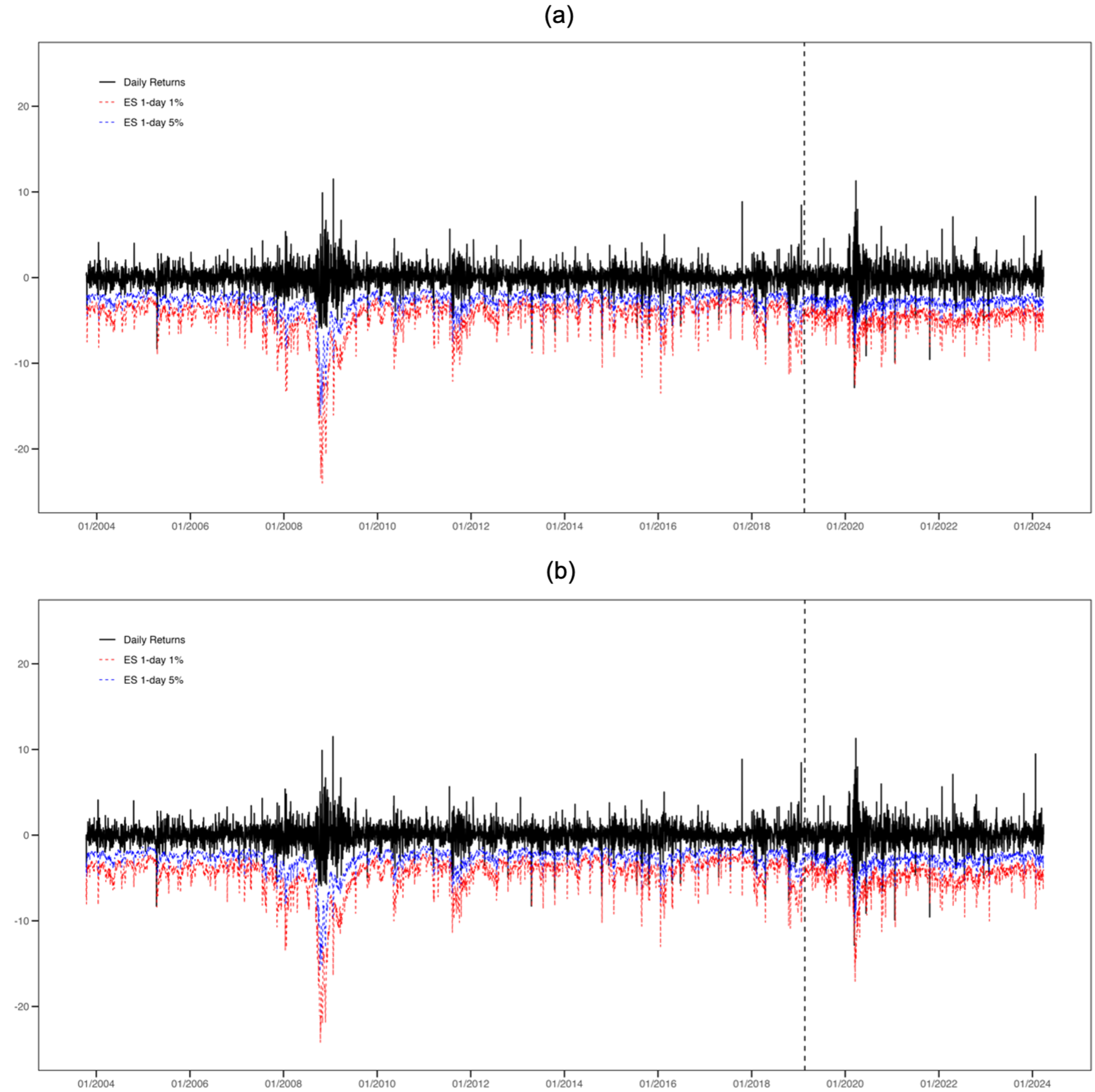


Figure 4: IBM One-day ahead Expected Shortfall (ES) forecasts and daily returns for the Realized GARCH model (upper panel) and Realized HAR-GARCH model (lower panel) at the 5% and 1% significance levels. The vertical dashed line indicates the beginning of the out-of-sample period, beginning February 19, 2019.

with just AAPL being significant at the 5% level for the 10-day horizon. For the remaining tickers, we find no statistically significant difference in the predictive accuracy of the two models.

