Graph-based Methods for Forecasting Realized Covariances

Chao Zhang*[1,2], Xingyue Pu*[3,4], Mihai Cucuringu[1,2,4,5], and Xiaowen Dong[3,4]

[1]Department of Statistics, University of Oxford, Oxford, UK

[2]Mathematical Institute, University of Oxford, Oxford, UK

[3]Department of Engineering Science, University of Oxford, Oxford, UK

[4]Oxford-Man Institute of Quantitative Finance, University of Oxford, Oxford, UK

[5]The Alan Turing Institute, London, UK

November 2022

Abstract

We forecast the realized covariance matrix of asset returns in the U.S. equity market by exploiting the predictive information of graphs in volatility and correlation. Specifically, we augment the Heterogeneous Autoregressive (HAR) model via neighborhood aggregation on these graphs. Our proposed method allows for the modeling of interdependence in volatility (also known as spillover effect) and correlation, while maintaining parsimony and interpretability. We explore various graph construction methods, including sector membership and graphical LASSO (for modeling volatility), and line graph (for modeling correlation). The results generally suggest that the augmented model incorporating graph information yields both statistically and economically significant improvements for out-of-sample performance over the traditional models. Such improvements remain significant over horizons up to one month ahead, but decay in time. The robustness tests demonstrate that the forecast improvements are obtained consistently over the different out-of-sample sub-periods, and are insensitive to measurement errors of volatilities.

Keywords: Realized covariance, HAR, Graphical LASSO, Line graph, Graph learning.

JEL Codes: C31, C53, C58, G17

*Equal contribution. Correspondence to: Chao Zhang <chao.zhang@stats.ox.ac.uk>. The authors thank the Oxford Suzhou Centre for Advanced Research for providing the computational facilities.

1 Introduction

Covariance matrix estimation is of particular importance in areas such as risk management, asset pricing and portfolio optimization. Traditional econometric approaches include the multivariate Generalized Autoregressive Conditional Heteroscedasticity (GARCH) type of models (for detailed reviews see Bauwens et al. (2006)) and the multivariate Stochastic Volatility (SV) type of models (see Asai et al. (2006)). In the multivariate GARCH and SV frameworks, the covariances are treated as latent variables and may suffer from the curse of dimensionality problem due to the use of low-frequency data (see Andersen et al. (2003)).

More recently, the increased availability of reliable high-frequency intraday asset prices has motivated researchers to model and forecast the nonparametric and ex-post Realized Covariance (RC) measures constructed from intraday data (see Andersen et al. (2003, 2011); Sheppard (2010); Barndorff-Nielsen et al. (2011); Varneskov and Voev (2013)). It has been extensively documented in the literature that the high-frequency-based realized covariances are superior to those constructed from low-frequency return data, for practical investment and portfolio allocation decisions (for example Hautsch et al. (2015); Bollerslev et al. (2018); Symitsi et al. (2018)).

There is a large volume of studies dedicated to forecasting realized covariance matrices. Chiriac and Voev (2011) extended the univariate Heterogeneous Autoregressive (HAR) model of Corsi (2009), to a multivariate setting via the Cholesky decomposition on the covariance matrix. Cholesky decomposition is also used in some papers for modeling RC (probably with some variations), such as Symitsi et al. (2018); Bucci (2020a); Fiszeder and Orzeszko (2021); Bollerslev et al. (2021). Inspired by the dynamic conditional correlation (DCC) model of Engle (2002), an alternative approach to forecasting realized covariances is based on the decomposition of the covariance matrix into the volatilities and correlations components, also know as the DRD decomposition (Lee and Long (2009); Oh and Patton (2016); Vassallo et al. (2021)). Bollerslev et al. (2018) reported that the covariance forecasts based on the DRD decomposition generally outperform the ones based on the Cholesky decomposition. In this paper, we pay more attention to the methods based on the DRD decomposition. We will briefly review the related literature on modeling and forecasting realized volatilities and correlations in the next paragraphs.

There have been numerous contributions made to the literature on the topic of forecasting daily realized volatility (RV). Andersen et al. (2003) proposed the AutoRegressive Fractionally Integrated Moving Average (ARFIMA) model for forecasting daily RVs. Corsi (2009) put forward the Heterogeneous Autoregressive (HAR) for predicting daily RVs using various realized volatility components over different time horizons. Izzeldin et al. (2019) compared the forecasting performance of ARFIMA and HAR, and concluded their performance is essentially indistinguishable. These methods may give important information about the

dynamics of volatilities, but ignored the interdependence of volatilities among assets, as pointed out by Bollerslev et al. (2014); Cubadda et al. (2017).

Several well-established studies have examined the interdependence of volatilities within the asset market or across various markets (see Baele (2005); Buncic and Gisler (2016); Degiannakis and Filis (2017); Wang et al. (2018)). This phenomenon is also known as *volatility spillover* effect, which implies that a large shock of a specific asset (or market) may not only increase its own subsequent volatility, but also that of other assets (or markets). The volatility spillover effect can be construed as a type of network effect. It is believed that the panic followed by a shock may easily spread through a network of interdependent financial assets (or markets) (see King and Wadhwani (1990); Schwert (1990); Glasserman and Young (2015)). Such a type of interdependence can be leveraged to improve the volatility prediction of the target assets or markets. For example, Buncic and Gisler (2016) revealed the cross-market volatility spillover effect from the U.S. equity market to 17 global foreign asset markets, and showed how VIX plays an important role in predicting the volatility in all of these 17 markets. Similarly, Wang et al. (2018) showed that accounting for spillover information from the U.S. market can significantly improve the forecasting accuracy of international stock price volatility. Degiannakis and Filis (2017) found evidence that different assets, such as stocks and foreign currencies, can improve the prediction of the realized volatility of crude oil price.

In literature, the volatility spillover effect can be modeled through multivariate GARCH, stochastic volatility models, Vector Autoregression (VAR), Wishart Autoregression (WAR) and etc. The BEKK-GARCH introduced by Engle and Kroner (1995) and the VAR-GARCH proposed by Ling and McAleer (2003) are probably two popular GARCH-type models used to analyze volatility spillover with low-frequency data. The standard WAR model proposed by Gouriéroux et al. (2009) is built on latent processes and unable to capture the long memory dependence in volatility. In terms of modeling realized volatility, Wilms et al. (2021) used VAR to obtain the multivariate volatility forecasts for stock market indices. However, Callot et al. (2017) pointed out that the aforementioned models may deliver poor out-of-sample forecasts due to the curse of dimensionality. This is mainly because the dependence of each pair of financial assets or markets is modeled as an individual feature, which leads to large computational burdens. We are thus interested into a parsimonious model that well integrate the complex interdependence.

