Forecasting Realized Covariance Matrices: New Methods to Improve Financial Decision Making

by

Zihao Chen

BFin

MA(AppEcmets)

Submitted in fulfilment of the requirement

for the degree of

Doctor of Philosophy

Faculty of Business and Law

Queensland University of Technology

2022

Abstract

The availability of high-frequency financial data recent has lead to substantial research on modeling and forecasting ex-post realized covariance measures constructed from intraday data. These realized covariance forecasts have been extensively used in empirical financial applications. This thesis consists of three studies that centre around forecasting the elements of the realized covariance matrix. The first study focuses on the diagonal element of the realized covariance, namely the realized variance (RV). The empirical analysis presented in this study illustrates that the state-of-the-art RV forecasting model accounts for both heteroskedastic measurement error and non-linear dependence in the conditional mean of the RV. More importantly, the study one shows the logarithmic transformation of RV is able to address these two issues simultaneously. The study two focuses on the realized correlation matrix, which is decomposed from the realized covariance matrix. We find that the dependence between the elements of the correlation matrix significantly impacts the accuracy of covariance matrix forecasts. We then propose a dynamic specification to incorporate this potential relationship and illustrate that the new model outperforms a series of benchmark models in an out-of-sample study. Study three presents a comprehensive analysis of forecasting realized portfolio weight and validates the previous findings within the context of portfolio selection. The empirical result shows, compared to the conventional approach of first forecasting the covariance matrix, direct forecasting the ex-post realized portfolio weight can generate more precise forecasts and effectively improve the economic gains.

Keywords

- Realized variance
- Realized covariance matrix
- Realized GMVP weights
- Forecasting
- HAR model
- Multivariate HAR model
- LASSO

Acknowledgments

There are many people that I need to acknowledge for their assistance and support with my research journey.

First and foremost, I am extremely grateful to my supervisors, Prof Adam Clements and Dr Stephen Thiele, for their invaluable advice, continuous support, and patience during my PhD studies. Their immense knowledge and plentiful experience have encouraged me all the time in my academic research and daily life. Without the guidance of Adam and Steve, this PhD dissertation would not have been achievable. I still remember Adam going through the MATLAB code with me patiently and explaining everything clearly in his office. I also remember Steve demonstrating to me the idea of LASSO in an early afternoon back in 2019. They are more like friendly friends who can offer you help anytime.

I would also like to express my appreciation to Prof Stan Hurn, who is the panel member of my confirmation and final seminar. Stan has provided me with heaps of invaluable advice on my research and career. He is also the author of my first econometrics textbook: Econometric Modelling with Time Series. Stan might be someone who encouraged me to study econometrics, even though he did not know it in 2016.

I gratefully acknowledge the financial support from the QUT Postgraduate Research Award (QUTPRA) from Queensland University of Technology. I would like to thank Karyn Gonano, who helped me with my thesis editing.

To all my friends at QUT, especially Dr. Lina Xu, Dr. Yang Liu, Justin Case, thank you for your constant encouragement and support. During my study, I have been a sessional teaching

staff in the School of Economics and Finance at QUT. I would like to thank all the staffs who have helped me with my teaching activities.

I also owe a great deal of thanks to my father and mother. I would not be able to pursue my studies in Australia without their unbelievable support and unconditional love. Finally, I would like to express my gratefulness to Edgar Barreto, who has been a pillar of strength through all the highs and lows.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

Over the last three decades, the concept of volatility permeates almost every facet of finance, and volatility forecasts have been used extensively in multiple financial applications such as option pricing, risk management, and portfolio allocation. In the early stages, most of the volatility forecasts were based on the daily or coarser frequency data. Nowadays, with the increased availability of high-frequency data for many financial assets, researchers have focused on investigating the rich information inherent in the high-frequency data and generating more reliable volatility forecasts. Andersen and Bollerslev (1998a) were the first to use the realized volatility computed from high-frequency data as an accurate proxy of the latent volatility. Realized volatility is a non-parametric measure of volatility based on high-frequency data, and is often used interchangeably with *realized variance* in a univariate setting and *realized covariance matrix* in a multivariate setting. Due to the non-latent feature, realized volatility can be observed and modeled directly via the time-series models.

Financial markets are related, now more than ever, due to the acceleration of economic globalization. Price fluctuations in one market may cause or be caused by the movement in other markets, making it legitimate to consider the dependence or correlation across multiple financial assets. Therefore, this research aimed to establish a new perspective to forecast the

realized covariance and improve financial decision-making.

The thesis consists of three studies that focused on forecasting the elements of the realized covariance matrix. Study one focuses on the diagonal element of the realized covariance, namely the realized variance (RV). It has been documented that RV has a heteroskedastic measurement error and the dynamic modeling of RV inherently faces an errors-in-variables problem. Based on the asymptotic theory of RV (Barndorff-Nielsen and Shephard, 2002) and the heterogeneous autoregressive (HAR) model (Corsi, 2009), the HARQ model (Bollerslev et al., 2016) was developed to address the issue of heteroskedastic measurement error. The HARQ model represents state-of-the-art in RV forecasting models due to its superior prediction performance. However, the empirical analysis presented in this study illustrates the success of the HARQ model can be attributed to capturing both heteroskedastic measurement error and non-linear dependence in the conditional mean of RV. In addition to this, study one shows the logarithmic transformation of RV could address those two issues simultaneously.

The second study focuses on the realized correlation matrix decomposed from the realized covariance matrix. The parametrization method of Archakov and Hansen (2021) was used to obtain a reliable realized correlation matrix forecast, and guarantee the positive definiteness of correlation matrix forecasts. The Least Absolute Shrinkage and Selection Operator (LASSO) technique of Tibshirani (1996) was applied to circumvent the curse of dimensionality. The empirical study in study two shows that the dependence between the elements of correlation matrix, often neglected during the modeling, significantly impacts the realized covariance matrix forecasts. We then propose a new dynamic specification to incorporate this potential relationship and illustrate that the new model outperforms a series of benchmark models in an out-of-sample study.

The precise realized covariance matrix estimates and forecasts can be used to compute the weight of realized Global Minimum Variance Portfolio (GMVP). The GMVP is the origin of the Markowitz mean-variance efficient frontier, which attracted significant attention from both academics and practitioners. Study three is a comprehensive analysis of forecasting realized GMVP weight and validates the previous findings of study two within the context of portfolio

selection. We show that, compared to the conventional approach of first forecasting the realized covariance matrix, direct modeling and forecasting the ex-post realized GMVP weight can generate more precise forecasts and effectively improve the economic gains. In particular, the empirical result indicates incorporating the dependence between the weight series may potentially enhance the economic performance of the GMVP.

The remainder of this chapter completes the introduction to the thesis. Section 1.2 lists the key research questions that will be examined in this thesis. Section 1.3 outlines the thesis structure, including the literature review, three research chapters and conclusion. Finally, Section 1.4 highlights the main contributions of this thesis.

1.2 Key Research Questions

Several research questions related to realized covariance forecasting are proposed in order to address the aims of the thesis.

Does the success of the HARQ model also attribute to addressing the non-linear dependence in the conditional mean of the RV?

Although RV is consistent for the true latent volatility under certain conditions, it is subject to the heteroskedastic measurement error in any given finite sample. Building on the asymptotic theory of RV, Bollerslev et al. (2016) augmented the standard HAR model with a term as a function of realized quarticity (RQ) and developed the HARQ model to address the heteroskedastic measurement error directly. Given that the square root of RQ is strongly correlated with RV, we replaced the square root of RQ in the HARQ model with RV, yielding a new dynamic specification with a quadratic term of RV. Surprisingly, this new model exhibits very similar forecasting performance to the HARQ model. The forecasting gains may be attributed to capturing the non-linear dependence in the conditional mean of RV. However, there is no empirical evidence to show if the HARQ model also accounts for the non-linear dependence

in the conditional mean of the RV. We applied the logarithmic function on RV to address this question since the log-transformation is suitable for data series whose variance exhibits heteroskedasticity. Suppose the heteroskedastic measurement error is the only issue that needs to be solved, the HAR model with log-transformed RV will be outperformed by the HARQ model due to the non-linear transformation imposed on RV.

Are the highly parsimonious multivariate volatility models too restrictive to optimize the forecast of the realized covariance matrix?

Due to the curse of dimensionality, the multivariate HAR model estimated via Ordinary Least Squares (OLS) fails to deliver reliable correlation matrix forecasts. The most common solution is to employ highly parsimonious models. Oh and Patton (2016) was the first modeled the realized correlation matrix using a scalar multivariate HAR model, which reduces the number of parameters and assumes the same parameter structure for all elements in the correlation matrix. Although the scalar multivariate HAR model avoids the curse of dimensionality, it neglects the potential dependence between the correlation matrix elements. However, evidence on whether these parsimonious models reduce the forecasting accuracy of the realized covariance matrix is limited. To address the question, we first applied the LASSO to reveal the optimal parameter structure of the multivariate HAR model and then assessed the predictive accuracy of the model that incorporates the dependence between the elements of the correlation matrix.

Will directly forecasting the realized GMVP weight be more beneficial than the conventional approach of first forecasting the realized covariance matrix?

The realized covariance matrix can be used to find the weight of realized Global Minimum Variance Portfolio (GMVP). The conventional approach to forecast the realized GMVP weight is based on the realized covariance matrix forecasts obtained from multivariate volatility models. The alternative approach is to model and forecast the ex-post realized GMVP weight directly with those multivariate volatility models. While forecasting GMVP weight has attracted

attention from both academics and practitioners, there is no clear comparison of which approach is more advantageous to provide better forecasts. In order to address this question, an out-of-sample analysis of multiple multivariate volatility models was therefore conducted using both conventional and direct forecasting approaches. The forecasting performance was evaluated based on the statistical and economic evaluation criteria derived from the Markowitz portfolio optimization theory.

1.3 Thesis Structure

The thesis is structured as follows. The literature review in Chapter 2 details the theoretical foundation of RV and realized covariance. In particular, this chapter reviews a range of models that can be used to generate the forecasts of realized covariance. The measures and methodologies to evaluate forecast accuracy are also introduced in this chapter.

Before turning to the full realized covariance matrix, Chapter 3 examines the realized variance, which is the diagonal element of the realized covariance. The empirical analysis presented in this chapter illustrates that the state-of-the-art HARQ model accounts for both heteroskedastic measurement error and non-linear dependence in the conditional mean of the RV. and that the logarithmic transformation of RV is able to address these two issues simultaneously.

Chapter 4 analyses the realized correlation matrix, which is decomposed from the realized covariance matrix. This chapter aims to improve the accuracy of the realized covariance forecast from a correlation matrix modeling perspective. An empirical study is presented in this chapter to demonstrate the importance of the dependence between the correlation matrix elements. Moreover, a dynamic specification is proposed to incorporate this potential relationship, and the new model is demonstrated to outperform a series of benchmark models in an out-of-sample study.

The realized covariance matrix estimates or forecasts can be used to find the weight of realized Global Minimum Variance Portfolio (GMVP). Chapter 5 therefore provides a comparison study between the conventional approach and the direct forecasting approach of forecasting

realized GMVP weight. In addition, the study in this chapter validated the findings from Chapter 4 within the context of portfolio selection.

Finally, the Chapter 6 concludes by summarising this thesis's findings and a number of avenues for future research.

1.4 Key Contributions

This thesis addresses the aforementioned research questions and makes three key contributions to the literature of realized covariance forecasting.

1. The empirical analysis presented in this thesis illustrates that the heteroskedastic measurement error and non-linear dependence in the conditional mean of RV must be considered. Therefore, the success of the HARQ model can be attributed to accounting for both heteroskedastic measurement error and non-linear dependence in the conditional mean of the RV. Since the estimation for realized quarticity is quite imprecise in many realistic scenarios, the HAR model with log-transformed RV can be a viable alternative to the HARQ model. The empirical analysis suggests the logarithmic transformation of RV is able to address both heteroskedastic measurement error and non-linearity simultaneously.

2. The empirical results suggest a rich parameter structure in the MHAR model is needed to forecast the correlation matrix. Meanwhile, the empirical result indicates that the potential relationship between the elements of the correlation matrix significantly impacts the accuracy of covariance matrix forecasts. The proposed model that incorporates this relationship outperforms a series of benchmark models in an out-of-sample study. Moreover, the empirical evidence shows that the correlation matrix parametrization method of Archakov and Hansen (2021) outperforms other re-parametrized methods for correlation matrix, such as Cholesky decomposition and the matrix logarithmic function.

3. The empirical analysis shows the direct forecasting realized GMVP weight outperforms the conventional approach in the sense of providing realized GMVP weight forecasts

that are significantly closer to the ex-post realized GMVP compositions. Compared to the conventional approach, the direct forecasting approach generally provides superior portfolio allocation performance, reduces portfolio turnover substantially, and effectively improves economic gains. The empirical result validates that incorporating the correlation between the portfolio weight series enhances the economic performance of the portfolio.

Chapter 2

Literature Review: Forecasting Realized Volatility

2.1 Overview

Andersen et al. (2006) defines volatility as *the variability of the random (unforeseen) component of a time series*. As a measure of variability, the term volatility is often used interchangeably with variance in the univariate setting and covariance in the multivariate setting. Over the last three decades, the Financial Econometrics literature has focused on modeling and forecasting volatility. Meanwhile, many financial applications have extensively used volatility forecasts as a risk measure, including asset allocation, risk management, and financial derivatives pricing. However, as noted throughout the literature, volatility of returns is latent, a volatility proxy is required, and volatility forecasts need to be generated to inform these applications. In the early stages, most of the volatility forecasts were generated from the (Generalized) Autoregressive Conditional Heteroskedasticity (GARCH) models (Bollerslev, 1986, Engle, 1982), and the stochastic volatility (SV) models. Those models were based on the volatility proxies constructed from daily, or coarser frequency data. Nowadays, with the increased availability of high-frequency data for many financial assets, researchers have moved their attention to exploiting the inherent information on such kind of data. The theoretical results indicate that high-frequency data should be helpful to construct more accurate volatility models (Merton, 1980, Nelson, 1992). A large number of research has shown the use of realized volatility measure

may effectively harness the information that inherent in the high-frequency data, and generate reliable volatility forecasts. Realized volatility is a non-parametric measure of volatility based on high-frequency data. To avoid confusion, throughout the thesis, *realized variance* refers only to the univariate realized volatility, and *realized covariance matrix* refers only to the multivariate realized volatility.

This chapter aims to provide a review of the theoretical foundation of realized variance and realized covariance, as well as the methodologies used for forecasting. Although other extensive reviews have been conducted, see, e.g., Andersen et al. (2013) and Bucci et al. (2017), the review in this chapter is essential for the rest of the thesis, since the terminologies and methodologies used throughout the thesis will be established.

To illustrated the developments in realized volatility forecasting, Chapter 2 is structured as follows. Section 2.2 introduces the basic definition of realized variance and realized covariance. Section 2.3 highlights the characteristics of realized volatility and describes the econometric methodologies of modeling realized volatility. In particular, this section introduces the range of models that can generate realized volatility forecasts. Section 2.4 provides a series of alternative methods for realized volatility modeling while Section 2.5 lists the techniques used to evaluate the performance of realized volatility forecasts. Finally, 2.6 concludes.

2.2 Realized volatility

2.2.1 Univariate realized volatility: realized variance

To formally define the realized variance concepts, let $dp(t)$ be the instantaneous returns, or logarithmic price increments, evolve continuously through time according to the stochastic volatility diffusion

$$dp(t) = \mu(t)dt + \sigma(t)d\omega(t), \tag{2.1}$$

where $\mu(t)$ and $\sigma(t)$ denote the instantaneous drift and volatility processes, respectively, and $\omega(t)$ is a standard Brownian motion. For simplicity, the instantaneous drift term in equation

(2.1) will henceforth assumed to be zero.

Following Andersen and Bollerslev (1998b), Andersen et al. (2001) and Barndorff-Nielsen and Shephard (2002), the realized variance (RV) on trading day t is formally defined by

$$RV_t \equiv \sum_{i=1}^{M} r_{t,i}^2, \tag{2.2}$$

where $M = 1/\Delta$ refers to the number of intraday equally spaced returns over the trading day, $r_{t,i}$ refers to the ith intraday return, which is defined by

$$r_{t,i} = p_{t-1+i\Delta} - p_{t-1+(i-1)\Delta}, \quad i = 1, 2, \ldots M. \tag{2.3}$$

For example, for 5-minute returns in a market that opens for six-and-half hours per day, as in the case of the U.S. equity markets, would correspond to $M = 78$.

Assume Δ goes to zero and the finer intraday returns defined by equation (2.3) are continuously observable, then the realized variance estimator approaches the integrated variance of the underlying continuous time stochastic volatility process on the day t, formally defined by

$$IV_t = \int_{t-1}^{t} \sigma^2(\tau) d\tau. \tag{2.4}$$

Hence, in contrast to the volatility estimates based on the lower frequency return, the true ex-post volatility of the day can be observed directly.

However, in reality, the asset prices are not observed continuously, and data limitations invariably put an upper bound on the number of intraday returns. Although the high-frequency data are recorded for many assets, those prices are usually observed at discrete and irregular intervals. The high-frequency returns are also affected by various market microstructure frictions, noise, etc (Brownlees and Gallo, 2006). The high-frequency data cleaning is therefore essential before the estimation of integrated variance.

Andersen and Bollerslev (1998a) proved that realized variance will be a consistent estimator

of integrated variance if the number of M goes infinity. Barndorff-Nielsen and Shephard (2002) developed the asymptotic distribution theory of realized variance, as the Δ goes to zero,

$$RV_t = IV_t + \eta_t, \quad \eta_t \sim MN\left(0, 2\Delta IQ_t\right), \tag{2.5}$$

where

$$IQ_t = \int_{t-1}^{t} \sigma^4(\tau) d\tau \tag{2.6}$$

denotes the univariate integrated quarticity (IQ), and MN stands for mixed normal distribution. In parallel to the integrated variance, the integrated quarticity may be consistently estimated by the realized quarticity (RQ),

$$RQ_t \equiv \frac{M}{3} \sum_{i=1}^{M} r_{t,i}^4. \tag{2.7}$$

There are multiple popular extentions to the RV framework. For example, the financial prices often present "jumps" over some short time intervals (Andersen et al., 2007). In particular, taking into account the possibility of jumps in the underlying price process, the realized variance measures discussed in equation (2.2) no longer converge to the integrated variance. The total ex-post quadratic variance, therefore, is given by

$$QV_t = IV_t + JV_t, \tag{2.8}$$

where IV_t still be the integrated variance that accounts for the variation coming from the continuous price increments over the day, and

$$JV_t = \sum_{j=1}^{\mathcal{J}_t} J_{t,j}^2 \tag{2.9}$$

measures the variation due to the $\mathcal{J}_t$ jumps that occurred on day t. Several alternative volatility estimators that consider the impact of jumps have been proposed in the literature. The most

famous being the bi-power variation estimator of Barndorff-Nielsen and Shephard (2004b),

$$BPV_t = \frac{\pi}{2}\frac{M}{M-1}\sum_{i=1}^{M-1}|r_{t.i}|\,|r_{t,i+1}|\,, \tag{2.10}$$

which offers a consistent estimator of the integrated variance in the presence of jumps.

2.2.2 Multivariate realized volatility: realized covariance matrix

The discussion in section 2.2.1 has focused on the realized variance for univariate returns. Financial markets are more related than ever due to the acceleration of economic globalization. Price fluctuations in one market may cause or be caused by the movement in other markets. Financial applications such as portfolio selection usually require estimates of multivariate volatility. The widely available intraday data for many financial assets are also useful to estimate realized covariance, just as the RV scenario. This section introduces the multivariate realized volatility for the full N-dimensional asset returns.

Generalizing the univariate setting in equation (2.1), assuming a $N \times 1$ logarithmic price vector $P(t)$ is governed by a multivariate stochastic volatility diffusion process that evolves continuously through time,

$$dP(t) = M(t)dt + \Sigma(t)^{1/2}dW(t), \tag{2.11}$$

where $M(t)$ and $\Sigma(t)^{1/2}$ denote the $N \times 1$ instantaneous drift vector and the $N \times N$ positive definite "square-root" of the covariance matrix, respectively. $W(t)$ is a N-dimensional vector of independent Brownian motions. $M(t)$ is set to zero.

The daily realized covariance matrix is defined by simply extending the univariate realized variance measure in equation (2.2),

$$RCov_t \equiv \sum_{j=1}^{M} R_{t-1+j\Delta}\,R'_{t-1+j\Delta}, \tag{2.12}$$

where, as before, $M = 1/\Delta$. Theoretically, if the intraday frequency Δ approaches to zero and the price vector in equation (2.11) is able to be observed continuously, then it enables us to compute the realized covariance matrix in equation (2.12). Therefore, the realized covariance converges to the integrated covariance matrix of the continuous time stochastic volatility process on day t, defined by,

$$ICov_t = \int_{t-1}^{t} \Sigma(\tau) d\tau. \tag{2.13}$$

Again, this expression shows that the true ex-post multivariate volatility for the day t becomes directly observable, even in the absence of modeling $\Sigma(\tau)$. The variance and covariance entries in the covariance matrix no longer need to be extracted from a complicated model and estimated via maximum likelihood procedures, such as the multivariate GARCH models (Bollerslev et al., 1988). In practice of course, realized covariance needs to be computed given the high-frequency returns. It has been proved that the realized covariance matrix is a consistent estimator of the integrated covariance matrix if M goes to zero (Andersen et al., 2003, Barndorff-Nielsen and Shephard, 2004a).

The integration symbol in equation (2.13) indicates that a higher sampling frequency is preferred. However, as discussed in section 2.2.1, trading activities are not continuous, i.e. data recorded irregularly at the discrete intervals, which may present further challenges for the estimation of the integrated covariance matrix, especially if the number of assets is large and the trading intensities of some assets are relatively low. The condition above may not be restrictive for a set of extremely actively traded assets since they are frequently generating thousands of prices per day. However, the critical point is that all assets must have traded within each sampling interval, otherwise this will generally lead to a downward bias in the realized covariance estimates due to the presence of zero returns caused purely by the absence of trades. This feature of non-synchronous price observations is labeled as the Epps effect (Epps, 1979), which can be a major concern since many assets experience no-trading periodically. Therefore, some data cleaning techniques need to be applied to consistently estimate the integrated covariance matrix, though not all of them guarantee the estimated integrated covariance matrix is positive definite or even positive semi-definite.

Andersen et al. (2003) found that the simple realized covariance matrix in equation (2.12) is positive definite by construction, as long as the asset returns in R_t are linearly independent and trading activity is sufficiently high. However, it is critical to sample fairly sparsely to alleviate this Epps effect. Another requirement is that the number of intraday observations M exceeds the number of assets N. For instance, for 5-minute returns in the U.S. equity markets, the realized covariance matrix is singular if the number of assets N is over 78.

To circumvent these complications, a number of alternative integrated covariance estimators have been proposed in the literature (Bannouh et al., 2012, Engle et al., 2019, Fan et al., 2011, Hautsch et al., 2012, Ledoit and Wolf, 2003, Lunde et al., 2016). The empirical analysis throughout this thesis relies on the multivariate kernel estimator from Barndorff-Nielsen et al. (2011). Technically, the requirement of price synchronicity implies that the integrated covariance matrix can only be estimated consistently if the sampling scheme is adapted to the trading intensity of the least active asset at any given time. This idea is part of the multivariate realized kernel approach where the kernel consists of a suitably chosen weight function for the lead and lag returns in the computation of the covariance matrix. This ensures consistency in the presence of general classes of microstructure noise and guarantees the estimate of the realized covariance matrix is positive semi-definite. Given $R_{i,t}$ is the N-dimensional vector returns for the ith intraday time interval on day t, the multivariate kernel estimator is defined as

$$MK_t = \sum_{h=-M}^{M} k\left(\frac{h}{H}\right)\Gamma_h, \tag{2.14}$$

where M is the number of observed synchronized intraday returns for each of the N assets, $\Gamma_h = \sum_{i=h+1}^{M} R_{i,t}R'_{i-h,t}$, $\Gamma_h = \Gamma'_{-h}$, and an appropriate kernel function $k(.)$ with bandwidth H. In order to facilitate the notation, S_t represents the multivariate kernel estimator of the integrated covariance matrix and Σ_t will denote true ex-post integrated covariance matrix. Let $\operatorname{vech}\Sigma_t$ denote the $N^* = N(N+1)/2$ dimensional vectorized version of the integrated covariance matrix. Similarly, let $\operatorname{vech} S_t$ denote the vectorized multivariate kernel estimator. According to Barndorff-Nielsen et al. (2011), the asymptotic theory of the multivariate kernel estimator

follows that

$$M^{1/5}\left(\operatorname{vech} S_t - \operatorname{vech} \Sigma_t\right) \rightarrow {}_{Ls}MN\left(\omega_t, \Pi_t\right). \tag{2.15}$$

ω_t is the negligible bias term. Π_t refers to the $N^* \times N^*$ covariance matrix for the corresponding measurement error vector $(\operatorname{vech} S_t - \operatorname{vech} \Sigma_t)$. Π_t is proportional to the integrated quarticity matrix,

$$\Pi_t \propto \int_{t-1}^{t} \Sigma(u) \otimes \Sigma(u) du, \tag{2.16}$$

where $\otimes$ denotes the Kronecker product. The integrated quarticity matrix may be consistently estimated using high frequency intraday data. Following the studies of Barndorff-Nielsen and Shephard (2004a), Π_t is estimated based on the pre-averaged data. Let $\bar{R}_{i,t}$ be the pre-averaged data and define $x_{i,t} \equiv \operatorname{vech}\left(\bar{R}_{i,t}\bar{R}'_{i,t}\right)$, the consistent estimator for integrated quarticity matrix is

$$\widehat{IQ_t} = M \sum_{i=1}^{M-H} x_{i,t}x'_{i,t} - \frac{M}{2} \sum_{j=1}^{M-H-1} \left(x_{j,t}x'_{j+1,t} + x_{j+1,t}, x'_{j,t}\right). \tag{2.17}$$

By treating the realized covariance matrix as a proxy of the true underlying integrated covariance, the standard time series techniques may be employed for modeling and forecasting, addressed in the following sections.

2.3 Econometric models for forecasting volatility

Empirical studies highlight that forecasting models accounting for the features of the volatility process tend to produce the most accurate forecasts. The characteristics of realized volatility will be outlined in this section before presenting the developments of volatility models, which will enable us to understand why plenty of models have been developed.

First, the realized volatility is non-negative which indicates the realized variance of a single series is always non-negative and the realized covariance matrix is always positive (semi-)definite. A model is fundamentally flawed if it fails to maintain this non-negative feature. Second, empirical evidence shows that realized volatility clusters. Realized volatility may be

high for certain time periods and low for others. Third, realized volatility exhibits long memory, it reflects long-run dependencies between very distant observations. Fourth, realized volatility does not diverge to infinity. Statistically speaking, this means that realized volatility is often stationary. Fifth, realized volatility reacts differently to a big price increase or a big price drop, referred to as the leverage effect.

There is no doubt that the features of realized volatility have played an essential role in the development of volatility models.

2.3.1 Time series models for realized variance

Although realized variance is ultimately only an estimate of the underlying true integrated variance, it is potentially highly accurate and thus presents an opportunity to construct the ex-ante volatility forecasts using standard autoregressive moving average (ARMA) time series tools. Let RV_t denote any of the high frequency based realized variance measures, a simple first-order AR model for the daily volatility series is

$$RV_t = \beta_0 + \beta_1 RV_{t-1} + \varepsilon_t. \tag{2.18}$$

This model, together with other higher-order AR models for realized variance, can easily be estimated by standard OLS regression. Although the simple short-memory AR(1) model may be adequate for short-horizon risk forecasts, when looking at longer forecast horizons, more accurate forecasts may be obtained by using richer dynamic models that better capture the long-range dependence associated with slowly-decaying autocorrelations. Oomen (2001), Andersen et al. (2003) investigated the persistence of the log-transformed realized variance via the fractionally integrated ARMA (ARFIMA) model, which can be expressed as

$$\Phi(L)(1-L)^d \left(\log RV_t - \mu\right) = \Theta(L)\varepsilon_t, \tag{2.19}$$

L is the lag operator, $\Phi(L) = 1 - \Phi_1 L \ldots \Phi_p L^p$, $\Theta(L) = 1 + \Theta_1 L + \ldots + \Theta_q L^q$, and the fractional difference operator $(1-L)^d$ is defined as

$$(1-L)^d = \sum_{k=0}^{\infty} \frac{\Gamma(k-d)L^k}{\Gamma(-d)\Gamma(d+1)}, \tag{2.20}$$

where $\Gamma(.)$ is the gamma function. The d can be assumed as any values between 0 and 1. For example, a $I(0)$ model is defined when $d = 0$ in the equation (2.19), and the equation (2.19) defines a integrated model $I(1)$ when $d = 1$.

Long-memory models can be difficult to implement. Instead, a simpler approach may be pursued by directly exploiting longer-run realized volatility regressors. Corsi (2009) introduced a heterogeneous autoregressive (HAR) model, which is an alternative to ARFIMA models to capture the long-term memory, expressed as

$$RV_t = \beta_0 + \beta_1 RV_{t-1}^{(d)} + \beta_2 RV_{t-1}^{(w)} + \beta_3 RV_{t-1}^{(m)} + \varepsilon_t \tag{2.21}$$

where the $\beta_j (j = 0, 1, 2, 3)$ are unknown parameters that need to be estimated, RV_t is the realized variance of day t, and $RV_{t-1}^d = RV_{t-1}$, $RV_{t-1}^w = \frac{1}{5}\sum_{i=1}^{5} RV_{t-i}$, $RV_{t-i}^m = \frac{1}{22}\sum_{i=1}^{22} RV_{t-i}$ denotes the daily, weekly and monthly lagged realized variance, respectively. As shown in Corsi (2009), the HAR framework may be viewed as an approximate long-memory model and the long-memory pattern is reproduced by a sum of volatility components constructed over different time horizons. The HAR model can be simply estimated by OLS, providing out-of-sample forecasts proved as good as long-memory ARFIMA models. Also, by including coarser lagged realized volatility in equation (2.21), say quarterly, a closer approximation to exact long-memory dependence can be obtained. Moreover, the leverage effect can be easily incorporated into the HAR model by including additional volatility terms that interacted with dummies indicating the sign of lagged return (Corsi and Renò, 2012).

To include the effect of jumps, given the bi-power variation measure, the HAR-with-Jumps

(HAR-J) model is defined as,

$$RV_t = \beta_0 + \beta_1 RV_{t-1}^{(d)} + \beta_2 RV_{t-1}^{(w)} + \beta_3 RV_{t-1}^{(m)} + \beta_J J_{t-1} + \varepsilon_t, \tag{2.22}$$

where $J_t = \max[RV_t - BPV_t, 0]$. HAR-J model includes a measure of the jump variation as an additional explanatory variable in the standard HAR model. In practice, the jump component J_t has been found to be unpredictable and so motivate the development of the Continuous-HAR (C-HAR) model, which only includes bi-power variation on the right hand side,

$$RV_t = \beta_0 + \beta_1 BPV_{t-1}^{(d)} + \beta_2 BPV_{t-1}^{(w)} + \beta_3 BPV_{t-1}^{(m)} + \varepsilon_t. \tag{2.23}$$

It is also claimed that the Semivariance-HAR (SV-HAR) model outperforms HAR-J and C-HAR models (Patton and Sheppard, 2015). On the basis of semi-variation measures of Barndorff-Nielsen et al. (2008), the SV-HAR model decomposes the total variation in the standard HAR model into the variation due to negative and positive intraday returns, defined as

$$RV_t = \beta_0 + \beta_1^+ RV_{t-1}^{+(d)} + \beta_1^- RV_{t-1}^{-(d)} + \beta_2 RV_{t-1}^{(w)} + \beta_3 RV_{t-1}^{(m)} + \varepsilon_t, \tag{2.24}$$

where $RV_t^- \equiv \sum_{i=1}^{M} r_{t,i}^2 \mathbb{I}\{r_{t,i} < 0\}$ and $RV_t^+ \equiv \sum_{i=1}^{M} r_{t,i}^2 \mathbb{I}_{\{r_{t,i}>0\}}$. As with the standard HAR model, HAR-J, C-HAR and SV-HAR models can all be estimated and implement easily.

In addition to considering the effect of jumps, researchers also attempted to incorporate the impact of measurement error. Under certain conditions, the realized variance is consistent with the true integrated variance. However, it is always subject to measurement error in any given finite sample. As shown in equation (2.5), the realized variance will be equal to the sum of the true latent integrated variance and a measurement error. Therefore, modeling the estimator of integrated variance, i.e. realized variance, will lead to an attenuation bias, which means the realized variance process will be less persistent than the latent integrated variance. Building on the asymptotic theory of realized variance (Barndorff-Nielsen and Shephard, 2002), Bollerslev et al. (2016) proposed the HARQ model by augmenting the standard HAR model with a variable

as a function of realized quarticity. The full HARQ model can be written as

$$\begin{aligned} RV_t =& \beta_0 + (\beta_1 + \alpha_1\sqrt{RQ^d_{t-1}})RV^d_{t-1} + (\beta_2 + \alpha_2\sqrt{RQ^w_{t-1}})RV^w_{t-1} \\ &+ (\beta_3 + \alpha_3\sqrt{RQ^m_{t-1}})RV^m_{t-1} + \varepsilon_t. \end{aligned} \tag{2.25}$$

Parallel to the HAR model, RQ^d_{t-1}, RQ^w_{t-1}, RQ^m_{t-1} denotes the daily, weekly and monthly lagged realized quarticity, respectively. Bollerslev et al. (2016) claimed that the simplified version of HARQ model,

$$RV_t = \beta_0 + (\beta_1 + \alpha\sqrt{RQ^d_{t-1}})RV^d_{t-1} + \beta_2 RV^w_{t-1} + \beta_3 RV^m_{t-1} + \varepsilon_t, \tag{2.26}$$

is useful as most of the attenuation bias in forecasts is due to the estimation error in RV^d_{t-1}. An important consequence is that the daily autoregressive coefficient is now time-varying and related to the realized quarticity. Bollerslev et al. (2016) showed the HARQ model improves the accuracy of forecasts, to not only outperform the standard HAR model but also some of its extensions.

While realized variance modeling continues to be an active area of research, it is difficult to choose the dominant forecasting model. Risk managers typically care not only about the dynamics of univariate volatility but also more generally about the dynamics of the multivariate scenario. Instead of further discussing techniques for modeling the univariate realized variance, we now turn to dynamic models for forecasting the realized covariance matrix.

2.3.2 Time series models for the realized covariance matrix

Given the characteristics of the realized covariance matrix, it is challenging to extend the univariate models to the multivariate volatility setting. First, models must produce a positive definite or even positive semi-definite matrix forecast. Second, models may fail to deliver reliable estimates due to the curse of dimensionality and the large computational burden of estimation. Building on the univariate models discussed in section 2.3.1, this section outlines various strategies for modeling and forecasting realized covariance matrices. To set up the

notation, let Σ_t denote the $N \times N$ integrated covariance matrix of the $N \times 1$ vector of asset returns R_t, S_t be the $N \times N$ realized covariance matrix estimate with $N^* = N(N+1)/2$ distinct elements.

The first multivariate volatility model presented here is the VARFIMA model of Chiriac and Voev (2011), which can capture the persistent behavior of volatility. More importantly, it employs the Cholesky decomposition to guarantee a semi-positive definite forecast and provide a unique square-root of a positive definite realized covariance matrix,

$$S_t = \chi_t \chi_t'. \tag{2.27}$$

The Cholesky factor χ_t is a $N \times N$ unique lower triangular matrix. Let $X_t = \operatorname{vech} \chi_t$ be a N^*-dimensional vector, then the VARFIMA model is defined as

$$\Phi(L)D(L)\left[X_t - BZ_t\right] = \Theta(L)\varepsilon_t \quad \varepsilon_t \sim N\left(0, \Omega\right) \tag{2.28}$$

where Z_t is a vector of exogenous variables of dimension $K \times 1$, B is the coefficient matrix of dimension $N^* \times K$, $\Phi(L) = I_{N^*} - \Phi_1 L - \Phi_2 L^2 - \ldots - \Phi_p L^p$, $\Theta(L) = I_{N^*} - \Theta_1 L - \Theta_2 L^2 - \ldots - \Theta_q L^q$ are matrix lag polynomials where Φ_i, for $i = 1, \ldots, p$, and Θ_j, for $j = 1, \ldots, q$, are the AR- and MA- coefficient matrices. $D(L) = \operatorname{diag}\left\{(1-L)^{d_1}, \ldots, (1-L)^{d_{N^*}}\right\}$, where $d_1, \ldots, d_{N^*}$ are the degrees of fractional integration of each element of the vector X_t, Ω is the covariance matrix of ε_t. The roots of $\Phi(L)$ and $\Theta(L)$ were assumed to be outside the unit circle, and X_t is stationary when $d_i < 0.5$ (Chiriac and Voev, 2011, Sowell, 1992). The quasi-maximum likelihood was used to limit the number of estimated parameters, and guarantee the unique representation,

$$(1 - \Phi L)D(L)\left[X_t - c\right] = (1 - \Theta L)\varepsilon_t \quad \varepsilon_t \sim N(0, \Omega), \tag{2.29}$$

where c is a $N^* \times 1$ vector.

Building on the Cholesky decomposition and the standard univariate HAR model, Chiriac

and Voev (2011) proposed a scalar version of the multivariate HAR (MHAR) model, defined as,

$$X_t = \boldsymbol{\beta}_0 + \beta_1 X_{t-1}^{(d)} + \beta_2 X_{t-1}^{(w)} + \beta_3 X_{t-1}^{(bw)} + \beta_4 X_{t-1}^{(m)} + \varepsilon_t. \tag{2.30}$$

Except for the daily, weekly, and monthly variables, the biweekly (10 days) term is also included. $\boldsymbol{\beta}_0$ is a $N^* \times 1$ intercept vector and β_i, $i = 1, \ldots, 4$, are scalar parameters. The most distinctive feature of this scalar version of the MHAR model is highly parsimonious. The scalar MHAR model assumes the same structure for all elements in X_t. The parameters can be estimated simply via the OLS regression.

Čech and Baruník (2017) suggested modeling the Cholskey factors through a generalized MHAR (GHAR) model,

$$X_t = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \odot X_{t-1}^{d} + \boldsymbol{\beta}_2 \odot X_{t-1}^{w} + \boldsymbol{\beta}_3 \odot X_{t-1}^{m} + \varepsilon_t, \tag{2.31}$$

where $\boldsymbol{\beta}_i$, , $i = 0, \ldots, 3$ are $N^* \times 1$ vector of parameters and $\odot$ denotes the Hadamard product operator. The GAHR model assumes the the dynamics of elements in the realized covariance matrix are explained by their own lag variables only. The parameters are estimated with the two-step generalized least square method.