4.4 Expected Shortfall for Portfolio Hedging

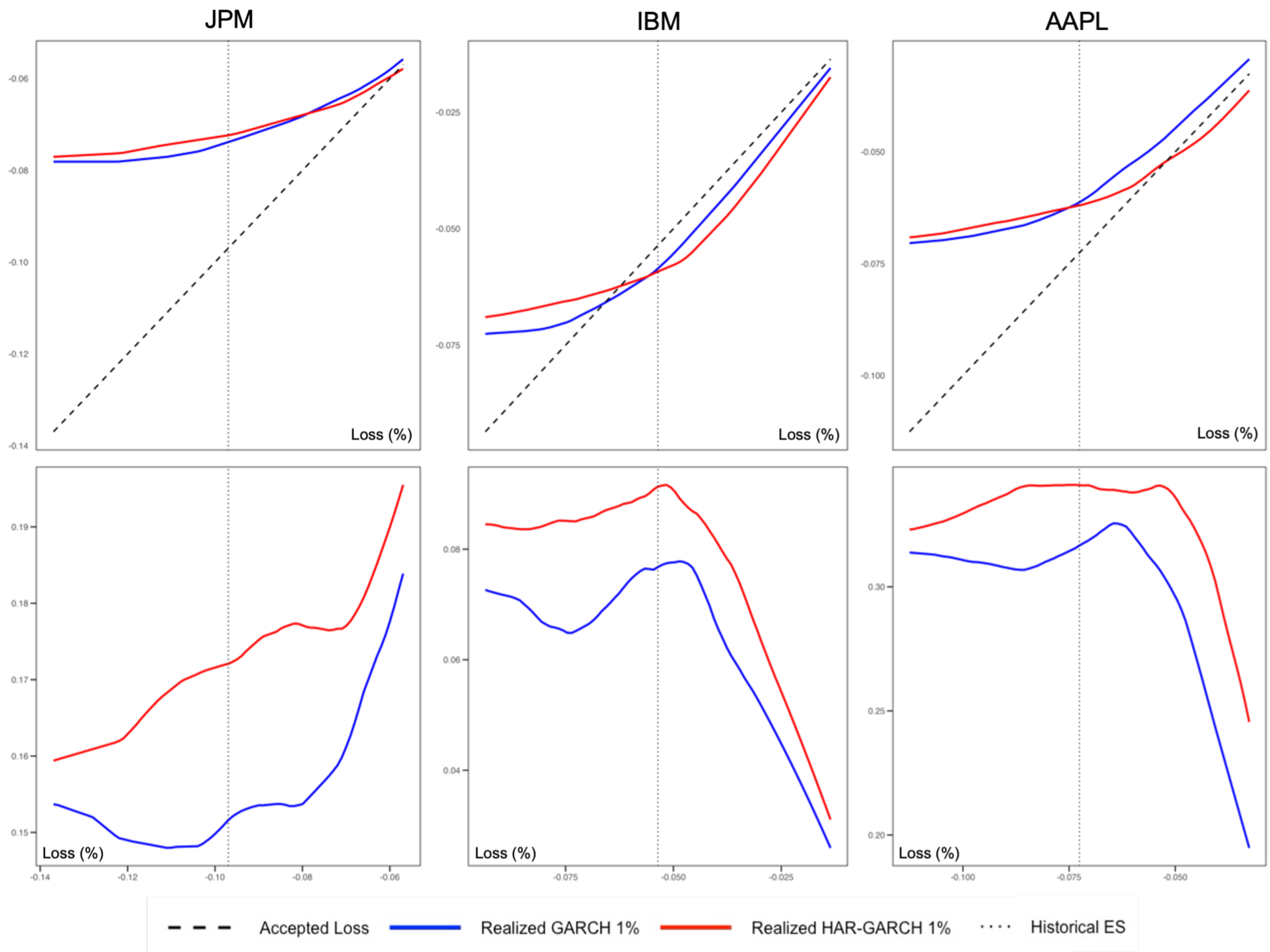
In this section, we evaluate our model performance within the context of a portfolio hedging exercise inspired by Lyócsa and Plíhal's (2024)[39] economic evaluation of expected shortfall. For simplicity, we consider an investor who holds a portfolio with a long position in one stock, where her goal is to minimize her risk. Given a certain significance level (in this we consider only at

Table 7: DM forecast evaluation for VaR and ES jointly for different stocks over different horizons

	Horizon 1		Horizon 10	
Ticker	**Statistic**	**p-value**	**Statistic**	**p-value**
AAPL	-2.723	**0.007**	-2.304	**0.021**
BA	-0.051	0.959	-1.191	0.234
IBM	-1.150	0.250	-0.192	0.848
JPM	-2.660	**0.008**	-1.267	0.205
MMM	-1.297	0.195	1.330	0.184
SPY	1.579	0.115	-0.584	0.559

Notes: Values in the table correspond to the two-sided Diebold Mariano (DM) test statistics and p-values for VaR & ES, at $h = 1$ and $h = 10$ day-ahead forecasts.

Figure 5: Evaluation of a hedging strategy using forecasts of expected shortfall



Notes: The top row of the figure shows the expected shortfall to which an investor is exposed after using either a hedging strategy from the Realized GARCH model (blue line) or the Realized HAR-GARCH model (red line), given an accepted level of loss. The y-axis denotes the expected shortfall. The x-axis represents the accepted loss of the investor. The vertical dotted line is the historical expected shortfall calculated from each stock's in-sample data. The bottom row displays the average annualized daily returns (%) from using either strategy at each level of accepted loss.

the 1% level) and her maximum accepted expected loss (which is the highest level of expected shortfall they are willing to be exposed to), the investor will use forecasts of expected shortfall from the Realized GARCH and Realized HAR-GARCH models to determine the number of shares to hedge.

Whenever the investor's forecasted expected shortfall exceeds her accepted level of risk, the hedge ratio is defined by:

$$H_t(L) = \frac{ES_{R_t}(\tau) - L}{ES_{R_t}(\tau)}$$

where $ES_{R_t}(\tau)$ is the expected shortfall at time t and significance level τ, and L is the accepted expected loss. If her predictions are lower than her accepted level of loss (risk level), there is no need to hedge the portfolio and thus $H_t(L) = 0$.

The return of the hedge position can then be computed as:

$$HR_t(L) = R_t \times (1 - H_t(L))$$

The blue and red lines indicate the level of expected shortfall the investor is exposed to, given that she has undertaken a hedging strategy inferred by either the Realized GARCH (blue) or Realized HAR-GARCH (red) model. Hedging should lead to a lower realized ES because we only enter a short position when the forecasted ES is greater than the accepted level of loss L. A successful hedging strategy results in the curved lines lying above the diagonal line, which we denote as a reduction in expected shortfall (ES).