On the other hand, correlation forecasting has not yet drawn much attention in literature, to the best of our knowledge. In the classic dynamic conditional correlation (DCC) model of Engle (2002), the temporal dependencies in the conditional correlations are characterized by a simple scalar GARCH(1, 1) model, where the dynamics are imposed to be equal for all the correlation pairs. In terms of realized correlation modeling, Andersen et al. (2001) studied the impact of past realized volatilities and past returns on forecasting realized daily correlations. Oh and Patton (2016); Bollerslev et al. (2018) employed a HAR-type model for forecasting

realized correlations. While these models are effective, but not flexible to jointly model all the elements in the correlation matrix. Some generalizations of the above models, e.g. allowing each unique correlation pair to follow its own dynamics, require much more free parameters, therefore are not feasible for high-dimensional applications.

Although there seems to be general agreement that pair correlations of asset returns vary over time, it is less clear whether and how correlations affect each other, i.e. the existence of interdependence of correlation pairs. For example, the correlation between two equities APPL and IBM, denoted as APPL-IBM, might have influence on other correlations that share an equity, e.g. IBM-MSFT and AMZN-JPM. In the literature, King and Wadhwani (1990) noticed a significant increase in the correlations during crisis periods. Berben and Jansen (2005) documented that correlations among the U.S. stock market have increased significantly during 1980-2000. Boyer et al. (1999); Forbes and Rigobon (2002) further revealed that the changes in correlations might be due to a simultaneous increase in the volatility of asset prices. Thus, we are interested to model such co-movement or interdependence between correlation pairs, and subsequently to examine whether such effects could improve the modeling of correlation matrices.

After seeing in the previous literature that the interdependence of volatility (i.e. volatility spillover effect) may help improve the forecasting accuracy of multivariate volatility, we are particularly interested into the following research questions: (i) How do we capture the time-varying interdependence in volatilities and correlation pairs? (ii) Are the two types of interdependence predictive for covariance matrices in the U.S. equity market? (iii) Is there a parsimonious and interpretable forecasting model for realized covariances that well incorporates these interdependence?

To answer the above research questions, we utilize graphs as the mathematical model to represent two types of interdependence in the U.S. equity market. In the first type of graph, each asset is modeled as a node and an edge connecting two nodes represents the existence of any dependence between their volatilities. In the second type of graph, each node is a correlation pair of two assets and an edge connecting two nodes represents the relations between one pair and another pair. Some related graph construction and estimation methods will be reviewed briefly in Section 3.1.

The perspective of representing the above interdependence as a graph allows us to adopt the modeling tools developed specifically for the analysis of graph-structured data. In particular, we are able to apply neighborhood[1] aggregation to generate a new input feature for every underlying asset and incorporate it into the traditional models for volatility forecasting. For correlation forecasting, we could generate a similar feature for every pair of assets. Our aggregated feature summarizes the influence of the connected assets

[1] Andrada-Félix et al. (2016) demonstrated the merit of the neighbor information for realized variances forecasting via the nearest neighbor approach.

(resp. pairs of assets) on a specific asset (resp. a specific pair correlation). Neighborhood aggregation[2] is an effective method to integrate the structural interdependence of all the node pairs into the features observed on every node while maintaining a low dimension. In fact, the aggregated features remain of the same dimension as the original features. Therefore, our model can avoid the curse of dimensionality and be applied to a large universe of assets.

To some extent, our framework may have some conceptual similarity to the work of Zhu et al. (2017), who proposed a network vector autoregressive model, assuming each node's response at a given time point as a linear combination of (a) its previous value, (b) the average of its connected neighbors, (c) a set of node-specific covariates and (d) an independent noise. However, our framework addresses the important challenges of constructing desirable graphs, not only for enhancing predictive power on covariances but also for understanding the dynamics of covariances. Our novel framework also allows for time-varying graphs, which can well capture the dynamics in graph effects of volatilities and correlations, while Zhu et al. (2017) assumed a constant graph. Moreover, moving to the covariance matrix forecasting problem opens up a host of interesting and practically relevant econometric applications, as illustrated in our following analysis.

The main contributions of this paper are summarized as follows. First, we put forward a methodology to model and forecast realized covariance matrices by leveraging the graph information. Our method may shed new light on the dynamic interdependence of realized volatility and correlation, through the lens of neighborhood aggregation. Second, we apply our method to forecasting realized covariance matrices over the daily, weekly, and monthly future horizons, in order to study the variation in the predictive power of graphs across time. Last, we examine the stability and robustness of our model through comprehensive experiments.

More specifically, using realized volatility data for the components of the Dow Jones Industrial Average, spanning the period from July 2007 to June 2021, we find that graph information of volatility and correlation can be used to improve the forecasts of realized covariance matrices. Our in-sample results first show that short-term volatility is the most important source for future volatility forecasting, while the correlation model attributes the most importance to the mid-term component. This highlights the different dependencies in volatilities and correlations. Moreover, our in-sample analysis shows that graph-based predictors for volatility (resp. correlation) modeling are highly significant, and including these graph components in the traditional models can substantially improve the in-sample fit accuracy of realized covariance matrices.

By performing the out-of-sample tests in a rolling window setup, our analysis provides strong evidence that the graph-based model yields superior 1-day-ahead forecasts over a strong set of traditional baselines. The MCS test (Hansen et al. (2011)) indicates that the augmented model with a data-driven graph determined by Graphical LASSO (Friedman et al. (2008)) is consistently included in the subset of best forecasting

[2]It was firstly proposed to predict chemical properties of molecules by Gilmer et al. (2017).

models irrespective of the loss functions. The graph analysis highlights the time-varying nature of graph effects both in volatilities and correlations. We also investigate the economic benefits of covariance forecasts from different models via the global minimum variance portfolio (GMVP). The results show that portfolios employing graph information are able to generate significantly lower out-of-sample variance compared to the traditional models without graph information or the naive equally-weighted portfolio (see DeMiguel et al. (2009)). The forecast improvements over longer horizons (1-week, 1-month) remain significant, but decay as the prediction horizon gets longer. We conclude the paper by several robustness tests, which demonstrate the forecast improvements are experienced consistently over the different out-of-sample sub-periods and are insensitive to measurement errors of volatilities (Bollerslev et al. (2016, 2018)).

The rest of the paper is structured as follows. We introduce the estimation of realized covariance matrices from high-frequency data and the baseline models in Section 2. We present the background of graph and propose a new methodology based on graphs for modeling and forecasting realized covariance matrices in Section 3. Section 4 provides a description of dataset, evaluation criterion, in-sample analysis and out-of-sample results. In Section 5, we perform several robustness tests. Finally, we conclude our analysis in Section 6 and highlight potential future research directions.

2 Traditional models

This section outlines the approach for the estimation of realized covariances, and two traditional models that we use as baselines for forecasting realized covariances in the U.S. equity market.