One drawback to using of Cholesky decomposition is that the estimates inevitably involve a bias, arising from forecasting a nonlinear transformation and then mapping the forecasts back into the covariance matrix. Chiriac and Voev (2011) provided approximate bias correction terms for this and claimed the extent of the bias to be relatively minor in their empirical application. Another strategy to ensure the (semi-)positive definiteness of the realized covariance matrix forecast is matrix logarithmic transformation (Bauer and Vorkink, 2011). Specifically, the matrix logarithmic transformation, when applied to the realized covariance matrix S_t,

$$A_t = \operatorname{logm}(S_t) \tag{2.32}$$

is a $N \times N$ logarithmic transformed matrix, where $\operatorname{logm}(.)$ is the matrix logarithmic transformation operator. The matrix logarithmic transformation is the inverse of the matrix exponential function. One may then proceed as before by specifying the dynamic models of $\operatorname{vech} A_t$, forecasting and re-constructing the realized covariance via the matrix exponential function,

$$\hat{S}_{t+1} = \exp\left(\hat{A}_{t+1}\right). \tag{2.33}$$

One of the innovations of Bauer and Vorkink (2011) was employing a series of exogenous variables as volatility determinants. They proposed a model that relies on a set of HAR regressors and macroeconomic/financial variables, which improved the forecasting accuracy. Let $a_t = \operatorname{vech} A_t^{BP}$ be the vectorized logarithmic transformed bi-power covariance matrix, the model is defined as

$$a_t = \boldsymbol{\gamma}_0 + \boldsymbol{\gamma}_1 a_{t-1}^d + \boldsymbol{\gamma}_2 a_{t-1}^w + \boldsymbol{\gamma}_3 a_{t-1}^m + \boldsymbol{\gamma} EV_{t-1} + \varepsilon_t, \tag{2.34}$$

where $\boldsymbol{\gamma}_0$ is a $N^* \times 1$ intercept vector and $\boldsymbol{\gamma}_i$, $i = 1, \ldots, 3$, are $N^* \times N^*$ parameters. EV_t represents the exogenous variables. However, the model involves estimating of a large number of parameters. In order to reduce the dimensions of predictors, Bauer and Vorkink (2011) suggested two methods. First, considering the a_t^d, a_t^w and a_t^m series is supposed to be dominated by a small number of factors, the hypothesis was then tested by estimating the principal components of a_t^d, a_t^w and a_t^m. For example, $a^d(i)_t$ is the ith principal component of a_t^d. The alternative method Bauer and Vorkink (2011) suggested was involving a latent factor approach. Given a series of explanatory variables

$$Z_t = \left[a^d(1)_t, \ldots, a^d(i)_t, a^w(1)_t, \ldots, a^w(i)_t, a^m(1)_t, \ldots, a^m(i)_t, EV_t\right], \tag{2.35}$$

which is related to the K unknown volatility factors, and the kth volatility factor, $v_{k,t}$, is denoted as

$$v_{k,t} = \theta_k Z_{t-1}. \tag{2.36}$$

$\theta_k = \left\{\theta_{k,(1)}, \ldots\ldots, \theta_{k,(P)}\right\}$ are the coefficients that combine those explanatory variables, then the volatility factor is a linear combination of the set of P variables Z_t. Each of the log-space volatilities a_t^i is a function of the K volatility factors, such that

$$a_t^i = \gamma_0^i + \beta^i \theta Z_{t-1} + \varepsilon_t^i \quad i = 1, \ldots, N^*, \tag{2.37}$$

where γ_0^i is the ith element of the intercept vector $\boldsymbol{\gamma}_0$, β^i is the $1 \times K$ vector of loadings of log-space volatility i on the K factors, and the $K \times P$ matrix θ contains the coefficients on the Z_{t-1} variables for the K factors. Aggregating for the N^* log-volatilities yields

$$\alpha_t = \boldsymbol{\gamma}_0 + \beta\theta Z_{t-1} + \varepsilon_t. \tag{2.38}$$

This latent factor approach can significantly reduce the number of parameters and combine the lagged volatility with exogenous explanatory variables.

Similar to the principal components method that reduces dimensionality, Luo and Chen (2020) applied the Bayesian random compressed approach (Koop et al., 2019) to compress the MHAR model predictors randomly, and therefore reduce the number of parameters. Callot et al. (2017) modeled and forecast large realized covariance matrix by penalized vector autoregressive (VAR) model with the LASSO technique. The LASSO was applied to each entry of the VAR model in which all variance-covariance elements were used as the predictors.

Just like the HARQ model that incorporates the effect of measurement error, Bollerslev et al. (2018) extended the HARQ model to a multivariate setting. Let $Y_t = \operatorname{vech} S_t$ denote the N^*-dimensional vector, based on the scalar version of the MHAR model and the asymptotic distribution by Barndorff-Nielsen et al. (2011), then the scalar version multivariate HARQ model takes the form

$$\begin{aligned} Y_t &= \boldsymbol{\beta}_0 + \boldsymbol{\beta}_{1,t} \odot Y_{t-1}^{(d)} + \beta_2 Y_{t-1}^{(w)} + \beta_3 Y_{t-1}^{(m)} + \epsilon_t, \\ \boldsymbol{\beta}_{1,t} &= \beta_1 \boldsymbol{\iota} + \beta_{1Q} \pi_{t-1}, \end{aligned} \tag{2.39}$$

where $\boldsymbol{\iota}$ is an N^*-dimensinal vector of ones, β_1 and β_{1Q} are scalar parameters, and $\pi_t \equiv$

$\sqrt{\text{diag}\,(\Pi_t)}$ is the vector of asymptotic standard deviations for each of the individual element in the S_t. The model in equation (2.39) can be estimated by OLS and easily guarantee the positive definite forecast[1].

All the methodologies discussed before treat the realized covariance matrix as a whole. However, decomposing the covariance matrix and modeling variances and a correlation matrix separately is another popular approach, enabling more general dynamic dependencies, while still ensuring positive definite covariance matrix forecasts. A conditional covariance matrix may always be decomposed into a conditional correlation matrix pre- and post-multiplied by the diagonal matrix of conditional standard deviations, such as

$$\Sigma_t = D_t C_t D_t, \tag{2.40}$$

where $D_t = \text{diag}(\sigma_{1,\text{t}}, \ldots, \sigma_{\text{N,t}})$ with $\sigma^2_{i,t} = [\Sigma_t]_{ii}$, $i = \{1, \ldots, N\}$, and $C_t = corr\,(R_t \mid \mathcal{F}_{t-1})$ is the conditional correlation matrix of R_t. Motivated by the decomposition, Engle (2002) first developed the Dynamic Conditional Correlation (DCC) GARCH model, which allows for dynamically varying conditional correlation matrix within a GARCH framework. Oh and Patton (2016) applied the matrix decomposition on the realized covariance matrix, where the individual variances in D_t are modeled via the univariate HAR model and the conditional correlation matrix is modeled using a scalar multivariate HAR model.

So far, the discussion above has focused on the modeling procedures based on the HAR framework. The next section will briefly introduce a series of alternative methods for realized volatility modeling.

[1]Due to the difficulties in estimating the integrated quarticity matrix, the multivariate HARQ model occasionally produces negative definite forecasts. Bollerslev et al. (2018) applied the "insanity filter" to replace the negative definite covariance forecast with the simple average realized covariance over the relevant estimation sample.

2.4 Other models

All the dynamic models introduced in section 2.3 are realized volatility-based procedures, given any realized volatility measure, those models can be implemented easily. Recently, researchers have combined the classical GARCH model and realized volatility to imply the latent volatility process, such as the Multiplicative Error Model (MEM) of Engle and Gallo (2006), the high-frequency-based volatility (HEAVY) model of Shephard and Sheppard (2010) and the Realized GARCH model of Hansen et al. (2012).

Let conditional variance, $h_t = \mathrm{Var}\left(r_{t-1} \mid I_{t-1}\right)$, be the variable of interest. The simplest way to combine GARCH and realized volatility is the GARCH-X model, which includes the realized variance as an additional explanatory variable in the GARCH equation, defined as

$$\begin{aligned} r_t &= \sqrt{h_t} z_t \\ h_t &= \omega + \alpha r_{t-1}^2 + \beta h_{t-1} + \gamma RV_{t-1}. \end{aligned} \tag{2.41}$$

The GARCH-X model typically results in a insignificant α coefficient. Visser (2011) showed the accuracy of the coefficient estimates in conventional daily GARCH models may be improved through the use of realized volatility measures in the estimation. The GARCH-X model offers one-day-ahead volatility forecasts directly but not for longer horizons forecasts. However, this can be accomplished by augmenting the GARCH-X model with any realized volatility based procedures discussed in the previous sections. This so-called Realized GARCH model is developed by Hansen et al. (2012). For example, under a univariate setting, the Realized GARCH model includes realized variance as a part of the h_t function,

$$\begin{aligned} r_t &= \sqrt{h_t} z_t \\ h_t &= \omega + \beta h_{t-1} + \gamma RV_{t-1} \\ RV_t &= \xi + \varphi h_t + \tau\left(z_t\right) + u_t, \end{aligned} \tag{2.42}$$

where u_t denotes a random error with the property that $u_t \sim$ i.i.d. $\left(0, \sigma_u^2\right)$, and $z_t \sim$ i.i.d. $(0, 1)$. The Realized GARCH model can be estimated through quasi-maximum likelihood and

extended to multiple assets (Hansen et al., 2014).

Alternatively, Shephard and Sheppard (2010) proposed the HEAVY model as an extension of the basic model in equation (2.41),

$$\begin{aligned} r_t &= \sqrt{h_t} z_t \\ h_t &= \omega + \beta h_{t-1} + \alpha RV_{t-1} \\ \mu_{RV,t} &= \omega_{RV} + \beta_{RV} \mu_{RV,t-1} + \alpha_{RV} RV_{t-1}, \end{aligned} \tag{2.43}$$

where $\mu_{RV,t} = E(RV_t | I_{t-1}^{HF})$ is the condition mean of realized variance. HEAVY models have the advantage of adapting to new information and market conditions much more quickly than the regular daily GARCH models, where the parameters are estimated via quasi-maximum likelihood. A multivariate extension of the HEAVY model is provided by Noureldin et al. (2012). Let $U_t = R_t R_t'$ be the realized daily return cross-product and Σ_t be a $N \times N$ realized covariance,

$$\begin{aligned} E\left[U_t \mid I_{t-1}^{HF}\right] &= H_t \\ E\left[\Sigma_t \mid I_{t-1}^{HF}\right] &= M_t, \end{aligned} \tag{2.44}$$

H_t denotes the covariance matrix of the daily return vector. The multivariate HEAVY model can be written as

$$\begin{aligned} H_t &= \bar{C}_H \bar{C}_H' + \bar{B}_H H_{t-1} \bar{B}_H' + \bar{A}_H \Sigma_{t-1} \bar{A}_H' \\ M_t &= \bar{C}_M \bar{C}_M' + \bar{B}_M M_{t-1} \bar{B}_M' + \bar{A}_M \Sigma_{t-1} \bar{A}_M', \end{aligned} \tag{2.45}$$

where $\bar{B}_H$, $\bar{A}_H$, $\bar{B}_M$ and $\bar{A}_M$ are $N \times N$ parameter matrices, while $\bar{C}_H$ and $\bar{C}_M$ are lower triangular matrices with N^* parameters. The model is designed for the daily realized return cross-product but the information set is given by corresponding high frequency observations. In the general form, the model inherits the curse of dimensionality from multivariate GARCH representations, so the empirical work focuses on parsimonious representations by imposing $\bar{B}_H$, $\bar{A}_H$, $\bar{B}_M$ and $\bar{A}_M$ to be scalars or diagonal matrices. The scalar HEAVY model may be estimated with standard maximum likelihood techniques. Noureldin et al. (2012) show that including the high frequency data can provide significant improvement over GARCH models utilizing only daily return observations. Also, the inclusion of high frequency data appear to

provide a similar boost to the predictive performance that was observed in the univariate case.

Another possible route to model the realized covariance is the Wishart Autoregressive (WAR)-HAR model proposed by Chiriac and Voev (2011). The WAR model was introduced based on the Wishart distribution of the covariance matrix (Gouriéroux et al., 2009). Let $X_{k,t}$, with $k = 1, \ldots, K$, be a vector of N independent Gaussian VAR(1) processes,

$$X_{k,t} = MX_{k,t-1} + \varepsilon_{k,t} \quad \varepsilon_{k,t} \sim N(0, \Omega). \tag{2.46}$$

The Wishart Autoregressive process of order 1 is defined as

$$Y_t = \sum_{k=1}^{K} X_{k,t} X'_{k,t} \quad Y_t \sim W_n(K, M, \Omega), \tag{2.47}$$

where K denotes the degrees of freedom. Substituting equation (2.46) into equation (2.47), Y_t can be expressed as

$$Y_t = MY_{t-1}M' + K\Omega + \eta_t, \tag{2.48}$$

where η_t is the heteroskedastic error term with zero mean and M is a $N \times N$ matrix. It is necessary that $K \geq N$ to ensure Y_t is positive definite and has a Wishart distribution, even if this does not happen often in practice (Chiriac, 2007). The WAR model was extended further by Bonato et al. (2008), who suggested combining the WAR model with HAR dynamics on the latent processes $X_{k,t}$. Following their ideas, Chiriac and Voev (2011) developed the WAR-HAR specification:

$$Y_t = M^{(d)} Y_{t-1}^{(d)} M^{(d)'} + M^{(w)} Y_{t-1}^{(w)} M^{(w)'} + M^{(m)} Y_{t-1}^{(m)} M^{(m)'} + K\Omega + \eta_{t-1}, \tag{2.49}$$

where Y_t is the realized covariance matrix and M is diagnoal $N \times N$ parameter matrix. The drawback of this model is the restrictive parametric assumptions, and the Bayesian Markov Chain Monte Carlo (MCMC) estimation technique is computationally very expensive.

The literature on modeling and forecasting the realized covariance matrix is growing rapidly

in many different directions. However, there is still no consensus on the relative merits of the approaches. In order to select the optimal volatility forecasting, it is inevitable to involve the volatility forecast evaluations, which will be reviewed in the next section.

2.5 Evaluating volatility forecasts

The optimal volatility forecasting model for a specific application is often selected by evaluating a large set of competing models. The measures and tests that evaluate the forecast performance can be broadly divided into two classes: methods relied on a statistical evaluation of the forecasts to obtain the ranking of compared models, and the methods relied on the economic value of forecasts, such as Value-at-Risk and portfolio variance. This section therefore provides a brief review of forecast evaluation methods.

2.5.1 Statistical evaluation

Mincer and Zarnowitz (1969) introduced a regression that evaluates the bias of point forecasts. Let $E\left[r_t^2 \mid I_{t-1}\right] = \sigma_t^2$ be the true conditional variance of the return r_t and h_t be a forecast from any model, the Mincer-Zarnowitz regression is defined as

$$\hat{\sigma}_t^2 = \alpha + \beta h_t + \varepsilon_t, \tag{2.50}$$

where $\hat{\sigma}_t^2$ is the proxy of true conditional variance, α is a constant value and β is the coefficient of the forecast at time t. The test is based on estimating the coefficients of linear regression. The null hypothesis is

$$H_0 : \alpha = 0 \text{ and } \beta = 1 \tag{2.51}$$

When the forecast is unbiased. The Mincer-Zarnowitz regression is applicable when an unbiased proxy is used. However, such an approach does not always lead to the same outcome if the true latent variables were used (Andersen et al., 2005, Hansen and Lunde, 2006). It can be noticed that the Mincer-Zarnowitz regression strongly depends on the accuracy of the

volatility proxy, which affects the coefficients estimates. For example, Patton and Sheppard (2009) suggested using realized volatility as an accuracy proxy and estimate via generalized least squares (GLS) rather than OLS. The standard Mincer-Zarnowitz regression might select the inaccurate forecasting models, which should be applied in conjunction with other measures and tests to evaluate the forecast accuracy.

When the forecasts from multiple competing models are available, the evaluation of the volatility forecasts is based on their ranking, which is determined by the loss functions, which evaluate the forecast performance by measuring the distance of the forecast to the actual value. For univariate volatility forecasts, let $E(r_t \mid I_{t-1}) = 0$, $E\left[r_t^2 \mid I_{t-1}\right] = \sigma_t^2$ be the true latent conditional variance of the return r_t, and h_t be a forecast from a forecasting model, the loss function can be specified as

$$L\left(\sigma_t^2, h_t\right), \tag{2.52}$$

where L represents a particular functional form. In the multivariate framework, the loss function is given by

$$L\left(\Sigma_t, H_t\right), \tag{2.53}$$

where $\Sigma_t = E\left[R_t R_t' \mid I_{t-1}\right]$ is the conditional covariance matrix. To discriminate between the competing forecasts, those that produce the smallest expected loss are deemed to be the best.

The loss function should rely on a volatility proxy, defined as $\hat{\sigma}_t^2$ in the univariate setting and $\hat{\Sigma}_t$ in the multivariate setting, since the true latent volatility is always unobservable. However, the research by Hansen and Lunde (2006) and Patton (2011) both showed that not all loss functions could rank competing volatility forecasts correctly after employing the volatility proxies. The loss function should be "robust", which means the ranking between the competing models based on the volatility proxy is the same as that obtained in the presence of the real conditional variance (Patton, 2011), such as

$$E\left[L\left(\sigma_t^2, h_{k,t}\right)\right] \leq E\left[L\left(\sigma_t^2, h_{j,t}\right)\right] \Leftrightarrow E\left[L\left(\hat{\sigma}_t^2, h_{k,t}\right)\right] \leq E\left[L\left(\hat{\sigma}_t^2, h_{j,t}\right)\right]. \tag{2.54}$$

Patton (2011) evaluated the robustness of nine loss functions, which include four mean-squared-error (MSE) based loss functions, four mean-absolute-error (MAE) based loss functions, and one likelihood-based measure (QLIKE). It is concluded that only the MSE and QLIKE provide robust ranking when the volatility proxy is noisy (Patton, 2011), as so is defined as

$$\text{MSE} : L\left(\hat{\sigma}_t^2, h_t\right) = \left(\hat{\sigma}_t^2 - h_t\right)^2 \tag{2.55}$$

$$\text{QLIKE} : L\left(\hat{\sigma}_t^2, h_t\right) = \log h_t + \frac{\hat{\sigma}_t^2}{h_t} \tag{2.56}$$

Although not all loss functions are robust, Patton (2011) showed the ability of these non-robust loss functions will improve as the volatility proxies become more precise. However, the precise multivariate volatility proxies are not freely available, making it essential that the multivariate loss function is robust to proxy noise. Patton and Sheppard (2009) showed that the multivariate version of MSE and QLIKE is robust to noise in the volatility proxy, specified as

$$\text{MSE} : L\left(\hat{\Sigma}_t, H_t\right) = \frac{1}{N^2}\left[\text{vec}\left(\hat{\Sigma}_t - H_t\right)' \text{vec}\left(\hat{\Sigma}_t - H_t\right)\right] \tag{2.57}$$

and

$$\text{QLIKE} : L\left(\hat{\Sigma}_t, H_t\right) = \ln|H_t| + \text{vec}\left(H_t^{-1} \odot \widehat{\Sigma}_t\right)' \mathbf{1}, \tag{2.58}$$

where vec is a column stacking operator, and $\mathbf{1}$ is a vector of ones.

Even though ranking the statistical measures helps distinguish the competing forecasts, the ranking may vary due to the different data samples. Therefore, several statistical tests have been developed in the literature to overcome the limitations associated with the ranking of forecasts.

Diebold and Mariano (1995), and West (1996) introduced the test of equal predictive ability (EPA), commonly known as the DM test. The DM test is based on the differential of the robust loss functions for the models k and j,

$$d_t = L\left(\hat{\sigma}_t^2, h_{k,t}\right) - L\left(\hat{\sigma}_t^2, h_{j,t}\right) \tag{2.59}$$

in the univariate setting and

$$d_t = L\left(\hat{\Sigma}_t, H_{k,t}\right) - L\left(\hat{\Sigma}_t, H_{j,t}\right) \tag{2.60}$$

in the multivariate setting. The null hypothesis of EPA can be expressed as

$$H_0 : \mathrm{E}\left(d_t\right) = 0. \tag{2.61}$$

The test statistic assumes the form

$$DM = \sqrt{T}\frac{\bar{d}}{\sqrt{w}} \xrightarrow{d} N(0,1), \tag{2.62}$$

where

$$\bar{d} = \frac{1}{T}\sum_{t=1}^{T} d_t \tag{2.63}$$

and

$$w = \lim_{t\to\infty} VAR(\sqrt{T}\bar{d}) \tag{2.64}$$

is an estimator of the asymptotic variance of $\bar{d}$ through the sample variance. The extensions of DM test have been significantly developed (see, e.g., Clark and West 2007, Giacomini and White 2006). However, the DM test cannot test the relative performance of multiple competing models.

In order to compare more than two models, the model confidence set (MCS) approach was developed by Hansen et al. (2003, 2005, 2011), allowing to compare all forecasts against each other. The MCS contains a set of models with the best out-of-sample forecasts for a given confidence level, and consists of a sequential procedure that allows testing the EPA of the compared models. The set of candidate models is trimmed by deleting models that are found to be inferior. The models in the final MCS include the optimal model with a given level of confidence and are not significantly different in terms of their forecast performance. Given a full set of candidate models $\mathcal{M}_0 = \{1, \ldots, \mathrm{m}_0\}$, all loss differentials between models i and j

are computed,

$$d_{i,j,t} = L\left(\hat{\sigma}_t^2, h_{i,t}\right) - L\left(\hat{\sigma}_t^2, h_{j,t}\right), \tag{2.65}$$

for all $i, j \in \mathcal{M}_0$ and $i \neq j$. The null hypothesis

$$H_0 : \mathrm{E}\left(d_{ij,t}\right) = 0, \quad \forall i, j \in \mathcal{M}_0 \tag{2.66}$$

is tested. If H_0 is rejected at the significance level α, the worst performing model is removed, and the process continues until non-rejection occurs. The remaining models are statistically equivalent models for a given loss function.

2.5.2 Economic evaluation

As an alternative to statistical evaluation discussed in Section 2.5.1, portfolio optimization and Value-at-Risk (VaR) evaluate the predictive ability through the economic benefit of the forecasts, which is defined as the indirect evaluation since the loss function does not directly measure accuracy (Patton and Sheppard, 2009).

Under the classic Markowitz (1952a) mean-variance portfolio framework, investors determine the optimal allocation of wealth by minimizing the expected variance of their portfolio, given a target return. The literature has been evaluating the multivariate volatility forecast in the context of portfolio optimization since then (Becker et al., 2015). The portfolio optimization problem can be written as

$$\begin{aligned} \min_w \quad & w_t' \hat{\Sigma}_t w_t \\ \text{s.t.} \quad & w_t' \hat{\mu}_t = \mu_p \\ & w_t' \iota = 1, \end{aligned} \tag{2.67}$$

where w is the portfolio weights vector and $\hat{\Sigma}_t$ is the covariance matrix forecast. These two constraints ensure that the investor achieves the target return and invests 100% of capital. Let $\hat{w}_t$ be the solution for the above optimization problem, the measure of economic benefit will

be determined by $\hat{w}_t$. Portfolio variance is commonly used to measure the economic benefit of investors, defined as

$$\hat{\sigma}^2_{p,t} = \hat{w}'_t \hat{\Sigma}_t \hat{w}_t. \tag{2.68}$$

Given that the aim of the portfolio optimization problem is to minimize the portfolio variance, the forecast, $\hat{\Sigma}_t$, that minimize $\hat{\sigma}^2_{p,t}$ would be deemed the optimal forecast.

VaR is another measure of economic benefit used extensively to evaluate volatility forecasts. The VaR is defined as the maximum portfolio loss at a given confidence level, α, over a given horizon, described as

$$P\left[R_{p,t} \leq -VaR_t(\alpha)\right] = \alpha, \tag{2.69}$$

where R_p is the portfolio return. There are several ways to calculate VaR, such as the mean-variance approach, the historical simulation approach, and the Monte Carlo simulations approach. Based on the mean-variance approach, a portfolio's VaR forecast at a daily horizon is given by

$$= VaR_t(\alpha) = \sqrt{h_t} F_i^{-1}(\alpha), \tag{2.70}$$

where F_i^{-1} is the inverse of the specified cumulative density function and $h_t = \hat{\sigma}^2_{p,t}$. The greater the VaR means the larger the risk of the portfolio. A risk-averse investor would prefer to minimize the VaR of the portfolio,

$$\begin{aligned} \min_w \quad & VaR_t(\alpha) \\ \text{s.t.} \quad & w'_t \iota = 1. \end{aligned} \tag{2.71}$$

2.6 Conclusion

This chapter highlighted the definition of realized volatility and identified the variety of approaches used to forecast realized volatility. Furthermore, it has also reviewed various loss

functions and evaluation methods that can be applied to discriminate between competing volatility forecasts. There have been a large number of significant achievements in forecasting realized volatility, and the developments associated with realized volatility models are extensive. Although all of them have been proved their forecasting ability, no one dominates all others. However, it is clear that using high-frequency intraday data and realized volatility measures would improve the accuracy of volatility forecasting.

There are several areas for potential developments in realized volatility forecasting that are open for investigating. First, while the recent contributions of applying the HARQ model can improve the forecasts accuracy of realized variance, the investigation of why the HARQ model outperforms the standard HAR model is not widely discussed in the literature, especially from an empirical perspective. Therefore, further analysis on the realized volatility dynamics will be addressed.

Second, while there are studies that proposed various competing multivariate HAR models to forecast the realized covariance, none of these have considered the cross-effect between the entries in the covariance matrix. However, this thesis shows that it will improve the accuracy of the forecast by incorporating the cross-effect in the standard multivariate HAR model.

Finally, forecasting the global minimum variance portfolio (GMVP) weight is ultimately important for financial investment decisions as it is the starting point of the Markowitz mean-variance efficient frontier. The conventional method is plugging the covariance prediction into the GMVP formula. While a lot of methodologies have been designed for forecasting realized covariance, there is little discussion on the ability of those methodologies to the GMVP prediction. These three areas are the main focus of the rest of the chapters.

Chapter 3

HARQ Model: Modeling with Measurement Error and Non-linear Dependence

---

3.1 Overview

RV is a non-parametric volatility measure built from the intraday returns to help exploit the inherent information in high-frequency data. Raw RV is a consistent estimator of integrated variance (IV) with a few distinctive features, such as spikes/outliers and non-Gaussianity. Moreover, RV is a noisy estimate of the true IV. According to equation (2.5), the RV will be equal to the sum of IV and a heteroskedastic measurement error in any given finite sample. Therefore, the dynamic modeling of RV faces an errors-in-variables problem, which leads to an attenuation bias where the RV process is less persistent than the latent IV. Particularly, the degree of attenuation bias depends on the variance of measurement error, with the greater variance comes the less persistent RV. Given the stylized features of RV, the standard HAR model of Corsi (2009), estimated using raw RV and under OLS, is far from ideal (Clements and Preve, 2021). In most situations, the measurement error of RV has been treated as homoskedastic. For example, Asai et al. (2012) estimated a series of state-space models for the realized variance with homoskedastic measurement error. Recently, based on the asymptotic theory of RV by Barndorff-Nielsen and Shephard (2002), the HARQ model of Bollerslev et al. (2016) was

designed to address the heteroskedastic measurement error directly by augmenting the standard HAR model with a term as a function of realized quarticity (RQ). It has been documented that the HARQ model outperforms the standard HAR model and some of its numerous extensions, representing the state-of-the-art univariate RV forecasting models.

This chapter aims to understand deeper why the HARQ model can provide improved RV forecasts from an empirical perspective. A dynamic specification was derived by replacing the square root of RQ (sqrt-RQ) in the HARQ model with RV, given the fact that sqrt-RQ is strongly correlated with RV. This dynamic model, denote as the HARP model[1], exhibits a very similar forecasting performance to the HARQ model. Because of the quadratic term, the squared RV, in the HARP model, we speculate that the forecasting gains provided by the HARP may be attributed to potential non-linear dependence in the conditional mean of RV, as captured by the quadratic term. Even though the HARQ model is designed to tackle the heteroskedastic measurement error of raw RV, it is possible that the interaction term (i.e. the product of sqrt-RQ and RV) in the HARQ model may capture both the effect of the heteroskedastic measurement error and the non-linear dependence in the conditional mean of RV.

This speculation was verified using the natural logarithmic transformation (log-transformation) on the RV series before dynamic modeling. The logarithmic transformation, a concave non-linear function, is a special case of the Box-Cox transformation (Box and Cox, 1964). In particular, it is suitable for data series whose variance exhibits heteroskedasticity. It can be demonstrated that the log-transformation could eliminate the heteroskedasticity in the measurement error if sqrt-RQ and RV are perfectly correlated. The intuition is, if the heteroskedastic measurement error is the only issue that needs to be addressed, the HAR model with log-transformed RV (log-HAR model) will be worse than the standard HARQ model in terms of forecasting ability, after the non-linear transformation is imposed on RV.

The out-of-sample analysis was implemented based on the HAR, HARQ, HARP models and their logarithmic versions, providing strong evidence that both heteroskedastic measurement

[1] Cipollini et al. (2021) investigated the empirical relationship between sqrt-RQ and RV, proposing a similar dynamic specification.

error and non-linear dependence in the conditional mean of RV must be considered before the forecasting modeling. Therefore, the success of the HARQ model can be attributed to these two roles. While the HARQ model is developed for the heteroskedastic measurement error, it also helps with the non-linearity in the conditional mean of RV. Similarly, the HARP model captures the non-linearity via the squared RV but also handles the heteroskedasticity in the measurement error.

More importantly, the log-transformation of RV is able to deal with both issues simultaneously. First, regardless of forecasting schemes and horizons, the out-of-sample forecasts show the log-HAR model outperforms the standard HARQ model, especially when sqrt-RQ and RV are highly correlated. It indicates the non-linear transformation on RV further improves the forecasts accuracy and the conditional mean of RV exhibits the non-linear dependence. Next, the log-HARQ model presents a very similar forecasting performance to the log-HAR model, adding an extra quarticity variable to the log-HAR model will not be able to enhance the forecast accuracy significantly. In other words, the log-transformation itself is enough to deal with most of the heteroskedastic measurement error. Finally, compared to the log-HAR model, the log-HARP model works poorly when sqrt-RQ and RV are highly correlated. This implies that, after controlling for the majority of the heteroskedasticity, the non-linear dependence in the conditional mean of RV can be captured via the log-transformation, adding extra variables with a higher exponent (e.g. quadratic term, cubic term, etc.) cannot help with the non-linear dependence.

There are a few studies that are related to our work. Buccheri and Corsi (2021) extended the HAR model by applying the Kalman Filter technique to accounting for the misspecification caused by heteroskedastic measurement error and/or non-linearity dynamics of RV. Buccheri and Corsi (2021) established significant forecasting gains based on their approach. Building on the asymmetric multiplicative measurement error representation by Engle and Gallo (2006), Cipollini et al. (2021) proposed a new framework to deal with the effect of the heteroskedasticity of measurement error.

The remainder of this chapter is organized as follows: Section 3.2 describes the econometric

specifications, and Section 3.3 reports the results of its empirical study. Section 3.4 concludes.

3.2 Methodology

This section introduces the proxies and models employed to construct realized volatility forecasts, then presents the logarithmic transformation and the methods of assessing forecasts accuracy employed in the literature.

3.2.1 The relationship between RV and quarticity

As discussed in Section 2.2.1, given that prices evolve continuously through time as a stochastic volatility diffusion process, the integrated variance is not directly observable, but can be consistently measured by realized variance, using high frequency data (Andersen and Bollerslev, 1998a,b). Barndorff-Nielsen and Shephard (2002) later developed the asymptotic distribution theory of RV, implying the measurement relationship between RV and IV, shown as

$$RV_t = IV_t + \eta_t. \tag{3.1}$$

The measurement error η_t is conditionally heteroskedastic and depends on the level of integrated quarticity (IQ), following a mixed normal distribution, defined as

$$\eta_t \sim MN\left(0,\, 2\Delta IQ_t\right), \tag{3.2}$$

where Δ is the intraday interval for sampling returns. For practical purposes, IQ can be consistently estimated by RQ, which is built from the fourth power of intraday returns (see equation (2.7)). Bollerslev et al. (2016) discussed the attenuation bias in the coefficient induced by measurement errors in the HAR model (Corsi, 2009) and proposed the HARQ model where a component as a function of the sqrt-RQ is used to capture the measurement error. The HARQ

model in equation (2.26) can be rewritten as

$$RV_t = \beta_0 + \beta_1 RV_{t-1}^d + \beta_2 RV_{t-1}^w + \beta_3 RV_{t-1}^m + \alpha_1 \sqrt{RQ_{t-1}^d} \times RV_{t-1}^d + \varepsilon_t. \tag{3.3}$$

Compared to the standard HAR model, the extra interaction term between sqrt-RQ and RV allows for more weight to be placed on the RV when the measurement error is lower. Theoretically, although sqrt-RQ is different from the RV, the two series are proven to be closely related. From an empirical point of view, sqrt-RQ proves to be proportional to RV. Figure 3.1 shows the RV and the sqrt-RQ plots of the S&P 500 index over the period May 24, 2000, to June 27, 2019. It can be observed that RV and sqrt-RQ share a similar dynamic pattern. Figure 3.2 presents the histogram of the correlation coefficient between RV and sqrt-RQ for fifty U.S. individual stocks, showing the correlation coefficient is centered on 0.92, with the smallest number greater than 0.85. This high correlation between sqrt-RQ and RV motivates us to replace sqrt-RQ in the HARQ model with RV, leading to a standard HAR model with an extra squared RV item, denoted the HARP model, where P stands for Polynomial. The standard HARP model is defined as

$$RV_t = \beta_0 + \beta_1 RV_{t-1}^d + \beta_2 RV_{t-1}^w + \beta_3 RV_{t-1}^m + \alpha_1 \left(RV_{t-1}^d\right)^2 + \varepsilon_t. \tag{3.4}$$

Both in-sample and out-of-sample analysis, between equations (3.3) and (3.4), suggests the two models are similar. The α_1 parameter of the HARP model is negative and significant, which is consistent with the α_1 of the HARQ model. Cipollini et al. (2021) claimed the curvature term of the HARP model explained the idea that volatility dynamics are linkable to market behavior. In this study, given the features of RV, we speculate the quadratic term in the HARP model is related to the non-linear dependence in the conditional mean of RV. To further investigate the non-linearity of RV, based on the HARP model, an extra cubic daily lagged RV term is included,

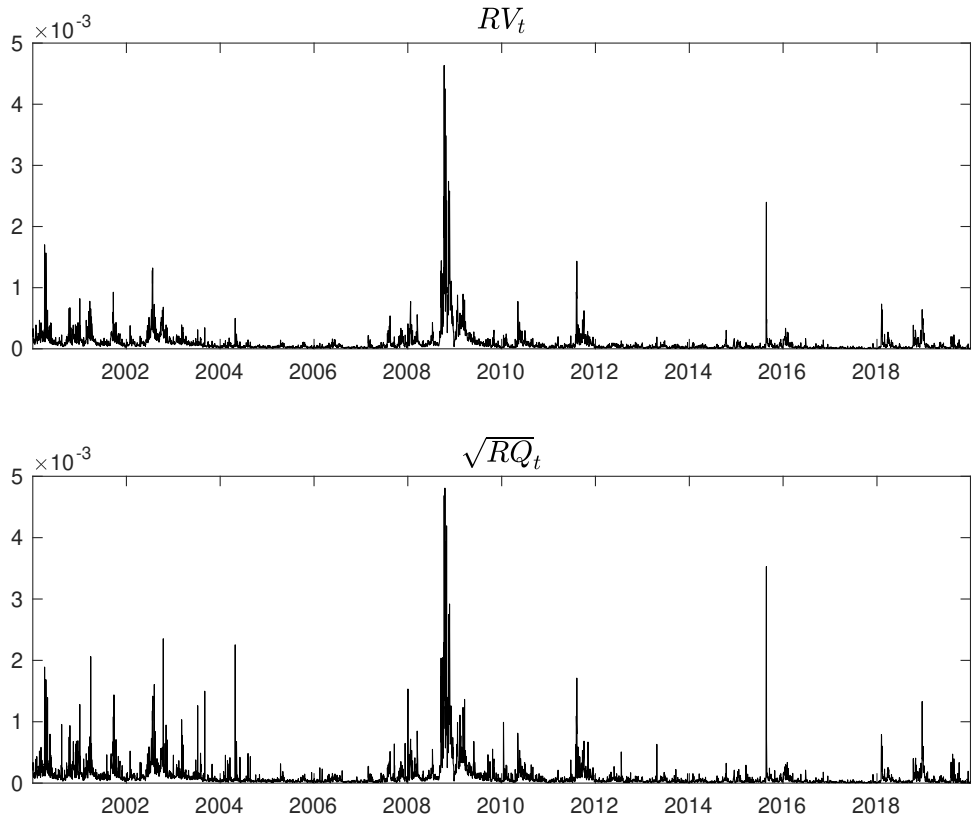

Figure 3.1: RV versus the sqrt-RQ. The figure shows RV and sqrt-RQ of the S&P 500 index from May 24, 2000, to June 27, 2019.

labeled as HARP3 model[2]. The HARP3 model is shown as

$$RV_t = \beta_0 + \beta_1 RV_{t-1}^d + \beta_2 RV_{t-1}^w + \beta_3 RV_{t-1}^m + \alpha_1 \left(RV_{t-1}^d\right)^2 + \alpha_2 \left(RV_{t-1}^d\right)^3 + \varepsilon_t. \tag{3.5}$$

However, based on the out-of-sample analysis, adding higher powers of RV does not improve the forecasting performance of the HARP model.

[2]We also included the higher power of RV, such as the fourth power and fifth power, in the original HARP model. Based on the empirical analysis, the higher power items are insignificant and the effects are negligible.