We can observe from the top row of Figure 5 that across all three stocks, there is a greater reduction in ES for those with a high accepted level of loss, regardless of which models' strategy they follow. In the case of JPM, these gains in reduced ES diminish as the level of accepted loss falls (moving rightwards). For IBM, the actual ES following either model falls below the level of accepted loss for many investors with more conservative risk profiles, as seen by the red and blue line crossing below the diagonal line. Interestingly, in the case of AAPL, only the strategy using Realized GARCH leads to a reduction in ES across all levels of accepted loss.

The bottom row of Figure 5 shows the average annualized returns under each strategy. In our specific test-period and chosen stocks, it happens that hedging actually leads to positive returns. However, the key point to focus on here is the difference in returns between the two strategies, not whether they are positive or negative, as this is highly dependent on the stock selection and sample period chosen. We find the strategy that utilizes forecasts from the Realized HAR-GARCH model to yield higher returns across all three stocks than the Realized GARCH-based strategy.

Investors with a more conservative risk profile (and thus a lower level of accepted loss) will hedge more often than those with a higher level of accepted loss. Because JPM, IBM, and AAPL all experienced gains of over 45% during our sample period, repeatedly hedging against stocks trending upwards might explain why returns go down.

A point worth considering here is that of trading costs and the influence they are likely to have on realized returns. This is due to the predictions of Expected Shortfall (ES) determining the size and frequency of the hedging positions taken. We therefore expect this to impact our analysis and we leave thorough consideration of this for further research. This being said stocks listed on the Dow Jones 30 are some of the most liquid equities in the world and feature very low bid/ask spreads, so we anticipate the effects to be somewhat negligible.

We calculate an approximation of trading costs considering a portfolio size of \$1 across each of the three stocks above (JPM, IBM, AAPL). Assuming a trading fee of \$0.0035 per share, we find the total trading costs across the entire out-of-sample period to be \$0.60, \$0.87, and \$0.55 for the Realized GARCH and \$0.04, \$1.17, and \$0.23 for the Realized HAR-GARCH model. We find the HAR-GARCH leads to a reduction in total hedging costs across these selected stocks.

5. Conclusion

In this paper, we examined the out-of-sample forecasting performance of the benchmark Realized GARCH model against the Realized HAR-GARCH model. We found that the Realized GARCH model tends to predict higher volatility during periods of high and spiking volatility compared to the Realized HAR-GARCH model. For single-period forecasts, the Realized HAR-GARCH model outperforms Realized GARCH both in terms of Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) for the majority of the series studied. Both models display very similar results when evaluated through QL. As we extend the forecast horizon, the Realized HAR-GARCH model consistently outperforms Realized GARCH across all series and for all performance metrics. This performance gap widens with increasing horizon lengths, underscoring the Realized HAR-GARCH model's capacity to capture the long-memory characteristics of volatility through its incorporation of longer-term RV components.

Mixed results were found for the Value at Risk (VaR) and Expected Shortfall (ES) evaluation exercises. The Realized HAR-GARCH model exhibits higher VaR violation rates compared to those of the Realized GARCH model at the 1-day horizon for most series, except IBM, where the RHG model is notably accurate. For the 10-day ahead forecasts at the 5% level RG outperforms RHG in both the violation rate and Unconditional Coverage (UC) test; yet, the converse is true at the 1% level, with the Realized HAR-GARCH model passing the UC backtest for 3 tickers compared to just 1 for that of the Realized GARCH. Both models fail the Dynamic Quantile (DQ) test more often than not across the board - indicating clustering amongst violations. Both models generally fail to reject the null of correct specification for joint VaR and ES backtests.

Joint evaluation of VaR and ES forecasts through the FZ^0 Loss found the Realized GARCH model to be the more accurate of the two across all tickers and horizons. However, Diebold Mariano tests classified this outperformance as statistically significant in just 3 out of the 12 ticker/horizon combinations.

This study has extended the work of Huang et al. (2016) by application of the Realized HAR-GARCH model to multi-step forecasting of Value at Risk (VaR) and Expected Shortfall (ES). Whilst the model's superiority over traditional GARCH models and its effectiveness for one-day VaR forecasting are known, our study is the first to extend its use to multi-step VaR and ES forecasting. Additionally, we demonstrate how investors can reduce their expected shortfall through hedging strategies across a spectrum of risk profiles. The results also show that following a hedging strategy built on the Realized HAR-GARCH model can lead to a reduction in hedging costs. Using data up to 2024, our empirical analysis also provided insights into the performance of the model during the COVID-19 period. We leave an idea for future studies. The sparse research surrounding the Realized HAR-GARCH model has been applied predominantly to single equities, thus, it would be interesting to extend the VaR and ES evaluation to mutli-asset portfolios.

5. Bibliography

[1] Yamai, Y. & Yoshiba, T. Value-at-risk versus expected shortfall: A practical perspective. *Journal of Banking & Finance* **29**, 997–1015 (2005).

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

[3] Engle, R. F. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica* **50**, 987 (1982).

[4] Bollerslev, T. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**, 307–327 (1986).

[5] Nelson, D. B. Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* **59**, 347 (1991).

[6] Engle, R. F. & Ng, V. K. Measuring and testing the impact of news on volatility. *The Journal of Finance* **48**, 1749–1778 (1993).