2.1 Problem set-up

Let us consider a vector of asset returns $\boldsymbol{r}_t = (r_{1,t}, \ldots, r_{N,t})'$ as follows

$$\boldsymbol{r}_t = \boldsymbol{\mu}_t + \boldsymbol{\epsilon}_t, \tag{1}$$

where N is the number of assets, and $\boldsymbol{\mu}_t$ is the conditional mean vector given the information set $\mathcal{F}_{t-1}$. $\boldsymbol{\epsilon}_t$ is a vector of innovations, which can be expressed as $\boldsymbol{\epsilon}_t = \boldsymbol{\Sigma}_t^{1/2}\boldsymbol{z}_t$, where $\boldsymbol{z}_t$ is a random vector that follows a multivariate standard normal distribution, i.e. $\boldsymbol{z}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_N)$ ($\boldsymbol{I}_N$ is an $N \times N$ identity matrix). Therefore, $\boldsymbol{\Sigma}_t$ is the conditional covariance matrix of $\boldsymbol{r}_t$. The square root of $\boldsymbol{\Sigma}_t$ is not unique, as long as the operator satisfies the condition that $\boldsymbol{\Sigma}_t^{1/2}\boldsymbol{\Sigma}_t^{1/2} = \boldsymbol{\Sigma}_t$.

Note that $\boldsymbol{\Sigma}_t$ is unobservable. One consistent estimator of $\boldsymbol{\Sigma}_t$ is the multivariate realized covariance

(RC), which is constructed from the intra-day returns, e.g. 5-min returns.[3] Specifically, we denote $\boldsymbol{r}_{t(l)}$ as the l-th vector of non-overlapping Δ-min log-returns during day t, i.e. $\boldsymbol{r}_{t(l)} = \log \boldsymbol{p}_{t(l\Delta)} - \log \boldsymbol{p}_{t((l-1)\Delta)}$, where $\boldsymbol{p}_{t(l\Delta)}$ is the price at time $l\Delta$ of day t. Then the RC matrix of day t can be computed as

$$\boldsymbol{H}_t := \sum_{l=1}^{M} \boldsymbol{r}_{t(l)} \boldsymbol{r}'_{t(l)} \tag{2}$$

where $M = 390/\Delta$ is the number of non-overlapping Δ-min slots during day t. The general properties of the RC estimator are summarized by Sheppard (2010).

In addition, we adopt the subsampling averaging method (see Sheppard (2010); Andersen et al. (2011); Varneskov and Voev (2013)) to improve the above RC estimation, which uses all Δ-minute returns, not just non-overlapping ones. For example, suppose prices are sampled every minute. The standard realized covariance uses 5-min returns constructed from prices sampled at 09:30, 09:35, 09:40, ..., 15:55, 16:00. The subsampled realized covariance uses returns computed from all 5-minute windows, i.e., 09:30 and 09:35, 09:31 and 09:36, 09:32 and 09:37, and so on. In the end, the subsampled realized covariance is computed by the average of all standard realized covariances. Covariances over longer horizons of k days are computed by the sum of daily realized covariances (see Symitsi et al. (2018)).

The daily realized volatility for a particular asset i at day t is $RV_{i,t} = \boldsymbol{H}_t[i,i] = \sum_{l=1}^{M} r^2_{i,t(l)}$. We refer to $\boldsymbol{RV}_t = (RV_{1,t}, \ldots, RV_{N,t})'$ as the vector of cross-sectional realized volatilities.

2.2 HAR-Cholesky

Corsi (2009) proposed a Heterogeneous Autoregressive Regression (HAR) model for modeling and forecasting the realized volatility where the lagged daily, weekly and monthly volatility components are incorporated as predictors. For a given asset i, its realized volatility of day t is modeled as

$$RV_{i,t} = \alpha + \beta_d RV_{i,t-1} + \beta_w RV_{i,t-5:t-2} + \beta_m RV_{i,t-22:t-6} + u_{i,t}, \tag{3}$$

where $RV_{i,t-5:t-2} = \frac{1}{4}\sum_{k=2}^{5} RV_{i,t-k}$, $RV_{i,t-22:t-6} = \frac{1}{17}\sum_{k=6}^{22} RV_{i,t-k}$ denote the weekly and monthly lagged realized volatility, respectively. The choice of a daily, weekly and monthly lag is aiming to capture the long-memory dynamic dependencies observed in most realized volatility series.

Chiriac and Voev (2011) generalized the univariate HAR model to estimate the realized covariances as a linear combination of past daily, weekly and monthly realized covariances, in the form

[3] Liu et al. (2015) demonstrate that no sub-sampling frequency significantly outperforms a 5-min interval in terms of forecasting daily RVs, making it a widely accepted time interval in the literature.

$$\boldsymbol{y}_t = \alpha^{(C)} + \beta_d^{(C)}\boldsymbol{y}_{t-1} + \beta_w^{(C)}\boldsymbol{y}_{t-5:t-2} + \beta_m^{(C)}\boldsymbol{y}_{t-22:t-6} + \boldsymbol{u}_t^{(C)}, \tag{4}$$

where $\boldsymbol{y}_t = \text{vech}_{Chol}(\boldsymbol{H}_t)$ denotes the $N^* = N(N+1)/2$ dimensional vectorized version of the Cholesky decomposition of $\boldsymbol{H}_t$ and $\boldsymbol{y}_{t-5:t-2}$ (resp. $\boldsymbol{y}_{t-22:t-6}$) is computed as $\frac{1}{4}\sum_{k=2}^{6}\boldsymbol{y}_{t-k}$ (resp. $\frac{1}{17}\sum_{k=6}^{22}\boldsymbol{y}_{t-k}$), following Symitsi et al. (2018); Bollerslev et al. (2018). The intercept α is a N^*-dimensional vector, while the β_d, β_w and β_m parameters are all assumed to be scalar. $\boldsymbol{u}_t^{(C)}$ is the error term with conditional mean zeros.

2.3 HAR-DRD

Andersen et al. (2006) demonstrated that the dynamic dependencies in the correlations are different from the volatilities, also known as *correlation breakdown*. Oh and Patton (2016) proposed the HAR-DRD model, which is based on the decomposition of the covariance matrix into the diagonal matrix of realized volatilities and the correlation matrix

$$\boldsymbol{H}_t = \boldsymbol{D}_t\boldsymbol{R}_t\boldsymbol{D}_t, \tag{5}$$

where $\boldsymbol{D}_t$ is the diagonal matrix with the elements of the square roots of $\boldsymbol{RV}_t$ on the main diagonal, i.e. $\boldsymbol{D}_t[i,i] = \sqrt{RV_{i,t}}, \forall i$, and $\boldsymbol{D}_t[i,j] = 0, \forall i \neq j$. $\boldsymbol{R}_t$ is the correlation matrix.