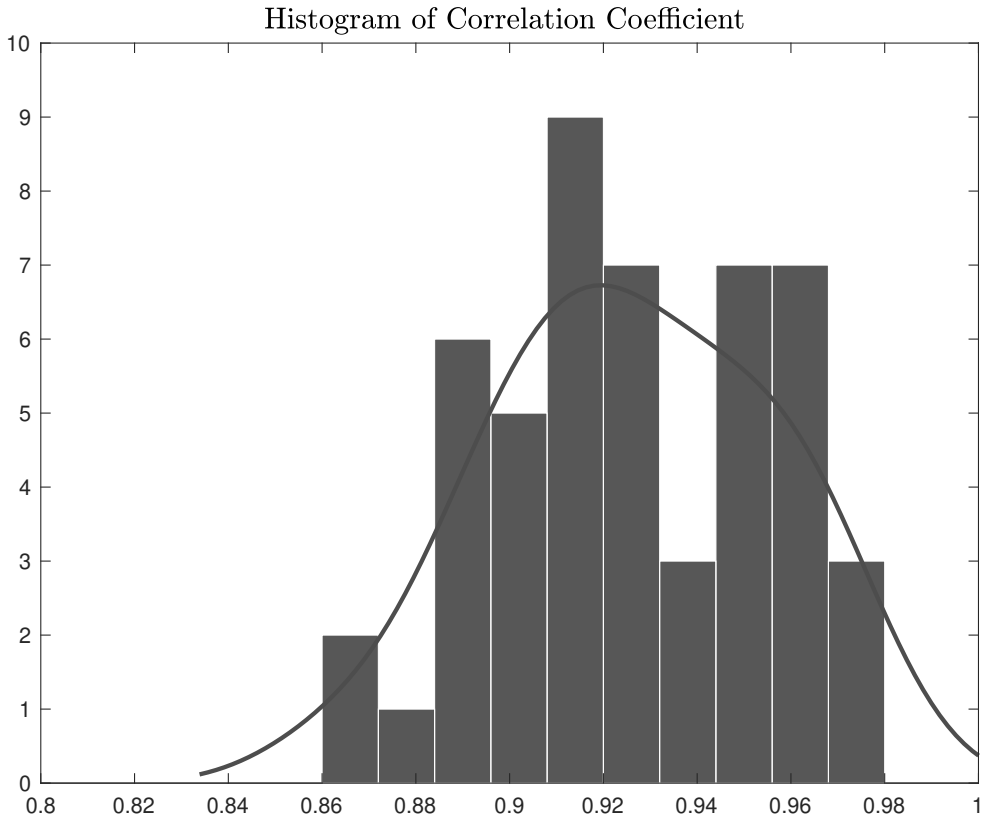

Figure 3.2: Histogram of the correlation coefficient between RV and sqrt-RQ for fifty U.S. market stocks. Solid line is kernel density plot of the histogram.

3.2.2 The logarithmic transformation

The forecasting gains provided by the HARQ model may be attributed to capturing both the effect of heteroskedastic measurement error and the non-linear dependence in the conditional mean of RV, due to the close relationship between the HARQ and HARP models. To verify this speculation, the log-transformation is applied on the RV series before the dynamic modeling. The logarithmic transformation is a special case of the Box-Cox transformation (Box and Cox, 1964), which is a family of variance-stabilizing transformations, suitable particulary for data series where the standard deviation increases linearly with the mean (Brockwell and Davis,

2009). The Box-Cox transformation of a time series variable y_t is

$$y_t(\lambda) = \begin{cases} \frac{y_t^\lambda - 1}{\lambda}, & \lambda \neq 0, \\ \log y_t, & \lambda = 0, \end{cases} \tag{3.6}$$

where λ is the power parameter. The log-transform on RV has been used for reasons of positiveness, substantial skewness, excess kurtosis and potential heteroskedasticity. Barndorff-Nielsen and Shephard (2006) showed in a simulation study that the standard errors of log-transformed RV tend to be smaller than those of the RV estimator in equation (3.2). Using the delta method, the asymptotic distribution of log-transformed RV is

$$\log\left(RV_t\right) = \log\left(IV_t\right) + \xi_t, \quad \xi_t \sim \mathrm{MN}\left(0, 2\Delta\frac{\mathrm{IQ}_t}{\mathrm{IV}_t^2}\right) \tag{3.7}$$

as the intraday period Δ goes to zero (Barndorff-Nielsen and Shephard, 2001, Corsi et al., 2008). Given the consistent estimator of IV and IQ in equation (2.2) and (2.7), a consistent estimator of the variance of ξ_t can be obtained, denoted as V_t, where

$$V_t = \frac{2}{3}\frac{\sum_{i=1}^{M} r_{i,t}^4}{\left(\sum_{i=1}^{M} r_{i,t}^2\right)^2}. \tag{3.8}$$

Figure 3.3 presents a comparison of variances between the raw RV and the log-transformed RV of the S&P 500 index over the period May 24, 2000, to June 27, 2019. The variance of log-transformed RV tends to be stable and less heteroskedastic, compared to the variance of raw RV.

Applying the HAR framework on the log-transformed RV, the log-HAR model can be written as

$$\log RV_t = \beta_0 + \beta_1 \log RV_{t-1}^d + \beta_2 \log RV_{t-1}^w + \beta_3 \log RV_{t-1}^m + u_t, \tag{3.9}$$

where the daily, weekly and monthly lagged log-transformed RV are defined as $\log RV_{t-1}^d =$

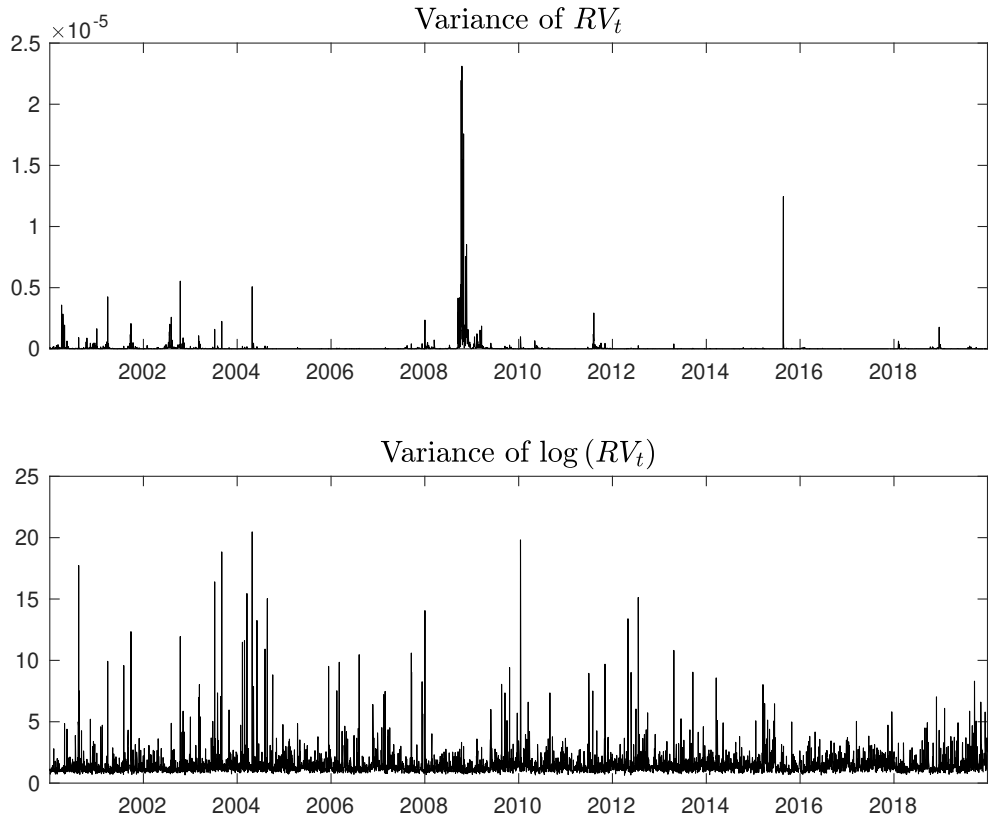

Figure 3.3: Variance of raw RV versus variance of log-transformed RV. Figure shows a comparison of variances between raw RV and log-transformed RV of the S&P 500 index over period May 24, 2000, to June 27, 2019.

$\log RV_{t-1}$, $\log RV_{t-1}^{w} = \frac{1}{5}\sum_{i=1}^{5} \log RV_{t-i}$, $\log RV_{t-i}^{m} = \frac{1}{22}\sum_{i=1}^{22} \log RV_{t-i}$, respectively. Following the standard HARQ and HARP model, the log-HARQ model is defined as

$$\log RV_t = \beta_0 + \beta_1 \log RV_{t-1}^{d} + \beta_2 \log RV_{t-1}^{w} + \beta_3 \log RV_{t-1}^{m} + \alpha \left(\frac{\sqrt{RQ}_{t-1}}{RV_{t-1}} \right) \log RV_{t-1}^{d} + u_t, \tag{3.10}$$

and the log-HARP model is

$$\log RV_t = \beta_0 + \beta_1 \log RV_{t-1}^{d} + \beta_2 \log RV_{t-1}^{w} + \beta_3 \log RV_{t-1}^{m} + \alpha \left(\log RV_{t-1}^{d} \right)^2 + u_t. \tag{3.11}$$

The empirical evidence clearly shows the log-transformation results in symmetric, approximately Gaussian, distributions. One issue with using logarithmic transformation is that the back-transformed RV point forecast will be biased. A bias correction factor is applied to obtain unbiased predictions of RV. Let $\log \widehat{RV}_t$ be the RV point forecast from any logarithmic model, a bias-corrected RV forecast is

$$\widehat{RV}_t = \exp\left(\log \widehat{RV}_t + \frac{\sigma_{ut}^2}{2}\right), \tag{3.12}$$

where σ_{ut}^2 is the conditional variance of the errors u_t. The point forecast in equation (3.12) is optimal if the transformed data is normally distributed. For simplicity, the sample variance of the residuals from the estimation window is used to estimate the σ_{ut}^2.

3.2.3 Insanity filter

The HARQ and HARP-type models may sometimes generate unbelievably large or small forecasts due to the interaction or quadratic RV terms which can be addressed by the methodology of Bollerslev et al. (2016) where an "insanity filter" is applied to all forecasts as in this study. This filter will replace any forecast greater than the maximum or less than the minimum of the dependent variable with the sample average over the estimation period.

3.3 Empirical Analysis

3.3.1 Data

The empirical study presented here is based on two stock market indices, the Standard & Poor's 500 (SPX) and the Dow Jones Industrial Average (DJI), and a set of fifty stocks from the U.S. market. The 5-minute prices for the indices and stocks are obtained from Thomson Reuters Tick History Database. The whole sample period of fifty stocks starts on May 24, 2000, and ends on June 27, 2019, yielding 4612 daily RV observations for each stock. The sample period

of stock market indices spans from January 3, 2000, to December 31, 2019, representing 4624 daily RV observations for DJI and 4925 daily RV observations for SPX. The standard summary statistics of daily RV for each stock are provided in Table A.1 (Supplementary Appendix A.1). All of the RVs are based on 5-minute returns. The first-order autocorrelation coefficients and the correlation coefficients between RV and sqrt-RQ are also reported in Table A.1.

3.3.2 In-sample estimation results

Table 3.1: In-sample estimation results for S&P 500 index.

	HAR	HARQ	HARP	HARP3	log-HAR	log-HARQ	log-HARP	log-HARP3
β_0	0.000	0.000	0.000	0.000	-0.491	-0.579	0.097	-1.739
	(0.000)	(0.000)	(0.000)	(0.000)	(0.079)	(0.078)	(0.414)	(1.922)
β_1	0.281	0.676	0.615	0.449	0.459	0.461	0.585	-0.006
	(0.017)	(0.026)	(0.026)	(0.032)	(0.017)	(0.016)	(0.089)	(0.610)
β_2	0.488	0.412	0.427	0.435	0.314	0.273	0.312	0.312
	(0.027)	(0.027)	(0.027)	(0.027)	(0.024)	(0.024)	(0.024)	(0.024)
β_3	0.150	0.052	0.064	0.070	0.184	0.177	0.183	0.182
	(0.023)	(0.023)	(0.023)	(0.023)	(0.019)	(0.019)	(0.019)	(0.019)
α_1		-118.145	-111.555	83.123		0.031	0.006	-0.056
		(5.974)	(6.563)	(24.026)		(0.003)	(0.004)	(0.064)
α_2				-39499.384				-0.002
				(4692.125)				(0.002)
R^2	0.596	0.626	0.619	0.624	0.604	0.609	0.599	0.607
MSE	0.160	0.148	0.151	0.149	0.157	0.155	0.159	0.156
AIC	-74035.27	-74409.88	-74314.26	-74382.70	7748.32	7610.17	7748.22	7749.26
BIC	-74009.28	-74377.39	-74281.77	-74343.72	7774.30	7642.66	7780.71	7788.24

Notes: Table provides in-sample parameter estimates and measures of fit for the various HAR-type models and their logarithmic versions. Top panel reports the actual parameter estimates with robust standard errors in parentheses. Bottom panel summarizes the in-sample R^2 and Mean Squared Error (MSE), together with the model selection criterion values. The MSE values for all models have been scaled up by 1×10^7.

In order to provide insights into the log-transformed RV, this section outlines the in-sample estimation results of standard HAR-type models and their logarithmic versions. The in-sample estimation results in this section are based on the full sample size, all estimated by OLS.

Table 3.2: In-sample estimation results for Dow Jones Industrial Average index.

	HAR	HARQ	HARP	HARP3	log-HAR	log-HARQ	log-HARP	log-HARP3
β_0	0.000	0.000	0.000	0.000	-0.644	-0.527	-0.296	4.124
	(0.000)	(0.000)	(0.000)	(0.000)	(0.095)	(0.090)	(0.478)	(2.294)
β_1	0.292	0.718	0.523	0.326	0.323	0.429	0.397	1.836
	(0.018)	(0.029)	(0.025)	(0.031)	(0.018)	(0.017)	(0.102)	(0.737)
β_2	0.466	0.383	0.428	0.412	0.419	0.306	0.418	0.418
	(0.028)	(0.028)	(0.028)	(0.028)	(0.027)	(0.026)	(0.027)	(0.027)
β_3	0.168	0.107	0.119	0.130	0.201	0.170	0.201	0.201
	(0.024)	(0.024)	(0.024)	(0.024)	(0.022)	(0.021)	(0.022)	(0.022)
α_1		-137.440	-91.299	171.108		0.029	0.004	0.157
		(7.423)	(7.261)	(25.200)		(0.001)	(0.005)	(0.078)
α_2				-57084.833				0.005
				(5255.489)				(0.003)
R^2	0.615	0.642	0.628	0.638	0.622	0.635	0.623	0.609
MSE	0.162	0.151	0.157	0.153	0.159	0.154	0.159	0.165
AIC	-69500.66	-69829.70	-69654.29	-69768.93	8410.63	7914.67	8412.07	8410.19
BIC	-69474.92	-69797.53	-69622.11	-69730.32	8436.36	7946.84	8444.25	8448.80

Notes: Table provides in-sample parameter estimates and measures of fit for the various HAR-type models and their logarithmic versions. Top panel reports the actual parameter estimates with robust standard errors in parentheses. Bottom panel summarizes the in-sample R^2 and Mean Squared Error (MSE), together with the model selection criterion values. The MSE values for all models have been scaled up by 1×10^7.

Table 3.1 shows the results for the SPX index, and Table 3.2 shows the results for the DJI index. Both tables provide the parameter estimates and measures of fit. The top panel reports the actual parameter estimates with robust standard errors in parentheses. The bottom panel summarizes the in-sample R^2, MSE, and QLIKE losses from the regressions, together with the model selection criterion values. The model parameter estimates for the individual stocks are presented in Appendix A.2.

The standard HARQ model (Bollerslev et al., 2016) places greater weight on the daily lag and less weight on the weekly and monthly lags, compared to the HAR model, which is reflected in both Table 3.1 and 3.2, with an increase in β_1 and decrease in β_2 and β_3. It can be observed that the standard HARQ model has the highest in-sample R^2 and lowest in-sample

MSE values. The in-sample results show that the standard HARP model has a similar effect to HARQ regression, with an increase in β_1 and corresponding decreases in β_2 and β_3 estimates relative to the HAR model. Both standard HARQ and HARP models outperform the standard HAR model.

The second half of the table reports the in-sample estimation results for the logarithmic models. Just like the standard HARQ and HARP models, the log-HAR, log-HARQ, and log-HARP models place greater weight on more recent RV, with β_1 and β_2 parameter estimates being larger than β_3 estimate. Even if the log-HARQ model has the highest in-sample R^2 and lowest in-sample MSE, the log-HARQ model only shows a small improvement in measures of fit over the log-HAR model. Meanwhile, it is worth mentioning that the log-HARP model behaves differently across the market indices. The results in Table 3.1 show that the log-HARP model performs poorly in terms of the measures of fit, with the lowest in-sample R^2 and highest in-sample MSE, while the in-sample performance of the log-HARP model in Table 3.2 is only slightly different from the log-HAR model. The summary statistics in Table A.1 show the correlation coefficient between RV and sqrt-RQ is 0.951 for the SPX index and 0.871 for the DJI index. Compared to the log-HAR model, the log-HARP model works poorly when sqrt-RQ and RV are highly correlated. This implies the non-linear dependence in the conditional mean of RV can be captured via the log-transformation, which means adding extra variables with higher powers (e.g. quadratic RV, cubic RV, etc.) cannot help with the non-linear dependence.

The results of this section show that the standard HARP model performs similarly to the standard HARQ model. The results for in-sample fit are mixed. However, the following out-of-sample forecasting analysis will clarify the practical benefits.

3.3.3 Out-of-sample analysis

We focus our discussion on the out-of-sample forecasting results in this section. The results reported here are based on two forecasting schemes, a rolling scheme with a 1000 day rolling window and a recursive (increasing window) forecasting scheme that uses the first 1000 trading

days to estimate the models. Following Bollerslev et al. (2016), the "insanity filter" is applied to all the RV forecasts since the HARQ and HARP models may produce implausibly large or small forecasts. Figure 3.4 shows the total number of "insanity filter" that triggered during the out-of-sample analysis for both rolling and increasing windows. It is noticed that, compared to the standard HAR model, the HARQ and HARP models have applied "insanity filter" more frequently, especially when the correlation coefficient is large.

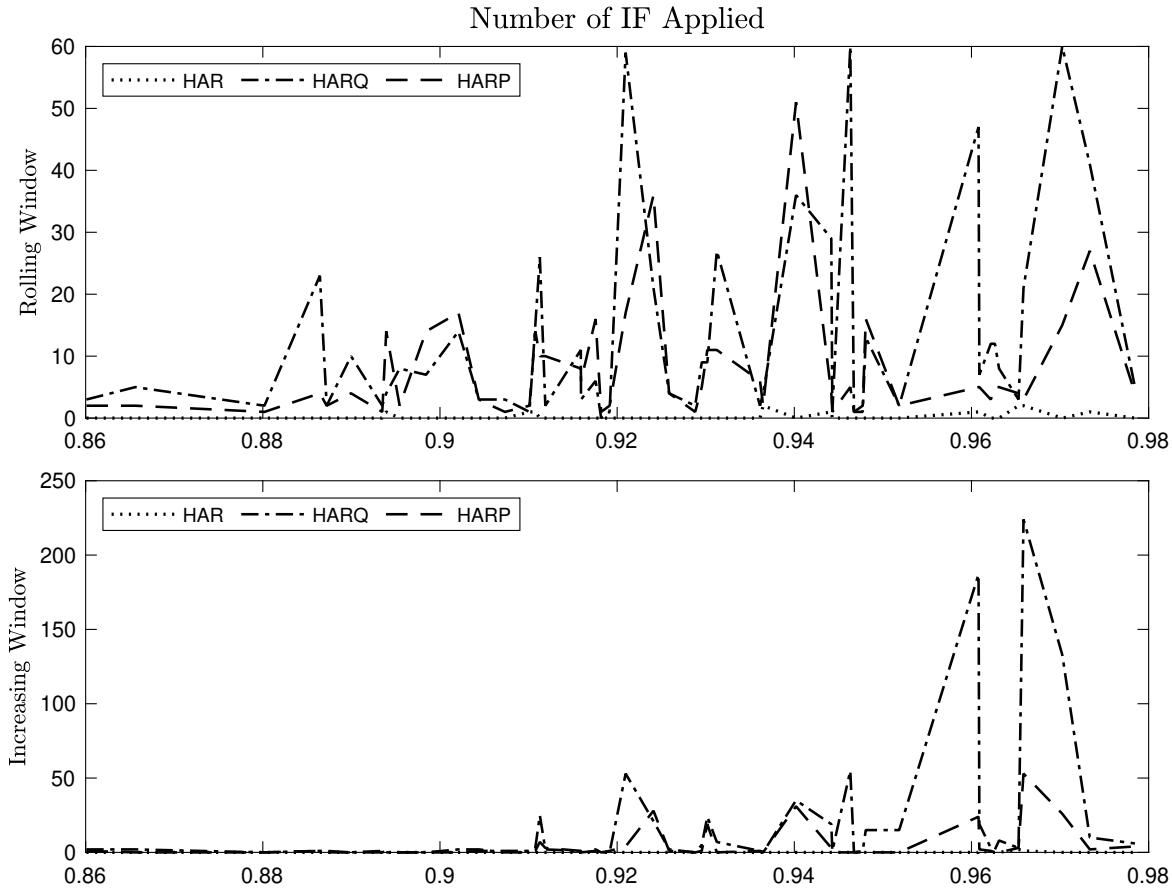

Figure 3.4: Number of "insanity filter" that applied for standard HAR (dotted), standard HARQ (dot-dash) and standard HARP (dashed) models for fifty stocks at 1-day-ahead forecast horizon. Top figure shows the total number based on the rolling window while the bottom figure shows results from the increasing window. Horizontal axis presents correlation coefficients between RV and sqrt-RQ that are sorted in ascending order.

The out-of-sample forecasting results for the 1-day-ahead forecast horizon are reported in Table 3.3, and the out-of-sample results for longer horizons are reported in Table 3.4 and Table

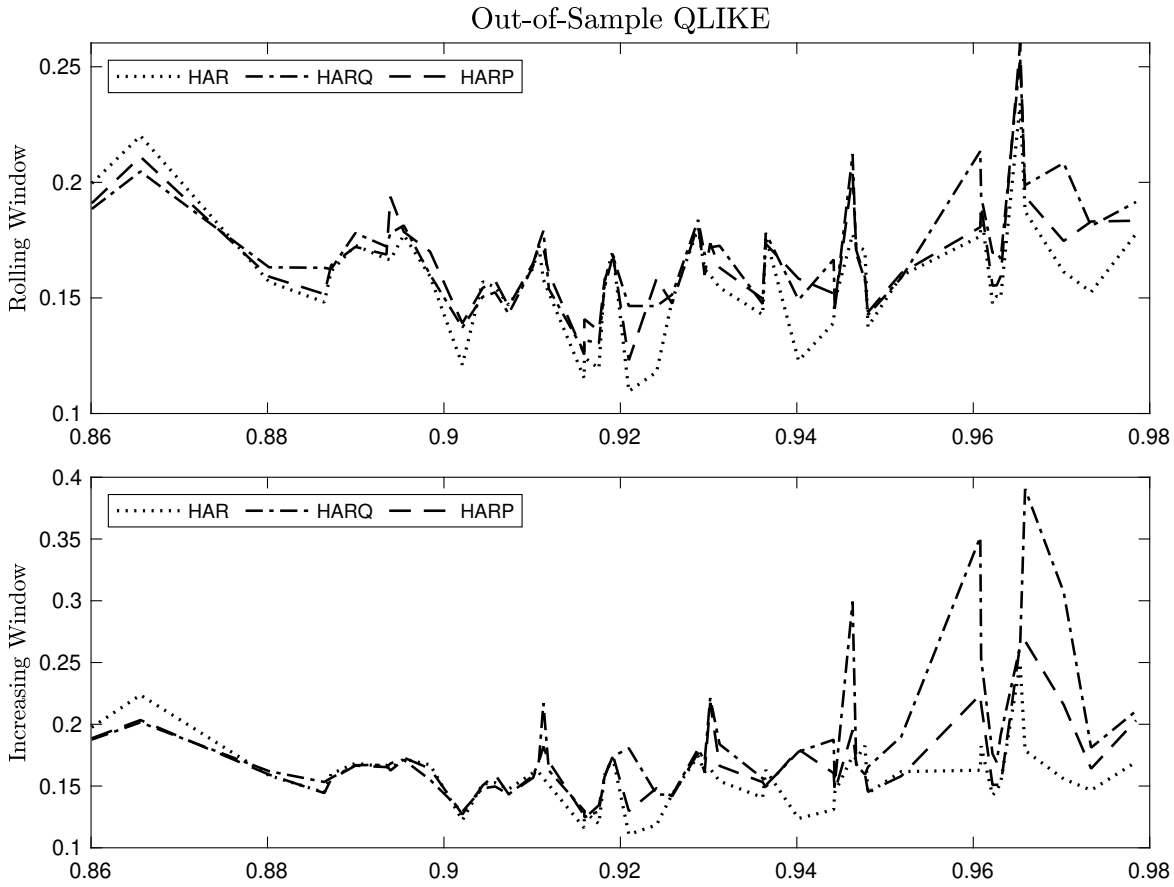

Figure 3.5: Out-of-sample QLIKE losses of standard HAR (dotted), standard HARQ (dot-dash) and standard HARP (dashed) models for fifty stocks at 1-day-ahead forecast horizon. Top figure shows the QLIKE loss based on the rolling window while the bottom figure shows results from the increasing window. Horizontal axis presents correlation coefficients between RV and sqrt-RQ that are sorted in ascending order.

3.5. All three tables present the QLIKE and MSE loss ratios of the alternative approaches relative to the standard HAR model, for both rolling and increasing window forecasting schemes. These ratios will be used to compare the forecasting performance of the alternative models to the benchmark model, which is the standard HAR model. The model with a loss ratio that is less than 1 outperforms the standard HAR model in terms of forecasting accuracy. The model confidence set (MCS) test highlights the best set of performing models (Hansen et al., 2003, 2005) is employed and the results are recorded in Table 3.6. The table reports the MCS results at different forecasting horizons under two forecasting schemes. The result for market indices

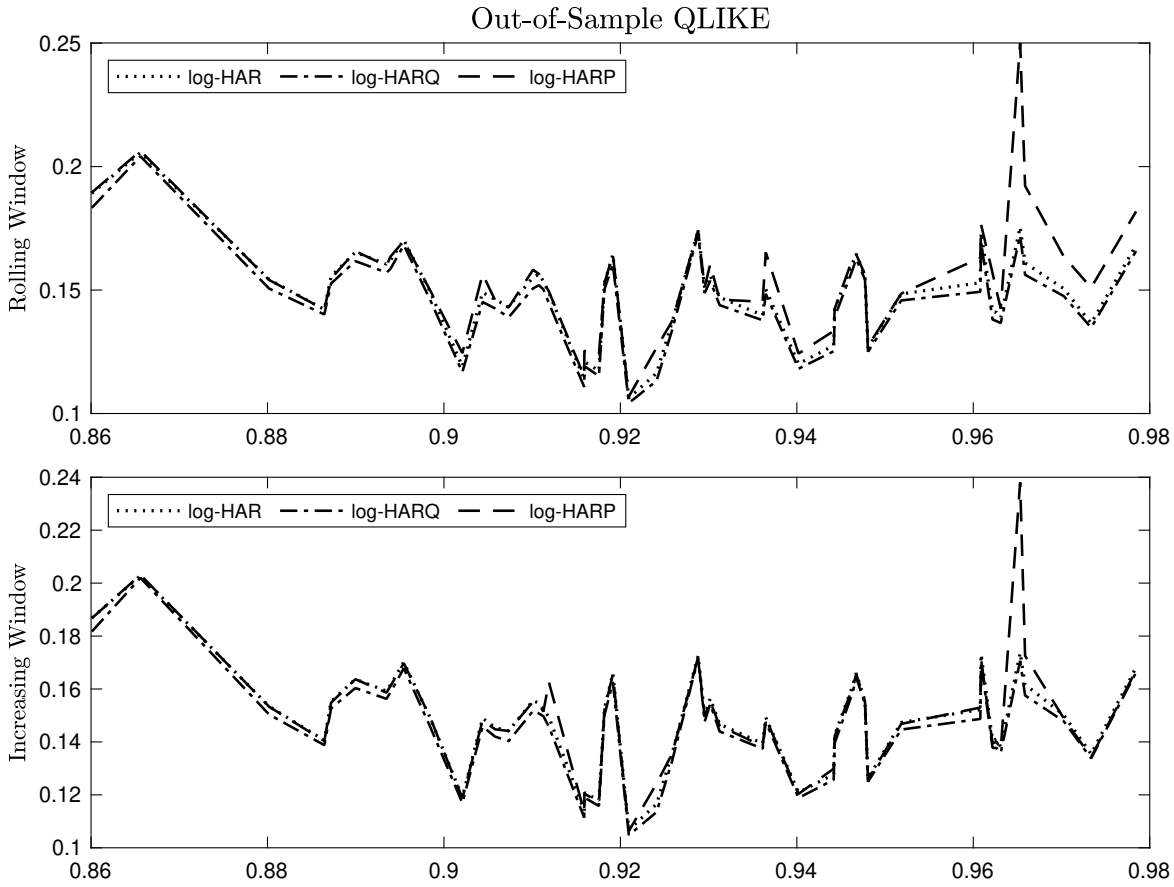

Figure 3.6: Out-of-sample QLIKE losses of log-HAR (dotted), log-HARQ (dot-dash) and log-HARP (dashed) models for fifty stocks at 1-day-ahead forecast horizon. Top figure shows the QLIKE loss based on the rolling window while the bottom figure shows results from the increasing window. Horizontal axis presents correlation coefficients between RV and sqrt-RQ that are sorted in ascending order.

reflects the $p-$value of a 90% MCS while the result for stocks reflects the frequency a model is included in a 90% MCS among the 50 individual stocks. Meanwhile, the 1-day-ahead out-of-sample QLIKE for forecasting models is visualized in Figure 3.5 and 3.6. In both figures, stocks are sorted in ascending order by the correlation coefficient between RV and sqrt-RQ. Figure 3.5 shows the QLIKE loss of standard HAR, HARQ and HARP models while Figure 3.5 shows the QLIKE loss of logarithmic models. The daily out-of-sample forecasting performance difference between the log-HAR and standard HARQ/HARP model is presented in Figure 3.7. The figures of QLIKE loss at longer forecasting horizons and figures of MSE loss at daily

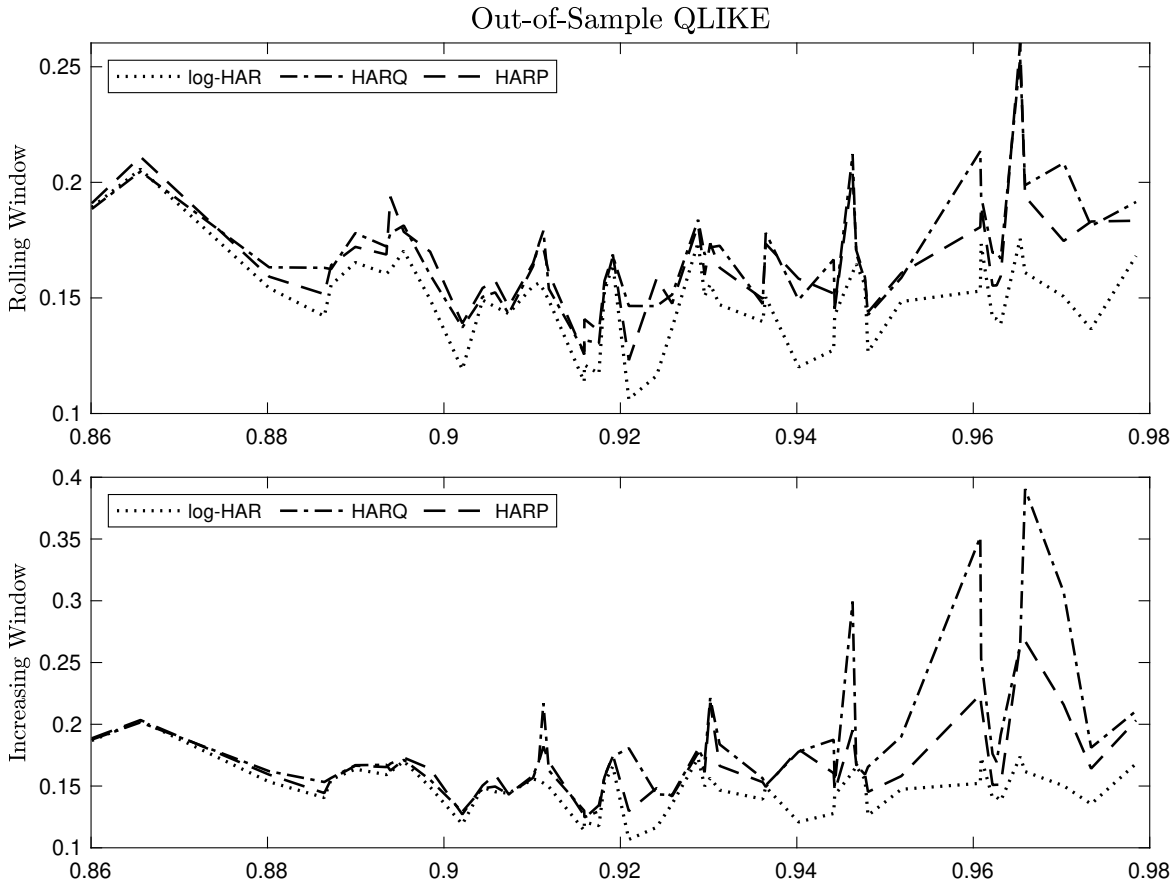

Figure 3.7: Out-of-sample QLIKE losses of log-HAR (dotted), standard HARQ (dot-dash) and standard HARP (dashed) models for fifty stocks at 1-day-ahead forecast horizon. Top figure shows the QLIKE loss based on the rolling window while the bottom figure shows results from the increasing window. Horizontal axis presents correlation coefficients between RV and sqrt-RQ that are sorted in ascending order.

forecasting horizons are provided in the Supplementary Appendix A.3. The weekly out-of-sample QLIKE plots are presented in Figure A.4 and Figure A.5, and the monthly plots are presented in Figure A.7 and Figure A.8. The daily out-of-sample MSE plots are presented in Figure A.1, A.2 and A.3.

This chapter aims to reach a deeper understanding of why the HARQ model can provide improved RV forecasts. More specifically, to investigate whether the heteroskedastic measurement error and non-linear dependence in the conditional mean of RV affect the predictive accuracy. We compare the daily forecast loss ratios reported in columns 1-4 of Table 3.3. Given that

Table 3.3: Daily out-of-sample forecast losses.

		HAR	HARQ	HARP	HARP3	log-HAR	log-HARQ	log-HARP	log-HARP3
DJI									
QLIKE	*rw*	1.000	1.372	1.166	1.026	0.931	0.856	0.933	0.938
	iw	1.000	1.849	1.116	1.021	0.952	0.887	0.952	0.953
MSE	*rw*	1.000	0.917	0.909	0.947	0.871	0.846	1.037	0.824
	iw	1.000	0.952	0.951	0.973	0.934	0.896	0.956	0.992
SPX									
QLIKE	*rw*	1.000	1.317	1.232	1.016	0.861	0.846	0.868	0.881
	iw	1.000	1.673	1.380	0.907	0.877	0.859	0.876	0.877
MSE	*rw*	1.000	0.844	0.841	0.883	0.802	0.796	0.858	0.750
	iw	1.000	1.072	0.917	1.037	0.900	0.891	0.923	0.924
Stocks									
$\overline{\text{QLIKE}}$	*rw*	1.000	1.067	1.052	1.126	0.942	0.924	0.973	0.966
	iw	1.000	1.173	1.075	1.157	0.950	0.933	0.956	0.963
$\overline{\text{MSE}}$	*rw*	1.000	0.965	0.965	1.020	0.891	0.884	0.951	0.913
	iw	1.000	0.966	0.979	1.002	0.929	0.922	0.999	0.952

Notes: Out-of-sample forecast losses at 1-day-ahead forecast horizon. Table reports the ratio of the losses for the different models relative to the losses of the HAR model. The top two panels show the results for the two market indices. Bottom panel reports the average loss ratios across all individual stocks.

the sqrt-RQ is strongly correlated with RV, the HARP model is obtained by replacing the sqrt-RQ variable in the HARQ model with an RV variable. Results indicate the HARP model has the lowest MSE loss ratio, regardless of the forecasting scheme, and the MSE loss ratio of the HARQ model is not far behind. While the out-of-sample MSE loss ratios show that the standard HARQ and HARP model outperforms the standard HAR model, the out-of-sample QLIKE loss reveals a completely different picture. Surprisingly, compared to the HARQ and HARP models, the standard HAR model (estimated via OLS) has a better forecasting performance in terms of QLIKE, and this might be because the application of "insanity filter" when RV is highly correlated with sqrt-RQ. Table 3.3 identifies the HARQ model has the highest QLIKE loss ratio for both rolling and increasing windows, and again, the QLIKE of HARP model followed closely behind. This comparable out-of-sample performance between HARQ and HARP can

Table 3.4: Weekly out-of-sample forecast losses.

		HAR	HARQ	HARP	HARP3	log-HAR	log-HARQ	log-HARP	log-HARP3
DJI									
QLIKE	*rw*	1.000	0.979	0.961	0.932	0.783	0.727	0.967	0.841
	iw	1.000	0.947	0.932	0.951	0.894	0.840	0.959	0.893
MSE	*rw*	1.000	0.963	1.074	1.068	0.824	0.804	1.213	0.987
	iw	1.000	0.920	1.015	1.107	0.914	0.878	1.068	0.944
SPX									
QLIKE	*rw*	1.000	0.913	0.897	0.985	0.776	0.766	0.977	0.821
	iw	1.000	0.841	0.836	0.857	0.814	0.802	0.846	0.810
MSE	*rw*	1.000	0.966	1.098	1.151	0.907	0.894	1.396	1.046
	iw	1.000	0.983	0.998	1.082	0.869	0.850	1.013	0.932
Stocks									
$\overline{\text{QLIKE}}$	*rw*	1.000	0.962	0.970	1.041	0.864	0.852	0.994	0.913
	iw	1.000	1.000	0.979	1.064	0.881	0.869	0.971	0.908
$\overline{\text{MSE}}$	*rw*	1.000	0.988	1.041	1.106	0.945	0.935	1.145	1.010
	iw	1.000	0.981	1.016	1.056	0.934	0.924	1.072	0.982

Notes: This table reports the same loss ratios for the weekly forecasting models previously reported for the 1-day-ahead forecasts in Table 3.3. The top two panels show the results for the two market indices. The bottom panel reports the average loss ratios across all of the individual stocks.

be observed directly in Figure 3.5 where it is clear to see the dot-dashed line of HARQ and the dashed line of HARP coincide most of the time, despite a few discrepancies when RV is highly correlated with sqrt-RQ. Overall, given the empirical fact that the HARP model has a consistent forecasting performance compared with the standard HARQ model, and the squared RV in the HARP model is more related to the potential non-linear dependence in the (conditional mean of) RV, it is possible that the interaction term of the HARQ model captures both the effect of heteroskedastic measurement error and the non-linear dependence in the conditional mean of RV. Compared to the estimation of RV, RQ estimation is more challenging and RQ estimation is quite imprecise in many realistic scenarios, which means the HARP model is a viable alternative to HARQ specification.