[7] Glosten, L. R., Jagannathan, R. & Runkle, D. E. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance* **48**, 1779–1801 (1993).

[8] Baillie, R. T., Bollerslev, T. & Mikkelsen, H. O. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **74**, 3–30 (1996).

[9] Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. Modeling and forecasting realized volatility. *Econometrica* **71**, 579–625 (2003).

[10] Andersen, T. G. & Bollerslev, T. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review* **39**, 885 (1998).

[11] Barndorff-Nielsen, O. E. & Shephard, N. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society. Series B, Statistical Methodology* **64**, 253–280 (2002).

[12] Corsi, F. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics* **7**, 174–196 (2008).

[13] Andersen, T. G., Bollerslev, T. & Diebold, F. X. Roughing it up: including jump components in the measurement, modeling, and forecasting of return volatility. *The Review of Economics and Statistics* **89**, 701–720 (2007).

[14] Hua, J. & Manzan, S. Forecasting the return distribution using high-frequency volatility measures. *Journal of Banking & Finance* **37**, 4381–4403 (2013).

[15] Wilms, I., Rombouts, J. & Croux, C. Multivariate volatility forecasts for stock market indices. *International Journal of Forecasting* **37**, 484–499 (2021).

[16] Engle, R. Dynamic conditional correlation. *Journal of Business & Economic Statistics* **20**, 339–350 (2002).

[17] Engle, R. F. & Gallo, G. M. A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics* **131**, 3–27 (2006).

[18] Shephard, N. & Sheppard, K. Realising the future: forecasting with high-frequency-based volatility (heavy) models. *Journal of Applied Econometrics* **25**, 197–231 (2010).

[19] Brownlees, C. T. & Gallo, G. M. Comparison of volatility measures: a risk management perspective. *Journal of Financial Econometrics* **8**, 29–56 (2009).

[20] Christoffersen, P., Jacobs, K. & Mimouni, K. Volatility dynamics for the s&p500: Evidence from realized volatility, daily returns, and option prices. *The Review of Financial Studies* **23**, 3141–3189 (2010).

[21] Corsi, F., Fusari, N. & La Vecchia, D. Realizing smiles: Options pricing with realized volatility. *Journal of Financial Economics* **107**, 284–304 (2013).

[22] Huang, Z., Liu, H. & Wang, T. Modeling long memory volatility using realized measures of volatility: A realized har garch model. *Economic Modelling* **52**, 812–821 (2016).

[23] Hansen, P. R., Huang, Z. & Shek, H. H. Realized garch: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics* **27**, 877–906 (2011).

[24] Tian, S. & Hamori, S. Modeling interest rate volatility: A realized garch approach. *Journal of Banking & Finance* **61**, 158–171 (2015).

[25] Gerlach, R. & Wang, C. Forecasting risk via realized garch, incorporating the realized range. *Quantitative Finance* **16**, 501–511 (2016).

[26] Huang, Z., Wang, T. & Hansen, P. R. Option pricing with the realized garch model: An analytical approximation approach. *The Journal of Futures Markets* **37**, 328–358 (2016).

[27] Hansen, P. R., Lunde, A. & Voev, V. Realized beta garch: A multivariate garch model with realized measures of volatility. *Journal of Applied Econometrics* **29**, 774–799 (2014).

[28] Hansen, P. R. & Huang, Z. Exponential garch modeling with realized measures of volatility. *Journal of Business & Economic Statistics* **34**, 269–287 (2016).

[29] Chen, C. W. & Watanabe, T. Bayesian modeling and forecasting of value-at-risk via threshold realized volatility. *Applied Stochastic Models in Business and Industry* **35**, 747–765 (2018).

[30] Patton, A. J. Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics* **160**, 246–256 (2011).

[31] Barndorff-Nielsen, O. E. & Shephard, N. Estimating quadratic variation using realized variance. *Journal of Applied Econometrics* **17**, 457–477 (2002).

[32] Kupiec, P. H. Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives* **3**, 73–84 (1995).

[33] Engle, R. F. & Manganelli, S. Caviar. *Journal of Business & Economic Statistics* **22**, 367–381 (2004).

[34] Bayer, S. & Dimitriadis, T. Regression-based expected shortfall backtesting. *Journal of Financial Econometrics* **20**, 437–471 (2020).

[35] Fissler, T. & Ziegel, J. F. Higher order elicitability and osband's principle. *Annals of Statistics* **44**, 197–231 (2016).

[36] Creal, D., Koopman, S. J. & Lucas, A. Generalized autoregressive score models with applications. *Journal of Applied Econometrics* **28**, 777–795 (2012).

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

[38] Diebold, F. X. & Mariano, R. S. Comparing predictive accuracy. *Journal of Business & Economic Statistics* **13**, 253–263 (1995).

[39] Lyócsa, , Plíhal, T. & Výrost, T. Forecasting day-ahead expected shortfall on the eur/usd exchange rate: The (i)relevance of implied volatility. *International Journal of Forecasting* (2024).