In the HAR-DRD model, the realized variance vector is modeled by vectorized HAR model as in Eqn (6). The realized correlation matrix is modeled using a scalar multivariate HAR model as in Eqn (7), in consistent with Bollerslev et al. (2018).

$$\boldsymbol{RV}_t = \alpha_0^{(D)} + \beta_d^{(D)}\boldsymbol{RV}_{t-1} + \beta_w^{(D)}\boldsymbol{RV}_{t-5:t-2} + \beta_m^{(D)}\boldsymbol{RV}_{t-22:t-6} + \boldsymbol{u}_t^{(D)}, \tag{6}$$

$$\boldsymbol{x}_t = \alpha_0^{(R)} + \beta_d^{(R)}\boldsymbol{x}_{t-1} + \beta_w^{(R)}\boldsymbol{x}_{t-5:t-2} + \beta_m^{(R)}\boldsymbol{x}_{t-22:t-6} + \boldsymbol{u}_t^{(R)}, \tag{7}$$

where $\boldsymbol{x}_t = \text{vech}_{Tri}(\boldsymbol{R}_t)$ is the $N^{\#} = N(N-1)/2$ dimensional vectorized version of the lower triangular of $\boldsymbol{R}_t$ and $\boldsymbol{x}_{t-5:t-2}$ (resp. $\boldsymbol{x}_{t-22:t-6}$) is computed as $\frac{1}{4}\sum_{k=2}^{6}\boldsymbol{x}_{t-k}$ (resp. $\frac{1}{17}\sum_{k=6}^{22}\boldsymbol{x}_{t-k}$).

3 Proposed models

Before introducing our new forecasting models based on graphs, we first provide some graph-related definitions for better understanding.

3.1 Background on graph construction and estimation

Definition 3.1 (Graph). *A graph $\mathcal{G}$ is defined as $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, where $\mathcal{V} = \{v_1, \ldots, v_N\}$ is a set of N nodes and $\mathcal{E}$ is a set of edges, where $e_{ij} = (v_i, v_j) \in \mathcal{E}$ denotes an edge connecting node v_i and node v_j.*

Definition 3.2 (Adjacency Matrix). *An adjacency matrix $\mathbf{A}$ is a square matrix whose dimension is $N \times N$, where $\mathbf{A}[i,j]$ represents the connection between v_i and v_j in the graph $\mathcal{G}$. An adjacency matrix can be weighted, where $\mathbf{A}[i,j] \geq 0,\ \forall i,j$ represents the strength/intensity of the connection between nodes v_i and v_j. If $\mathbf{A}[i,j] \in \{0,1\},\ \forall i,j$, the graph is a binary graph. In this paper, we mainly consider the **binary** graph.*

Example 1 (Two Trivial Graphs)

The adjacency matrix of a **complete** graph (often referred to as K in the literature) contains all ones except along the diagonal where there are only zeros. The adjacency matrix of an **empty** graph is a zero matrix.

Example 2 (The GICS Sector Graph)

The Global Industry Classification Standard (GICS) is an industry taxonomy developed for classifying major public companies. A GICS sector graph considers an edge between two asset nodes if they belong to the same GICS sector. Table 2 will introduce the corresponding sector of each stock under consideration.

Graphical LASSO (GLASSO) is a sparsity-penalized maximum likelihood estimator for the precision matrix (i.e. inverse of the covariance matrix) proposed by Friedman et al. (2008). Assume observations of $\boldsymbol{\epsilon}_t$ in Eqn (1) are drawn from a multivariate Gaussian distribution $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$. If the ij-th component of the precision matrix is zero, then the i-th and j-th variables are conditionally independent, given the other variables (for further details, see Meinshausen and Bühlmann (2006); Friedman et al. (2008)). Alternatively, zeros in the precision matrix indicate conditional independence. Thus we are interested in estimating the precision matrix $\boldsymbol{\Theta} = \boldsymbol{\Sigma}^{-1}$. The graphical LASSO algorithm aims to obtain an estimator $\hat{\boldsymbol{\Theta}}$ that minimizes the following objective function

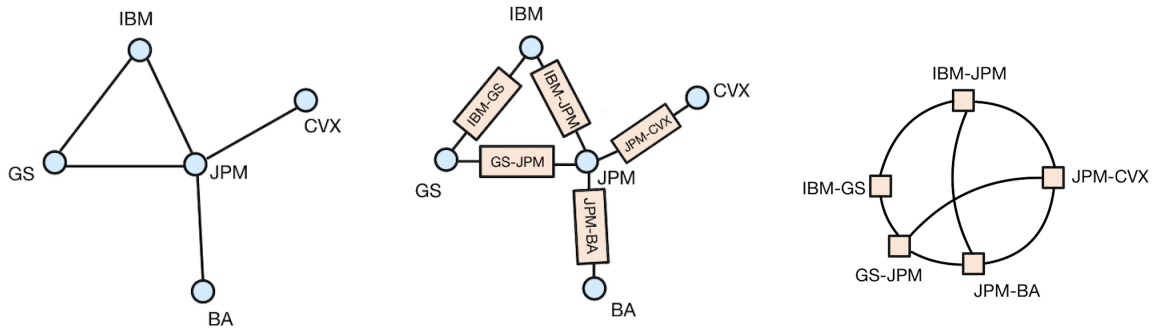
$$\hat{\boldsymbol{\Theta}} = \text{argmin}_{\boldsymbol{\Theta} \succeq 0} \left(\text{Tr}(\boldsymbol{S}\boldsymbol{\Theta}) - \log\det(\boldsymbol{\Theta}) + \lambda \sum_{j \neq k} |\boldsymbol{\Theta}_{jk}| \right),$$

where $\boldsymbol{S}$ is the sample covariance of $\boldsymbol{\epsilon}_t$, and λ is the penalizing trade-off parameter that controls the sparsity of $\hat{\boldsymbol{\Theta}}$. In this work, we use the standard cross-validation procedure to choose λ. The estimated precision matrix is a graph defined in a Gaussian Markov random field. We remove these zero edges in the estimated precision matrix (i.e. those connecting conditionally-independent node pairs) and consider the non-zero elements as edges. In other words, the adjacency matrix $\boldsymbol{A}$ from GLASSO is defined as $\boldsymbol{A}[i,j] = 1$ if $\hat{\boldsymbol{\Theta}}[i,j] \neq 0$; otherwise $\boldsymbol{A}[i,j] = 0$.