Furthermore, the daily forecast loss ratios of logarithmic models in Table 3.3 provided

Table 3.5: Monthly out-of-sample forecast losses.

		HAR	HARQ	HARP	HARP3	log-HAR	log-HARQ	log-HARP	log-HARP3
DJI									
QLIKE	*rw*	1.000	0.958	0.967	0.991	0.839	0.834	0.942	0.914
	iw	1.000	0.922	0.931	0.924	0.875	0.853	0.938	0.867
MSE	*rw*	1.000	0.989	1.020	1.072	0.850	0.868	1.053	1.148
	iw	1.000	0.942	0.972	0.978	0.830	0.826	1.044	0.880
SPX									
QLIKE	*rw*	1.000	0.927	0.926	0.925	0.769	0.764	0.884	0.871
	iw	1.000	0.920	0.875	0.839	0.787	0.782	0.814	0.785
MSE	*rw*	1.000	1.018	1.031	1.070	0.893	0.894	1.123	1.205
	iw	1.000	1.000	1.000	1.049	0.904	0.905	1.045	0.957
Stocks									
$\overline{\text{QLIKE}}$	*rw*	1.000	0.964	0.966	0.964	0.890	0.885	0.932	0.915
	iw	1.000	0.948	0.951	0.950	0.846	0.842	0.866	0.858
$\overline{\text{MSE}}$	*rw*	1.000	1.015	1.010	1.030	1.085	1.083	1.210	1.141
	iw	1.000	0.990	1.003	1.015	0.972	0.970	1.084	1.025

Notes: This table reports the same loss ratios for the monthly forecasting models previously reported for the 1-day-ahead forecasts in Table 3.3. The top two panels show the results for the two market indices. The bottom panel reports the average loss ratios across all of the individual stocks.

strong evidence both heteroskedastic measurement error and non-linear dependence in the conditional mean of RV must be considered before the forecasting modeling. The concave non-linear logarithmic transformation is a special case of the Box-Cox transformation, particularly suitable for data series where variance exhibits heteroskedasticity. We first compare the loss ratios reported under the log-HAR and benchmark models in Table 3.3. After applying the log-transform on the RV series, the QLIKE and MSE loss ratios of the log-HAR model are all less than 1, which indicates the log-HAR model outperforms both the standard HAR model and the HARQ and HARP models. Figure 3.7 highlights the QLIKE loss difference between log-HAR and standard HARQ, HARP models. Regardless of the forecasting schemes, the dotted line (QLIKE loss of log-HAR) is beneath the other two most of the time. It is clear that non-linear transformation on RV improves the accuracy of the forecast especially when sqrt-RQ and

Table 3.6: Model confidence set (MCS) results.

		HAR	HARQ	HARP	log-HAR	log-HARQ	log-HARP
Daily							
DJI	*rw*	0.000	0.000	0.000	0.000	1.000	0.000
	iw	0.000	0.000	0.000	0.000	1.000	0.000
SPX	*rw*	0.000	0.000	0.000	0.094	1.000	0.094
	iw	0.000	0.000	0.000	0.112	1.000	0.112
Stocks	*rw*	1	3	2	7	50	7
	iw	0	2	2	6	50	15
Weekly							
DJI	*rw*	0.003	0.003	0.003	0.003	1.000	0.003
	iw	0.000	0.198	0.198	0.198	1.000	0.198
SPX	*rw*	0.002	0.005	0.005	0.060	1.000	0.060
	iw	0.000	0.000	0.005	0.005	1.000	0.005
Stocks	*rw*	3	7	8	14	50	12
	iw	3	8	8	12	50	15
Monthly							
DJI	*rw*	0.129	0.350	0.350	0.764	1.000	0.350
	iw	0.057	0.358	0.358	0.358	1.000	0.358
SPX	*rw*	0.051	0.207	0.207	0.224	1.000	0.224
	iw	0.002	0.308	0.308	0.394	1.000	0.394
Stocks	*rw*	21	26	26	41	50	41
	iw	9	16	16	31	50	43

Notes: The table reports the Model Confidence Set (MCS) results at different forecasting horizons. The result for the market index reflects the $p-$value of a 90% MCS while the result for stocks reflects the frequency that a model being included in a 90% MCS among 50 stocks.

RV are highly correlated, and the conditional mean of RV exhibits the non-linear dependence because if the heteroskedastic measurement error is the only issue that needs to be addressed, the log-HAR model will be worse than the standard HARQ model in terms of forecasting ability, after the non-linear transformation is imposed on RV. Next, Figure 3.6 shows the log-HARQ model presents a very similar forecasting performance to the log-HAR model for 50 stocks where the dotted line of log-HAR and dot-dashed line of log-HARQ are almost identical. The same conclusion also can be reached in Table 3.3, where the log-HARQ model has the lowest QLIKE, MSE loss ratios, which are slightly lower than the loss ratios of the log-HAR model. Adding an extra quarticity variable to the log-HAR model will not significantly enhance the forecast accuracy, which means the log-transformation itself is enough to deal with most of the heteroskedastic measurement error. We also compare the daily forecast performance between the log-HARP and log-HAR models. From the Table 3.3, the log-HARP model has the highest QLIKE and MSE loss ratios among those logarithmic models. The QLIKE loss comparison in Figure 3.6 also graphically confirms the poor forecasting performance of log-HARP, which implies that after the log-transformation controls most of the heteroskedasticity, the non-linear dependence in the conditional mean of RV can also be captured via the log-transformation, adding extra variables with higher a exponent (e.g. quadratic term, cubic term, etc.) cannot help with the non-linear dependence.

The out-of-sample analysis to this point has focused on a 1-day-ahead forecasting horizon, and is now extended to longer weekly and monthly horizons. The forecasts at longer horizons are based on direct projection by replacing the daily RV on the left-hand side of the different models with the weekly and monthly RV. The loss ratios reported in Table 3.4 and Table 3.5 are very similar to those reported in Table 3.3. For example, the HARP and HARQ have a similar forecasting performance at longer horizons where both outperform the benchmark HAR model in terms of the QLIKE ratios. Also, the log-HARQ has the lowest loss ratios across all the models, and the log-HAR model is not far behind, reported similarly in Figure A.5 and A.8 in the Appendix A.3

The log-HARP model still works poorly at longer horizons, with adding an extra quadratic

term in the log-HAR not helping with the non-linear dependence. In the end, Table 3.6 presents the Model Confidence Set (MCS) results at different forecasting horizons. The market index results reflect the $p-$value of a 90% MCS while the result for stocks reflects the frequency of a model being included in a 90% MCS among the 50 stocks. The standard HAR, HARQ, and HARP models have a much lower p-value and are often excluded from the confidence set. Notably, the log-HARQ model has always been included in the MCS at all three forecast horizons. The MCS results suggest the logarithmic models have been included in MCS more frequently.

Overall, the out-of-sample results provide strong evidence that both the heteroskedastic measurement error and non-linear dependence in the conditional mean of RV must be considered for forecasting. The HARQ model is developed to deal with the heteroskedastic measurement error, but it also helps with the non-linearity in the conditional mean of RV. The empirical results suggest that the HARP model is a viable alternative to HARQ specification. More importantly, the log-transformed RV is able to deal with both issues simultaneously.

3.4 Conclusion

It is important to understand the dynamics of univariate realized volatility before investigating the full realized covariance matrix. The goal of this chapter is to reach a deeper understanding of why the HARQ model can provide improved RV forecasts. Given the empirical fact that sqrt-RQ is strongly correlated with RV, we first derived the HARP model which exhibits a comparable forecasting performance to the HARQ model. We then employed the natural logarithmic transformation to verify the speculation that the HARQ model might capture both the effect of heteroskedastic measurement error and the non-linear dependence in the conditional mean of RV. An out-of-sample study covering the S&P 500 index, the Dow Jones Industrial Average index, and 50 frequently traded U.S. market stocks found that both heteroskedastic measurement error and non-linear dependence in the conditional mean of RV are required to be considered before the forecasting modeling. Therefore, the success of the HARQ model can

be attributed to the actions of the dual role, which means while the HARQ model is developed for the heteroskedastic measurement error, it also helps deal with the non-linearity in the conditional mean of RV. Similarly, the HARP model captures the non-linearity via the squared RV but also handles the heteroskedasticity of measurement error. Since the RQ estimation is relatively imprecise in many realistic scenarios, the HARP model can be a viable alternative to the HARQ specification. More importantly, the empirical analysis suggests that the implementation of log-transformation on RV can deal with both heteroskedastic measurement error and non-linearity simultaneously.

All of the findings here are based on the univariate setting. The increased availability of high-frequency data has inspired the forecasting of realized covariance. It is valuable to extend the data transformation to a multivariate setting. However, adapting the log-transformation in the multivariate setting is challenging, due to the dimensionality issue and the positive definiteness requirement of the covariance matrix forecast. The extension will be explored in later studies.

Chapter 4

Lassoing the Realized Correlation Matrix

4.1 Overview

Multivariate volatility is typically quantified via the conditional covariance matrix of returns. Its forecasting has received extensive attention in the literature due to its importance in financial risk management, portfolio management, and asset pricing. In the early stages, the vast majority of the forecasts were based on the multivariate GARCH and multivariate stochastic volatility models (e.g., Bauwens et al. 2006, Asai et al. 2006), designed for the latent covariance matrix constructed from the low-frequency returns. The widespread availability of high-frequency financial data today enables the realized covariance to be used as a precise estimate of multivariate volatility. Many studies have shown the use of realized covariance may effectively harness the rich information inherent in the high-frequency data and generate reliable volatility forecasts. Chapter 3 focused on the diagonal element of the realized covariance. This chapter moves to the realized correlation matrix, which is decomposed from the realized covariance matrix. In particular, this chapter attempts to extend the work towards a multivariate setting and offers a new perspective to improve the realized covariance forecasting.

The empirical results of Chapter 3 have established that log-transformed RV can deal with heteroskedastic measurement error and non-linear dependence simultaneously. While conceptually straightforward, it is infeasible to apply the simple logarithmic transformation to

realized covariance directly due to the negative covariance elements. To incorporate the effect of measurement errors and time-varying attenuation biases into the covariance forecasts, Bollerslev et al. (2018) developed the multivariate HARQ model as an extension of the univariate HARQ model of Bollerslev et al. (2016). More importantly, based on the decomposition of the covariance matrix in equation (2.40), Bollerslev et al. (2018) proposed the HARQ-DCD[1] approach, which uses the univariate HARQ model to forecast the individual standard deviation elements and forecasting the correlation matrix with a scalar multivariate HAR (MHAR) model. The empirical results showed that the matrix decomposition approach, modeling the variances and correlation matrix separately, actually outperforms the multivariate HARQ model, improves the realized covariance forecasting, and obtains significant economic gains. While decomposing a conditional covariance matrix into a conditional correlation matrix pre- and post-multiplied by the diagonal matrix of conditional standard deviations, has been a popular approach to model and forecast the covariance matrix, the approach still has two main challenges. First, the covariance matrix properties require that models produce a positive (semi) definite correlation matrix forecast. Second, the multivariate autoregressive type models cannot deliver reliable estimates due to the curse of dimensionality. The scalar MHAR model from the HARQ-DCD ensures the positive definiteness of the correlation matrix forecast and avoids the curse of dimensionality by limiting the number of parameters.

The scalar MHAR model reduces the number of parameters, assuming the same parameter structure for all the elements in the correlation matrix. While it is highly parsimonious, it neglects the cross-effect between the matrix elements. The research undertaken in this chapter aimed to relax this restrictive scalar MHAR model and incorporate the effect of potential dependence between the elements of the correlation matrix.

In this study, all the out-of-sample realized covariance forecasts were obtained from the matrix decomposition approach. To incorporate the effect of measurement errors efficiently, the elements of standard deviation were predicted via the simple log-HAR model since the

[1] Bollerslev et al. (2018) used HARQ-DRD to denote this approach since R stands for the correlation matrix in their study. In order to apply the consistent terminologies for the entire thesis, we denote this approach as HARQ-DCD, where C stands for the correlation matrix.

log-HAR model was proven to outperform the HARQ model in Chapter 3. In terms of correlation matrix modeling, three models are considered: the scalar MHAR of Chiriac and Voev (2011), the generalized multivariate HAR (GHAR) model of Čech and Baruník (2017) and the original MHAR model. The OLS method is not suitable to estimate the original MHAR mode. Inspired by the study of Callot et al. (2017), the Least Absolute Shrinkage and Selection Operator (LASSO) technique was applied to estimate the original MHAR model. As a parameter selection device, LASSO solves the issue of dimensionality and investigates the optimal parameter structure of the correlation matrix modeling. The novel correlation matrix parametrization method of Archakov and Hansen (2021) was also implemented to ensure the positive definiteness of the correlation matrix forecast.

Our goal in this study is to illustrate the importance of the dependence within the correlation matrix elements and improve the realized covariance forecast from a correlation matrix modeling perspective. The contributions of this study are as follows. First, the empirical results showed that the cross-effect between the entries in the correlation matrix significantly impact the realized covariance forecast. The model that includes this relationship outperforms the restrictive scalar MHAR model. Next, building on the scalar MHAR model, we developed an easily implemented approach that incorporates the impact of plausible dependence between the correlation matrix elements, outperforming the standard scalar MHAR model. In the end, the empirical results showed that the correlation matrix parametrization method of Archakov and Hansen (2021) not only ensures the positive definiteness of the forecast but also outperforms other matrix re-parametrized methods, such as Cholesky decomposition and the matrix logarithmic function.

Apart from the HAR-DCD model of Oh and Patton (2016) and the HARQ-DCD model of Bollerslev et al. (2016), there are some papers related to this study. Dong and Tse (2020) forecast the realized covariance matrix via the matrix decomposition approach. The factor approach is applied for the large dimensional correlation matrix but the approach fails to guarantee the positive definiteness of the correlation matrix forecast. Archakov et al. (2020) applied the correlation matrix parametrization method of Archakov and Hansen (2021), and forecast

diagonal standard deviation matrix and correlation matrix via the GARCH framework with a block equicorrelation structure. However, studies related to multivariate volatility forecasting tend to model and forecast the covariance matrix directly without matrix decomposition. For example, Chiriac and Voev (2011) developed the scalar multivariate HAR model on Cholesky decomposed realized covariance matrix, and Bauer and Vorkink (2011) applied the matrix logarithmic function to transform the realized covariance matrix and then forecast via a principal component method.

The rest of the study is organized as follows. Section 4.2 describes the notations, the parametrization method for correlation matrix, the LASSO technique, and several dynamic models that were employed for this study. Section 4.3 discussed the dataset and the out-of-sample forecasting results. The robustness check of the linkage function is discussed in Section 4.4. Section 4.5 concludes.

4.2 Methodology

Based on the discussed in section 2.2, the integrated realized covariance of N-dimensional high-frequency financial assets on day t is defined as

$$\Sigma_t = \int_{t-1}^{t} \Sigma(\tau) d\tau. \tag{4.1}$$

As the integrated realized covariance is latent, Andersen et al. (2003) claimed it can be estimated approximately by the sum of products of intraday return vector, which is shown as

$$\hat{\Sigma}_t = \sum_{j=1}^{M} R_{t-1+j\Delta} R'_{t-1+j\Delta}, \tag{4.2}$$

where $M = 1/\Delta$ refers to the number of intraday returns over the trading day.

However, the realized covariance estimator in equation (4.2) might be biased and inconsistent due to the microstructure noise. In this study, the realized covariance estimator is based on

the multivariate realized kernel approach of Barndorff-Nielsen et al. (2011), defined in equation (2.14). The multivariate realized kernel estimator ensures consistency in the presence of general classes of microstructure noise and guarantees the estimate of the realized covariance matrix is positive semi-definite. To facilitate the notation in this study, S_t denotes any consistent realized covariance matrix estimator of the integrated covariance matrix.

A realized covariance matrix S_t can be decomposed into a realized correlation matrix C_t pre- and post-multiplied by a diagonal matrix of realized standard deviations D_t, such as

$$S_t = D_t \times C_t \times D_t. \tag{4.3}$$

For example, the decomposition of a 3×3 realized covariance matrix is

$$\begin{bmatrix} \sigma_1^2 & \sigma_{1,2} & \sigma_{1,3} \\ \sigma_{2,1} & \sigma_2^2 & \sigma_{2,3} \\ \sigma_{3,1} & \sigma_{3,2} & \sigma_3^2 \end{bmatrix} = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{bmatrix} \times \begin{bmatrix} 1 & \rho_{1,2} & \rho_{1,3} \\ \rho_{2,1} & 1 & \rho_{2,3} \\ \rho_{3,1} & \rho_{3,2} & 1 \end{bmatrix} \times \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{bmatrix}. \tag{4.4}$$

Having decomposed the realized covariance matrix into the correlation matrix and the standard deviation matrix, the next step was to forecast the correlation matrix and the standard deviation matrix separately. Usually, the individual standard deviation forecasts are from the univariate forecasting models, such as the univariate log-HAR model in this study. The realized correlation matrix forecasts to produce a positive (semi-)definite forecast and consider the issue of high-dimensionality. The following sections will introduce the approaches employed in this study that guarantee the positive (semi-)definiteness and curse of dimensionality.

4.2.1 Parametrization of the correlation matrix

It is convenient to reparametrize a covariance matrix before the modeling and forecasting, as parametrization offers a new route to build multivariate volatility models and ensures positive definiteness. These methods include the Cholesky decomposition and methods based on the spectral representation, such as the matrix logarithm function (e.g., Andersen et al. 2003, Chiriac and Voev 2011, Bauer and Vorkink 2011, Luo and Chen 2020). Recently, Archakov and Hansen (2021) proposed a method to parametrize a correlation matrix that ensures positive definiteness without imposing additional restrictions. Instead of applying the parametrization to the covariance matrix directly, Archakov and Hansen (2021) applied the matrix logarithm to the correlation matrix, where the lower off-diagonal elements of parametrized correlation matrix are stacked into the vector. Archakov and Hansen (2021) claimed the new parametrization method could be used in dynamic models of multivariate realized volatility.

Considering a non-singular $N \times N$ correlation matrix, C_t, the parametrization of Archakov and Hansen (2021) is based on the following mapping:

$$\boldsymbol{y}_t = \mathrm{g}(C_t) = \mathrm{vecl}\left(\mathrm{logm}\, C_t\right), \tag{4.5}$$

where $\mathrm{logm}\, C_t$ is the matrix logarithmic transformed correlation matrix and $\mathrm{vecl}(.)$ extracts and vectorizes the lower off-diagonal elements. To illustrate this parametrization, considering an arbitrary correlation matrix,

$$C_t = \begin{bmatrix} 1.0 & \bullet & \bullet \\ 0.5 & 1.0 & \bullet \\ 0.1 & 0.2 & 1.0 \end{bmatrix}, \tag{4.6}$$

and the unique vector of the parametrized correlation matrix is

$$\boldsymbol{y}_t = \text{vecl}\left(\text{logm}\begin{bmatrix} 1.0 & \bullet & \bullet \\ 0.5 & 1.0 & \bullet \\ 0.1 & 0.2 & 1.0 \end{bmatrix}\right) = \text{vecl}\begin{bmatrix} -0.14 & \bullet & \bullet \\ 0.55 & -0.16 & \bullet \\ 0.05 & 0.19 & -0.02 \end{bmatrix} = \begin{bmatrix} 0.55 \\ 0.05 \\ 0.19 \end{bmatrix}. \quad (4.7)$$

Any non-singular $N \times N$ correlation matrix can always be represented as a unique vector $\boldsymbol{y}_t$ in $\mathbb{R}^{n(n-1)/2}$. Archakov and Hansen (2021) proved that $\text{g}(C_t)$ is a one-to-one mapping between C_t and $\mathbb{R}^{n(n-1)/2}$. However, the inverse mapping, $C_t = \text{g}^{-1}(\boldsymbol{y}_t)$ is not given in closed form when $N > 2$ expect in some special cases. Instead, we follow the inverse mapping algorithm that provided by Archakov and Hansen (2021). The inverse mapping algorithm is based on the matrix exponential function and allows the reconstructed correlation matrix to be unique for all vector y_t. The software implementations of the inverse mapping are presented in the Online Supplemental Material of Archakov and Hansen (2021).

4.2.2 Dynamic models for the correlation matrix

This research aims to illustrate the importance of the cross-effect of the correlation matrix elements and improve the realized covariance forecast from a correlation matrix modeling perspective. All out-of-sample realized covariance forecasts are obtained from the matrix decomposition approach. In order to incorporate the effect of measurement errors efficiently, the elements of standard deviation are predicted via the simple log-HAR model since the log-HAR model has proven to outperform the HARQ model in Chapter 3. In terms of correlation matrix modeling, three models are considered. The scalar MHAR of Chiriac and Voev (2011), the generalized multivariate HAR (GHAR) model of Čech and Baruník (2017) and the original MHAR model. To facilitate the discussion, let C_t be the $N \times N$ realized correlation matrix at day t and $\boldsymbol{y}_t = \text{vecl}\,(\text{logm}\, C_t)$ be the $K = N(N-1)/2$ dimensional unique vector of the parametrized correlation matrix of Archakov and Hansen (2021), where $\boldsymbol{y}_t = [y_{1,t}, y_{2,t}, \ldots, y_{K,t}]'$.

The multivariate HAR model

As discussed in Chapter 2.3.2, Bauer and Vorkink (2011) was the first to propose a multivariate HAR model specification on realized covariance matrix, see equation (2.34). The model relies on a set of HAR regressors and macroeconomic variables which have been proved to improve volatility forecasting performance. The regressors of the multivariate HAR model include the variance-covariance elements from the lagged daily, weekly and monthly realized covariance matrices. In order to model the parametrized correlation matrix, the multivariate HAR model in this study exclude the macroeconomic/financial variables and includes only the lagged daily, weekly and monthly correlation matrix elements. The multivariate HAR model is defined as

$$\boldsymbol{y}_t = \boldsymbol{\varphi}_0 + \boldsymbol{\varphi}_1 \boldsymbol{y}_{t-1}^d + \boldsymbol{\varphi}_2 \boldsymbol{y}_{t-1}^w + \boldsymbol{\varphi}_3 \boldsymbol{y}_{t-1}^m + \boldsymbol{\varepsilon}_t, \tag{4.8}$$

where $\boldsymbol{\varphi}_0$ is a $K \times 1$ intercept vector, $\boldsymbol{\varphi}_1$, $\boldsymbol{\varphi}_2$ and $\boldsymbol{\varphi}_3$ are $K \times K$ parameter matrices for the lagged daily, weekly or monthly predictor vectors. The multivariate HAR model will result in an enormous number of predictors as the dimension increases. The traditional estimation methods like OLS fail to deliver reliable estimates due to the curse of dimensionality and large computational burden.

Expressing the multivariate HAR model in equation (4.8) in a matrix form will visually illustrate the dimensional issues. Let $\boldsymbol{y}_i = [y_{i,1}, y_{i,2}, \ldots, y_{i,T}]$, $i = \{1, 2, \ldots, K^*\}$, be a $1 \times T$ vector that represents all observations of the ith element throughout the time and then,

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_K \end{bmatrix} = \begin{bmatrix} y_{1,1} & y_{1,2} & \cdots & y_{1,T} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,T} \\ \vdots & \vdots & \vdots & \vdots \\ y_{K,1} & y_{K,2} & \cdots & y_{K,T} \end{bmatrix} \tag{4.9}$$

as a $K \times T$ matrix of the dependent variable.

Similarly, let $\boldsymbol{y}_i^d$, $\boldsymbol{y}_i^w$ and $\boldsymbol{y}_i^m$ be the lagged daily, weekly and monthly vectors of $\boldsymbol{y}_i$, and $\boldsymbol{x}_i = \left[\boldsymbol{y}_i^{d\prime}, \boldsymbol{y}_i^{w\prime}, \boldsymbol{y}_i^{m\prime}\right]'$ be the $3 \times T$ regressors matrix, which can then be stacked into

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_K \end{bmatrix} = \begin{bmatrix} y_{1,0}^d & y_{1,1}^d & \cdots & \cdots & y_{1,T-1}^d \\ y_{1,0}^w & y_{1,1}^w & \cdots & \cdots & y_{1,T-1}^w \\ y_{1,0}^m & y_{1,1}^m & \cdots & \cdots & y_{1,T-1}^m \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ y_{K,0}^d & y_{K,1}^d & \cdots & \cdots & y_{K,T-1}^d \\ y_{K,0}^w & y_{K,1}^w & \cdots & \cdots & y_{K,T-1}^w \\ y_{K,0}^m & y_{K,1}^m & \cdots & \cdots & y_{K,T-1}^m \end{bmatrix}. \tag{4.10}$$

The multivariate HAR model then can be expressed as

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_K \end{bmatrix} = \begin{bmatrix} \varphi_0^{(1)} & \varphi_{1,1}^{(1)} & \varphi_{2,1}^{(1)} & \varphi_{3,1}^{(1)} & \cdots & \varphi_{1,K}^{(1)} & \varphi_{2,K}^{(1)} & \varphi_{3,K}^{(1)} \\ \varphi_0^{(2)} & \varphi_{1,1}^{(2)} & \varphi_{2,1}^{(2)} & \varphi_{3,1}^{(2)} & \cdots & \varphi_{1,K}^{(2)} & \varphi_{2,K}^{(2)} & \varphi_{3,K}^{(2)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \varphi_0^{(K)} & \varphi_{1,1}^{(K)} & \varphi_{2,1}^{(K)} & \varphi_{3,1}^{(K)} & \cdots & \varphi_{1,K}^{(K)} & \varphi_{2,K}^{(K)} & \varphi_{3,K}^{(K)} \end{bmatrix} \times \begin{bmatrix} \boldsymbol{\iota} \\ \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_K \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_K \end{bmatrix}, \tag{4.11}$$

where φ_0^i represents the intercept coefficient of the HAR model for the ith element $\boldsymbol{y}_i$ and $\boldsymbol{\iota}$ is a $1 \times T$ vector of 1. Let $\boldsymbol{\Phi}^{(i)} = \left[\varphi_{1,1}^{(i)}, \varphi_{2,1}^{(i)}, \varphi_{3,1}^{(i)}, \ldots, \varphi_{1,K}^{(i)}, \varphi_{2,K}^{(i)}, \varphi_{3,K}^{(i)}\right]$ be a $1 \times 3K$ coefficient vector for the ith element $\boldsymbol{y}_i$, we have

$$\boldsymbol{y}_i = \varphi_0^{(i)} \times \boldsymbol{\iota} + \boldsymbol{\Phi}^{(i)} \times \boldsymbol{x} + \boldsymbol{\varepsilon}_i. \tag{4.12}$$

As we can see the number of parameters in the vector $\boldsymbol{\Phi}^{(i)}$ increases rapidly and may be very

large, even for a moderate dimension. The standard estimation techniques such as OLS may not provide the precise estimates. However, the LASSO could unveil the optimal parameter structure for the best forecasting performance, which will be introduced in the next section.

The scalar multivariate HAR model

The scalar version of the multivariate HAR model (Chiriac and Voev, 2011) is a simple approach to solve the dimensionality issue. While highly parsimonious, it can be estimated via the OLS and still ensure the positive definiteness of the matrix. Equation (2.30) shows the original scalar MHAR model of Chiriac and Voev (2011). In this study, to be consistent with the multivariate HAR model, we excluded the lagged biweekly variable and included only the lagged daily, weekly, monthly variables. The scalar MHAR model is defined as

$$\boldsymbol{y}_t = \boldsymbol{\varphi}_0 + \varphi_1 \boldsymbol{y}_{t-1}^d + \varphi_2 \boldsymbol{y}_{t-1}^w + \varphi_3 \boldsymbol{y}_{t-1}^m + \boldsymbol{\varepsilon}_t. \tag{4.13}$$

The intercept $\boldsymbol{\varphi}_0$ is a $K \times 1$ vector while the φ_1, φ_2 and φ_3 parameters are all assumed to be scalar, assuming the same structure for all the elements in $\boldsymbol{y}_t$.

Let $\boldsymbol{\varphi} = [\varphi_1, \varphi_2, \varphi_3]$ be a 1×3 vector that includes all the scalar parameters, the matrix form of the scalar MHAR model can be expressed as

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_K \end{bmatrix} = \begin{bmatrix} \varphi_0^{(1)} & \boldsymbol{\varphi} & 0 & \dots & 0 \\ \varphi_0^{(2)} & 0 & \boldsymbol{\varphi} & & \vdots \\ \vdots & \vdots & & \ddots & 0 \\ \varphi_0^{(K)} & 0 & \dots & 0 & \boldsymbol{\varphi} \end{bmatrix} \times \begin{bmatrix} \boldsymbol{\iota} \\ \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_K \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_K \end{bmatrix}. \tag{4.14}$$

For the ith element $\boldsymbol{y}_i$,

$$\boldsymbol{y}_i = \varphi_0^{(i)} \times \boldsymbol{\iota} + \boldsymbol{\varphi} \times \boldsymbol{x}_i + \boldsymbol{\varepsilon}_i. \tag{4.15}$$

Equation (4.15) above explains why the scalar MHAR is highly restrictive. It focuses only on the relationship between ith element $\boldsymbol{y}_i$ and its own lagged variables, ignoring the potential relationship between $\boldsymbol{y}_i$ and the lagged variables of other elements. Moreover, the scalar MHAR model assumes the same parameter $\boldsymbol{\varphi}$ for the elements of the correlation matrix, which is highly parsimonious.

The generalized multivariate HAR model

Building on the MHAR framework and the scalar MHAR model, Čech and Baruník (2017) proposed the generalized multivariate HAR (GHAR) model, which uses a system by assuming the dynamics of elements in the realized covariance matrix are only explained by their own lag variables. The GHAR model is defined as

$$\boldsymbol{y}_t = \boldsymbol{\varphi}_0 + \boldsymbol{\varphi}_1 \odot \boldsymbol{y}_{t-1}^d + \boldsymbol{\varphi}_2 \odot \boldsymbol{y}_{t-1}^w + \boldsymbol{\varphi}_3 \odot \boldsymbol{y}_{t-1}^m + \boldsymbol{\varepsilon}_t, \tag{4.16}$$

where $\boldsymbol{\varphi}_1$, $\boldsymbol{\varphi}_2$ and $\boldsymbol{\varphi}_3$ are the $K \times 1$ vectors of coefficient for the lagged daily, weekly and monthly variables. $\odot$ denotes the Hadamard product operator. The matrix form of GHAR model can be shown as

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_K \end{bmatrix} = \begin{bmatrix} \varphi_0^{(1)} & \boldsymbol{\varphi}^{(1)} & 0 & \dots & 0 \\ \varphi_0^{(2)} & 0 & \boldsymbol{\varphi}^{(2)} & & \vdots \\ \vdots & \vdots & & \ddots & 0 \\ \varphi_0^{(K)} & 0 & \dots & 0 & \boldsymbol{\varphi}^{(K)} \end{bmatrix} \times \begin{bmatrix} \boldsymbol{\iota} \\ \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_K \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_K \end{bmatrix}, \tag{4.17}$$

where $\boldsymbol{\varphi}^{(i)} = \left[\varphi_1^{(i)}, \varphi_2^{(i)}, \varphi_3^{(i)}\right]$ is a 1×3 vector that includes all the scalar parameters for the ith element $\boldsymbol{y}_i$. For the individual element above, the GHAR model can be rewritten as

$$\boldsymbol{y}_i = \varphi_0^{(i)} \times \boldsymbol{\iota} + \boldsymbol{\varphi}^{(i)} \times \boldsymbol{x}_i + \boldsymbol{\varepsilon}_i. \tag{4.18}$$

Although the GHAR model assumes the different parameter estimates for each element in the parametrized correlation matrix, the model still neglects the plausible dependence between the correlation matrix entries.

The MHAR model with the effect of general dependence

This study proposes a new model to incorporate the potential cross-effect between the correlation matrix elements. This new dynamic specification, named the MHAR-X model, is developed based on the scalar MHAR model. Instead of completely ignoring the relationship between the ith element and other elements in the correlation matrix, the MHAR-X model assumes a simple, fixed relationship existed between the correlation matrix elements $\boldsymbol{y}_i$ and $\boldsymbol{y}_j$, where $i \neq j$.

To illustrate the MHAR-X model, let $\sum \boldsymbol{y}_t = y_{1,t} + y_{2,t} + \ldots + y_{K,t}$ be the sum of all elements in $\boldsymbol{y}_t$ and

$$\underline{\boldsymbol{y}_t} = \begin{bmatrix} \sum \boldsymbol{y}_t - y_{1,t} \\ \sum \boldsymbol{y}_t - y_{2,t} \\ \vdots \\ \sum \boldsymbol{y}_t - y_{K,t} \end{bmatrix} = \begin{bmatrix} y_{2,t} + y_{3,t} + \ldots + y_{K,t} \\ y_{1,t} + y_{3,t} + \ldots + y_{K,t} \\ \vdots \\ y_{1,t} + y_{2,t} + \ldots + y_{K-1,t} \end{bmatrix} \tag{4.19}$$

be the vector that excludes the ith element on the ith row. Following the definition of $\underline{\boldsymbol{y}_t}$, we can then define the vectors $\underline{\boldsymbol{y}_t^d}$, $\underline{\boldsymbol{y}_t^w}$ and $\underline{\boldsymbol{y}_t^m}$ from $\boldsymbol{y}_t^d$, $\boldsymbol{y}_t^w$ and $\boldsymbol{y}_t^m$. The MHAR-X can be be written as

$$\boldsymbol{y}_t = \boldsymbol{\varphi}_0 + \varphi_1 \boldsymbol{y}_{t-1}^d + \varphi_2 \boldsymbol{y}_{t-1}^w + \varphi_3 \boldsymbol{y}_{t-1}^m + \theta_1 \underline{\boldsymbol{y}_{t-1}^d} + \theta_2 \underline{\boldsymbol{y}_{t-1}^w} + \theta_3 \underline{\boldsymbol{y}_{t-1}^m} + \boldsymbol{\varepsilon}_t, \tag{4.20}$$

where θ_1, θ_2 and θ_3 are the scalar coefficients that incorporate the cross-effect between the correlation matrix elements $\boldsymbol{y}_i$ and $\boldsymbol{y}_j$, where $i \neq j$. The matrix form of the MHAR-X can be

easily shown as

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_K \end{bmatrix} = \begin{bmatrix} \varphi_0^{(1)} & \boldsymbol{\varphi} & \boldsymbol{\theta} & \dots & \boldsymbol{\theta} \\ \varphi_0^{(2)} & \boldsymbol{\theta} & \boldsymbol{\varphi} & & \vdots \\ \vdots & \vdots & & \ddots & \boldsymbol{\theta} \\ \varphi_0^{(K)} & \boldsymbol{\theta} & \dots & \boldsymbol{\theta} & \boldsymbol{\varphi} \end{bmatrix} \times \begin{bmatrix} \boldsymbol{\iota} \\ \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_K \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_K \end{bmatrix}, \tag{4.21}$$

where $\boldsymbol{\theta} = [\theta_1, \theta_2, \theta_3]$ is a 1×3 coefficient vector. For the individual element $\boldsymbol{y}_i$, the MHAR-X model is

$$\boldsymbol{y}_i = \varphi_0^{(i)} \times \boldsymbol{\iota} + \boldsymbol{\varphi} \times \boldsymbol{x}_i + \boldsymbol{\theta} \times \underline{\boldsymbol{x}_i} + \boldsymbol{\varepsilon}_i, \tag{4.22}$$

where $\underline{\boldsymbol{x}_i} = \sum_{i=1}^{K} \boldsymbol{x}_i - \boldsymbol{x}_i$. The model can be estimated easily by OLS. Compared to the scalar MHAR model in equation (4.15) and the GHAR model in equation (4.18), the MHAR-X model incorporates the impact of dependence between the correlation matrix elements, by assuming a simple, fixed relationship θ between $\boldsymbol{y}_i$ and $\boldsymbol{y}_j$, where $i \neq j$. The empirical results of MHAR-X model are examined in the next section 4.3.

A combination of the scalar MHAR model and GHAR model

The final dynamic model investigated in this study combines the scalar MHAR model and the GHAR model, and simply expressed as

$$\boldsymbol{y}_t = \boldsymbol{\varphi}_0 + \varphi_1 \boldsymbol{y}_{t-1}^d + \varphi_2 \boldsymbol{y}_{t-1}^w + \varphi_3 \boldsymbol{y}_{t-1}^m + \boldsymbol{\varphi}_1 \odot \boldsymbol{y}_{t-1}^d + \boldsymbol{\varphi}_2 \odot \boldsymbol{y}_{t-1}^w + \boldsymbol{\varphi}_3 \odot \boldsymbol{y}_{t-1}^m + \boldsymbol{\varepsilon}_t \tag{4.23}$$

and the matrix form of this combined model is given as

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_K \end{bmatrix} = \begin{bmatrix} \varphi_0^{(1)} & \boldsymbol{\varphi} & 0 & \dots & 0 & \boldsymbol{\varphi}^{(1)} & 0 & \dots & 0 \\ \varphi_0^{(2)} & 0 & \boldsymbol{\varphi} & & \vdots & 0 & \boldsymbol{\varphi}^{(2)} & & \vdots \\ \vdots & \vdots & & \ddots & 0 & \vdots & & \ddots & 0 \\ \varphi_0^{(K)} & 0 & \dots & 0 & \boldsymbol{\varphi} & 0 & \dots & 0 & \boldsymbol{\varphi}^{(K)} \end{bmatrix} \times \begin{bmatrix} \boldsymbol{\iota} \\ \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_K \\ \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_K \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_K \end{bmatrix}. \tag{4.24}$$

For the individual element $\boldsymbol{y}_i$, the model is defined as

$$\boldsymbol{y}_i = \varphi_0^{(i)} \times \boldsymbol{\iota} + \boldsymbol{\varphi} \times \boldsymbol{x}_i + \boldsymbol{\varphi}^{(i)} \times \boldsymbol{x}_i + \boldsymbol{\varepsilon}_i. \tag{4.25}$$

From equation (4.25), the same regressor $\boldsymbol{x}_i$ has been included twice, which means the model can not be identified and the traditional OLS regression cannot be used here to estimate the model. However, the LASSO regression allows for parameter selection and model estimation, which will be discussed in the next section. In a nutshell, the Frisch-Waugh-Lovell theorem allows the LASSO regression to shrink the partial parameters $\boldsymbol{\varphi}^{(i)}$, in other words, shrink towards the scalar MHAR model. In fact, the Combined Model can be treated as another form of the GHAR model, a common structure represented by scalar MHAR and a deviation part represented by GHAR. This model, estimated via LASSO, is designed as a reference to examine the adequacy of the scalar MHAR model and the validity of the GHAR model. The details of partial shrinking is addressed in the next section.