A. Appendix

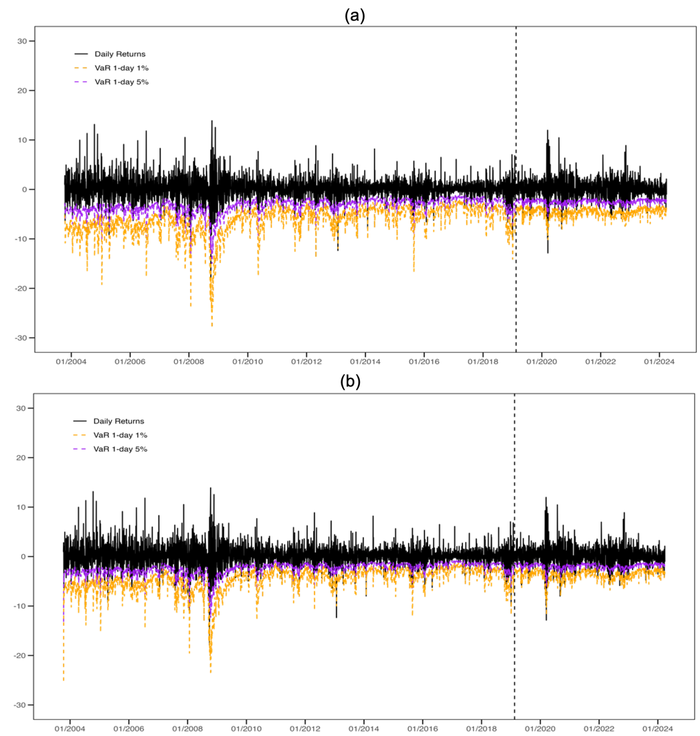


Figure 6: AAPL One-day ahead VaR forecasts for the Realized GARCH model (upper panel) and Realized HAR-GARCH model (lower panel).

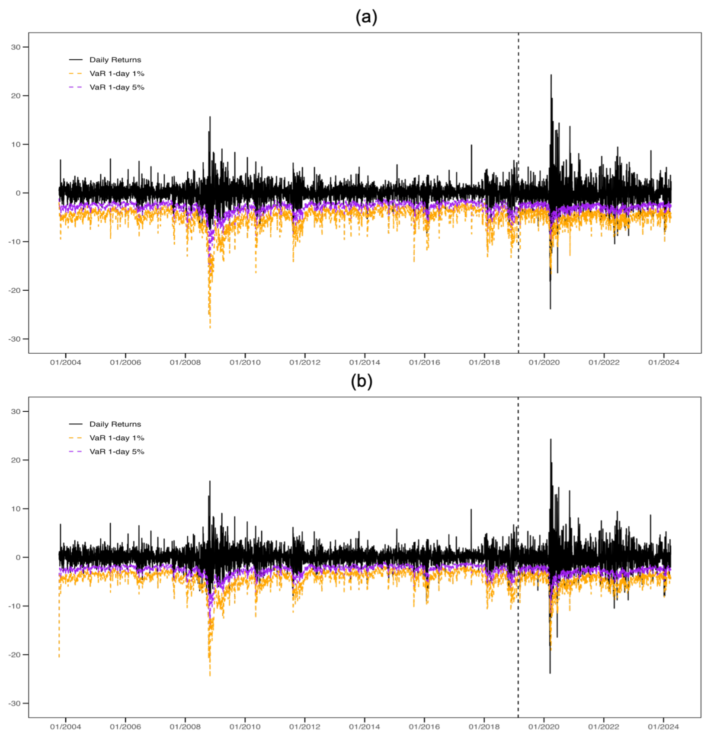


Figure 7: BA One-day ahead VaR forecasts for the Realized GARCH model (upper panel) and Realized HAR-GARCH model (lower panel).

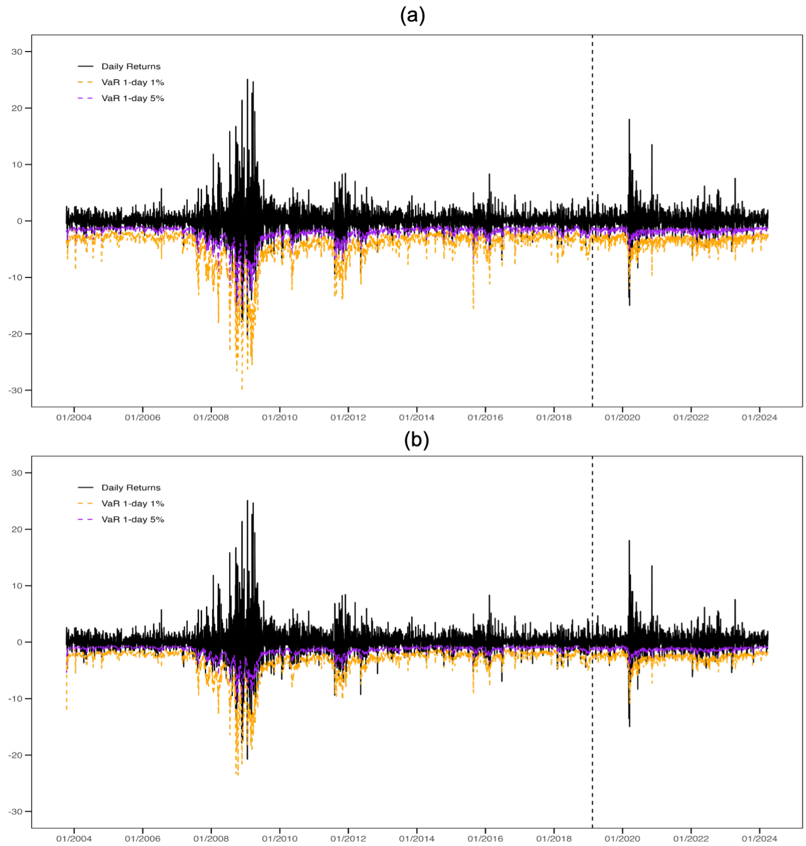


Figure 8: JPM One-day ahead VaR forecasts for the Realized GARCH model (upper panel) and Realized HAR-GARCH model (lower panel). The vertical dashed line indicates the beginning of the out-of-sample period.