Definition 3.3 (Line Graph)**.** *Given a graph $\mathcal{G}$, its line graph $L(\mathcal{G})$ is a graph such that*

- *each node of $L(\mathcal{G})$ represents an edge of $\mathcal{G}$;*
- *two nodes of $L(\mathcal{G})$ are adjacent if and only if their corresponding edges share a common endpoint in $\mathcal{G}$.*

Figure 1 shows an illustration of graphs defined in our framework. Figure 1(a) models the interdependence between the volatility of assets, where each node represents an asset. Their edges, as shown in Figure 1(b), constitute a node in the line graph shown in Figure 1(c). The edges in the line graph capture the interdependence between two correlation pairs that have a asset in common.



(a) Graph $\mathcal{G}$ for volatility (b) Labeled edges in $\mathcal{G}$ (c) Line graph $L(\mathcal{G})$ for correlation

Figure 1: An illustration of the process of building the line graph $L(\mathcal{G})$ for $N = 5$ assets.

3.2 Forecasting the realized covariance matrix with graphs

Based on the HAR-DRD model in Eqn (5), we propose to forecast the realized covariance matrix by modeling the realized volatilites and correlation matrix separately. For each sub-task, we incorporate new features that represent the structural information contained in the aforementioned graphs via neighborhood aggregation.

3.2.1 Forecasting realized volatilities with graphs

The HAR model (Eqn (3)) assumes that the future volatility of a specific asset only relies on its own past volatilities. Considering that volatilities of connected assets may have predictive power on the volatility of the target asset, we propose the following model to aggregate the spillover effects from neighbors (i.e. connected assets), which we denote as the Graph-HAR model (or in short GHAR)

$$\begin{aligned}\textbf{GHAR}(\boldsymbol{A}): \quad \boldsymbol{RV}_t = \alpha^{(D)} &+ \underbrace{\beta_d^{(D)}\boldsymbol{RV}_{t-1} + \beta_w^{(D)}\boldsymbol{RV}_{t-5:t-2} + \beta_m^{(D)}\boldsymbol{RV}_{t-22:t-6}}_{Self} \\ &+ \underbrace{\gamma_d^{(D)}\boldsymbol{W}\cdot\boldsymbol{RV}_{t-1} + \gamma_w^{(D)}\boldsymbol{W}\cdot\boldsymbol{RV}_{t-5:t-2} + \gamma_m^{(D)}\boldsymbol{W}\cdot\boldsymbol{RV}_{t-22:t-6}}_{Graph} + \boldsymbol{u}_t^{(D)},\end{aligned} \tag{8}$$

where $\boldsymbol{W} = \boldsymbol{O}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{O}^{-\frac{1}{2}}$ is the normalized adjacency matrix. Specifically, $\boldsymbol{A}$ is a $N \times N$ adjacency matrix indicating the connections between assets with diagonal elements as 0, and $\boldsymbol{O} = \text{diag}\{n_1, \ldots, n_N\}$, where $n_i = \sum_j \boldsymbol{A}[i,j]$, $\forall i$. Therefore $\boldsymbol{W}\cdot\boldsymbol{RV}_{t-1}, \boldsymbol{W}\cdot\boldsymbol{RV}_{t-5:t-2}, \boldsymbol{W}\cdot\boldsymbol{RV}_{t-22:t-6}$ represent the **neighborhood aggregation** over daily, weekly and monthly horizons. γ_d, γ_w, γ_m represent the effects from connected neighbors over different horizons. For simplicity, we denote the effect from a stock's own historical events as *Self*, and that from a graph as *Graph*, as shown in Eqn (8). We use the Ordinary Least Squares (OLS) to obtain the estimates of Eqn (8).

In terms of the choice of adjacency matrices for volatility modeling, we first consider an empty graph, i.e. the elements of $\boldsymbol{A}$ are all 0s. In this case, Eqn (8) reduces to the standard vectorized HAR model (6) for modeling volatilities. When the off-diagonal elements of $\boldsymbol{A}$ are all 1s, i.e. a complete graph, $\boldsymbol{W}\cdot\boldsymbol{RV}_{t-1}$ represents the global volatility as studied in Bollerslev et al. (2018). Given that companies from the same sector tend to behave similarly and also influence each other, the GICS sector is another frequently-used knowledge-based graph for the construction of $\boldsymbol{A}$, e.g. Fan et al. (2016). In addition, we employ a data-driven approach, i.e. graphical LASSO, to compute the adjacency matrix (see Hallac et al. (2017)).

3.2.2 Forecasting realized correlations with graphs

Moreover, we apply the idea of the graph effect to modeling correlations according to the model

$$\begin{aligned}\textbf{GHAR}(\widetilde{\boldsymbol{A}}): \quad \boldsymbol{x}_t = \alpha^{(R)} &+ \underbrace{\beta_d^{(R)}\boldsymbol{x}_{t-1} + \beta_w^{(R)}\boldsymbol{x}_{t-5:t-2} + \beta_m^{(R)}\boldsymbol{x}_{t-22:t-6}}_{Self} \\ &+ \underbrace{\gamma_d^{(R)}\widetilde{\boldsymbol{W}}\boldsymbol{x}_{t-1} + \gamma_w^{(R)}\widetilde{\boldsymbol{W}}\boldsymbol{x}_{t-5:t-2} + \gamma_m^{(R)}\widetilde{\boldsymbol{W}}\boldsymbol{x}_{t-22:t-6}}_{Graph} + \boldsymbol{u}_t^{(R)},\end{aligned} \tag{9}$$

where $\widetilde{\boldsymbol{W}} = \widetilde{\boldsymbol{O}}^{-\frac{1}{2}}\widetilde{\boldsymbol{A}}\widetilde{\boldsymbol{O}}^{-\frac{1}{2}}$ is the normalized adjacency matrix. Specifically, $\widetilde{\boldsymbol{A}}$ is a $N^{\#} \times N^{\#}$ ($N^{\#} = N(N-1)/2$) adjacency matrix indicating the connections between pairwise correlations with diagonal elements as 0, and $\widetilde{\boldsymbol{O}} = \text{diag}\{\tilde{n}_1, \ldots, \tilde{n}_{N^{\#}}\}$, where $\tilde{n}_i = \sum_j \widetilde{\boldsymbol{A}}[i,j]$, $\forall i$.

In terms of adjacency matrices of correlation networks, in addition to the empty graph (all 0's) and complete graph (all 1's), we also adopt the line graph, where the correlation $\boldsymbol{R}[i,j]$ between asset i and

j $(i \neq j)$ is connected to another correlation $\boldsymbol{R}[k,l]$ between k and l $(k \neq l)$, iff $\{i,j\} \cap \{k,l\} \neq \varnothing$ and $\{i,j\} \neq \{k,l\}$. It is not difficult to obtain that the ratio between the degree of a line graph and the degree of a complete graph is $\frac{N^{\#}(N-2)}{N^{\#}(N^{\#}-1)/2} = \frac{4(N-2)}{N(N-1)-2}$. To the best of our knowledge, this is the first study to explore the predictive information of line graphs for forecasting realized correlation matrices.