4.2.3 The LASSO

The LASSO regression is used for the model estimation to circumvent the dimensionality issue in the multivariate HAR model. LASSO, the Least Absolute Shrinkage and Selection Operator, is a regularization technique introduced by Tibshirani (1996), used frequently in the field of computational statistics and machine learning. LASSO employs a highly efficient algorithm and performs both variable selection and regularization to enhance prediction accuracy. LASSO regression encourages simple, sparse models by shrinking the parameters to a certain value (usually zero). Recently, the LASSO technique has been applied in realized volatility forecasting (e.g., Audrino and Knaus 2016, Callot et al. 2017). A brief introduction of LASSO regression is given in this section.

For any vector $\boldsymbol{a} \in \mathbb{R}^m$, let $\|\boldsymbol{a}\|_{\ell_1} = \sum_{i=1}^m |a_i|$ and $\|\boldsymbol{a}\|_{\ell_2} = \sqrt{\sum_{i=1}^m a_i^2}$ denote the ℓ_1 and ℓ_2 norms, respectively. For the multivariate HAR model given in equation (4.12), the LASSO estimator $\widehat{\boldsymbol{\Phi}}^{(i)\,LASSO}$ is obtained by solving

$$\widehat{\boldsymbol{\Phi}}^{(i)\,LASSO} = \underset{\varphi_0^{(i)}, \boldsymbol{\Phi}^{(i)}}{\operatorname{argmin}} \quad \frac{1}{2T} \left\| \boldsymbol{y}_i - \varphi_0^{(i)} - \boldsymbol{\Phi}^{(i)} \boldsymbol{x} \right\|_{\ell_2}^2 + \lambda \left\| \boldsymbol{\Phi}^{(i)} \right\|_{\ell_1}, \tag{4.26}$$

where λ is a non-negative regularization parameter that is estimated either by cross-validation or using an information criterion. Equation (4.26) can be viewed as a least squares objective function with an extra term penalizing coefficients of $\boldsymbol{\Phi}^{(i)}$. The Lasso performs estimation and variable selection simultaneously. In this study, the LASSO regression is applied to each element $\boldsymbol{y}_i$ separately without penalizing the intercepts, since we assume that none of the intercepts in $\boldsymbol{\varphi}_0$ is sparse.

The LASSO regression discussed above penalizes all the coefficients in $\boldsymbol{\Phi}^{(i)}$. However, under some circumstances, it is required to penalize only part of the coefficients. To illustrate how the LASSO technique penalizes part of the coefficients, we use the combined model in equation (4.25) as an example. For the ith element $\boldsymbol{y}_i$, the matrix form of the combined model

can be defined as

$$\begin{bmatrix} y_{i,1} \\ y_{i,2} \\ \vdots \\ y_{i,T} \end{bmatrix} = \begin{bmatrix} 1 & x_{i,0} & x_{i,0} \\ 1 & x_{i,1} & x_{i,1} \\ \vdots & \vdots & \vdots \\ 1 & x_{i,T-1} & x_{i,T-1} \end{bmatrix} \times \begin{bmatrix} \varphi_0^{(i)} \\ \boldsymbol{\varphi} \\ \boldsymbol{\varphi}^{(i)} \end{bmatrix} + \begin{bmatrix} \varepsilon_{i,1} \\ \varepsilon_{i,2} \\ \vdots \\ \varepsilon_{i,T} \end{bmatrix}. \tag{4.27}$$

For simplicity, let

$$\boldsymbol{X}_1 = \begin{bmatrix} 1 & x_{i,0} \\ 1 & x_{i,1} \\ \vdots & \vdots \\ 1 & x_{i,T-1} \end{bmatrix} \quad \text{and} \quad \boldsymbol{X}_2 = \begin{bmatrix} x_{i,0} \\ x_{i,1} \\ \vdots \\ x_{i,T-1} \end{bmatrix} \tag{4.28}$$

be the two partitions of the explanatory variables and

$$\boldsymbol{\Phi}_1 = \begin{bmatrix} \varphi_0^{(i)} \\ \boldsymbol{\varphi} \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Phi}_2 = \left[\boldsymbol{\varphi}^{(i)}\right] \tag{4.29}$$

be the two partitions of the parameter matrix, where $\boldsymbol{\Phi}_1$ includes the intercept and the parameter of scalar MHAR model, $\boldsymbol{\Phi}_2$ contains the parameters that will be shrunk by LASSO. The combined model in equation (4.25) can be expressed as

$$\boldsymbol{y}_i' = \boldsymbol{X}_1\boldsymbol{\Phi}_1 + \boldsymbol{X}_2\boldsymbol{\Phi}_2. \tag{4.30}$$

Since the parameter $\boldsymbol{\Phi}_2$ is required to be penalized only under the LASSO regression. The

problem becomes

$$\widehat{\boldsymbol{\Phi}}_2^{LASSO} = \underset{\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2}{\operatorname{argmin}} \quad \frac{1}{2T} \left\| \boldsymbol{y}_i' - \boldsymbol{X}_1\boldsymbol{\Phi}_1 - \boldsymbol{X}_2\boldsymbol{\Phi}_2 \right\|_{\ell_2}^2 + \lambda \left\| \boldsymbol{\Phi}_2 \right\|_{\ell_1} . \tag{4.31}$$

Based on the Frisch-Waugh-Lovell theorem (Frisch and Waugh, 1933; Lovell, 1963), Tibshirani and Taylor (2011) pointed out the problem in equation (4.31) can be rewritten as

$$\widehat{\boldsymbol{\Phi}}_2^{LASSO} = \underset{\boldsymbol{\Phi}_2}{\operatorname{argmin}} \quad \frac{1}{2T} \left\| \boldsymbol{M}_{\boldsymbol{X}_1} \boldsymbol{y}_i' - \boldsymbol{M}_{\boldsymbol{X}_1} \boldsymbol{X}_2\boldsymbol{\Phi}_2 \right\|_{\ell_2}^2 + \lambda \left\| \boldsymbol{\Phi}_2 \right\|_{\ell_1} , \tag{4.32}$$

where $\boldsymbol{M}_{\boldsymbol{X}_1} = \boldsymbol{I}_T - \boldsymbol{X}_1 \left(\boldsymbol{X}_1' \boldsymbol{X}_1 \right)^{-1} \boldsymbol{X}_1'$ is an orthogonal projection matrix and $\boldsymbol{I}_T$ denotes an identity matrix of order T. Yamada (2017) demonstrated that the Frisch-Waugh-Lovell theorem remains valid for the LASSO penalized least-squares regression. After the estimated parameter $\widehat{\boldsymbol{\Phi}}_2^{LASSO}$ is obtained from the LASSO regression, the parameter $\boldsymbol{\Phi}_1$ can be obtained from $\widehat{\boldsymbol{\Phi}}_1 = \left(\boldsymbol{X}_1' \boldsymbol{X}_1 \right)^{-1} \boldsymbol{X}_1' \left(\boldsymbol{y}_i' - \boldsymbol{X}_2 \widehat{\boldsymbol{\Phi}}_2^{LASSO} \right)$.

4.3 Empirical Analysis

4.3.1 Data

Instead of employing all 50 stocks in Chapter 3, the sample data used here in this study considers a portfolio of five individual stocks only, to investigate the potential relationship within the correlation matrix. The five stocks are selected based on alphabetical order. The 5-minute prices of each stock were obtained from Thomson Reuters Tick History Database with the sample period starting 22 July 2002 and ending 27 June 2019, yielding 4193 daily realized covariance observations.

Table 4.1 presents the summary statistics of the realized covariance observations. Particularly, Table 4.1 reports the statistics for the variance-covariance entries of the realized covariance matrix, and the correlation entries after matrix decomposition. The average covariance and average correlation are calculated by averaging each stock's pair covariance and

Table 4.1: Summary statistics

Stock	Mean	SD	Skewness	Kurtosis	Max	Min
Variance						
AAPL	3.15×10^{-4}	0.0005	8.0784	105.7763	0.0093	7.06×10^{-6}
AXP	3.07×10^{-4}	0.0008	7.9955	102.2618	0.0157	8.66×10^{-6}
BA	2.21×10^{-4}	0.0003	7.0365	76.0317	0.0060	9.49×10^{-6}
CAT	2.70×10^{-4}	0.0005	8.3099	117.2925	0.0110	1.56×10^{-5}
CSCO	2.51×10^{-4}	0.0004	7.6915	92.6693	0.0074	1.34×10^{-5}
Average Covariance						
AAPL	1.08×10^{-4}	0.0003	9.9155	141.0272	0.0054	-1.16×10^{-4}
AXP	1.10×10^{-4}	0.0003	9.2518	128.2277	0.0059	-1.60×10^{-4}
BA	9.55×10^{-5}	0.0002	9.6299	136.1636	0.0052	-1.35×10^{-4}
CAT	1.11×10^{-4}	0.0003	9.7194	141.7130	0.0061	-1.10×10^{-4}
CSCO	1.06×10^{-4}	0.0003	10.0252	149.3395	0.0057	-1.23×10^{-4}
Average Correlation Coefficient						
AAPL	0.3353	0.2011	-0.2902	3.0186	0.8747	-0.4121
AXP	0.3361	0.2007	-0.2992	2.9515	0.8821	-0.4339
BA	0.3231	0.2044	-0.2459	2.9201	0.8790	-0.4111
CAT	0.3471	0.2057	-0.3185	2.9905	0.8891	-0.4230
CSCO	0.3422	0.2031	-0.3558	3.1414	0.8750	-0.4768

Notes: Summary statistics. Table reports descriptive statistics of the variance, average covariance, and average correlation. Average covariance and the average correlation coefficient are computed by averaging each stock's pair covariance and correlations.

correlation entries. 4.1 identifies the means of average covariance and average correlation as all positive, suggesting the positive relationship between stocks dominates the sample period. The average correlation among the assets is around 0.34, indicating the diversification benefits based on accurate covariance forecasts. The variance and covariance series present right-skew and leptokurtic properties, while the correlation series exhibit the opposite.

4.3.2 Out-of-sample analysis

The out-of-sample forecasting results in this study are based on a step-by-step approach. Following the matrix decomposition principle in equation (4.3), the first step is to decompose the

realized covariance matrix observations S_t into the diagonal realized standard deviations matrix D_t and the realized correlation matrix C_t. The next step is to forecast the diagonal entries of D_t with the univariate log-HAR model and obtain the h-step-ahead standard deviations matrix forecast $\widehat{D}_{t+h}$. The unique vector $\boldsymbol{y}_t$ of parametrized correlation matrix $\mathrm{g}(C_t)$ was forecast with a series of multivariate volatility models. Once obtaining the forecast $\widehat{\boldsymbol{y}}_{t+h}$, the inverse mapping algorithm (Archakov and Hansen, 2021) was used to receive the realized correlation forecast $\widehat{C}_{t+h} = \mathrm{g}^{-1}(\widehat{\boldsymbol{y}}_{t+h})$. In the end, the realized covariance forecast can be calculated via

$$\widehat{S}_{t+h} = \widehat{D}_{t+h} \times \widehat{C}_{t+h} \times \widehat{D}_{t+h}. \tag{4.33}$$

The principal benchmark to forecast the correlation matrix in this study is the scalar MHAR model, applied not only to the parametrized correlation matrix $\boldsymbol{y}_t$ but also the raw correlation matrix. In addition to the multivariate volatility forecasting models introduced in the previous section, the Random Walk forecast is also included as a benchmark. The scalar MHAR, GHAR and MHAR-X models were estimated using the method of OLS while the multivariate HAR and the combined models in equation (4.23) were estimated via the LASSO (MHAR-LASSO and Combined Model-LASSO). The GHAR model has also been estimated using the LASSO (GHAR-LASSO) for comparison purposes.

All the forecasts in this study from a rolling window forecasting scheme with a 1000 day rolling window. Both the correlation matrix and covariance matrix forecasts are evaluated using the two statistical loss functions. The first is the Frobenius norm, commonly used to measure the distance between two matrices. Let $\widehat{S}_{t+h}$ be the $N \times N$ realized covariance matrix forecast at h-day-ahead horizon and S_{t+h} be the ex-post realized covariance matrix, the Frobenius norm is defined as

$$L_{t+h}^{\text{Frobenius}} = \sqrt{\mathrm{Tr}\left[\left(\widehat{S}_{t+h} - S_{t+h}\right)\left(\widehat{S}_{t+h} - S_{t+h}\right)'\right]}. \tag{4.34}$$

Another statistical criterion employed to evaluate the forecast performances is the Stein loss (James and Stein, 1992), which is treated as a multivariate extension of the traditional QLIKE

loss function. The Stein function is

$$L_{t+h}^{\text{Stein}} = \text{Tr}\left(\widehat{S}_{t+h}^{-1} S_{t+h}\right) - \log\left|\widehat{S}_{t+h}^{-1} S_{t+h}\right| - N. \tag{4.35}$$

The forecast losses are recorded in Table 4.2, including results from the short-horizon (1-day-ahead) and the longer horizons (5-day-ahead, 22-day-ahead). The correlation matrix forecasts of the longer horizons are based on the "direct" forecast approach, which replaces the 1-day-ahead correlation matrix with the correlation matrix over the relevant horizon of interest. It has been documented that, compared to the traditional "iterated" forecast approach, the "direct" forecast procedure often worked better in practice if the model is misspecified (see, e.g., Marcellino et al. 2006, Sizova 2011, Bollerslev et al. 2018). To highlight the best set of performing models, the model confidence set (MCS) test (Hansen et al., 2003, 2005) was conducted on the forecast losses. The MCS contained a number of models of equal predictive ability. From the Table 4.2, the value in boldface indicates that the model is included in the 90% MCS.

This section aims to study whether the plausible dependence of the correlation matrix entries will positively impact the forecasting performance. We start by inspecting the number of non-zero parameters for the MHAR model in Figure 4.1. Since the OLS is unsuitable for estimating the MHAR model, the LASSO regression is applied to the MHAR model to conduct the variable selection and reveal the potential optimal parameter structure for the MHAR model. Since the number of parameters of the MHAR model increases rapidly when the dimension increases then theoretically, for a K dimensional vector of parametrized correlation matrix $\boldsymbol{y}$, there will be $3K$ parameters that need to be estimated for each equation of $\boldsymbol{y}_i$. For instance, the empirical analysis in this study is based on the 5×5 realized covariance matrix series, where the parametrized correlation matrix has $K = 10$ unique elements in the vector and $3K = 30$ parameters for each equation. The LASSO can effectively shrink the parameter size and keep only the essential ones. Figure 4.1 presents the average number of non-zero parameters for the MHAR-LASSO model at different forecasting horizons over the whole out-of-sample period. The average parameter size is computed from 10 individual equations of $\boldsymbol{y}$. Figure 4.1 demonstrates that regardless of

Table 4.2: Results of out-of-sample forecasts

Model	1-day-ahead		5-day-ahead		22-day-ahead	
	Cov.Mat.	Corr.Mat.	Cov.Mat.	Corr.Mat.	Cov.Mat.	Corr.Mat.
Frobenius Norm						
Random Walk	4.899	0.998	3.944	0.558	4.184	0.422
Scalar MHAR (raw)	**4.613**	0.787	**3.830**	0.472	**4.090**	**0.364**
Scalar MHAR	4.634	0.790	3.853	0.479	4.121	0.373
MHAR-X	4.633	**0.771**	**3.845**	**0.461**	**4.115**	**0.362**
GHAR	4.683	0.852	3.902	0.552	**4.124**	0.438
Combined Model-LASSO	4.636	0.791	3.856	0.480	4.124	0.372
GHAR-LASSO	4.637	0.792	3.856	0.480	4.124	0.372
MHAR-LASSO	4.635	0.778	**3.849**	**0.460**	**4.109**	**0.355**
Stein Loss Function						
Random Walk	1.646	0.845	0.813	0.229	0.716	0.111
Scalar MHAR (raw)	1.264	0.451	0.735	0.147	0.672	**0.076**
Scalar MHAR	1.265	0.452	0.734	0.148	**0.665**	0.078
MHAR-X	**1.254**	**0.440**	**0.726**	**0.139**	**0.668**	**0.075**
GHAR	1.322	0.506	0.789	0.194	0.720	0.114
Combined Model-LASSO	1.266	0.452	0.734	0.148	**0.664**	0.078
GHAR-LASSO	1.267	0.454	0.734	0.148	**0.664**	0.078
MHAR-LASSO	1.257	0.445	**0.723**	**0.139**	**0.666**	**0.071**

Notes: Table reports out-of-sample forecast losses (realized covariance and realized correlation matrix) with various multivariate volatility models based on the Frobenius Norm and the Stein loss functions. The Scalar MHAR (raw) represents the Scalar MHAR model with raw data. Table includes out-of-sample losses at different forecast horizons. All Frobenius norm losses of the covariance matrix are scaled up by 10^4. Results in boldface indicates that the model is included in the 90% MCS.

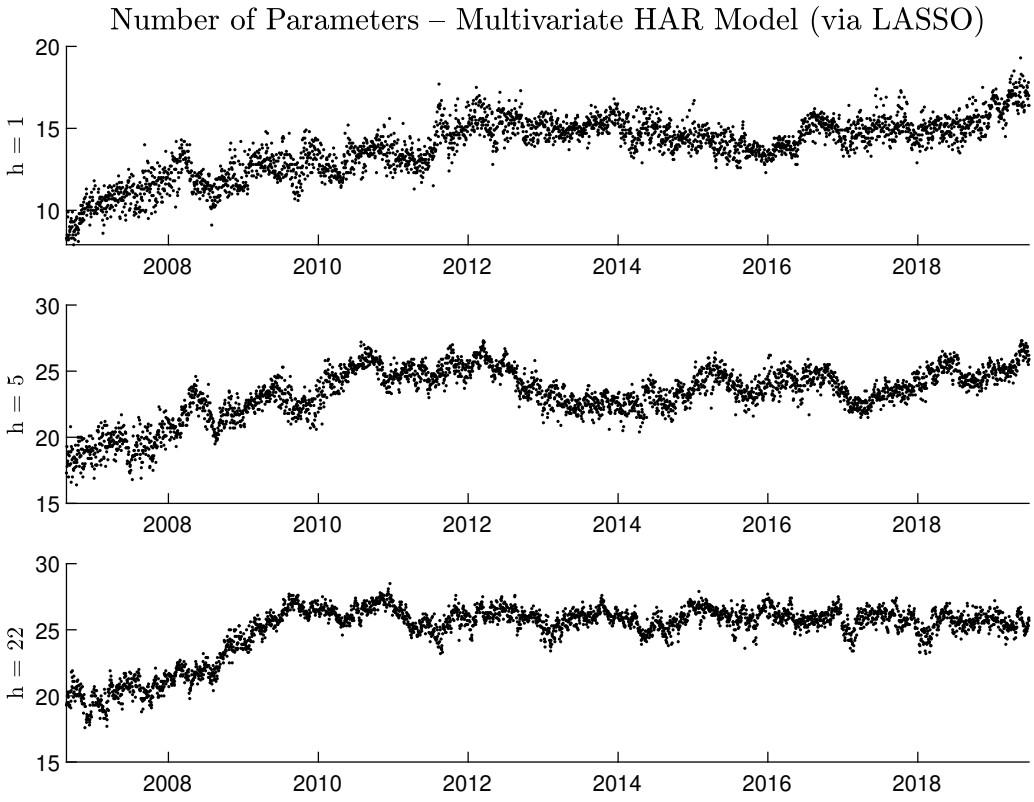

Figure 4.1: Average parameter size of the multivariate HAR model (estimated via LASSO) at 1-day-ahead, 5-day-ahead, and 22-day-ahead forecast horizons. Plots show average number of parameters of the multivariate HAR model during out-of-sample period, where the model is estimated by the LASSO regression.

the forecasting horizons, the number of non-zero parameters for the MHAR-LASSO model is well beyond the parameter number of those parsimoniously scalar MHAR and GHAR models. Table 4.2 also shows the MHAR-LASSO model, which incorporates the effect of correlation between the elements, outperforms the scalar MHAR and GHAR models for both loss functions and all forecasting horizons. The empirical results presented in Figure 4.1 and Table 4.2 suggest the scalar MHAR and GHAR models, which restrict the element $\boldsymbol{y}_i$ to be explained by its own lagged variables with only three parameters, should be far from ideal. It is important to include the dependence impact of the correlation matrix elements to improve the forecast accuracy.

Furthermore, the comparison between the MHAR-LASSO and GHAR-LASSO in Table 4.2 also indicates the importance of potential dependence between the matrix entries. Based

on equation (4.17), the GHAR-LASSO model of $\boldsymbol{y}_i$ can be viewed as all the parameters are shrunk to 0 except for the parameters of its own lagged variables. Suppose cross-effect between the correlation matrix elements have a negligible effect. In that case, LASSO will shrink those parameters of MHAR to 0, and the forecasting performance of the MHAR-LASSO will be similar to the GHAR-LASSO. As shown in Table 4.2, the GHAR-LASSO works poorly compared to the MHAR-LASSO in the correlation matrix forecasting at all the forecasting horizons.

The empirical results of the MHAR-LASSO model encourage us to explore a richer parameter structure for the MHAR model, with the MHAR-X model then proposed to address the potential dependence between the entries of the correlation matrix. Although the parameter structure of MHAR-X is somehow as restrictive as the scalar MHAR model, Table 4.2 shows the MHAR-X model still provides the best forecasting performance for the correlation matrix, followed by the MHAR-LASSO model. Besides, the MHAR-X model is included in the MCS with 90% confidence at all forecasting horizons while the MHAR-LASSO only in the 90% MCS for the longer horizons. These results are not surprising since both models include the variables that reflect these dependencies between the matrix entries.

Table 4.2 shows the GHAR model is less capable of forecasting performance than the scalar MHAR model, an unexpected result since the GHAR model of Čech and Baruník (2017) is designed to relax the restrictiveness of the scalar MHAR model, it is supposed to outperform, at least, the scalar MHAR model. To investigate why the GHAR model exhibits the worst performance, we proposed a model, defined as equation (4.23), that combines the scalar MHAR and the GHAR models, and used the LASSO to shrink the partial parameters. More specifically, we applied LASSO on the GHAR part only and shrunk the whole combined model towards the scalar MHAR model. In fact, the Combined Model can be treated as another form of the GHAR model, a common structure represented by scalar MHAR and a deviation part that GHAR represents. This Combined Model-LASSO can be used as a reference model to examine the scalar MHAR model's adequacy and the GHAR model's validity. Table 4.2 shows the Combined Model-LASSO and the scalar MHAR models have similar forecast losses at

all forecast horizons. Based on the Stein Loss function, the Combined Model-LASSO has the same forecast loss as the scalar MHAR model. This suggests that, without including the variables that reflect the cross-effect, increasing the parameter flexibility like the GHAR model cannot contribute to the forecasting performance. The common parameter structure like the scalar MHAR is enough.

The last finding in this chapter is the standard deviations matrix forecast $\widehat{D}_{t+h}$ might significantly impact the forecast of the covariance matrix. Table 4.2 reports the forecast losses and MCS results for both covariance matrix and correlation matrix. However, even if all standard deviations matrix forecasts are from the univariate log-HAR model, there is still an MCS inconsistency between the correlation matrix forecast and the covariance matrix forecast. For example, the Scalar MHAR model on raw data is excluded from the 90% MCS for correlation matrix forecasting but included in the MCS for covariance matrix forecasting. Also, for the 22-day-ahead forecasting, more models are included in the MCS for covariance matrix forecasting. The empirical results indicate that the RV forecast is vital to covariance matrix forecasting if the matrix decomposition approach is adopted.

4.4 Robustness Check

The out-of-sample results reported in the previous section are based on the correlation matrix parametrization method of Archakov and Hansen (2021), which offers a new way to structure multivariate volatility models without imposing additional restrictions and guarantees the positive definiteness of the covariance matrix forecast. This section presents the robustness check for the parametrization method of Archakov and Hansen (2021), the forecast performance will be compared to other parametrization methods previously suggested in the literature. This section considers two methods: the matrix logarithmic function and the Cholesky decomposition, are both applied to the correlation matrix in this study, even though it is more common to reparametrize the covariance matrix directly.

Table 4.3 reports the results from the matrix logarithmic function transformation. Compared

Table 4.3: Out-of-sample results – Matrix logarithm function

Model	1-day-ahead		5-day-ahead		22-day-ahead	
	Cov.Mat.	Corr.Mat.	Cov.Mat.	Corr.Mat.	Cov.Mat.	Corr.Mat.
Frobenius Norm						
Random Walk	4.899	0.998	3.899	0.555	4.113	0.421
Scalar MHAR (raw)	4.613	0.787	3.830	**0.472**	4.090	**0.364**
Scalar MHAR	4.584	0.790	**3.794**	0.480	**4.038**	0.375
MHAR-X	**4.582**	0.789	**3.793**	0.478	**4.037**	0.375
GHAR	4.592	0.805	**3.796**	0.493	**4.041**	0.382
Combined Model-LASSO	4.589	0.803	**3.795**	0.493	**4.036**	0.383
GHAR-LASSO	4.594	0.807	**3.797**	0.493	**4.037**	0.383
MHAR-LASSO	4.591	**0.781**	**3.828**	**0.467**	**4.065**	**0.357**
Stein Loss Function						
Random Walk	1.646	0.845	0.821	0.238	0.725	0.119
Scalar MHAR (raw)	**1.264**	**0.451**	**0.735**	**0.147**	**0.672**	**0.076**
Scalar MHAR	1.272	0.457	0.739	0.153	**0.672**	0.084
MHAR-X	1.271	0.456	**0.738**	0.153	**0.673**	0.084
GHAR	1.274	0.458	0.741	0.154	**0.672**	0.084
Combined Model-LASSO	1.273	0.457	0.741	0.154	**0.671**	0.084
GHAR-LASSO	1.274	0.458	0.741	0.154	**0.672**	0.084
MHAR-LASSO	**1.265**	**0.450**	**0.735**	**0.148**	**0.687**	**0.078**

Notes: Table reports out-of-sample forecast losses (realized covariance and realized correlation matrix) with various multivariate volatility models based on the Frobenius Norm and the Stein loss functions. The **matrix logarithm function** is applied to the correlation matrix before modeling. Table includes out-of-sample losses at different forecast horizons. All Frobenius norm losses of the covariance matrix are scaled up by 10^4. Result in boldface indicates the model is included in the 90% MCS.

Table 4.4: Out-of-sample results – Cholesky decomposition

Model	1-day-ahead		5-day-ahead		22-day-ahead	
	Cov.Mat.	Corr.Mat.	Cov.Mat.	Corr.Mat.	Cov.Mat.	Corr.Mat.
Frobenius Norm						
Random Walk	4.899	0.998	3.877	0.565	4.062	0.445
Scalar MHAR (raw)	4.613	**0.787**	3.830	**0.472**	4.090	**0.364**
Scalar MHAR	**4.569**	0.816	**3.779**	0.515	**4.000**	0.420
MHAR-X	**4.565**	0.811	**3.777**	0.512	**4.004**	0.421
GHAR	**4.568**	0.816	**3.776**	0.514	**4.001**	0.420
Combined Model-LASSO	**4.569**	0.816	**3.779**	0.514	**4.001**	0.420
GHAR-LASSO	**4.568**	0.816	**3.777**	0.514	**4.001**	0.421
MHAR-LASSO	**4.581**	0.802	**3.783**	0.495	**4.014**	0.400
Stein Loss Function						
Random Walk	1.646	0.845	0.834	0.250	0.739	0.132
Scalar MHAR (raw)	**1.264**	**0.451**	**0.735**	**0.147**	**0.672**	**0.076**
Scalar MHAR	1.291	0.472	0.758	0.169	0.691	0.100
MHAR-X	1.291	0.474	0.758	0.171	0.696	0.101
GHAR	1.288	0.470	0.756	0.167	0.695	0.100
Combined Model-LASSO	1.288	0.471	0.756	0.167	0.695	0.100
GHAR-LASSO	1.288	0.471	0.756	0.167	0.695	0.100
MHAR-LASSO	1.283	0.464	0.751	0.159	0.702	0.095

Notes: Table reports out-of-sample forecast losses (realized covariance and realized correlation matrix) with various multivariate volatility models based on the Frobenius Norm and the Stein loss functions. The **Cholesky decomposition** is adopted on the correlation matrix before the modeling. Table includes out-of-sample losses at different forecast horizons. All Frobenius norm losses of the covariance matrix are scaled up by 10^4. Result in boldface indicates the model is included in the 90% MCS.

to the results in Table 4.2 based on the parametrization method of Archakov and Hansen (2021), it is clear this parametrization method of Archakov and Hansen (2021) performs better in terms of correlation matrix forecasting. Treating the scalar MHAR (raw) as a reference model, the correlation matrix forecasts in Table 4.2 have lower losses. Table 4.3 indicates the MHAR-LASSO is the best model for correlation matrix forecasting. The results from the Cholesky decomposition in Table 4.4 shows a similar picture to Table 4.3. The parametrization method of Archakov and Hansen (2021) also outperforms the Cholesky decomposition in correlation matrix forecasting. Surprisingly, the scalar MHAR model with raw data beats all other models with the Cholesky decomposed correlation matrix. Overall, the empirical results suggest the parametrization method of Archakov and Hansen (2021) is the optimal option if the matrix decomposition approach is adopted, as it ensures the positive definiteness of the forecast with less computational cost.

Moreover, the robustness check results in this section also prove that the standard deviations matrix forecast $\widehat{D}_{t+h}$ might significantly impact the forecast of the covariance matrix. Both Table 4.3 and Table 4.4 show the MCS inconsistency, particularly under the Frobenius Norm section.

4.5 Conclusion

Matrix decomposition has been a popular approach to model and forecast the covariance matrix as Bollerslev et al. (2018) recently proposed the HARQ-DCD approach to incorporate the effect of measurement errors into the covariance forecasts. However, the parsimonious scalar MHAR model of the HARQ-DCD neglects the possible dependence between the matrix elements, somehow becoming the main limitation of the HARQ-DCD approach.

In this study, we demonstrated the importance of the cross-effect between the elements of correlation matrix and improving the realized covariance forecast from a correlation matrix modeling perspective. By applying the the LASSO on the MHAR model to reduce the dimensionality and select the variable, the number of parameters for MHAR-LASSO exceeded

the number of parameters for scalar MHAR, and the MHAR-LASSO outperformed the scalar MHAR in correlation matrix forecasting. This suggested a richer parameter structure for the MHAR model is needed. By incorporating the impact of dependence between the correlation matrix elements, the MHAR-X provides the best forecasting performance for the correlation matrix in an out-of-sample study, followed by the MHAR-LASSO. Overall, the empirical results imply that the potential dependence between the entries of the correlation matrix significantly impacts the covariance forecast, if the covariance forecast is obtained from a matrix decomposition approach. In addition to this, the study also presents empirical evidence that the correlation matrix parametrization method of Archakov and Hansen (2021) outperforms other re-parametrized methods for the correlation matrix, such as Cholesky decomposition and the matrix logarithmic function.

The increased availability of high-frequency data has inspired the forecasting of realized covariance. One possible direction for future research is to validate those multivariate volatility forecasts via a financial application, and we will explore those applications in later studies.

Chapter 5

Forecasting Realized GMVP Weight

5.1 Overview

The studies in Chapter 3 and Chapter 4 mainly focused on forecasting the elements of the realized covariance matrix. In this chapter, instead of building a new methodology to generate the volatility forecast, the study moves to a financial application that requires the forecast of the realized covariance matrix, aiming to provide a comprehensive analysis of forecasting portfolio weight and validate the previous findings within the context of portfolio selection.

The mean-variance portfolio has attracted significant attention from both academics and practitioners since the pioneering work of Markowitz (1952b). To find the optimal portfolio composition, investors are required to estimate and forecast both the mean vector and the covariance matrix of the assets returns. Unfortunately, it is difficult to obtain reliable forecasts of expected asset returns, using an inaccurate estimator of returns could lead the optimal portfolio weight to unreasonable positions (Jagannathan and Ma, 2003). For this reason, the global minimum variance portfolio (GMVP) weight is preferable in practical portfolio selection.

The global minimum variance portfolio is the origin of the Markowitz mean-variance efficient frontier. The weight of GMVP only depends on the covariance matrix, which can be estimated and forecast for short and medium investment horizons. Nowadays, the widespread

availability of high-frequency data allows us to compute precise and consistent ex-post estimates of the realized covariance matrix (Barndorff-Nielsen and Shephard, 2004a), which can then be used for computing the GMVP weight. The GMVP weight from the realized covariance matrix is labeled as *realized GMVP weight*. There are two approaches to forecast the realized GMVP weight. The conventional approach which is based on the realized covariance matrix forecasts obtained from volatility models. The realized GMVP weight forecast is then calculated using the covariance matrix forecast in the GMVP formula. The conventional approach has contributed numerous methods to forecast the realized covariance matrix, and some of them have been provided exceptional forecasting performance. The alternative approach is to forecast the ex-post realized GMVP weight directly, with those same multivariate volatility models. The empirical evidence suggests the realized GMVP weight has the same long memory feature as realized covariance (e.g. Golosnoy et al. 2020,Golosnoy and Gribisch 2022). The availability of realized GMVP weight measures opens new possibilities to directly predict the GMVP composition with those forecasting models of the realized covariance matrix. Even though forecasting GMVP weight has attracted attention from both academics and practitioners, there is no clear comparison of which approach is more advantageous.

This chapter, therefore provides a comprehensive analysis of forecasting realized GMVP weight, and aims to validate the findings in Chapter 4 within the context of portfolio selection. To compare the forecasting performance between two approaches, we first estimated the quarticity of realized GMVP weight based on the asymptotic distribution developed by Golosnoy et al. (2020). The quarticity of GMVP weight enables modeling the effect of the heteroskedastic measurement error. Following the study of Bollerslev et al. (2018), we adopted the scalar multivariate HARQ (scalar-MHARQ) model and proposed a generalized multivariate HARQ (GHARQ) for realized GMVP weight forecasting. Next, following the study in Chapter 4, we amended the MHAR-X model in equation (4.20) to incorporate the potential dependence between the GMVP weight series. In the end, the study provided a comparison to show the different forecasting performances between two approaches. The forecasting performance is evaluated based on the statistical and economic evaluation criterion, the latter derived from the

Markowitz portfolio optimization theory.

The empirical analysis involves ten portfolios of five stocks, from 50 stocks described in the Chapter 3. Based on statistical loss, we show that the direct forecasting approach outperforms the conventional approach in the sense of providing realized GMVP weight forecasts that are significantly closer to the ex-post realized weight. The empirical result also indicates that, compared to the conventional approach, the direct forecasting approach generally provides superior portfolio allocation performance, reduces the portfolio turnover substantially, and effectively improves the economic gains. We further document that incorporating the potential dependence between the portfolio weight enhances the economic performance of the portfolio significantly. The MHAR-X of Chapter 4 outperforms a series of models within the direct forecasting approach.

There is a vast econometrics literature focused on the use of high-frequency data for realized GMVP weight forecasting. Most studies adopt the conventional approach, seeking to develop a novel method for the realized covariance matrix forecasting (e.g., Callot et al. 2017, Chiriac and Voev 2011, De Nard et al. 2021, Golosnoy et al. 2012, Luo and Chen 2020, Noureldin et al. 2012). However, the conventional approach is subject to the curse of dimensionality and it is challenging to guarantee the positive definiteness for the realized covariance matrix forecasts. Instead, forecasting the realized GMVP weight directly alleviates the issue of dimensionality and gives greater flexibility without needing to guarantee the positive definiteness of the forecast. For example, Golosnoy et al. (2019) used the exponential smoothing approach to predict the realized GMVP weight directly, and Golosnoy and Gribisch (2022) proposed a GLS-type estimator to directly minimize the loss of the variance of GMVP forecasts. It should be emphasized that the purpose of this study is not to replicate all the existing methodologies and apply them to the direct forecasting approach. Instead, we only focus on the popular multivariate HAR-type models and investigate how to make the most out of them on direct realized GMVP weight forecasting.

The rest of the chapter is organized as follows: Section 5.2 describes the econometric specifications, and Section 5.3 reports the results of its empirical study. Section 5.4 concludes.

Additional results to complement the main paper are contained in the Appendix B.

5.2 Methodology

This section introduces the realized GMVP weight and the notations, then presents the approach for directly forecasting the realized GMVP weight. The statistical evaluation and economic evaluation criteria are then explained.

5.2.1 The realized GMVP weight

Let Σ_t be a N-dimensional integrated realized covariance matrix and S_t be any consistent realized covariance matrix estimator of the integrated covariance matrix. Barndorff-Nielsen and Shephard (2004a) showed the S_t is a consistent estimator of the Σ_t if the number of intraday returns $M \to \infty$. According to Barndorff-Nielsen et al. (2011), the asymptotic theory of S_t is

$$M^{1/5} \left(\operatorname{vech} S_t - \operatorname{vech} \Sigma_t \right) \to \mathcal{N} \left(\delta_t, \Pi_t \right), \tag{5.1}$$

where $\operatorname{vech} \Sigma_t$ denotes the $N^* = N(N+1)/2$ dimensional vectorized version of the integrated covariance matrix and $\operatorname{vech} S_t$ denotes the vectorized realized covariance matrix. δ_t is the negligible bias term. Π_t refers to the $N^* \times N^*$ covariance matrix for the corresponding measurement error vector $(\operatorname{vech} S_t - \operatorname{vech} \Sigma_t)$.