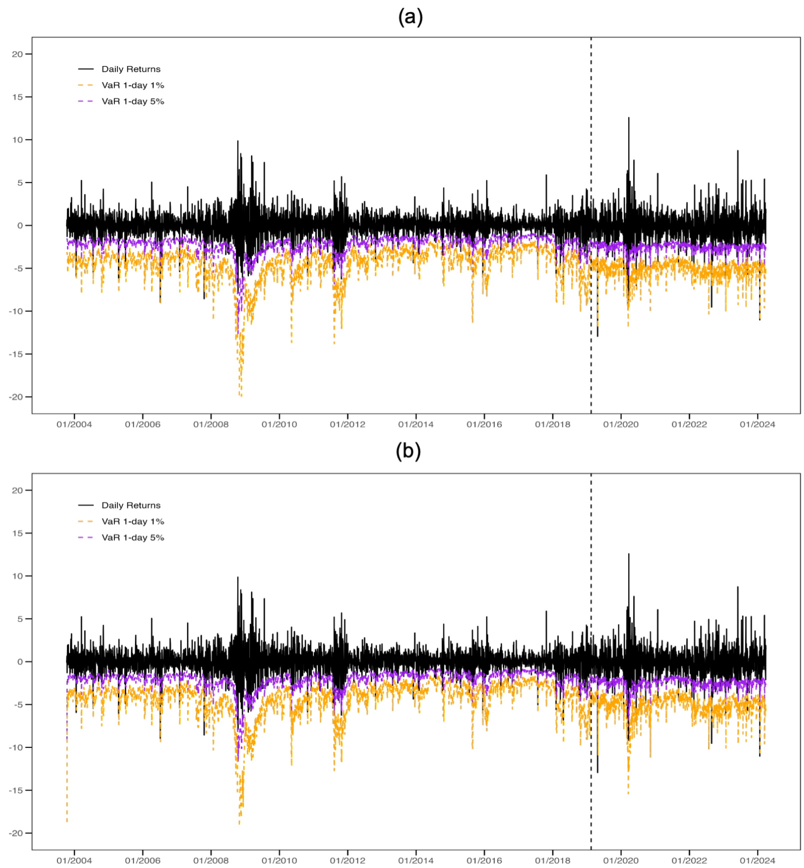


Figure 9: MMM One-day ahead VaR forecasts for the Realized GARCH model (upper panel) and Realized HAR-GARCH model (lower panel). The vertical dashed line indicates the beginning of the out-of-sample period.

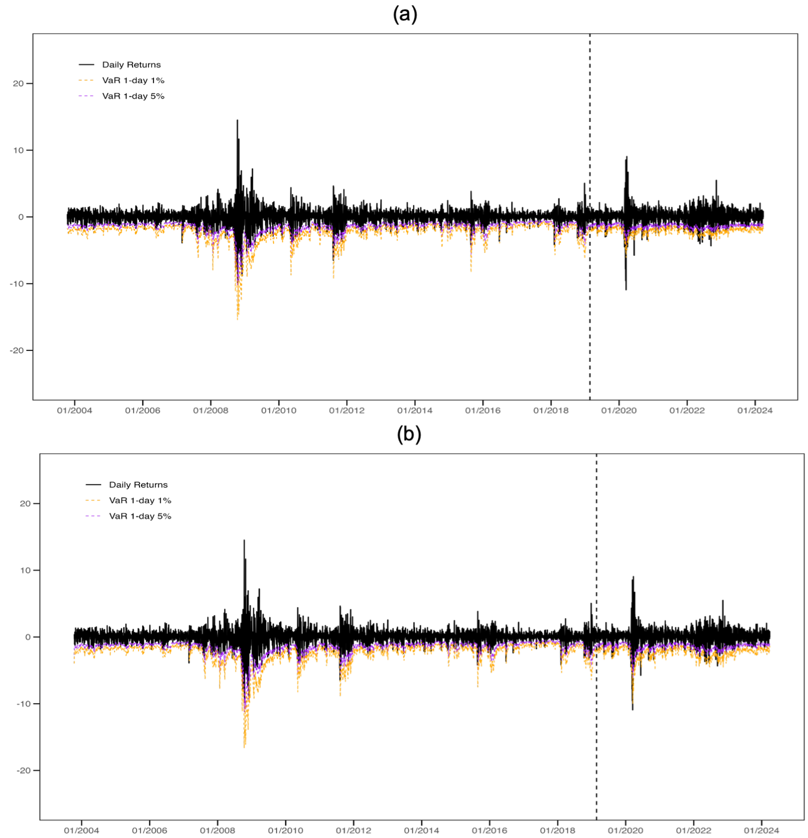


Figure 10: SPY One-day ahead VaR forecasts for the Realized GARCH model (upper panel) and Realized HAR-GARCH model (lower panel). The vertical dashed line indicates the beginning of the out-of-sample period.