Table 1: Choices of adjacency matrices for modeling volatilities and correlations, and the joint models for forecasting realized covariances.

$\boldsymbol{A}$ \ $\widetilde{\boldsymbol{A}}$	Empty	Complete	Line
Empty	HAR-DRD	GHAR(-, $\widetilde{K}$)	GHAR(-, $\widetilde{L}$)
Complete	GHAR(K, -)	GHAR(K, $\widetilde{K}$)	GHAR(K, $\widetilde{L}$)
Sector	GHAR(S, -)	GHAR(S, $\widetilde{K}$)	GHAR(S, $\widetilde{L}$)
GLASSO	GHAR(GL, -)	GHAR(GL, $\widetilde{K}$)	GHAR(GL, $\widetilde{L}$)

As a notational convention, we use $\boldsymbol{A}$ (resp. $\widetilde{\boldsymbol{A}}$) to denote the adjacency matrix for volatility (resp. correlation). Additionally, we refer to GHAR($\boldsymbol{A}$) (resp. GHAR($\widetilde{\boldsymbol{A}}$)) as the augmented HAR model with graph information $\boldsymbol{A}$ (resp. $\widetilde{\boldsymbol{A}}$) for forecasting volatilities (resp. correlations). In the end, GHAR($\boldsymbol{A}, \widetilde{\boldsymbol{A}}$) is referred to as the model for forecasting realized covariances, which combines the forecasts of realized volatilities based on $\boldsymbol{A}$ and correlations based on $\widetilde{\boldsymbol{A}}$. Table 1 lists the adjacency matrices for realized volatilities and correlations, and the corresponding joint models.

4 Empirical analysis

4.1 Data

We obtain intraday data on the components of Dow Jones Industrial Average (DJIA) from LOBSTER[4] for the period from July 1, 2007 to July 1, 2021.[5] Following Bollerslev et al. (2016), only stocks that traded continuously from the start to the end of our sample are maintained. As a result, 27 Dow Jones constituents are in the final sample and their ticker symbols are listed in Table 2, where we also provide summary statistics for the volatility estimates and the corresponding sector of each stock.

Figure 2 reports the average daily realized correlations across the 27 stocks. We calculate the realized correlation matrix of stock returns for each day and then compute an average across all trading days. All cross-correlations are positive and range between 0.17 and 0.63. The correlation matrix displays a block-diagonal structure that corresponds to GICS sector classification, which is in line with previous studies (e.g.

[4]https://lobsterdata.com/
[5]LOBSTER database contains data from the 27th of June 2007 up to the day before yesterday.

Benzaquen et al. (2017)).

In addition to the usual summary measures, we also report the autocorrelations for realized volatilities and correlations. Figure 3 reveals that the autocorrelation coefficients of volatilities are consistently larger than those of correlations, directly highlighting the importance of individual models for volatilities and correlations (see Andersen et al. (2006)).

Table 2: Summary statistics of realized volatilities.

	Mean	Std	Min	25%	50%	75%	Max	Sector
AAPL	2.30	3.39	0.07	0.70	1.25	2.46	38.30	Information Technology
MSFT	1.82	2.51	0.11	0.67	1.09	1.92	30.64	Information Technology
INTC	2.29	3.12	0.14	0.86	1.39	2.44	42.90	Information Technology
CSCO	1.98	2.92	0.14	0.70	1.13	2.09	43.74	Information Technology
CRM	4.00	4.93	0.22	1.44	2.41	4.64	61.67	Information Technology
IBM	1.38	2.33	0.11	0.47	0.75	1.34	30.22	Information Technology
DIS	1.89	3.04	0.12	0.60	1.01	1.88	40.56	Communication Services
VZ	1.40	2.36	0.10	0.50	0.77	1.33	34.19	Communication Services
HD	2.11	3.59	0.15	0.62	1.02	2.01	48.22	Consumer Discretionary
MCD	1.17	2.15	0.08	0.39	0.61	1.13	37.57	Consumer Discretionary
NKE	2.03	3.00	0.14	0.74	1.15	2.02	47.87	Consumer Discretionary
PG	1.00	1.76	0.09	0.38	0.58	0.98	31.60	Consumer Staples
KO	0.99	1.68	0.07	0.37	0.58	1.00	25.00	Consumer Staples
WMT	1.18	1.76	0.11	0.45	0.67	1.18	27.18	Consumer Staples
JNJ	0.92	1.56	0.06	0.35	0.54	0.90	24.74	Health Care
UNH	2.70	4.34	0.16	0.78	1.35	2.57	52.54	Health Care
MRK	1.65	2.45	0.12	0.58	0.92	1.74	30.99	Health Care
AMGN	1.91	2.34	0.16	0.82	1.27	2.14	33.44	Health Care
JPM	3.46	7.04	0.15	0.74	1.36	2.82	108.17	Financials
AXP	3.19	6.32	0.12	0.64	1.15	2.67	91.45	Financials
GS	3.24	6.27	0.19	0.92	1.49	2.81	112.41	Financials
TRV	2.04	4.09	0.11	0.49	0.81	1.76	57.95	Financials
BA	2.69	5.00	0.13	0.78	1.35	2.60	90.65	Industrials
HON	1.85	3.25	0.10	0.52	0.97	1.84	49.64	Industrials
MMM	1.43	2.25	0.08	0.46	0.81	1.49	31.11	Industrials
CAT	2.79	4.00	0.15	0.94	1.58	2.89	45.26	Industrials
CVX	2.03	3.51	0.13	0.61	1.07	2.04	48.07	Energy

Note: The table reports summary statistics for the daily realized volatility and the corresponding sector of 27 stocks in DJIA. The statistics are averaged across each trading day.