The optimal portfolio composition helps investors decide how much to invest in each risky asset. The Markowitz-type portfolio relies heavily on the forecasts of the mean vector and the covariance matrix. Using an inaccurate estimator of returns could lead to an unreasonable portfolio weight in terms of investment. For this reason, the global minimum variance portfolio (GMVP) weight appears to be preferable in a practical portfolio selection. The GMVP plays a prominent role since it is the starting point of the Markowitz mean-variance efficient frontier. The GMVP is based on the optimization problem similar to the Markowitz setup in equation

(2.67),

$$\begin{aligned} \min_w \quad & w_t' \Sigma_t w_t \\ \text{s.t.} \quad & w_t' \iota = 1, \end{aligned} \tag{5.2}$$

where ι is the N dimensional vector of ones. The GMVP weight can be solved analytically,

$$w_t = \frac{\Sigma_t^{-1} \cdot \iota}{\iota' \cdot \Sigma_t^{-1} \cdot \iota}. \tag{5.3}$$

Given S_t is the consistent estimator of Σ_t, the $N \times 1$ realized GMVP weight vector is defined as

$$\mathrm{w}_t = \frac{S_t^{-1} \cdot \iota}{\iota' \cdot S_t^{-1} \cdot \iota}. \tag{5.4}$$

Golosnoy et al. (2020) derive the asymptotic distribution of realized GMVP weight and show the realized GMVP weight could be seen as the consistent estimator of the true GMVP proportions for the current day. For $M \to \infty$,

$$M^{1/2} \left(\mathrm{w}_t - w_t \right) \to \mathcal{N} \left(0, \Omega_t \right), \tag{5.5}$$

where

$$\Omega_t = G_t' \Pi_t G_t \tag{5.6}$$

is the $N \times N$ covariance matrix for the corresponding measurement error $\mathrm{w}_t - w_t$ and Π_t refers to the $N^* = N(N+1)/2$ dimensional covariance matrix for the measurement error of realized covariance matrix. The gradient matrix G_t is defined as

$$G_t = \frac{\partial w_t}{\partial \operatorname{vech} \left(\Sigma_t \right)} \tag{5.7}$$

and

$$G_t' = \frac{1}{2 \left(\iota' \Sigma_t^{-1} \iota \right)^2} \left\{ \Sigma_t^{-1} \iota \left(\iota' \otimes \iota' \right) - \left(\iota' \Sigma_t^{-1} \iota \right) \left(\iota' \otimes I \right) \right\} \left(I_{N^2} + K \right) \left(\Sigma_t^{-1} \otimes \Sigma_t^{-1} \right) D, \tag{5.8}$$

where D is a unique $N^2 \times N^*$ duplication matrix and K is a $(N^2 \times N^2)$ commutation matrix. The detailed derivation can be found in Appendix B.1. Following Bollerslev et al. (2018), the quarticity of realized GMVP weight w_t can be estimated from $G_t' \Pi_t G_t$.

5.2.2 Dynamic models

This study considers two approaches to forecast the realized GMVP weight. The conventional approach is based on the realized covariance matrix predictions which are used to compute the GMVP weight. Another approach is based on the ex-post realized GMVP weight which are forecast directly.

As for the direct forecasting approach, we implemented those widely-used multivariate volatility models on the realized GMVP weight forecasting. All out-of-sample realized GMVP weight forecasts are based on the original and untransformed data without the concern of positive definiteness. Let S_t be the realized covariance estimator, the direct forecasting approach for GMVP weight forecast $\widehat{\mathrm{w}}_{t+h}$ is obtained via the following process:

$$S_t \longrightarrow \mathrm{w}_t \longrightarrow \widehat{\mathrm{w}}_{t+h}. \tag{5.9}$$

In particular, we consider the fully parameterized multivariate HAR and multivariate HARQ models, defined as

$$\mathrm{w}_t = \boldsymbol{\varphi}_0 + \boldsymbol{\varphi}_1 \mathrm{w}_{t-1}^d + \boldsymbol{\varphi}_2 \mathrm{w}_{t-1}^w + \boldsymbol{\varphi}_3 \mathrm{w}_{t-1}^m + \boldsymbol{\varepsilon}_t \tag{5.10}$$

and

$$\mathrm{w}_t = \boldsymbol{\varphi}_0 + \boldsymbol{\varphi}_1 \mathrm{w}_{t-1}^d + \boldsymbol{\varphi}_2 \mathrm{w}_{t-1}^w + \boldsymbol{\varphi}_3 \mathrm{w}_{t-1}^m + \boldsymbol{\varphi}_{1Q} (\pi_{t-1}^d \odot \mathrm{w}_{t-1}^d) + \boldsymbol{\varepsilon}_t, \tag{5.11}$$

where w_t^d, w_t^w and w_t^m denotes the daily, weekly and monthly lagged realized GMVP weight vectors, respectively. $\pi_t \equiv \sqrt{\mathrm{diag}\,(\Omega_t)}$ is the vector of asymptotic standard deviations for each of the individual elements in the w_t. Both MHAR and MHARQ models will result in an enormous number of predictors with the growth of the dimensions. The traditional estimation method like OLS fails to deliver reliable estimates due to the curse of dimensionality and large

computational burdens.

There are some popular approaches to reduce the number of parameters. We employed some of the widely-used techniques to overcome the issue of dimensionality, including the sparsely parameterized scalar version models, the Generalized multivariate HAR models that use a system of seemingly unrelated HAR regressors, the Least Absolute Shrinkage and Selection Operator (LASSO), and Principle Component Analysis (PCA).

Scalar version models

Chiriac and Voev (2011) first proposed a scalar version of the MHAR model, defined as,

$$\mathrm{w}_t = \boldsymbol{\varphi}_0 + \varphi_1 \mathrm{w}_{t-1}^d + \varphi_2 \mathrm{w}_{t-1}^w + \varphi_3 \mathrm{w}_{t-1}^m + \boldsymbol{\varepsilon}_t. \tag{5.12}$$

The intercept $\boldsymbol{\varphi}_0$ is a $N \times 1$ vector while the φ_1, φ_2 and φ_3 parameters are all assumed to be scalar. Similarly, the scalar version of MHARQ (Bollerslev et al., 2018) can be written as

$$\mathrm{w}_t = \boldsymbol{\varphi}_0 + \varphi_1 \mathrm{w}_{t-1}^d + \varphi_2 \mathrm{w}_{t-1}^w + \varphi_3 \mathrm{w}_{t-1}^m + \varphi_{1Q}(\pi_{t-1}^d \odot \mathrm{w}_{t-1}^d) + \boldsymbol{\varepsilon}_t, \tag{5.13}$$

where φ_{1Q} is also a scalar parameter. Both scalar-MHAR and scalar-MHARQ models in this study are estimated via the OLS regression.

The Generalized MHAR and Generalized MHARQ

Čech and Baruník (2017) proposed a generalized multivariate HAR (GHAR) model that uses a system of seemingly unrelated HAR regressors. The GHAR model is constructed by assuming the elements in the vector are only explained by their own lag variables. The GHAR model is defined as

$$\mathrm{w}_t = \boldsymbol{\varphi}_0 + \boldsymbol{\varphi}_1 \odot \mathrm{w}_{t-1}^d + \boldsymbol{\varphi}_2 \odot \mathrm{w}_{t-1}^w + \boldsymbol{\varphi}_3 \odot \mathrm{w}_{t-1}^m + \boldsymbol{\varepsilon}_t, \tag{5.14}$$

where $\boldsymbol{\varphi}_1$, $\boldsymbol{\varphi}_2$ and $\boldsymbol{\varphi}_3$ are the $N \times 1$ vectors of the coefficient for the lagged daily, weekly and

monthly variables. $\odot$ denotes the Hadamard product operator. Following the specification of GHAR, the Generalized MHARQ (GHARQ) model is

$$\mathrm{w}_t = \boldsymbol{\varphi}_0 + \boldsymbol{\varphi}_1 \odot \mathrm{w}_{t-1}^d + \boldsymbol{\varphi}_2 \odot \mathrm{w}_{t-1}^w + \boldsymbol{\varphi}_3 \odot \mathrm{w}_{t-1}^m + \boldsymbol{\varphi}_{1Q} \odot \left(\pi_{t-1}^d \odot \mathrm{w}_{t-1}^d\right) + \boldsymbol{\varepsilon}_t. \tag{5.15}$$

Both GHAR and GHARQ models are estimated with the two-step generalized least square (GLS) based on Čech and Baruník (2017).

The LASSO

Let $\mathrm{w}_{i,t}$, $i = \{1, 2, \ldots, N\}$, be the ith element in w_t, we have the dynamic specification

$$\mathrm{w}_{i,t} = \varphi_0 + \boldsymbol{Z}_{t-1}\boldsymbol{\Phi} + \varepsilon_{i,t}, \tag{5.16}$$

where the independent variable $\boldsymbol{Z}_t$ is defined as

$$\boldsymbol{Z}_t = \left[\mathrm{w}_{1,t}^d, \cdots, \mathrm{w}_{N,t}^d, \mathrm{w}_{1,t}^w, \cdots, \mathrm{w}_{N,t}^w, \mathrm{w}_{1,t}^m, \cdots, \mathrm{w}_{N,t}^m\right] \tag{5.17}$$

for the MHAR model. For the MHARQ model, $\boldsymbol{Z}_t$ is defined as

$$\boldsymbol{Z}_t = \left[\mathrm{w}_{1,t}^d, \cdots, \mathrm{w}_{N,t}^d, \mathrm{w}_{1,t}^w, \cdots, \mathrm{w}_{N,t}^w, \mathrm{w}_{1,t}^m, \cdots, \mathrm{w}_{N,t}^m, \pi_{1,t}^d, \cdots, \pi_{N,t}^d\right]. \tag{5.18}$$

The LASSO estimator $\widehat{\boldsymbol{\Phi}}^{LASSO}$ is obtained by solving

$$\widehat{\boldsymbol{\Phi}}^{LASSO} = \underset{\varphi_0, \boldsymbol{\Phi}}{\operatorname{argmin}} \quad \frac{1}{2T} \left\|\mathrm{w}_{i,t} - \varphi_0 - \boldsymbol{Z}_{t-1}\boldsymbol{\Phi}\right\|_{\ell_2}^2 + \lambda \left\|\boldsymbol{\Phi}\right\|_{\ell_1}, \tag{5.19}$$

where λ is a non-negative regularization parameter that often chosen either by cross-validation or using an information criterion.

Principle Component Analysis

Following Bauer and Vorkink (2011), the simple principle component method was used to avoid the issue of dimensionality that existed in the MHAR model (MHAR-PCA). Considering the w_t^d, w_t^w and w_t^m series are driven by a few factors, for instance, $\mathrm{w}^d(k)_t$ is the kth principal component of w_t^d. Given

$$\boldsymbol{Z}_t = \left[\mathrm{w}^d(1)_t, \cdots, \mathrm{w}^d(k)_t, \mathrm{w}^w(1)_t, \cdots, \mathrm{w}^w(k)_t, \mathrm{w}^m(1)_t, \cdots, \mathrm{w}^m(k)_t\right], \tag{5.20}$$

the MHAR model in equation (5.10) can be written as

$$\mathrm{w}_t = \boldsymbol{Z}_{t-1}\boldsymbol{\Phi} + \boldsymbol{\varepsilon}_t. \tag{5.21}$$

The factors are selected when their total explanatory power is over 90% in the empirical study. After estimating the principle components, it is easy to estimate the coefficient of the MHAR-PCA model by OLS.

The MHAR-X model

The MHAR-X model of Chapter 5 was also used to incorporate the impact of plausible dependence between the GMVP weight, in addition to the MHAR and MHARQ models. The MHAR-X model is

$$\mathrm{w}_t = \boldsymbol{\varphi}_0 + \varphi_1 \mathrm{w}_{t-1}^d + \varphi_2 \mathrm{w}_{t-1}^w + \varphi_3 \mathrm{w}_{t-1}^m + \boldsymbol{\theta_1}\underline{\mathrm{w}_{t-1}^d} + \boldsymbol{\theta_2}\underline{\mathrm{w}_{t-1}^w} + \boldsymbol{\theta_3}\underline{\mathrm{w}_{t-1}^m} + \boldsymbol{\varepsilon}_t, \tag{5.22}$$

where $\underline{\mathrm{w}_t^d}$, $\underline{\mathrm{w}_t^w}$ and $\underline{\mathrm{w}_t^m}$ are the variables that represent the potential dependence of the GMVP weight. φ_1, φ_2 and φ_3 are scalar coefficients that same as the scalar MHAR model. $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$ and $\boldsymbol{\theta}_3$ are the $N \times N$ parameter matrices for the lagged dependence variables. Due to the constraint $w_t'\iota = 1$, $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$ and $\boldsymbol{\theta}_3$ cannot be assumed to be scalars. Therefore, the LASSO was applied to shrink the equation (5.22) towards the scalar MHAR model in equation (5.12).

There are multiple routes to forecast the realized GMVP weight within the conventional approach. The first is to model and forecast the realized covariance matrix directly. Let S_t still be the realized covariance estimator, the GMVP weight forecast $\widehat{\mathrm{w}}_{t+h}$ is then obtained via the following process:

$$S_t \longrightarrow \widehat{S}_{t+h} \longrightarrow \widehat{\mathrm{w}}_{t+h}. \tag{5.23}$$

In order to have a valid positive definite covariance matrix forecast, the non-linear transformation on S_t is required most of the time. Another is to employ the matrix decomposition before the dynamic modeling. We continue the method in Chapter 4, decomposing the covariance matrix into a realized correlation matrix C_t pre- and post-multiplied by a diagonal matrix of realized standard deviations D_t. By forecasting the variances and correlation matrix separately, the GMVP weight forecast w_{t+h} is computed from:

$$S_t \longrightarrow D_t C_t D_t \longrightarrow \widehat{D}_{t+h}\widehat{C}_{t+h}\widehat{D}_{t+h} \longrightarrow \widehat{S}_{t+h} \longrightarrow \widehat{\mathrm{w}}_{t+h}. \tag{5.24}$$

To be consistent with Chapter 4, the matrix decomposition was adopted as the primary conventional approach. The elements of the standard deviation matrix are predicted via the simple log-HAR model, and the forecasts of the realized correlation matrix are from the scalar MHAR, GHAR, MHAR-Lasso, MHAR-PCA, and MHAR-X models. At the same time, the scalar MHAR and scalar MHARQ models are also used to direct forecast the original realized covariance matrix for comparison purposes.

5.2.3 Re-scale the realized GMVP weight forecast

In most situations, the summation of the portfolio weight forecasts from the direct forecasting approach is not equal to 1, defined as

$$\sum_{i=1}^{N} \widehat{\mathrm{w}}_{i,t+h} \neq 1. \tag{5.25}$$

(Golosnoy and Gribisch, 2022) only forecast the first $(N-1)$ weight series and calculate the last weight series from

$$\widehat{\mathrm{w}}_{N,t+h} = 1 - \sum_{i=1}^{N-1} \widehat{\mathrm{w}}_{i,t+h}. \tag{5.26}$$

However, the problem of using the method above is that there is no universal standard to decide which weight series should be excluded. This chapter, therefore, re-scales the weight forecast with the magnitude based on the total summation:

$$\sum_{i=1}^{N} \widehat{\mathrm{w}}_{i,t+h} = \kappa, \ \kappa \neq 1. \tag{5.27}$$

The re-scaled GMVP weight can be calculated by

$$\sum_{i=1}^{N} \frac{\widehat{\mathrm{w}}_{i,t+h}}{\kappa} = 1. \tag{5.28}$$

5.2.4 Forecast evaluation

We evaluated the forecast precision and economic significance for various multivariate volatility models in conventional and direct forecasting approaches. In particular, the unbiasedness of the realized GMVP weight forecasts was evaluated using the Root Mean Square Error (RMSE) statistical loss function. For out-of-sample window $t = \tau + 1, ..., T$,

$$\mathcal{L}_{Statistical} = \sqrt{\frac{1}{T-\tau} \sum_{t=\tau+1}^{T} \left(\mathrm{w}_{t+h} - \widehat{\mathrm{w}}_{t+h}\right)' \left(\mathrm{w}_{t+h} - \widehat{\mathrm{w}}_{t+h}\right)}, \tag{5.29}$$

where w_{t+h} is the ex-post realized GMVP weight proxy for asset i, and $\widehat{\mathrm{w}}_{t+h}$ is the h-step ahead realized GMVP weight forecast. The best GMVP weight forecast should provide the smallest realized GMVP variance from the economic perspective (Golosnoy and Gribisch, 2022). Therefore, another criterion to evaluate the GMVP weight forecast is from the minimum

variance investor's perspective,

$$\begin{aligned}\mathcal{L}_{Economic} &= \frac{1}{T-\tau}\sum_{t=\tau+1}^{T}\left(\widehat{\mathrm{w}}_{t+h}' S_{t+h}\widehat{\mathrm{w}}_{t+h} - \mathrm{w}_{t+h}' S_{t+h}\mathrm{w}_{t+h}\right) \\ &= \frac{1}{T-\tau}\sum_{t=\tau+1}^{T}\left(\widehat{\mathrm{w}}_{t+h}' S_{t+h}\widehat{\mathrm{w}}_{t+h} - \frac{1}{\iota' \cdot S_{t+h}^{-1} \cdot \iota}\right).\end{aligned} \tag{5.30}$$

The Model Confidence Set (MCS) approach of Hansen et al. (2011) was also emplyed to assess the statistical significance of ex-post losses. The MCS is able to identify the set of models with the best forecasting performance.

The investment improvements of different approaches are also assessed based on the Markowitz mean-variance optimization. Given the optimal weight in the realized GMVP, a series of evaluation criteria were constructed to assess the performances of the GMVP based on the various volatility models in different approaches:

Average turnover ratio

According to Bollerslev et al. (2018), Callot et al. (2017), DeMiguel et al. (2014), the average portfolio turnover over the out-of-sample period is calculated by

$$\text{Average } TO_t = \frac{1}{N\left(T-\tau\right)}\sum_{t=\tau+1}^{T}\sum_{i=1}^{N}\left|\widehat{\mathrm{w}}_{i,t} - \widehat{\mathrm{w}}_{i,t}^{\text{hold}}\right|, \tag{5.31}$$

where

$$\widehat{\mathrm{w}}_{i,t}^{\text{hold}} = \widehat{\mathrm{w}}_{i,t-1}\cdot\frac{\left(1+r_{i,t-1}\right)}{1+r_{p\,t-1}}. \tag{5.32}$$

The turnover measures the average change in the portfolio weight and $\widehat{\mathrm{w}}_{i,t}^{\text{hold}}$ is the weight of the hold portfolio. The hold portfolio at period t is defined as the portfolio that keeps the stocks from period $t-1$. Higher turnover ratios mean increased portfolio expenses, which can reduce the portfolio's overall performance. The weight forecasts with a lower turnover ratio indicate a more stable hold portfolio.

Average return

Assuming the fixed transaction costs are proportional to the turnover ratio in the portfolios, then the average portfolio return is defined as

$$\mu_p = \frac{1}{(T-\tau)} \sum_{t=\tau+1}^{T} r_{pt} = \frac{1}{(T-\tau)} \sum_{t=\tau+1}^{T} \left(\widehat{\mathrm{w}}_t' r_t - cTO_t\right), \tag{5.33}$$

where the TO_t is the turnover rate at period t. As all the stocks considered are very liquid in this study, we set the value of c to be 0.1%.

Accumulated return

According to Callot et al. (2017), the accumulated portfolio return over the out-of-sample period is

$$r_p^{Accum} = \prod_{t=\tau+1}^{T} \left(1 + r_{pt}\right). \tag{5.34}$$

Standard deviation

Given the portfolio return r_{pt}, the standard deviation of portfolio return over the out-of-sample period is

$$\sigma_p = \sqrt{\frac{1}{(T-\tau)} \sum_{t=\tau+1}^{T} \left(r_{pt} - \mu_p\right)^2} \tag{5.35}$$

Average diversification ratio

The average diversification ratio is defined as the average ratio of weighted standard deviations of the individual stocks over portfolio standard deviations, expressed as

$$\frac{1}{(T-\tau)} \sum_{t=\tau+1}^{T} \frac{\sum_{i=1}^{N} \widehat{\mathrm{w}}_{i,t}\sigma_{it}}{\sigma_{pt}}, \tag{5.36}$$

where

$$\sigma_{pt} = \widehat{\mathrm{w}}_t' S_t \widehat{\mathrm{w}}_t. \tag{5.37}$$

More diversification within a portfolio decreases the denominator and leads to a higher diversification ratio.

Economic value

The utility-based framework of Fleming et al. (2003) was considered to evaluate the economic significance of the forecasting models in different approaches. The utility function is used to measure the satisfaction of an investor that derives from the portfolio investment. In this study, we assume the investor has quadratic utility and the daily utility generated by the portfolio is expressed as

$$U\left(r_{pt},\right) = \left(1 + r_{pt}\right) - \frac{\gamma}{2(1+\gamma)}\left(1 + r_{pt}\right)^2, \tag{5.38}$$

where γ is the risk aversion coefficient γ. The larger the value of γ, the more risk-averse is the investor. A quadratic utility function is widely applied as an objective function for various optimization problems in finance due to the advantage of its tractability (e.g. Callot et al. 2017, Chiriac and Voev 2011). The economic value of the different portfolios is determined by solving for Δ in

$$\sum_{t=\tau+1}^{T} U\left(r_{p1\,t}\right) = \sum_{t=\tau+1}^{T} U\left(r_{p2\,t} - \Delta\right), \tag{5.39}$$

where Δ is the maximum return an investor with risk aversion γ would be willing to sacrifice to switch from portfolio $p1$ to portfolio $p2$. In this study, the benchmark portfolio, $p1$, will always be from the scalar MHAR model on the raw covariance matrix, and the value of Δ is reported as an annualized basis point fee.

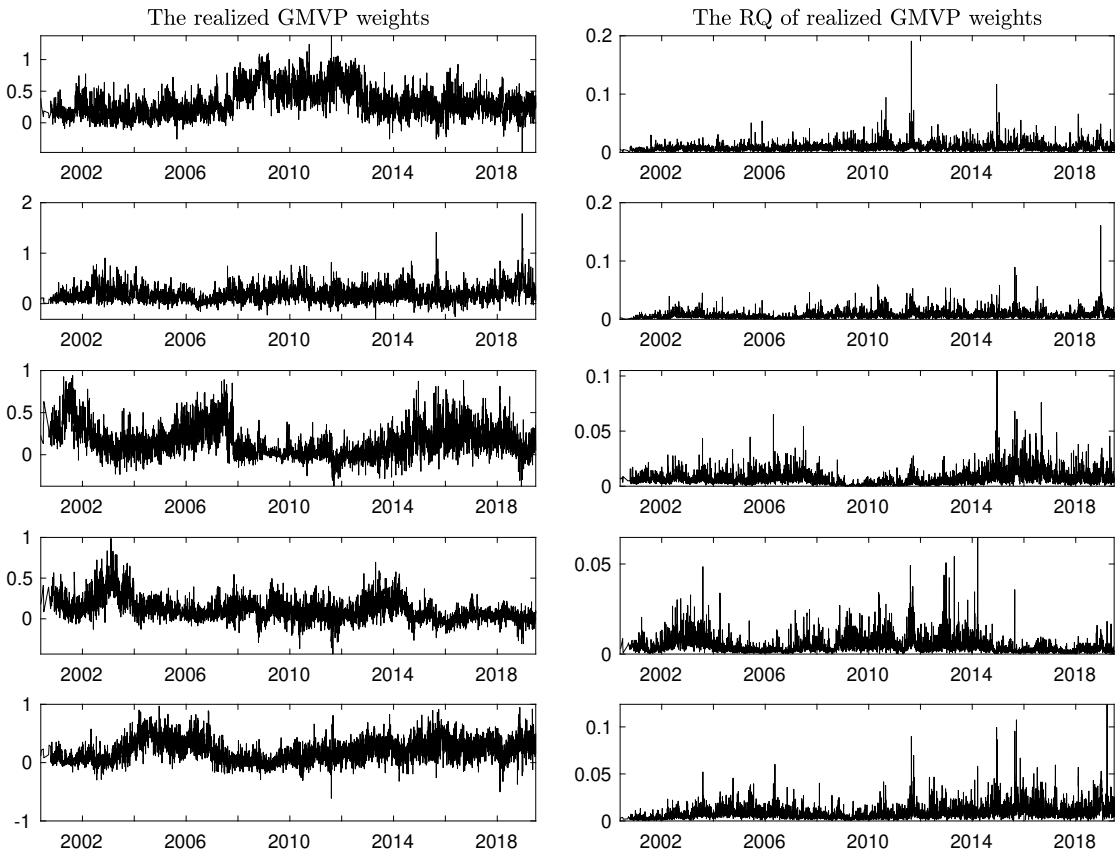

Figure 5.1: Realized GMVP weight (left panel) and RQ of realized GMVP weight (right panel) for 5-stock portfolio **A** over period May 24, 2000, and ends on June 27, 2019.

5.3 Empirical Results

5.3.1 Data

The empirical analysis is based on the data introduced in Chapter 3. The data are obtained from the Thomson Reuters Tick History Database, including the 5-minute prices of fifty stocks from the U.S. market. The whole sample period of fifty stocks starts on May 24, 2000, and ends on June 27, 2019, yielding 4193 daily realized covariance observations that were constructed based on the multivariate kernel estimator of Barndorff-Nielsen et al. (2011). The information of the stocks is given in Table A.1 of Appendix A.1. To make the empirical results more convincing, rather than modeling and forecasting one portfolio with fifty stocks, we construct ten portfolios

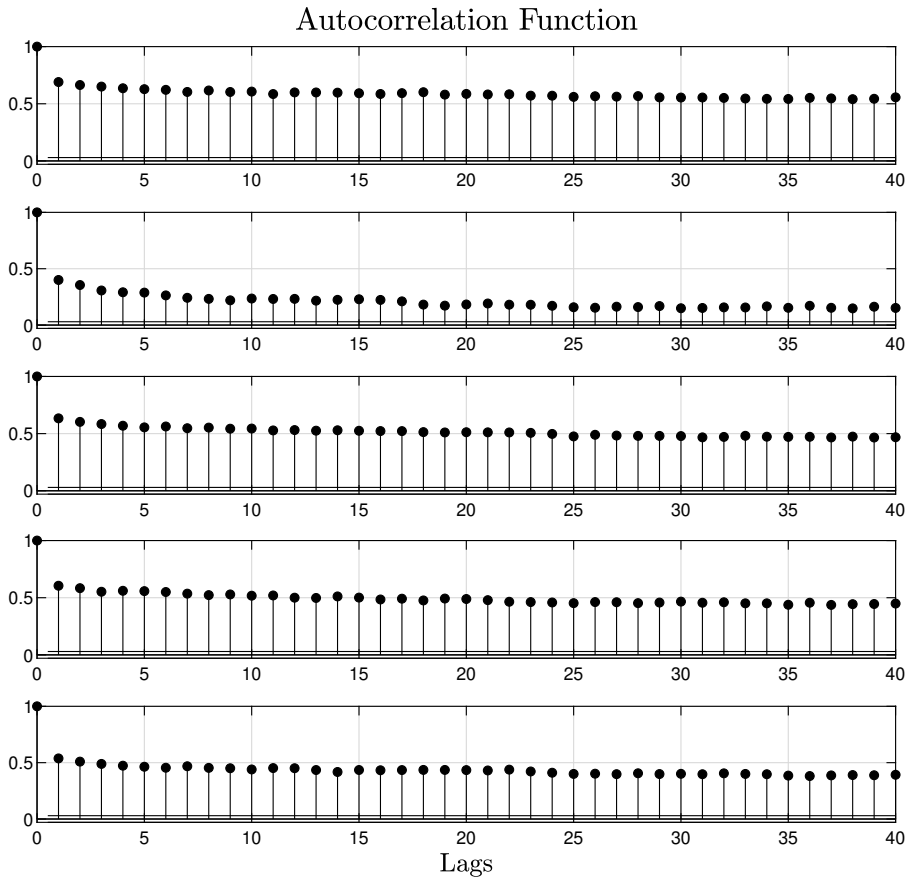

Figure 5.2: ACF plot of realized GMVP weight for 5-stock portfolio **A**.

with five stocks each, named in alphabetical order **A** to **J**. The out-of-sample results are reported based on the forecasts from those ten portfolios.

The five stocks in portfolio **A** visualize the dynamics of the realized GMVP weight. The left panel in Figure 5.1 presents the time series plot of the realized GMVP weight series. The right panel in Figure 5.1 shows the associated estimated realized quarticity (RQ) of GMVP weight. Figure 5.2 shows the sample autocorrelation functions (ACFs) of the realized GMVP weight for portfolio **A**. A slowly decaying serial correlation pattern in Figure 5.2 indicates the realized GMVP weight also has the long memory dependence feature. The RQ plots in Figure 5.1 show that the realized GMVP weight is more stable with much less heteroskedasticity than the variance-covariance series. Hence, modeling and forecasting of the realized GMVP weight directly is expected to lead to smaller forecasting errors, higher predictive power, and

less turnover than conventional procedures.

5.3.2 Out-of-sample analysis

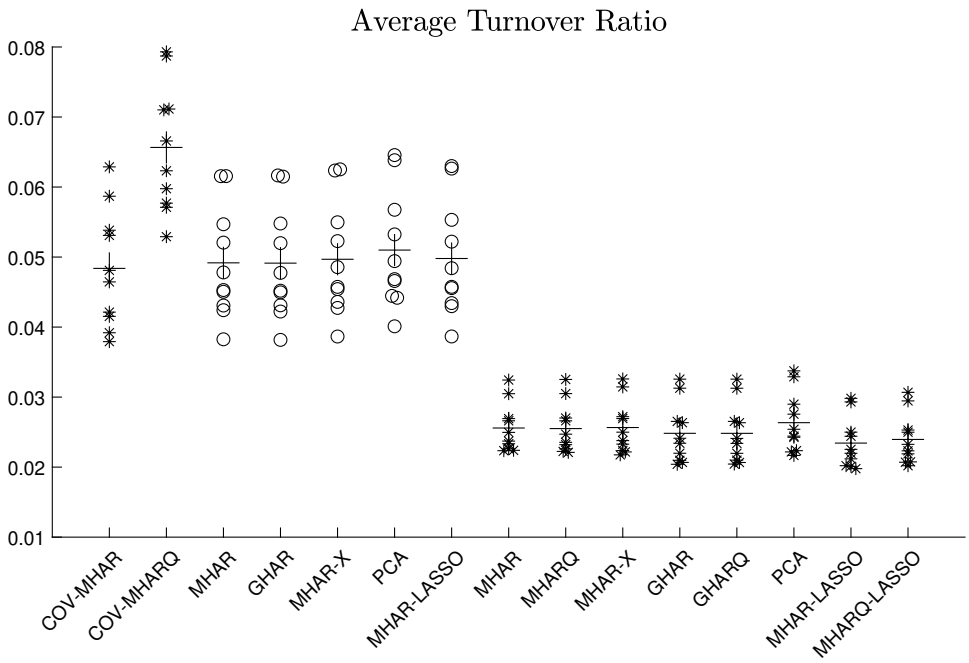

Figure 5.3: Average turnovers of GMVP based on various multivariate volatility forecast models of ten 5-stock portfolios. "$\circ$" represents average turnover ratio from the matrix decomposition method. The first two "$*$" columns represent average turnovers from the scalar-MHAR on the covariance matrix and scalar-MHARQ on the covariance matrix. The rest of "$*$" represents average turnovers from the direct forecasting approach. "$+$" represents mean of average turnovers across ten portfolios.

In this study, we compared the forecasting performance of different approaches, investigated the potential benefits of directly forecasting the realized GMVP weight, and demonstrated the validity of the findings in chapter 4. In order to directly highlight the forecasting performance of different approaches, the study focused on a daily investment horizon. The 1-day-ahead realized GMVP weight forecasts are based on the 5-stock portfolios and a set of volatility models introduced in the previous section. All models are re-estimated daily based on a rolling sample window of the past 1000 days. The GMVP is re-balanced given the realized GMVP weight forecasts. The out-of-sample forecasts of realized GMVP weight are evaluated based on

Table 5.1: Evaluation of out-of-sample forecasting accuracy

	$\widehat{S}_{t+1} \to \widehat{\mathrm{w}}_{t+1}$		$\widehat{D}_{t+1}\widehat{C}_{t+1}\widehat{D}_{t+1} \to \widehat{\mathrm{w}}_{t+1}$					$\mathrm{w}_t \to \widehat{\mathrm{w}}_{t+1}$							
Model:	MHAR (Scalar)	MHARQ (Scalar)	MHAR (Scalar)	GHAR	MHAR-X (Scalar)	MHAR (PCA)	MHAR (LASSO)	MHAR (Scalar)	MHARQ (Scalar)	MHAR-X (LASSO)	GHAR	GHARQ	MHAR (PCA)	MHAR (LASSO)	MHARQ (LASSO)
							$\mathcal{L}_{Statistical}$								
A	1.000	1.020	0.987	0.989	0.985	0.988	0.987	**0.965**	**0.965**	**0.965**	0.965	0.965	0.978	0.968	0.969
B	1.000	1.028	0.988	0.990	0.986	0.991	0.989	**0.969**	**0.969**	**0.969**	**0.968**	**0.968**	0.986	0.971	0.972
C	1.000	1.020	0.992	0.994	0.991	0.996	0.994	**0.967**	**0.967**	**0.967**	**0.967**	**0.967**	0.972	0.969	0.969
D	1.000	1.037	1.001	1.002	0.998	1.000	0.999	**0.981**	**0.981**	**0.979**	0.982	0.982	0.986	0.983	0.984
E	1.000	1.022	1.007	1.007	1.005	1.006	1.005	**0.977**	**0.977**	**0.977**	**0.977**	**0.977**	0.984	**0.979**	0.979
F	1.000	1.036	0.992	0.995	0.991	0.996	0.994	**0.967**	**0.967**	**0.967**	**0.966**	**0.966**	0.972	0.969	0.970
G	1.000	1.025	1.006	1.009	1.003	1.004	1.005	**0.980**	**0.980**	**0.979**	**0.980**	**0.980**	0.985	0.981	0.982
H	1.000	1.052	0.988	0.989	0.986	0.990	0.988	0.963	0.963	**0.963**	**0.963**	**0.962**	0.968	0.965	0.965
I	1.000	1.034	0.989	0.990	0.986	0.989	0.988	**0.966**	**0.966**	**0.967**	**0.966**	**0.966**	0.977	0.970	0.970
J	1.000	1.025	0.999	1.001	0.998	1.001	0.999	**0.974**	**0.974**	**0.974**	**0.974**	**0.974**	0.983	0.978	0.979
							$\mathcal{L}_{Economic}$								
A	1.000	1.298	**0.948**	**0.956**	**0.952**	**0.957**	**0.953**	**0.966**	**0.966**	**0.967**	**0.969**	**0.969**	1.035	0.981	0.983
B	1.000	1.149	**0.948**	**0.951**	**0.956**	**0.962**	**0.953**	0.977	0.977	0.981	0.974	**0.974**	1.037	0.991	0.992
C	1.000	1.062	**0.943**	**0.948**	**0.959**	0.970	**0.959**	**0.973**	0.973	**0.964**	**0.971**	**0.971**	0.982	0.974	0.980
D	1.000	1.266	0.995	1.001	**0.982**	**0.988**	**0.986**	1.001	1.000	**0.981**	1.011	1.011	0.997	0.995	0.998
E	**1.000**	1.050	**1.005**	**1.006**	**1.007**	**1.005**	**0.997**	**0.991**	**0.995**	**0.984**	**0.996**	**0.996**	**1.010**	**0.990**	**1.001**
F	1.000	1.108	0.961	0.980	**0.957**	0.978	0.971	**0.946**	**0.948**	**0.937**	**0.944**	**0.944**	0.955	**0.942**	**0.944**
G	1.000	1.234	0.933	0.938	0.935	0.937	0.931	**0.901**	**0.901**	**0.896**	**0.894**	**0.894**	0.910	**0.895**	**0.900**
H	1.000	1.325	**1.000**	**1.005**	**0.989**	**0.989**	**0.990**	**0.958**	**0.960**	**0.951**	**0.957**	**0.956**	**0.959**	**0.953**	**0.952**
I	1.000	1.073	**0.965**	0.974	0.968	**0.966**	**0.965**	**0.967**	**0.967**	**0.958**	0.967	**0.967**	1.031	0.970	0.978
J	1.000	1.089	**0.955**	**0.959**	**0.959**	0.971	**0.960**	**0.946**	**0.946**	**0.947**	**0.948**	**0.948**	0.986	**0.965**	**0.967**

Notes: Table reports the evaluation of out-of-sample forecasting accuracy at 1-day-ahead horizon for ten different portfolios (**A** to **J**). Top panel shows RMSE loss ratio ($\mathcal{L}_{Statistical}$, see equation (5.29)) while the bottom panel shows the ratio of the average GMVP variance forecast error ($\mathcal{L}_{Economic}$, see equation (5.30)). Loss ratios for the different models are relative to the scalar MHAR model on the covariance matrix. Result in boldface indicates that the model is included in the 90% MCS.

Table 5.2: Evaluation for daily GMVP forecasts

	$\widehat{S}_{t+1} \to \widehat{\mathrm{w}}_{t+1}$		$\widehat{D}_{t+1}\widehat{C}_{t+1}\widehat{D}_{t+1} \to \widehat{\mathrm{w}}_{t+1}$					$\mathrm{w}_t \to \widehat{\mathrm{w}}_{t+1}$							
Model:	MHAR (Scalar)	MHARQ (Scalar)	MHAR (Scalar)	GHAR	MHAR-X (Scalar)	MHAR (PCA)	MHAR (LASSO)	MHAR (Scalar)	MHARQ (Scalar)	MHAR-X (LASSO)	GHAR	GHARQ	MHAR (PCA)	MHAR (LASSO)	MHARQ (LASSO)
Average Turnover	0.048	0.066	0.049	0.049	0.050	0.051	0.050	0.026	0.026	0.026	0.025	0.025	0.026	0.023	0.024
Average Return (%)	-0.014	-0.016	-0.014	-0.014	-0.014	-0.015	-0.014	-0.014	-0.014	-0.013	-0.014	-0.014	-0.013	-0.013	-0.013
Accumulated Return	0.598	0.563	0.600	0.601	0.600	0.592	0.596	0.624	0.623	0.631	0.626	0.626	0.628	0.631	0.634
Standard Deviation (%)	0.923	0.935	0.924	0.925	0.922	0.923	0.923	0.927	0.927	0.925	0.927	0.927	0.927	0.926	0.927
Diversification ratio	1.459	1.446	1.454	1.452	1.453	1.452	1.452	1.490	1.491	1.487	1.490	1.490	1.487	1.491	1.492
Economic value (%)															
$\gamma = 1$	-	-0.446	0.120	0.166	0.112	-0.020	0.072	0.159	0.150	0.246	0.175	0.178	0.278	0.261	0.298
$\gamma = 5$	-	-0.554	0.110	0.143	0.120	-0.019	0.066	0.114	0.103	0.220	0.128	0.132	0.231	0.224	0.256
$\gamma = 10$	-	-0.688	0.098	0.113	0.131	-0.019	0.058	0.057	0.046	0.188	0.071	0.074	0.172	0.178	0.203

Notes: Table shows results for GMVP based on various multivariate volatility forecast models and different forecasting approaches.

the loss functions from equations (5.29) and (5.30). The descriptive measures on corresponding global minimum variance portfolio forecasts are also reported.

In Table 5.1 we evaluated the 1-day-ahead out-of-sample forecasting accuracy based on the RMSE loss function in equation (5.29) and the average GMVP variance forecast error in equation (5.30) across all ten 5-stock portfolios. The values presented in Table 5.1 are the loss ratios of the alternative approaches relative to the scalar-MHAR model of the first column. The evaluation results reported in the left and middle panels are for the conventional approach: forecasting the realized covariance directly, $\widehat{S}_{t+1} \rightarrow \widehat{\mathrm{w}}_{t+1}$, and the matrix decomposition method, $\widehat{D}_{t+1}\widehat{C}_{t+1}\widehat{D}_{t+1} \rightarrow \widehat{\mathrm{w}}_{t+1}$. The right panel records the evaluation results from the direct forecasting approach, $\mathrm{w}_t \rightarrow \widehat{\mathrm{w}}_{t+1}$.