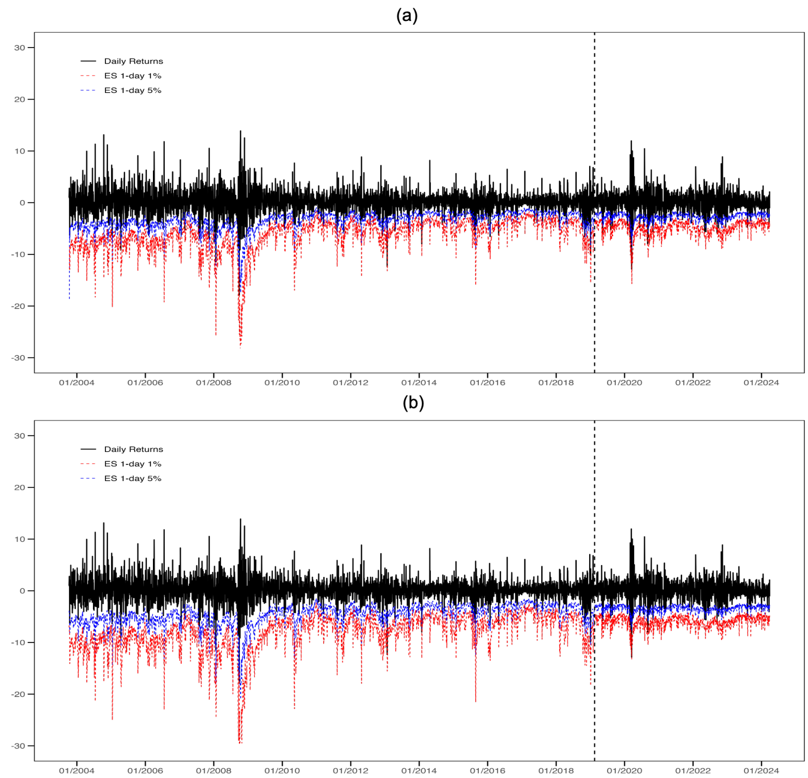


Figure 11: AAPL One-day ahead ES forecasts for the Realized GARCH model (upper panel) and Realized HAR-GARCH model (lower panel). The vertical dashed line indicates the beginning of the out-of-sample period.

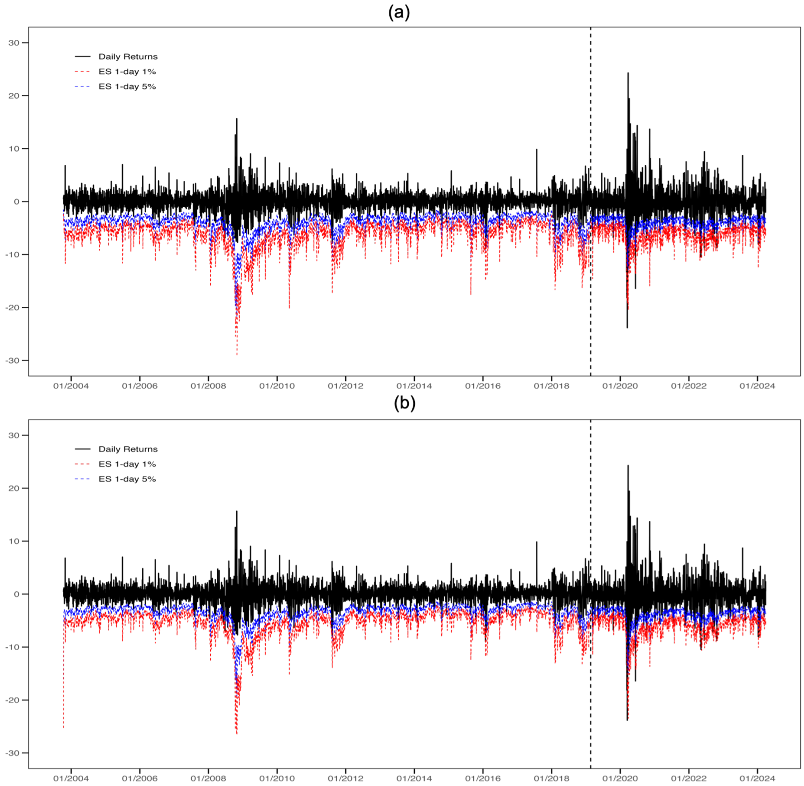


Figure 12: BA One-day ahead ES forecasts for the Realized GARCH model (upper panel) and Realized HAR-GARCH model (lower panel). The vertical dashed line indicates the beginning of the out-of-sample period.