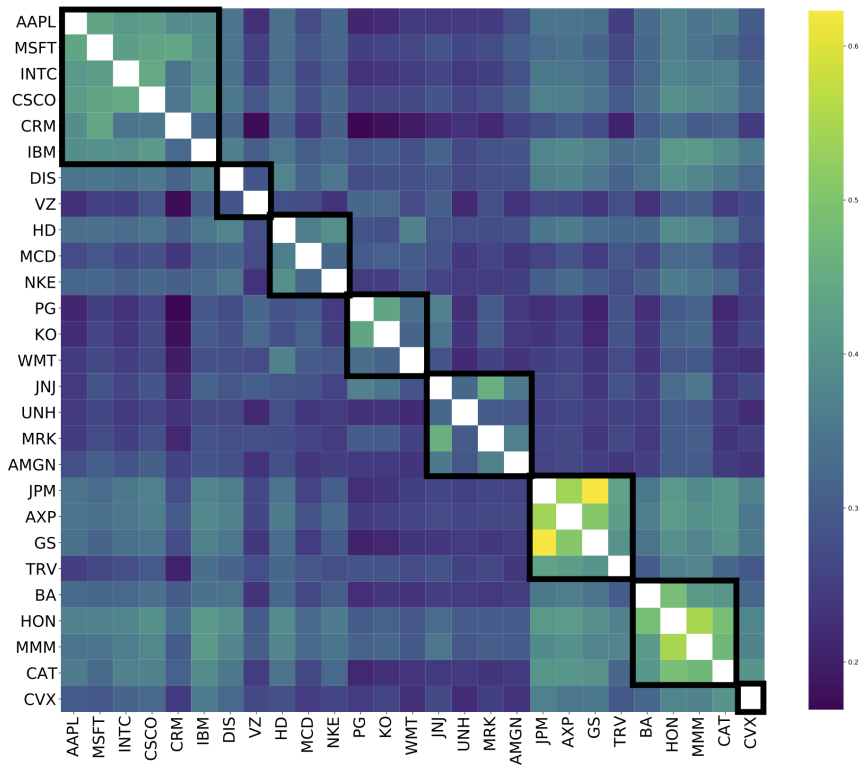


Figure 2: Realized correlation matrix. The figure shows the daily realized cross-correlations of 27 stocks in DJIA, averaged across each trading day. Stocks are sorted by their industrial sectors, as listed in Table 2.

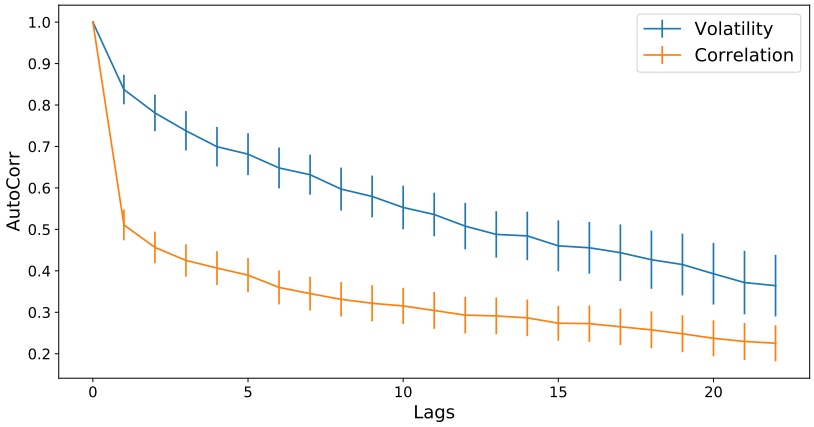


Figure 3: Autocorrelation of realized volatilities and correlations. The figure provides the standard autocorrelation coefficients and their standard deviations (vertical error bars) of the daily realized volatilities and correlations. The autocorrelation statistics for volatilities (resp. correlations) are averaged across trading days and stocks (resp. stock pairs).

4.2 Forecast evaluation metrics

Following the convention in the literature, we use hats to denote forecasts. For example, we refer to the forecast of the conditional covariance matrix on day t as $\widehat{\boldsymbol{\Sigma}}_t$. Consistent with previous studies (e.g. Symitsi et al. (2018); Bollerslev et al. (2018)), we consider the following loss functions to measure the average distance between predicted covariances and realized covariances for the model comparisons

$$\begin{aligned}
\mathcal{L}_t^E &= \sqrt{\text{vech}_{Tri}\left(\boldsymbol{H}_t - \widehat{\boldsymbol{\Sigma}}_t\right)' \text{vech}_{Tri}\left(\boldsymbol{H}_t - \widehat{\boldsymbol{\Sigma}}_t\right)}, \\
\mathcal{L}_t^F &= \sqrt{\text{Tr}\left[\left(\boldsymbol{H}_t - \widehat{\boldsymbol{\Sigma}}_t\right)'\left(\boldsymbol{H}_t - \widehat{\boldsymbol{\Sigma}}_t\right)\right]}, \\
\mathcal{L}_t^Q &= \log\det(\widehat{\boldsymbol{\Sigma}}_t) + \text{Tr}\left[\widehat{\boldsymbol{\Sigma}}_t^{-1}\boldsymbol{H}_t\right].
\end{aligned} \tag{10}$$

Here $\mathcal{L}_t^E$ represents the Euclidean distance between the vectorization of forecast covariances and ex-post realized covariances. $\mathcal{L}_t^F$ is the Frobenius distance between the two matrices. $\mathcal{L}_t^Q$ is the quasi-likelihood (QLIKE) loss based on the negative of the log-likelihood of a multivariate normal. All of these functions measure losses, therefore *lower values are preferred.*

Model Confidence Set. Hansen et al. (2011) proposed a non-parametric statistical test to formally compare the forecasting accuracy of different models, denoted as Model Confidence Set (MCS). The MCS identifies a set of models that contains the best forecasting model(s) with a given level of confidence. Specifically, given a collection of models $\mathcal{M}_0$, the test sequentially discards models with inferior predictive ability, and in the end determines a subset containing the best model(s) $\mathcal{M}^*$. The iterative elimination is based on sequentially testing the following hypothesis

$$H_0 : \mathbb{E}\left(\Delta\mathcal{L}_{ij,t}\right) = 0, \text{for all } i, j \in \mathcal{M}^*, \tag{11}$$

where $\mathcal{L}_{ij,t}$ is the loss difference between models i and j at day t in terms of a specific loss function $\mathcal{L}$ (from Eqn (10)). The MCS procedure makes it possible to make statements about statistical significance from multiple pairwise comparisons. For additional details, we refer to the studies of Hansen et al. (2003, 2011). Consistent with Symitsi et al. (2018), we implement the MCS procedure using a block bootstrap procedure with 10,000 replications and a block length of 2 observations.

4.3 In-sample results

Following the workflow in Bollerslev et al. (2016, 2018), we start the empirical analysis with an in-sample test of each model under consideration, followed by a graph structure study.

4.3.1 Model evaluation

Table 3 reports the estimated coefficients for modeling the 1-day-ahead realized covariances under the Cholesky decomposition, and the estimated coefficients for modeling 1-day-ahead realized volatilities and correlations under the DRD decomposition. The HAR-Cholesky model attributes the most importance to mid-term RV, followed by the monthly and daily components.[6] Additionally, our results support a stronger effect of the first lag on volatilities ($\beta_d = 0.609$) than the respective one on correlations ($\beta_d = 0.289$), consistent with the existing literature (e.g. Bollerslev et al. (2018)).

When examining the graph effect, we find the GHAR model for forecasting volatilities redistributes some weights from a stock's own daily lag to the daily lag of its connected neighbors. We observe a similar shift for correlation forecasting. Interestingly, the estimated coefficients of the weekly and monthly lags from neighbors are negative and statistically significant in the in-sample test. Future studies on the negative graph effects of longer lags are therefore recommended.