We start with comparing the RMSE loss ratios ($\mathcal{L}_{Statistical}$) between the two forecasting approaches. The lowest RMSE loss is obtained from the MHAR-X (via LASSO) of the direct forecasting approach, followed by the scalar-MHAR(Q) and GHAR(Q) models. The MHAR-X (via LASSO) is the best since it had been included in the 90% MCS for all ten portfolios. The scalar-MHAR and scalar-MHARQ models are included in the 90% MCS nine times while the GHAR and GHARQ are included in the MCS eight times. Remarkably the 90% MCS contains not a single model from the conventional approach. Table 5.1 shows that all RMSE losses from the direct forecasting approach are lower than those of the conventional approach. The results are consistent with our expectations since the realized GMVP weight exhibited much less heteroskedasticity than the variance-covariance series. Surprisingly, the MHAR-LASSO and MHARQ-LASSO models under the direct forecasting approach do not appear competitive, although these models account for the potential dependence between the GMVP weight series.

Turning to the RMSE losses for the conventional approach, we find the scalar-MHARQ on the covariance matrix produces the highest RMSE loss. In terms of forecasting the GMVP weight, the matrix decomposition method performs slightly better than direct forecasting the covariance matrix. Within the matrix decomposition method, the MHAR-X outperformed all other models and this is consistent with the findings in the Chapter 4.

The bottom panel of Table 5.1 reports ratios of the average GMVP variance forecast error

($\mathcal{L}_{Economic}$). While the statistical loss indicates the direct forecasting approach has advantages in GMVP weight prediction, the results in $\mathcal{L}_{Economic}$ somehow are mixed. For the direct forecasting approach, the MHAR-X and the GHARQ offer the most accurate GMVP variance forecasts, with both models included in the 90% MCS nine times. The GHAR and scalar-MHAR(Q) have similar forecasting performances as all of them have been included in the MCS multiple times. Similar to the RMSE loss results, the MHAR-PCA and MHAR(Q)-LASSO appear less competitive in GMVP variance forecasting.

Moving to the conventional approach, we find the scalar-MHARQ on covariance matrix has the highest loss across all the models, and this is consistent with the results in the panel $\mathcal{L}_{Statistical}$. The method of direct forecasting the covariance matrix is clearly outperformed by matrix decomposition. The empirical results show that the MHAR-X and MHAR-LASSO under the matrix decomposition method have the best forecasting performance since they are included in the 90% MCS eight times. The scalar-MHAR of matrix decomposition is slightly behind, included seven times in the MCS. Unlike the clear contrast in $\mathcal{L}_{Statistical}$, there is little difference between the matrix decomposition method and direct forecasting approach in terms of GMVP variance prediction. The direct forecasting approach performs slightly better due to the MCS result.

Table 5.2 reports the evaluation of GMVP based on the realized GMVP weight forecasts from a daily investment horizon. The results are averaged over the ten datasets. We assume all the stocks are considered very liquid and set the cost of the transactions to be 0.1%. As shown in Table 5.2, the MHAR-LASSO of the direct forecasting approach has the lowest transaction cost as it has the lowest average turnover. Compared to the conventional approach, the direct forecasting approach results in more stable realized GMVP weight forecasts and therefore reduces the spurious turnover induced by forecast error. One potentially important feature of the direct forecasting approach manifests in systematically lower portfolio turnover and reduced transaction costs. Figure 5.3 shows the comparison of the average turnover of GMVP between two approaches. From the Figure 5.3, we can observe that the average turnover of the direct forecasting approach is reduced substantially compared to the positions for the conventional

approach.

Table 5.2 also reports the accumulated return and diversification ratio of GMVP for both forecasting approaches. All the LASSO-based models under the direct forecasting approach deliver higher accumulated returns and diversification ratios than OLS-based models. Particularly, the MHARQ-LASSO has the highest accumulated return and diversification ratio, with the MHAR-X (LASSO) and MHAR-LASSO not far behind. Overall, the direct forecasting approach outperforms the conventional approach in terms of accumulated return and diversification ratio.

The bottom panel of Table 5.2 reports the economic values of various multivariate volatility models against the benchmark – scalar-MHAR on the covariance matrix. The positive economic values suggest the advantages of a specific model in terms of the investor's utility compared to the benchmark. Three types of investors are considered in this study, mild risk-aversion investor with $\gamma = 1$, moderate risk-aversion investor with $\gamma = 5$, and strong risk-aversion investor with $\gamma = 10$. As shown in Table 5.2, except for the scalar-MHARQ on the covariance and the MHAR-PCA of matrix decomposition, all other volatility models have positive economic values, ranging from 0.072% to 0.298% for the mild risk-aversion investors, 0.066% to 0.256% for the moderate risk-aversion investors, and 0.046% to 0.203% for the strong risk-aversion investors. The highest economic gains are from the MHARQ-LASSO of the direct forecasting approach. The economic gains of MHAR-X(LASSO) under the direct forecasting approach are slightly worse than those of the MHARQ-LASSO model. The economic gains from the direct forecasting approach are overall higher than the economic gains from the conventional approach.

5.4 Conclusion

The global minimum variance portfolio is the origin of the Markowitz mean-variance efficient frontier and the increased availability of high-frequency data encourages us to compute the precise and consistent ex-post estimates of the realized GMVP weight via the realized covariance

matrix. The conventional approach to forecast the GMVP weight is based on the covariance matrix forecasts obtained from multivariate volatility models. The alternative approach is to directly model and forecast the ex-post realized GMVP weight with the same multivariate volatility models. This chapter seeks to provide a comprehensive analysis of forecasting realized GMVP weight and aims to validate the previous findings within the context of portfolio selection.

The out-of-sample forecasts of the realized GMVP weight are based on ten portfolios of five stocks from the U.S. market. The forecasting performance is evaluated based on the statistical and economic evaluation criteria. The empirical analysis shows the direct forecasting approach outperforms the conventional approach in the sense of providing realized GMVP weight forecasts that are significantly closer to the ex-post realized GMVP composition. Based on the performance in portfolio allocation, the empirical result indicates that the direct forecasting approach provides superior GMVP weight forecasts. In particular, the direct approach could reduce the portfolio turnover substantially and lead to a higher economic gain. In the end, the empirical results validate the previous findings in Chapter 4, suggesting the value of incorporating the potential dependence between the portfolio weight to improve the economic performance of the portfolio. More specifically, we find that the MHAR-X of Chapter 4 outperforms a series of models within the direct forecasting approach.

All findings in this study are based on the short forecast horizon. The comparison of forecasting performance between the conventional approach and the direct forecasting approach at longer investment horizons is still unknown. It is interesting to extend current work to a longer forecast horizon, and we will leave this for a future study.

Chapter 6

Conclusions

The availability of high-frequency financial data facilitated research on the modeling and forecasting of ex-post realized covariance measures constructed from intraday data. These realized covariance forecasts have been extensively used in empirical financial applications. This thesis consists of three studies that centre around forecasting the elements of the realized covariance matrix. The first study focuses on the diagonal element of the realized covariance, namely the realized variance (RV). Although RV is consistent for the true latent volatility under certain conditions, it is subject to the heteroskedastic measurement error in any finite sample. Bollerslev et al. (2016) developed the HARQ model to address the heteroskedastic measurement error directly. The HARQ model exhibits superior prediction performance. However, the HARP model, which replaces the square root of RQ in the HARQ model with RV, shows very similar forecasting performance to the HARQ model. We speculate the forecasting gains from the HARP may be attributed to capturing the non-linear dependence in the conditional mean of RV, and the HARQ model may also account for this non-linearity. The empirical analysis presented in the first study illustrates that the heteroskedastic measurement error and non-linear dependence in the conditional mean of RV must be considered for RV forecasting. Therefore, the success of the HARQ model can be attributed to accounting for both heteroskedastic measurement error and non-linear dependence in the conditional mean of the RV. In addition, the empirical analysis suggests the logarithmic transformation of RV is able to address both

heteroskedastic measurement error and non-linearity simultaneously.

The second study focuses on the realized correlation matrix, which is decomposed from the realized covariance matrix. Due to the curse of dimensionality, the highly parsimonious multivariate volatility model becomes a popular approach to avoid the issue of dimensionality. However, these models reduce the number of parameters by assuming the same parameter structure for all the elements in the correlation matrix, neglecting the cross-effect between the correlation matrix elements. We apply the LASSO regression to reveal the optimal parameter structure for correlation matrix modeling to address this issue. The empirical evidence indicates the potential dependence between the correlation matrix elements significantly impacts the predictive accuracy of the realized covariance. In the end, we proposed a new dynamic model to incorporate the cross-effect within the correlation matrix elements, and we illustrate that the new model outperforms a series of benchmark models in an out-of-sample study.

The precise realized covariance matrix estimates and forecasts can be used to compute the weight of realized Global Minimum Variance Portfolio (GMVP). GMVP forecasting has attracted significant attention from both academics and practitioners. The final study provides an analysis of forecasting realized GMVP weight and seeks to validate the previous findings within the context of portfolio selection. The conventional approach to forecast the realized GMVP weight is based on the realized covariance matrix forecasts obtained from multivariate volatility models. The alternative approach is to directly model and forecast the ex-post realized GMVP weight with the same multivariate volatility models. The empirical analysis in the final study shows the direct forecasting approach outperforms the conventional approach in the sense of providing realized GMVP weight forecasts that are significantly closer to the ex-post realized GMVP compositions. Besides, compared to the conventional approach of forecasting realized covariance matrix first, the direct forecasting approach generally provides superior portfolio allocation performance, reduces the portfolio turnover substantially, and effectively improves the economic gains. More importantly, the empirical result validates the previous findings, suggesting incorporating the potential dependence between the portfolio weight to improve the economic performance of the portfolio. We find that the MHAR-X of Chapter 4 outperforms a

series of models within the direct forecasting approach.

Financial markets are more related than ever before due to economic globalization. Realized covariance, derived from the high-frequency data, is an accurate proxy to measure the common risk between financial assets. Financial Econometrics literature has focused on modeling and forecasting realized covariance over the past few decades. This thesis aims to establish an approach by improving the realized covariance forecast by resolving multiple issues simultaneously, such as the heteroskedastic measurement error of RV, the non-linearity in the conditional mean of RV, and the cross-effect between the entries of the correlation matrix. More importantly, the approach is relatively efficient and performs well in empirical financial application.

There remains a large scope for future work, and only the most pressing will be mentioned here. First, the study in Chapter 3 highlights the use of logarithmic transformation, a member of the Box-Cox transformation family, in terms of handling the heteroskedastic measurement error and possible non-linearity in the conditional mean of RV. While the log-transformation performs well in a univariate setting, it is less effective when it comes to the multivariate realized covariance matrix as the off-diagonal elements of the covariance matrix might be non-positive. There is a pressing need to investigate several flexible transformations that can be applied to a multivariate setting. For example, Tsai et al. (2017) introduced the hyperbolic power transformation, which can transform the random variable to close to normality, and the data can take both positive and negative values. Next, while the MHAR-X model in Chapter 4 has been proved to be effective in incorporating the cross-effects between the elements of the realized correlation matrix, there remains insufficient theoretical for this model, more investigation should be pursued. For instance, the forecasting performance on realized covariance matrix can be examined in future studies since the cross-effect between the covariance matrix entries has also been ignored. Finally, the empirical study in Chapter 5 provides a review on realized GMVP weight forecasting but it is somehow limited and the investigation can be extended to the longer horions and other multivariate volatility models. Finally, the empirical study in Chapter 5 provides a review of realized GMVP weight forecasting. However, the study only focuses on the HAR framework at a short forecasting horizon, which is somehow limited. The investigation of

realized GMVP weight can be extended in the future by including other multivariate volatility models at the longer forecasting horizons.

Appendix A

HARQ Model: Modeling with Measurement Error and Non-linear Dependence

---

A.1 Descriptive Statistics

Table A.1 provides a standard set of summary statistics for the daily RV. All of the RVs are based on 5-minute returns. In addition to the usual summary measures, the first-order autocorrelation coefficients and the correlation coefficients between RV and sqrt-RQ are also reported.

Table A.1: Summary Statistics

Company	Symbol	Min	Mean	Median	Max	AR	Corr
S&P 500	SPX	0.002	0.093	0.042	4.634	0.717	0.951
Dow Jones Industrial Average	DJI	0.002	0.104	0.051	4.450	0.732	0.871
Abbott Laboratories	ABT	0.010	0.172	0.102	6.867	0.639	0.887
Archer Daniels Midland	ADM	0.013	0.266	0.151	17.406	0.625	0.918
American International Group	AIG	0.011	0.655	0.145	39.698	0.851	0.965
Apache	APA	0.028	0.385	0.245	12.060	0.766	0.918
American Express	AXP	0.009	0.330	0.109	15.699	0.734	0.946
Boeing	BA	0.010	0.242	0.138	5.991	0.714	0.929

Continue on the next page

Company	Symbol	Min	Mean	Median	Max	AR	Corr
Bank of America	BAC	0.010	0.454	0.147	24.655	0.792	0.979
BB&T	BBT	0.012	0.326	0.112	15.194	0.795	0.944
Best Buy	BBY	0.024	0.428	0.260	7.873	0.654	0.895
Bank of New York Mellon	BK	0.014	0.370	0.141	25.804	0.724	0.961
Citigroup	C	0.015	0.534	0.172	27.344	0.794	0.973
Caterpillar	CAT	0.016	0.276	0.161	11.045	0.749	0.936
Carnival	CCL	0.015	0.246	0.122	9.226	0.677	0.919
Capital One Financial	COF	0.012	0.525	0.169	20.277	0.820	0.952
Deere	DE	0.011	0.303	0.164	11.731	0.766	0.930
Walt Disney	DIS	0.012	0.243	0.117	11.859	0.723	0.910
Duke Energy	DUK	0.005	0.192	0.094	10.499	0.704	0.948
EOG Resources	EOG	0.024	0.384	0.247	14.421	0.768	0.916
General Electric	GE	0.007	0.287	0.122	13.425	0.726	0.937
Corning	GLW	0.008	0.606	0.245	14.523	0.782	0.947
Gap	GPS	0.024	0.419	0.254	9.117	0.647	0.890
Halliburton	HAL	0.030	0.525	0.303	23.453	0.641	0.916
Home Depot	HD	0.012	0.255	0.127	10.028	0.684	0.931
Hartford Financial Services Group	HIG	0.013	0.576	0.144	36.550	0.832	0.966
Honeywell International	HON	0.007	0.240	0.125	8.591	0.727	0.898
International Business Machines	IBM	0.011	0.163	0.083	7.684	0.680	0.906
Johnson & Johnson	JNJ	0.007	0.113	0.060	4.987	0.642	0.894
JPMorgan Chase	JPM	0.011	0.387	0.139	20.247	0.787	0.963
KeyCorp	KEY	0.016	0.545	0.169	27.105	0.782	0.963
Coca-Cola	KO	0.005	0.119	0.068	5.969	0.651	0.926
Kohls	KSS	0.016	0.355	0.225	7.954	0.676	0.880
Lennar	LEN	0.020	0.699	0.319	25.259	0.792	0.944
Eli Lilly	LLY	0.013	0.174	0.106	8.377	0.644	0.894
Mcdonald's	MCD	0.008	0.164	0.087	12.882	0.497	0.948
3M	MMM	0.006	0.150	0.084	9.385	0.622	0.930
Altria Group	MO	0.006	0.161	0.090	5.726	0.477	0.866
Merck	MRK	0.013	0.196	0.113	10.044	0.561	0.860
Marathon Oil	MRO	0.029	0.442	0.253	13.049	0.785	0.924
Nike	NKE	0.013	0.210	0.118	6.805	0.651	0.904

Continue on the next page

Company	Symbol	Min	Mean	Median	Max	AR	Corr
Occidental Petroleum	OXY	0.017	0.308	0.178	15.543	0.784	0.940
Pfizer	PFE	0.015	0.189	0.114	6.622	0.661	0.886
Procter & Gamble	PG	0.009	0.115	0.066	6.572	0.592	0.911
Transocean	RIG	0.034	0.615	0.420	13.326	0.710	0.902
Schlumberger NV	SLB	0.023	0.372	0.231	16.319	0.728	0.921
SunTrust Banks	STI	0.011	0.452	0.140	19.940	0.825	0.962
State Street	STT	0.014	0.388	0.143	30.630	0.741	0.961
TJX Companies	TJX	0.016	0.249	0.145	5.480	0.669	0.907
Union Pacific	UNP	0.011	0.231	0.130	10.774	0.758	0.912
United Technologies	UTX	0.009	0.177	0.101	7.848	0.687	0.911
Wells Fargo	WFC	0.010	0.368	0.114	18.101	0.805	0.970

Notes: The table reports the descriptive statistics for the daily RV for each of the series. The column labelled AR reports the standard first order autocorrelation coefficients, the colum labelled Corr reports the correlation coefficient between RV and sqrt-RQ. The minimum, maximum, mean and median values have been scaled up by 10^3.

A.2 In-sample Estimation Results

This section provides the in-sample estimation results for the individual stocks. We first rank the fifty market stocks based on the correlation coefficient between RV and sqrt-RQ, choosing three particular stocks which have the minimum, maximum and average correlation coefficient values. Table A.2 shows the results for stock with the minimum correlation coefficient and Table A.3 shows the one with the maximum value. The correlation coefficient of stock *Coca-Cola* equals to the mean value of correlation coefficient among fifty stocks. Therefore, Table A.4 shows the in-sample estimation results for stock *Coca-Cola*.

Table A.2: In-sample estimation results for stock with the **minimum** correlation coefficient (stock of *Merck*).

	HAR	HARQ	HARP	HARP3	log-HAR	log-HARQ	log-HARP	log-HARP3
β_0	0.000	0.000	0.000	0.000	-0.856	-0.887	0.008	-5.687
	(0.000)	(0.000)	(0.000)	(0.000)	(0.099)	(0.097)	(0.508)	(2.508)
β_1	0.196	0.591	0.418	0.547	0.371	0.391	0.575	-1.502
	(0.018)	(0.031)	(0.026)	(0.041)	(0.017)	(0.017)	(0.119)	(0.904)
β_2	0.395	0.280	0.354	0.344	0.256	0.227	0.254	0.252
	(0.032)	(0.032)	(0.032)	(0.032)	(0.026)	(0.026)	(0.026)	(0.026)
β_3	0.296	0.179	0.224	0.203	0.285	0.261	0.285	0.282
	(0.030)	(0.031)	(0.031)	(0.031)	(0.023)	(0.022)	(0.023)	(0.023)
α_1		-298.370	-43.410	-109.019		0.133	0.012	-0.238
		(19.509)	(3.763)	(16.475)		(0.012)	(0.007)	(0.108)
α_2				5659.192				-0.010
				(1383.676)				(0.004)
R^2	0.422	0.450	0.438	0.440	0.435	0.438	0.435	0.441
MSE	0.687	0.654	0.668	0.665	0.672	0.668	0.672	0.664
AIC	-62685.71	-62912.09	-62815.06	-62829.78	7193.78	7067.68	7192.78	7189.40
BIC	-62659.99	-62879.93	-62782.90	-62791.19	7219.51	7099.84	7224.94	7227.99

Notes: The table provides in-sample parameter estimates and measures of fit for the various HAR-type models and their logarithmic versions. The top panel reports the actual parameter estimates with robust standard errors in parentheses. The bottom panel summarizes the in-sample R^2, Mean Squared Error (MSE) and QLIKE losses from the regressions, together with the model selection criterion values. The MSE values for all the models have been scaled up by 1×10^7.

A.3 Out-of-Sample Analysis

This section provides the QLIKE loss comparisons between different models for 50 market stocks at the longer forecasting horizons. Figure A.4, A.5 and A.6 present the results at 5-day-ahead forecast horizon. Figure A.7, A.8 and A.9 present the results at 5-day-ahead forecast horizon.

Table A.3: In-sample estimation results for stock with the **maximum** correlation coefficient (stock of *Bank of America*).

	HAR	HARQ	HARP	HARP3	log-HAR	log-HARQ	log-HARP	log-HARP3
β_0	0.000	0.000	0.000	0.000	-0.437	-0.440	0.660	-0.586
	(0.000)	(0.000)	(0.000)	(0.000)	(0.063)	(0.062)	(0.241)	(0.922)
β_1	0.444	0.841	0.809	0.799	0.456	0.451	0.748	0.241
	(0.017)	(0.033)	(0.032)	(0.042)	(0.017)	(0.017)	(0.064)	(0.367)
β_2	0.263	0.164	0.167	0.166	0.250	0.235	0.240	0.239
	(0.027)	(0.027)	(0.027)	(0.027)	(0.026)	(0.026)	(0.026)	(0.026)
β_3	0.235	0.160	0.168	0.170	0.250	0.246	0.247	0.246
	(0.023)	(0.023)	(0.023)	(0.024)	(0.020)	(0.020)	(0.020)	(0.020)
α_1		-120.404	-20.951	-19.168		0.109	0.017	-0.050
		(8.622)	(1.539)	(5.192)		(0.013)	(0.004)	(0.048)
α_2				-66.478				-0.003
				(184.857)				(0.002)
R^2	0.675	0.689	0.688	0.688	0.673	0.674	0.665	0.675
MSE	5.967	5.724	5.735	5.735	6.027	6.002	6.157	5.966
AIC	-52760.73	-52949.92	-52940.68	-52938.81	6959.53	6895.72	6939.29	6939.33
BIC	-52735.01	-52917.76	-52908.52	-52900.22	6985.25	6927.88	6971.45	6977.92

Notes: The table provides in-sample parameter estimates and measures of fit for the various HAR-type models and their logarithmic versions. The top panel reports the actual parameter estimates with robust standard errors in parentheses. The bottom panel summarizes the in-sample R^2, Mean Squared Error (MSE) and QLIKE losses from the regressions, together with the model selection criterion values. The MSE values for all the models have been scaled up by 1×10^7.

Table A.4: In-sample estimation results for stock with the **average** correlation coefficient (stock of *Coca-Cola*).

	HAR	HARQ	HARP	HARP3	log-HAR	log-HARQ	log-HARP	log-HARP3
β_0	0.000	0.000	0.000	0.000	-0.635	-0.713	0.767	-3.531
	(0.000)	(0.000)	(0.000)	(0.000)	(0.093)	(0.092)	(0.533)	(2.705)
β_1	0.190	0.537	0.498	0.573	0.359	0.378	0.677	-0.795
	(0.018)	(0.026)	(0.027)	(0.035)	(0.017)	(0.017)	(0.120)	(0.916)
β_2	0.507	0.427	0.435	0.444	0.301	0.274	0.296	0.296
	(0.030)	(0.029)	(0.030)	(0.030)	(0.027)	(0.027)	(0.027)	(0.027)
β_3	0.209	0.087	0.112	0.092	0.279	0.257	0.276	0.273
	(0.026)	(0.026)	(0.026)	(0.027)	(0.023)	(0.022)	(0.023)	(0.023)
α_1		-382.353	-87.380	-158.242		0.102	0.017	-0.150
		(21.605)	(5.795)	(21.970)		(0.010)	(0.006)	(0.103)
α_2				10573.415				-0.006
				(3162.369)				(0.004)
R^2	0.529	0.559	0.551	0.552	0.538	0.542	0.531	0.544
MSE	0.193	0.180	0.184	0.183	0.189	0.188	0.192	0.187
AIC	-68518.19	-68819.49	-68738.34	-68747.52	6231.83	6132.67	6226.69	6226.06
BIC	-68492.47	-68787.33	-68706.18	-68708.93	6257.55	6164.83	6258.85	6264.65

Notes: The table provides in-sample parameter estimates and measures of fit for the various HAR-type models and their logarithmic versions. The top panel reports the actual parameter estimates with robust standard errors in parentheses. The bottom panel summarizes the in-sample R^2, Mean Squared Error (MSE) and QLIKE losses from the regressions, together with the model selection criterion values. The MSE values for all the models have been scaled up by 1×10^7.

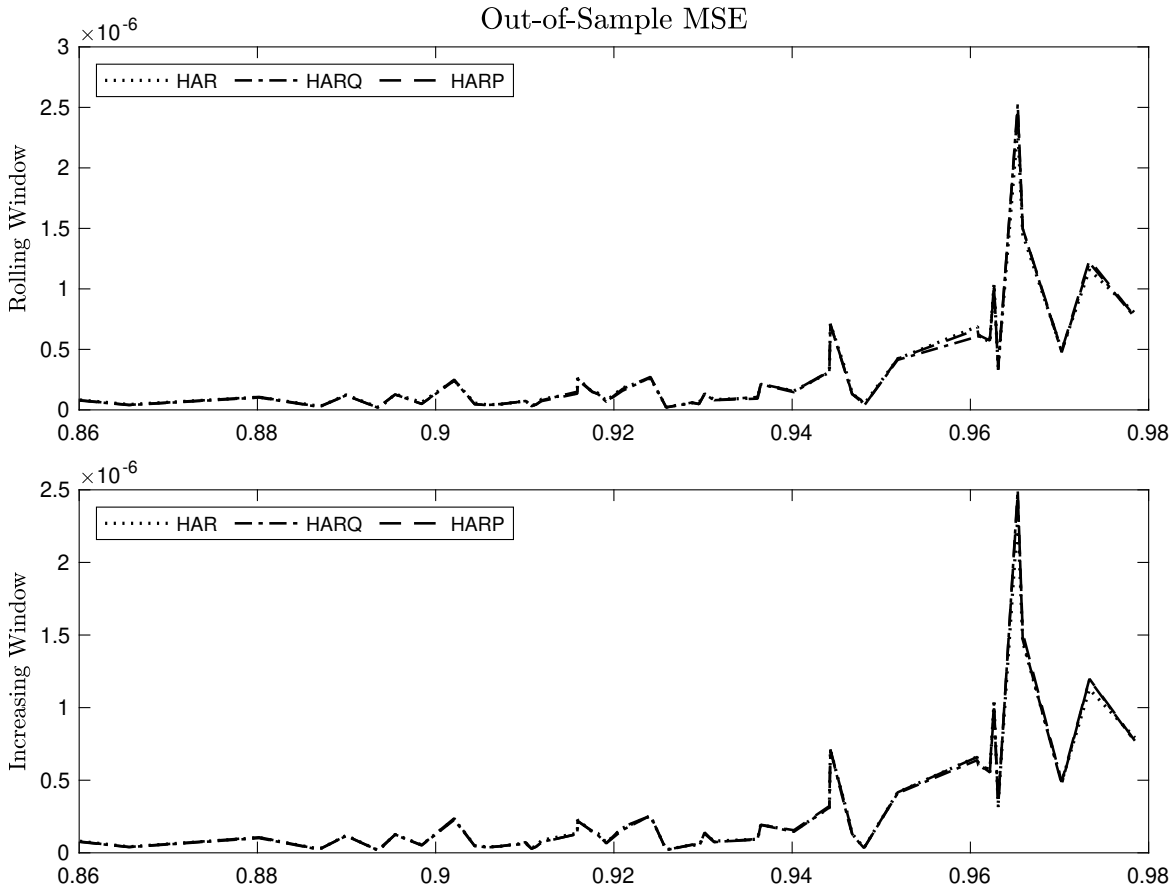

Figure A.1: The out-of-sample MSE losses of standard HAR (dotted), standard HARQ (dotdash) and standard HARP (dashed) models for fifty stocks at 1-day-ahead forecast horizon. The top figure shows the MSE loss based on the rolling window while the bottom figure shows the results from increasing window. The horizontal axis present the correlation coefficients between RV and sqrt-RQ that sorted in ascending order.

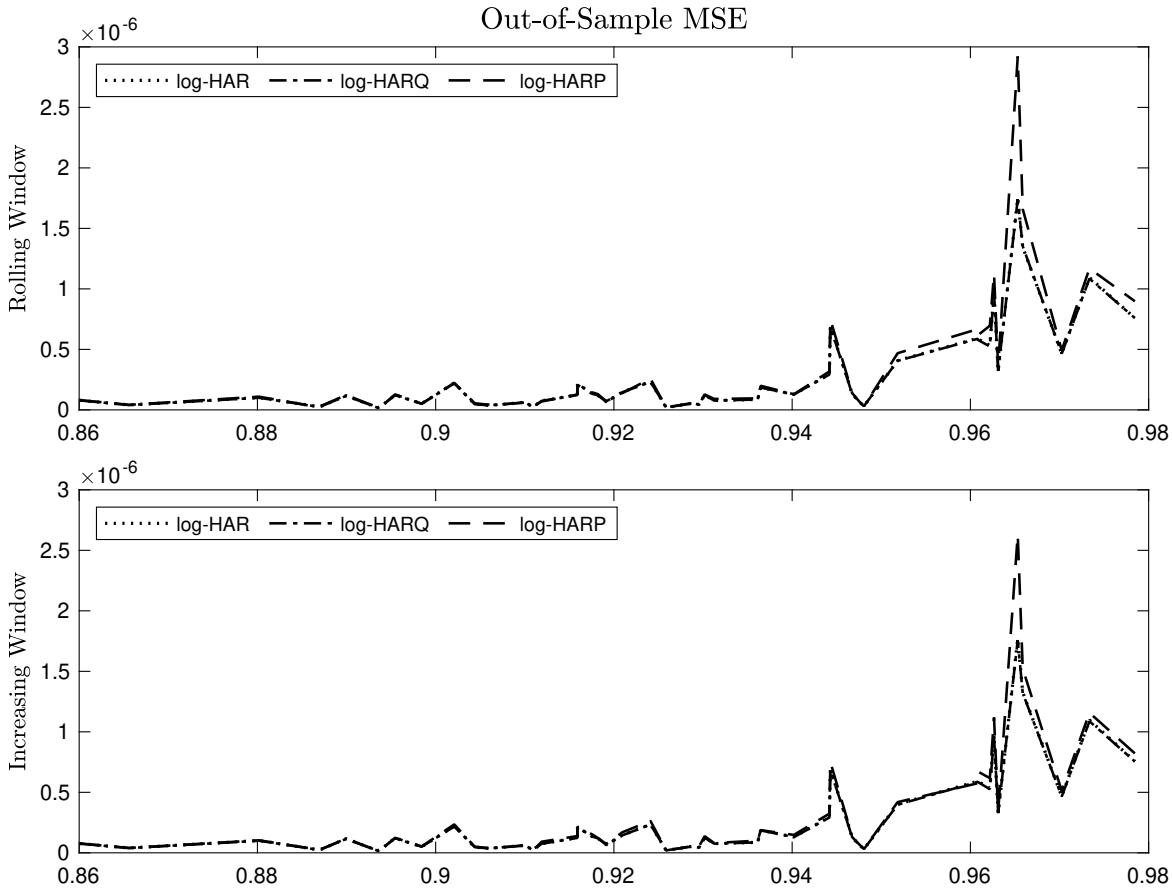

Figure A.2: The out-of-sample MSE losses of log-HAR (dotted), log-HARQ (dotdash) and log-HARP (dashed) models for fifty stocks at 1-day-ahead forecast horizon. The top figure shows the MSE loss based on the rolling window while the bottom figure shows the results from increasing window. The horizontal axis present the correlation coefficients between RV and sqrt-RQ that sorted in ascending order.

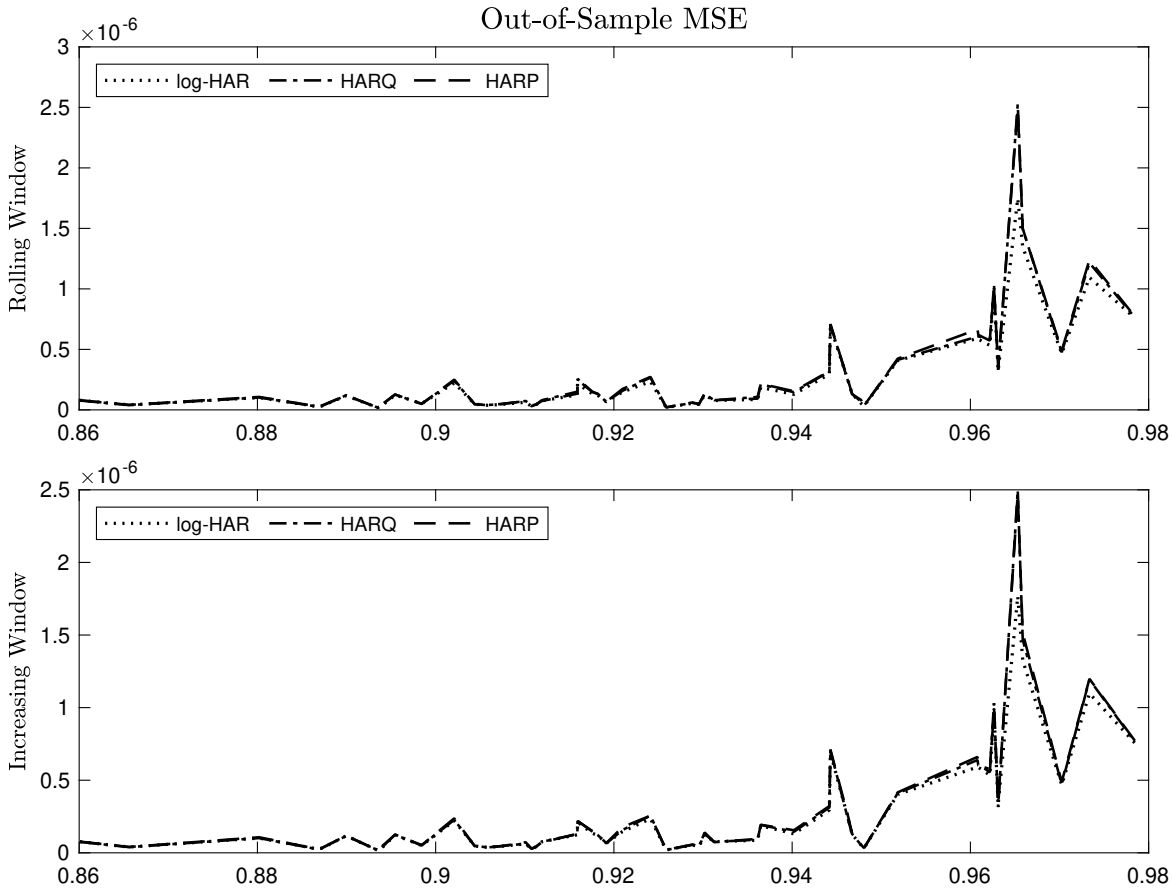

Figure A.3: The out-of-sample MSE losses of log-HAR (dotted), standard HARQ (dotdash) and standard HARP (dashed) models for fifty stocks at 1-day-ahead forecast horizon. The top figure shows the MSE loss based on the rolling window while the bottom figure shows the results from increasing window. The horizontal axis present the correlation coefficients between RV and sqrt-RQ that sorted in ascending order.

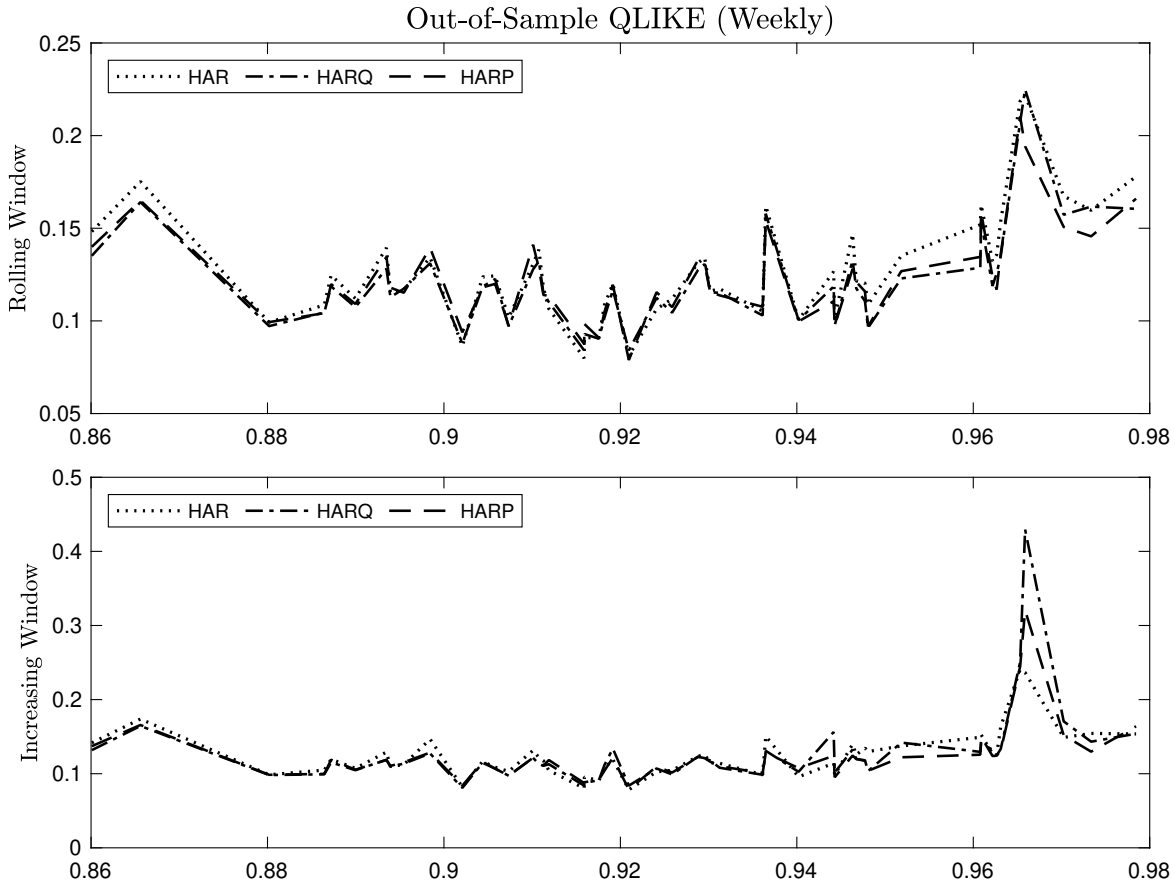

Figure A.4: The out-of-sample QLIKE losses of standard HAR (dotted), standard HARQ (dotdash) and standard HARP (dashed) models for fifty stocks at 5-day-ahead forecast horizon. The top figure shows the QLIKE loss based on the rolling window while the bottom figure shows the results from increasing window. The horizontal axis present the correlation coefficients between RV and sqrt-RQ that sorted in ascending order.