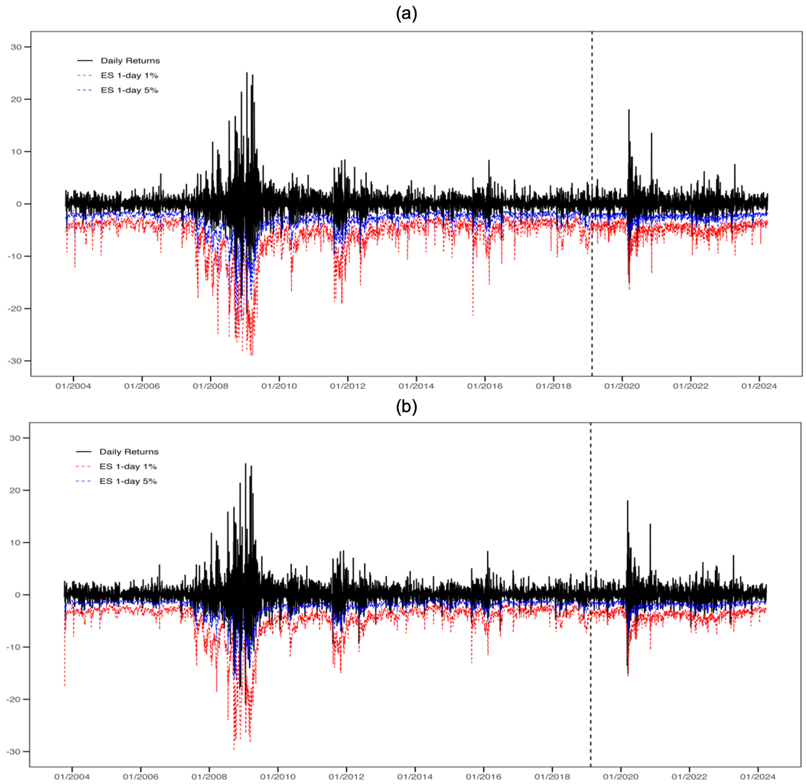


Figure 13: JPM One-day ahead ES forecasts for the Realized GARCH model (upper panel) and Realized HAR-GARCH model (lower panel). The vertical dashed line indicates the beginning of the out-of-sample period.

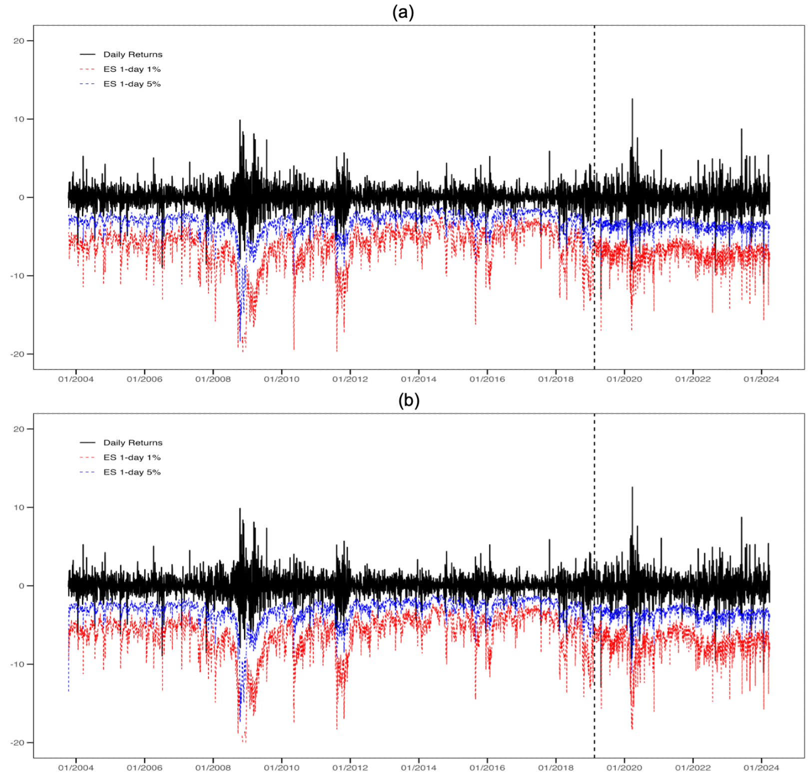


Figure 14: MMM One-day ahead ES forecasts for the Realized GARCH model (upper panel) and Realized HAR-GARCH model (lower panel). The vertical dashed line indicates the beginning of the out-of-sample period.

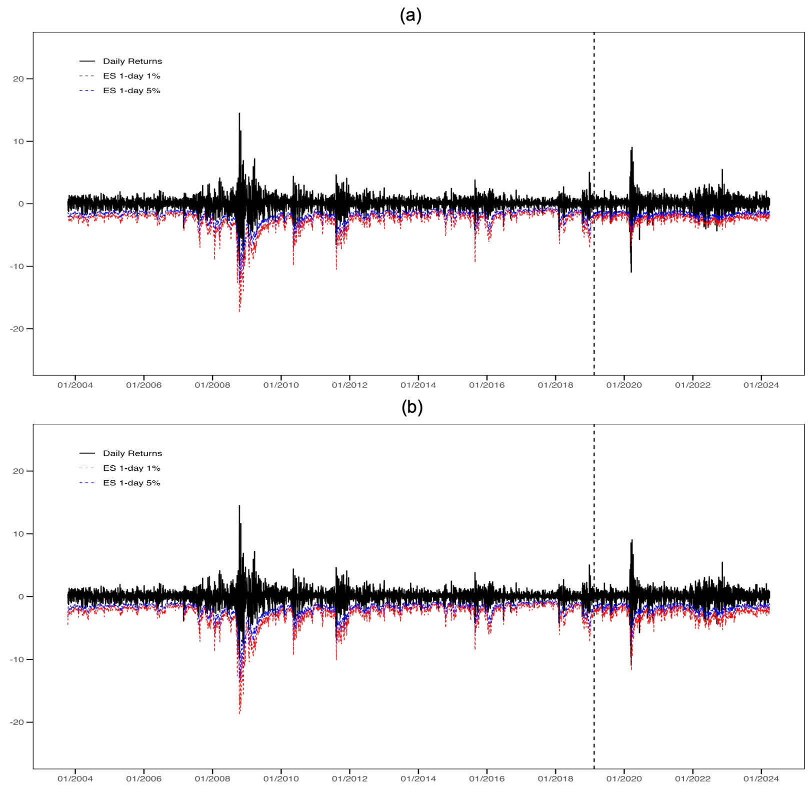


Figure 15: SPY One-day ahead ES forecasts for the Realized GARCH model (upper panel) and Realized HAR-GARCH model (lower panel). The vertical dashed line indicates the beginning of the out-of-sample period.