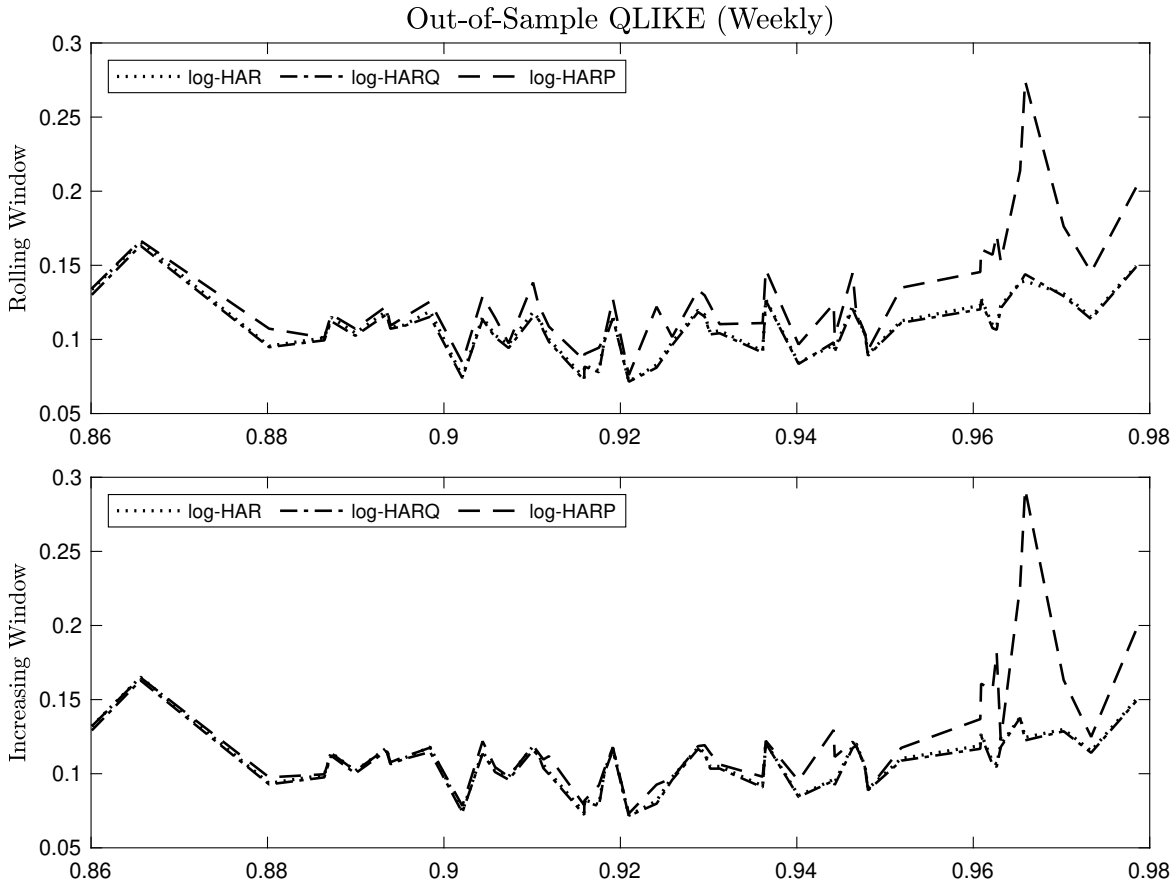

Figure A.5: The out-of-sample QLIKE losses of log-HAR (dotted), log-HARQ (dotdash) and log-HARP (dashed) models for fifty stocks at 5-day-ahead forecast horizon. The top figure shows the QLIKE loss based on the rolling window while the bottom figure shows the results from increasing window. The horizontal axis present the correlation coefficients between RV and sqrt-RQ that sorted in ascending order.

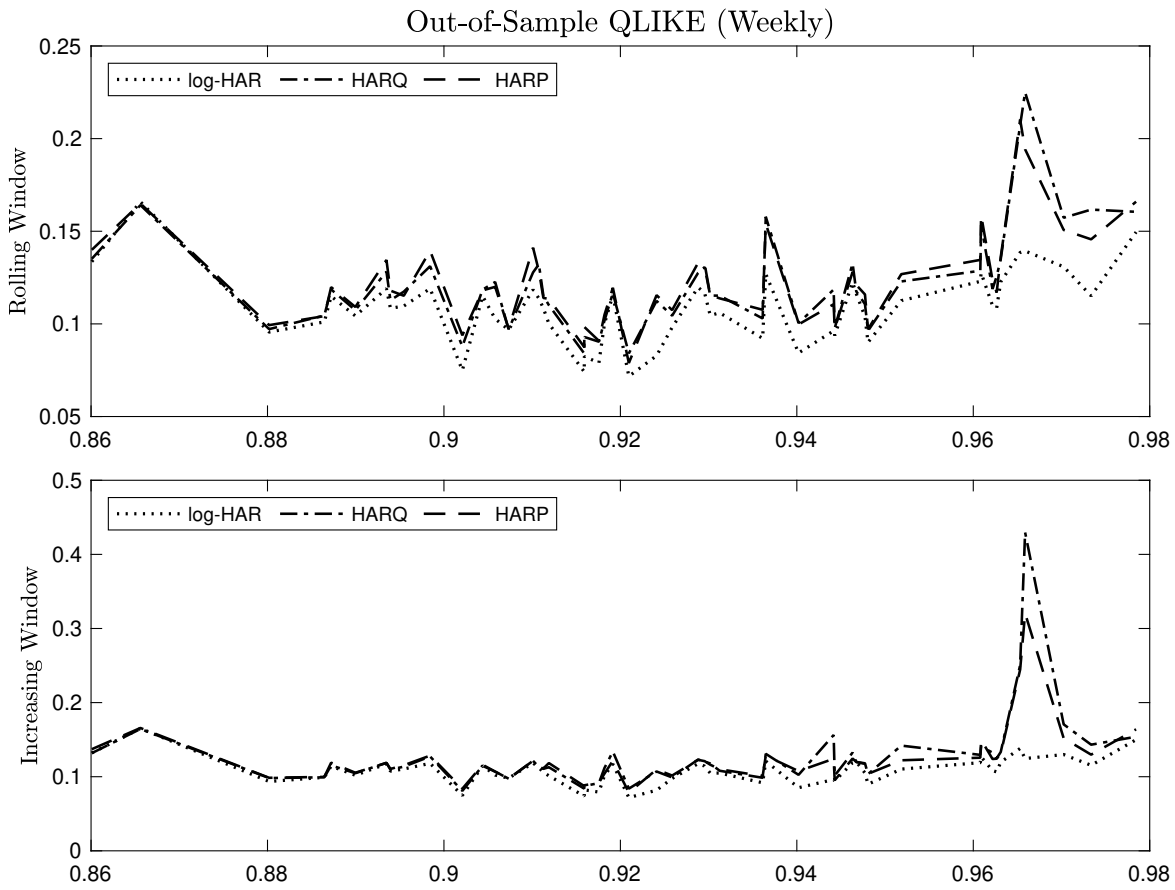

Figure A.6: The out-of-sample QLIKE losses of log-HAR (dotted), standard HARQ (dotdash) and standard HARP (dashed) models for fifty stocks at 5-day-ahead forecast horizon. The top figure shows the QLIKE loss based on the rolling window while the bottom figure shows the results from increasing window. The horizontal axis present the correlation coefficients between RV and sqrt-RQ that sorted in ascending order.

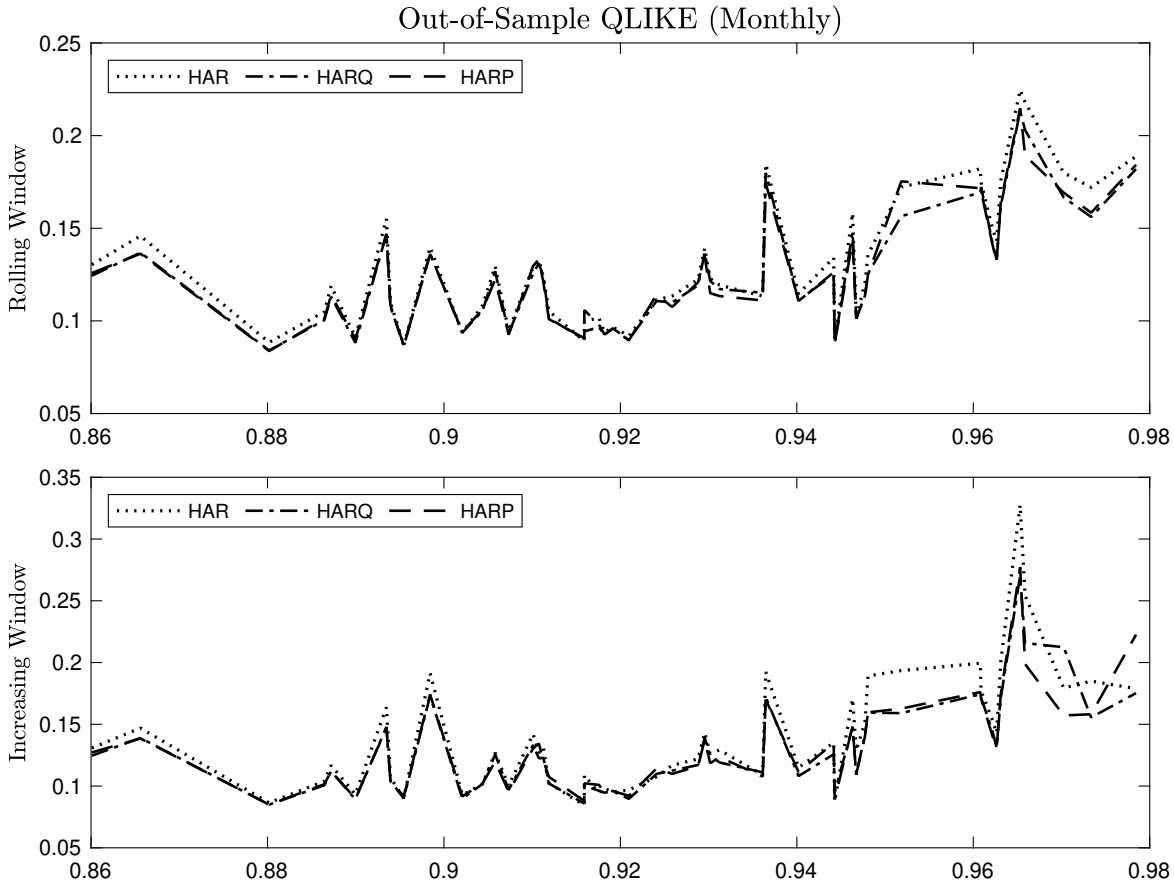

Figure A.7: The out-of-sample QLIKE losses of standard HAR (dotted), standard HARQ (dotdash) and standard HARP (dashed) models for fifty stocks at 22-day-ahead forecast horizon. The top figure shows the QLIKE loss based on the rolling window while the bottom figure shows the results from increasing window. The horizontal axis present the correlation coefficients between RV and sqrt-RQ that sorted in ascending order.

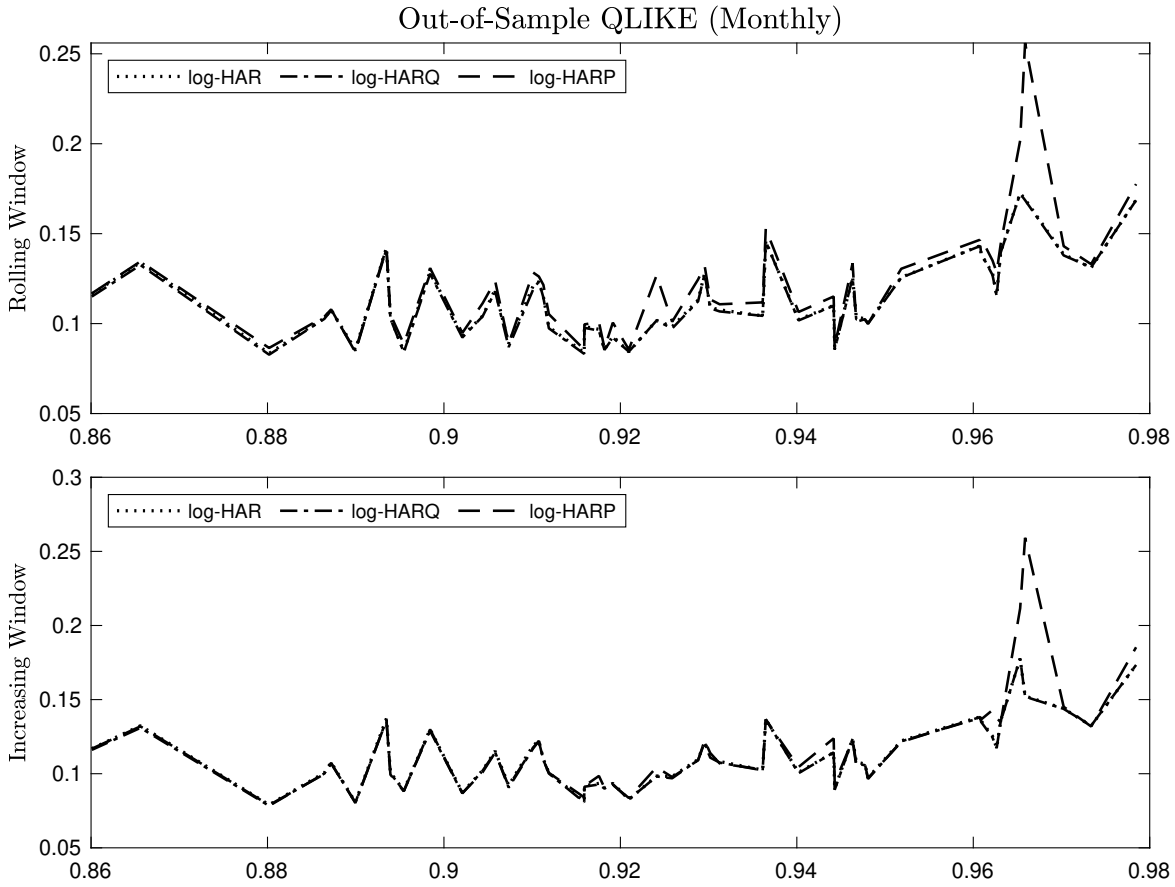

Figure A.8: The out-of-sample QLIKE losses of log-HAR (dotted), log-HARQ (dotdash) and log-HARP (dashed) models for fifty stocks at 22-day-ahead forecast horizon. The top figure shows the QLIKE loss based on the rolling window while the bottom figure shows the results from increasing window. The horizontal axis present the correlation coefficients between RV and sqrt-RQ that sorted in ascending order.

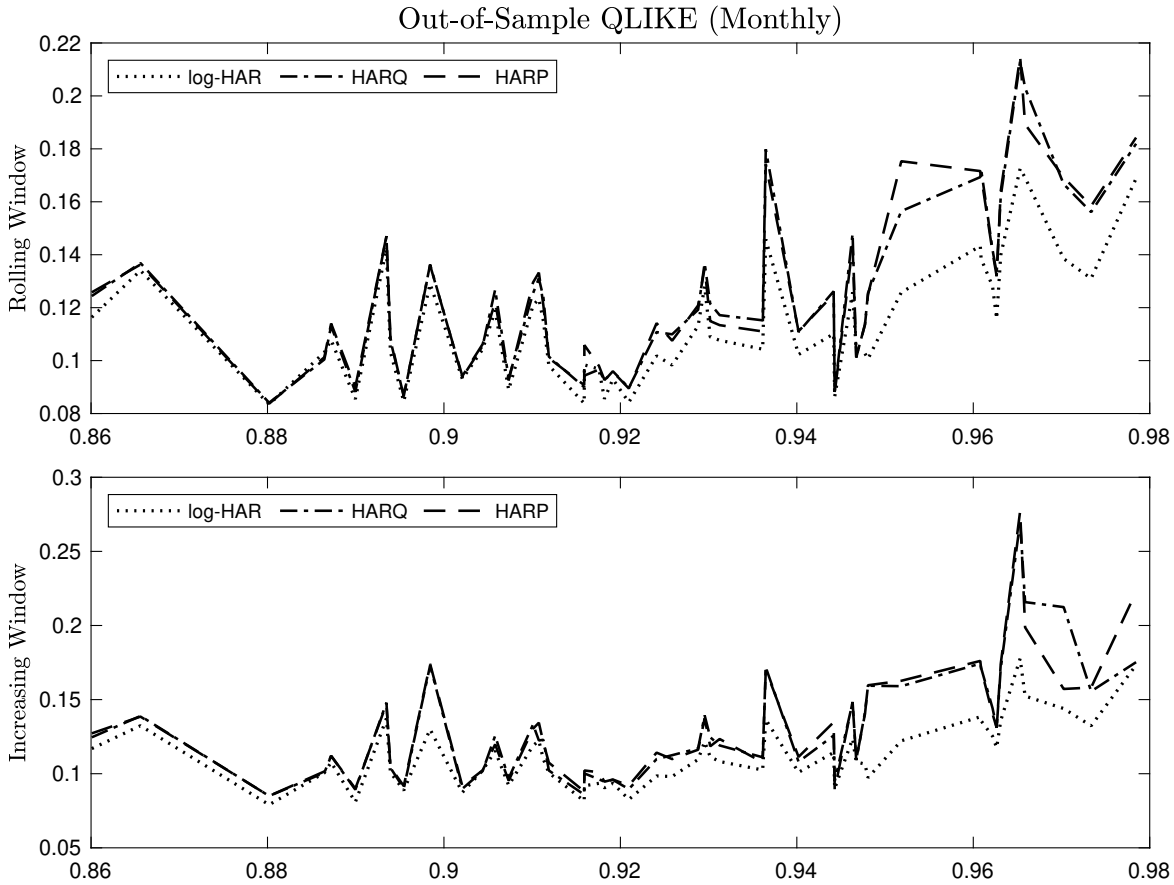

Figure A.9: The out-of-sample QLIKE losses of log-HAR (dotted), standard HARQ (dotdash) and standard HARP (dashed) models for fifty stocks at 22-day-ahead forecast horizon. The top figure shows the QLIKE loss based on the rolling window while the bottom figure shows the results from increasing window. The horizontal axis present the correlation coefficients between RV and sqrt-RQ that sorted in ascending order.

Appendix B

Forecasting Realized GMVP Weight

B.1 Asymptotic Distribution of Realized GMVP Weights

Golosnoy et al. (2020) derive the asymptotic distribution of realized GMVP weights. For $M \to \infty$, we have

$$M^{1/2}\left(\mathrm{w}_t - w_t\right) \to \mathcal{N}\left(0, \Omega_t\right), \tag{B.1}$$

where

$$\Omega_t = G_t' \Pi_t G_t. \tag{B.2}$$

Since

$$w_t = \frac{\Sigma_t^{-1} \cdot \iota}{\iota' \cdot \Sigma_t^{-1} \cdot \iota}, \tag{B.3}$$

from the product rule of differentiation, we have

$$G_t' = \frac{\partial w_t}{\partial \operatorname{vech}\left(\Sigma_t\right)'} = \frac{\partial\left(\Sigma_t^{-1} \cdot \iota\right)}{\partial \operatorname{vech}\left(\Sigma_t\right)'}\left(\iota' \cdot \Sigma_t^{-1} \cdot \iota\right)^{-1} + \left(\Sigma_t^{-1} \cdot \iota\right) \frac{\partial\left(\iota' \cdot \Sigma_t^{-1} \cdot \iota\right)^{-1}}{\partial \operatorname{vech}\left(\Sigma_t\right)'}. \tag{B.4}$$

Let D be the unique $N^2 \times N^*$ duplication matrix for Σ_t to transform $\operatorname{vech}\left(\Sigma_t\right)$ into $\operatorname{vec}\left(\Sigma_t\right)$ and D^+ be the $N^* \times N^2$ Moore-Penrose inverse of D, we then have

$$\begin{aligned}\frac{\partial\left(\Sigma_t^{-1}\cdot\iota\right)}{\partial\,\mathrm{vech}\left(\Sigma_t\right)'} &= \frac{\partial\left(\Sigma_t^{-1}\cdot\iota\right)}{\partial\,\mathrm{vech}\left(\Sigma_t^{-1}\right)}\cdot\frac{\partial\,\mathrm{vech}\left(\Sigma_t^{-1}\right)}{\partial\,\mathrm{vech}\left(\Sigma_t\right)'}\\ &= \frac{\partial\,\mathrm{vec}\left(\Sigma_t^{-1}\cdot\iota\right)}{\partial\,\mathrm{vech}\left(\Sigma_t^{-1}\right)}\cdot\frac{\partial\,\mathrm{vech}\left(\Sigma_t^{-1}\right)}{\partial\,\mathrm{vech}\left(\Sigma_t\right)'}\\ &= \frac{\partial\,\mathrm{vec}\left(\Sigma_t^{-1}\cdot\iota\right)}{\partial\,\mathrm{vec}\left(\Sigma_t^{-1}\right)}\cdot\frac{\partial\,\mathrm{vec}\left(\Sigma_t^{-1}\right)}{\partial\,\mathrm{vech}\left(\Sigma_t^{-1}\right)}\cdot\frac{\partial\,\mathrm{vech}\left(\Sigma_t^{-1}\right)}{\partial\,\mathrm{vech}\left(\Sigma_t\right)'}\\ &= \left(\iota'\otimes I\right)\cdot D\cdot\left(-D^{+}\left(\Sigma_t^{-1}\otimes\Sigma_t^{-1}\right)D\right)\end{aligned} \tag{B.5}$$

and

$$\begin{aligned}\frac{\partial\left(\iota'\cdot\Sigma_t^{-1}\cdot\iota\right)^{-1}}{\partial\,\mathrm{vech}\left(\Sigma_t\right)'} &= \frac{\partial\left(\iota'\cdot\Sigma_t^{-1}\cdot\iota\right)^{-1}}{\partial\left(\iota'\cdot\Sigma_t^{-1}\cdot\iota\right)}\cdot\frac{\partial\left(\iota'\cdot\Sigma_t^{-1}\cdot\iota\right)}{\partial\,\mathrm{vech}\left(\Sigma_t\right)'}\\ &= -\left(\iota'\cdot\Sigma_t^{-1}\cdot\iota\right)^{-2}\cdot\frac{\partial\left(\iota'\cdot\Sigma_t^{-1}\cdot\iota\right)}{\partial\,\mathrm{vech}\left(\Sigma_t\right)'}\\ &= -\left(\iota'\cdot\Sigma_t^{-1}\cdot\iota\right)^{-2}\cdot\frac{\partial\left(\iota'\cdot\Sigma_t^{-1}\cdot\iota\right)}{\partial\,\mathrm{vec}\left(\Sigma_t^{-1}\right)}\cdot\frac{\partial\,\mathrm{vec}\left(\Sigma_t^{-1}\right)}{\partial\,\mathrm{vech}\left(\Sigma_t^{-1}\right)}\cdot\frac{\partial\,\mathrm{vech}\left(\Sigma_t^{-1}\right)}{\partial\,\mathrm{vech}\left(\Sigma_t\right)}\\ &= -\left(\iota'\cdot\Sigma_t^{-1}\cdot\iota\right)^{-2}\cdot\left(\iota'\otimes\iota'\right)\cdot D\cdot\left(-D^{+}\left(\Sigma_t^{-1}\otimes\Sigma_t^{-1}\right)D\right).\end{aligned} \tag{B.6}$$

Let K be the $(N^2\times N^2)$ commutation matrix to transform the vectorized form of a matrix into the vectorized form of its transpose, i.e.,

$$K\cdot\mathrm{vec}\left(\Sigma_t\right)=\mathrm{vec}\left(\Sigma_t'\right). \tag{B.7}$$

Given

$$D\cdot D^{+}=\frac{1}{2}\left(I_{N^2}+K\right) \tag{B.8}$$

and equation (B.5) and (B.6), the equation (B.4) can be expressed as

$$\begin{aligned}\frac{\partial w_t}{\partial\,\mathrm{vech}\left(\Sigma_t\right)} &= \left(\iota'\otimes I\right)D\left(-D^{+}\left(\Sigma_t^{-1}\otimes\Sigma_t^{-1}\right)D\right)\cdot\left(\iota'\Sigma_t^{-1}\iota\right)^{-1}+\\ &\quad\left(\Sigma_t^{-1}\iota\right)\cdot\left[-\left(\iota'\Sigma_t^{-1}\iota\right)^{-2}\left(\iota'\otimes\iota'\right)D\left(-D^{+}\left(\Sigma_t^{-1}\otimes\Sigma_t^{-1}\right)D\right)\right]\\ &= \frac{1}{2\left(\iota'\Sigma_t^{-1}\iota\right)^{2}}\left\{\Sigma_t^{-1}\iota\left(\iota'\otimes\iota'\right)-\left(\iota'\Sigma_t^{-1}\iota\right)\left(\iota'\otimes I\right)\right\}\left(I_{N^2}+K\right)\left(\Sigma_t^{-1}\otimes\Sigma_t^{-1}\right)D.\end{aligned} \tag{B.9}$$

Given S_t is the consistent estimator of Σ_t, the estimated gradient can be calculated as

$$\widehat{G}_t' = \frac{1}{2\left(\iota' S_t^{-1}\iota\right)^2}\left\{S_t^{-1}\iota\left(\iota'\otimes\iota'\right) - \left(\iota' S_t^{-1}\iota\right)\left(\iota'\otimes I\right)\right\}\left(I_{N^2}+K\right)\left(S_t^{-1}\otimes S_t^{-1}\right)D.$$

Combining with the estimated $\widehat{\Pi}_t$, the quarticity of realized GMVP weight w_t can be expressed as $\widehat{G}_t\widehat{\Pi}_t\widehat{G}_t'$.

References

Andersen, T. G. and Bollerslev, T. (1998a). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International economic review*, pages 885–905.

Andersen, T. G. and Bollerslev, T. (1998b). Deutsche mark–dollar volatility: intraday activity patterns, macroeconomic announcements, and longer run dependencies. *the Journal of Finance*, 53(1):219–265.

Andersen, T. G., Bollerslev, T., Christoffersen, P. F., and Diebold, F. X. (2006). Volatility and correlation forecasting. *Handbook of economic forecasting*, 1:777–878.

Andersen, T. G., Bollerslev, T., Christoffersen, P. F., and Diebold, F. X. (2013). Financial risk measurement for financial risk management. In *Handbook of the Economics of Finance*, volume 2, pages 1127–1220. Elsevier.

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

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–625.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Vega, C. (2007). Real-time price discovery in global stock, bond and foreign exchange markets. *Journal of international Economics*, 73(2):251–277.

Andersen, T. G., Bollerslev, T., and Meddahi, N. (2005). Correcting the errors: Volatility

forecast evaluation using high-frequency data and realized volatilities. *Econometrica*, 73(1):279–296.

Archakov, I. and Hansen, P. R. (2021). A new parametrization of correlation matrices. *Econometrica*, 89(4):1699–1715.

Archakov, I., Hansen, P. R., and Lunde, A. (2020). A multivariate realized garch model. *arXiv preprint arXiv:2012.02708*.

Asai, M., McAleer, M., and Medeiros, M. C. (2012). Modelling and forecasting noisy realized volatility. *Computational Statistics & Data Analysis*, 56(1):217–230.

Asai, M., McAleer, M., and Yu, J. (2006). Multivariate stochastic volatility: a review. *Econometric Reviews*, 25(2-3):145–175.

Audrino, F. and Knaus, S. D. (2016). Lassoing the har model: A model selection perspective on realized volatility dynamics. *Econometric Reviews*, 35(8-10):1485–1521.

Bannouh, K., Martens, M., Oomen, R. C., and van Dijk, D. J. (2012). Realized mixed-frequency factor models for vast dimensional covariance estimation. *ERIM Report Series Reference No. ERS-2012-017-F&A*.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2011). Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics*, 162(2):149–169.

Barndorff-Nielsen, O. E., Kinnebrock, S., and Shephard, N. (2008). Measuring downside risk-realised semivariance. *CREATES Research Paper*, (2008-42).

Barndorff-Nielsen, O. E. and Shephard, N. (2001). Modelling by lévy processess for financial econometrics. *Lévy processes*, pages 283–318.

Barndorff-Nielsen, O. E. and Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(2):253–280.

Barndorff-Nielsen, O. E. and Shephard, N. (2004a). Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica*, 72(3):885–925.

Barndorff-Nielsen, O. E. and Shephard, N. (2004b). Power and bipower variation with stochastic volatility and jumps. *Journal of financial econometrics*, 2(1):1–37.

Barndorff-Nielsen, O. E. and Shephard, N. (2006). Econometrics of testing for jumps in financial economics using bipower variation. *Journal of financial Econometrics*, 4(1):1–30.

Bauer, G. H. and Vorkink, K. (2011). Forecasting multivariate realized stock market volatility. *Journal of Econometrics*, 160(1):93–101.

Bauwens, L., Laurent, S., and Rombouts, J. V. (2006). Multivariate garch models: a survey. *Journal of applied econometrics*, 21(1):79–109.

Becker, R., Clements, A. E., Doolan, M. B., and Hurn, A. S. (2015). Selecting volatility forecasting models for portfolio allocation purposes. *International Journal of Forecasting*, 31(3):849–861.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327.

Bollerslev, T., Engle, R. F., and Wooldridge, J. M. (1988). A capital asset pricing model with time-varying covariances. *Journal of political Economy*, 96(1):116–131.

Bollerslev, T., Patton, A. J., and Quaedvlieg, R. (2016). Exploiting the errors: A simple approach for improved volatility forecasting. *Journal of Econometrics*, 192(1):1–18.

Bollerslev, T., Patton, A. J., and Quaedvlieg, R. (2018). Modeling and forecasting (un) reliable realized covariances for more reliable financial decisions. *Journal of Econometrics*, 207(1):71–91.

Bonato, M., Caporin, M., and Ranaldo, A. (2008). Forecasting realized (co) variances with a block structure wishart autoregressive model. *Available at SSRN 1282254*.

Box, G. E. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2):211–243.

Brockwell, P. J. and Davis, R. A. (2009). *Time series: theory and methods*. Springer Science & Business Media.

Brownlees, C. T. and Gallo, G. M. (2006). Financial econometric analysis at ultra-high frequency: Data handling concerns. *Computational statistics & data analysis*, 51(4):2232–2245.

Buccheri, G. and Corsi, F. (2021). Hark the shark: Realized volatility modeling with measurement errors and nonlinear dependencies. *Journal of Financial Econometrics*, 19(4):614–649.

Bucci, A. et al. (2017). Forecasting realized volatility: a review. *Journal of Advanced Studies in Finance (JASF)*, 8(16):94–138.

Callot, L. A., Kock, A. B., and Medeiros, M. C. (2017). Modeling and forecasting large realized covariance matrices and portfolio choice. *Journal of Applied Econometrics*, 32(1):140–158.

Čech, F. and Baruník, J. (2017). On the modelling and forecasting of multivariate realized volatility: Generalized heterogeneous autoregressive (ghar) model. *Journal of Forecasting*, 36(2):181–206.

Chiriac, R. (2007). Nonstationary wishart autoregressive model. Technical report, Citeseer.

Chiriac, R. and Voev, V. (2011). Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics*, 26(6):922–947.

Cipollini, F., Gallo, G. M., and Otranto, E. (2021). Realized volatility forecasting: Robustness to measurement errors. *International Journal of Forecasting*, 37(1):44–57.

Clark, T. E. and West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of econometrics*, 138(1):291–311.

Clements, A. and Preve, D. P. (2021). A practical guide to harnessing the har volatility model. *Journal of Banking & Finance*, 133:106285.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196.

Corsi, F., Mittnik, S., Pigorsch, C., and Pigorsch, U. (2008). The volatility of realized volatility. *Econometric Reviews*, 27(1-3):46–78.

Corsi, F. and Renò, R. (2012). Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling. *Journal of Business & Economic Statistics*, 30(3):368–380.

De Nard, G., Ledoit, O., and Wolf, M. (2021). Factor models for portfolio selection in large dimensions: The good, the better and the ugly. *Journal of Financial Econometrics*, 19(2):236–257.

DeMiguel, V., Nogales, F. J., and Uppal, R. (2014). Stock return serial dependence and out-of-sample portfolio performance. *The Review of Financial Studies*, 27(4):1031–1073.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13(3):253–63.

Dong, Y. and Tse, Y.-K. (2020). Forecasting large covariance matrix with high-frequency data using factor approach for the correlation matrix. *Economics Letters*, 195:109465.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20(3):339–350.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the econometric society*, pages 987–1007.

Engle, R. F. and Gallo, G. M. (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics*, 131(1-2):3–27.

Engle, R. F., Ledoit, O., and Wolf, M. (2019). Large dynamic covariance matrices. *Journal of Business & Economic Statistics*, 37(2):363–375.

Epps, T. W. (1979). Comovements in stock prices in the very short run. *Journal of the American Statistical Association*, 74(366a):291–298.

Fan, J., Liao, Y., and Mincheva, M. (2011). High dimensional covariance matrix estimation in approximate factor models. *Annals of statistics*, 39(6):3320.

Fleming, J., Kirby, C., and Ostdiek, B. (2003). The economic value of volatility timing using "realized" volatility. *Journal of Financial Economics*, 67(3):473–509.

Frisch, R. and Waugh, F. V. (1933). Partial time regressions as compared with individual trends. *Econometrica: Journal of the Econometric Society*, pages 387–401.

Giacomini, R. and White, H. (2006). Tests of conditional predictive ability. *Econometrica*, 74(6):1545–1578.

Golosnoy, V. and Gribisch, B. (2022). Modeling and forecasting realized portfolio weights. *Journal of Banking & Finance*, page 106404.

Golosnoy, V., Gribisch, B., and Liesenfeld, R. (2012). The conditional autoregressive wishart model for multivariate stock market volatility. *Journal of Econometrics*, 167(1):211–223.

Golosnoy, V., Gribisch, B., and Seifert, M. I. (2019). Exponential smoothing of realized portfolio weights. *Journal of Empirical Finance*, 53:222–237.

Golosnoy, V., Schmid, W., Seifert, M. I., and Lazariv, T. (2020). Statistical inferences for realized portfolio weights. *Econometrics and Statistics*, 14:49–62.

Gouriéroux, C., Jasiak, J., and Sufana, R. (2009). The wishart autoregressive process of multivariate stochastic volatility. *Journal of Econometrics*, 150(2):167–181.

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

Hansen, P. R. and Lunde, A. (2006). Consistent ranking of volatility models. *Journal of Econometrics*, 131(1-2):97–121.

Hansen, P. R., Lunde, A., and Nason, J. M. (2003). Choosing the best volatility models: the model confidence set approach. *Oxford Bulletin of Economics and Statistics*, 65:839–861.

Hansen, P. R., Lunde, A., and Nason, J. M. (2005). Model confidence sets for forecasting models. Technical report, Working paper.

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

Hansen, P. R., Lunde, A., and Voev, V. (2014). Realized beta garch: A multivariate garch model with realized measures of volatility. *Journal of Applied Econometrics*, 29(5):774–799.

Hautsch, N., Kyj, L. M., and Oomen, R. C. (2012). A blocking and regularization approach to high-dimensional realized covariance estimation. *Journal of Applied Econometrics*, 27(4):625–645.

Jagannathan, R. and Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*, 58(4):1651–1683.

James, W. and Stein, C. (1992). Estimation with quadratic loss. In *Breakthroughs in statistics*, pages 443–460. Springer.

Koop, G., Korobilis, D., and Pettenuzzo, D. (2019). Bayesian compressed vector autoregressions. *Journal of Econometrics*, 210(1):135–154.

Ledoit, O. and Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of empirical finance*, 10(5):603–621.

Lovell, M. C. (1963). Seasonal adjustment of economic time series and multiple regression analysis. *Journal of the American Statistical Association*, 58(304):993–1010.

Lunde, A., Shephard, N., and Sheppard, K. (2016). Econometric analysis of vast covariance matrices using composite realized kernels and their application to portfolio choice. *Journal of Business & Economic Statistics*, 34(4):504–518.

Luo, J. and Chen, L. (2020). Realized volatility forecast with the bayesian random compressed multivariate har model. *International Journal of Forecasting*, 36(3):781–799.

Marcellino, M., Stock, J. H., and Watson, M. W. (2006). A comparison of direct and iterated multistep ar methods for forecasting macroeconomic time series. *Journal of econometrics*, 135(1-2):499–526.

Markowitz, H. (1952a). Portfolio selection. *The Journal of Finance*, 7(1):77–91.

Markowitz, H. (1952b). Portfolio selection. *The Journal of Finance*, 7(1):77–91.

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of financial economics*, 8(4):323–361.

Mincer, J. A. and Zarnowitz, V. (1969). The evaluation of economic forecasts. In *Economic forecasts and expectations: Analysis of forecasting behavior and performance*, pages 3–46. NBER.

Nelson, D. B. (1992). Filtering and forecasting with misspecified arch models i: Getting the right variance with the wrong model. *Journal of econometrics*, 52(1-2):61–90.

Noureldin, D., Shephard, N., and Sheppard, K. (2012). Multivariate high-frequency-based volatility (heavy) models. *Journal of Applied Econometrics*, 27(6):907–933.

Oh, D. H. and Patton, A. J. (2016). High-dimensional copula-based distributions with mixed frequency data. *Journal of Econometrics*, 193(2):349–366.

Oomen, R. C. (2001). Using high frequency stock market index data to calculate, model & forecast realized return variance. *European Univ., Economics Discussion Paper*, (2001/6).

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

Patton, A. J. and Sheppard, K. (2009). Evaluating volatility and correlation forecasts. In *Handbook of financial time series*, pages 801–838. Springer.

Patton, A. J. and Sheppard, K. (2015). Good volatility, bad volatility: Signed jumps and the persistence of volatility. *Review of Economics and Statistics*, 97(3):683–697.

Shephard, N. and Sheppard, K. (2010). Realising the future: forecasting with high-frequency-based volatility (heavy) models. *Journal of Applied Econometrics*, 25(2):197–231.

Sizova, N. (2011). Integrated variance forecasting: Model based vs. reduced form. *Journal of Econometrics*, 162(2):294–311.

Sowell, F. (1992). Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of econometrics*, 53(1-3):165–188.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

Tibshirani, R. J. and Taylor, J. (2011). The solution path of the generalized lasso. *The annals of statistics*, 39(3):1335–1371.

Tsai, A. C., Liou, M., Simak, M., and Cheng, P. E. (2017). On hyperbolic transformations to normality. *Computational Statistics & Data Analysis*, 115:250–266.

Visser, M. P. (2011). Garch parameter estimation using high-frequency data. *Journal of Financial Econometrics*, 9(1):162–197.

West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica: Journal of the Econometric Society*, pages 1067–1084.

Yamada, H. (2017). The frisch–waugh–lovell theorem for the lasso and the ridge regression. *Communications in Statistics-Theory and Methods*, 46(21):10897–10902.