Table 3: In-sample estimation.

		β_d	β_w	β_m	γ_d	γ_w	γ_m
Covariance	HAR-Cholesky	0.311	0.413	0.242			
		(0.001)	(0.001)	(0.001)			
Volatility	HAR	0.609	0.271	0.057			
		(0.003)	(0.004)	(0.003)			
	GHAR(K)	0.483	0.279	0.169	0.239	-0.075	-0.150
		(0.004)	(0.006)	(0.005)	(0.006)	(0.008)	(0.006)
	GHAR(S)	0.492	0.277	0.155	0.187	-0.048	-0.120
		(0.004)	(0.006)	(0.006)	(0.005)	(0.007)	(0.006)
	GHAR(GL)	0.471	0.285	0.144	0.224	-0.075	-0.103
		(0.004)	(0.006)	(0.006)	(0.005)	(0.007)	(0.006)
Correlation	HAR	0.289	0.386	0.213			
		(0.001)	(0.001)	(0.001)			
	GHAR($\widetilde{K}$)	0.144	0.268	0.484	0.348	0.068	-0.413
		(0.001)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
	GHAR($\widetilde{L}$)	0.076	0.222	0.595	0.407	0.121	-0.521
		(0.001)	(0.002)	(0.003)	(0.002)	(0.003)	(0.004)

Note: The table reports the in-sample estimates for coefficients and their standard errors (in parentheses) of every model, calculated from the entire sample period. The Covariance row reports the HAR-Cholesky model, for forecasting the Cholesky decomposition vector of the realized covariance matrix. The Volatility rows include different models for forecasting multivariate realized volatilities, while the Correlation rows include different models for forecasting the lower triangular part of the realized correlation matrix.

[6] The HAR-Cholesky in Bollerslev et al. (2018) model allocates roughly 0.247, 0.410, and 0.244 to the daily, weekly, and monthly components, respectively.

Table 4 reports the average in-sample predictive performance of each model under consideration. To conserve space, we present (i) the ratio of loss functions when comparing each model with HAR-DRD; (ii) the rank of loss values in 13 model candidates; (iii) p-values of the MCS test. In this manner, a value of ratio less than 1 indicates the predictive performance is better than the baseline model HAR-DRD; a rank of 1 indicates the model has the best in-sample fit; p-val greater than 5% indicates the model(s) that are superior to all other models identified by MCS.

From Table 4, we summarize the following findings. First, we conclude that HAR-Cholesky produces worse results than HAR-DRD, which is consistent with the conclusion of Bollerslev et al. (2018). The results further show that the GHAR(GL, $\widetilde{L}$) model achieves the best in-sample fit across all loss functions. For example, looking at the value of the $\mathcal{L}^E$ (resp. $\mathcal{L}^F$, $\mathcal{L}^Q$) loss function, we observe that the average loss of GHAR(GL, $\widetilde{L}$) is about 98.3% (resp. 98.4%, 98.2%) of the baseline, HAR-DRD. In general, all graph models outperform the baseline, providing strong evidence for the importance of leveraging graph information and existence of certain cross-asset interaction effects.

Table 4: In-sample losses.

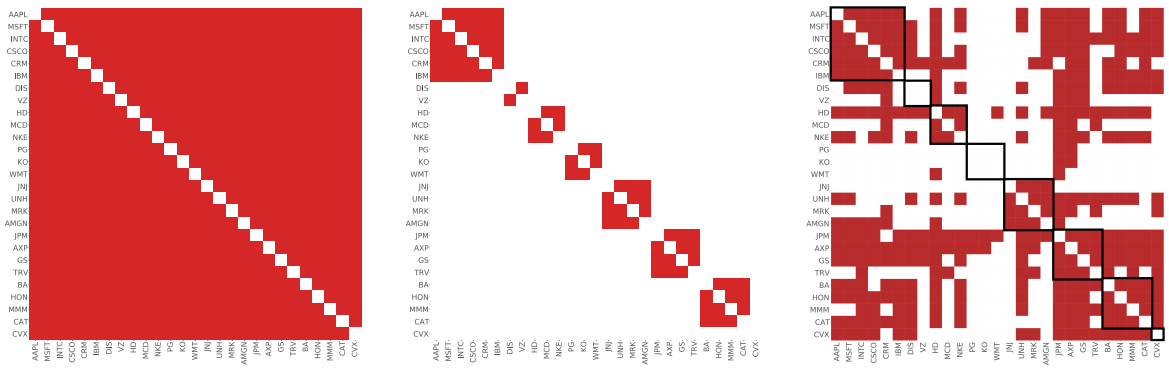
	$\mathcal{L}^E$			$\mathcal{L}^F$			$\mathcal{L}^Q$		
	Ratio	Rank	p-val	Ratio	Rank	p-val	Ratio	Rank	p-val
HAR-Cholesky	1.023	13	<0.001	1.032	13	<0.001	1.051	13	<0.001
HAR-DRD	1.000	12	<0.001	1.000	12	<0.001	1.000	10	<0.001
GHAR(-, $\widetilde{K}$)	0.997	11	<0.001	0.997	11	<0.001	0.987	6	<0.001
GHAR(-, $\widetilde{L}$)	0.995	10	<0.001	0.994	10	<0.001	0.985	4	<0.001
GHAR(S, -)	0.990	9	<0.001	0.992	9	<0.001	1.001	11	<0.001
GHAR(S, $\widetilde{K}$)	0.988	8	<0.001	0.989	8	<0.001	0.987	5	<0.001
GHAR(S, $\widetilde{L}$)	0.986	7	<0.001	0.987	5	<0.001	0.985	3	<0.001
GHAR(K, -)	0.986	5	<0.001	0.987	6	<0.001	1.006	12	<0.001
GHAR(K, $\widetilde{K}$)	0.985	4	<0.001	0.986	4	<0.001	0.991	8	<0.001
GHAR(K, $\widetilde{L}$)	0.983	2	0.557	0.984	2	0.745	0.989	7	<0.001
GHAR(GL, -)	0.986	6	<0.001	0.988	7	<0.001	0.998	9	<0.001
GHAR(GL, $\widetilde{K}$)	0.984	3	<0.001	0.986	3	<0.001	0.985	2	<0.001
GHAR(GL, $\widetilde{L}$)	0.983	1	1.000	0.984	1	1.000	0.982	1	1.000

Note: The table reports the in-sample losses of various models over the 1-day forecast horizons, averaged over the entire sample. $\mathcal{L}^E$ is the Euclidean distance, $\mathcal{L}^F$ is the Frobenius distance, and $\mathcal{L}^Q$ is the Quasi-Likelihood loss function.

4.3.2 Graph analysis

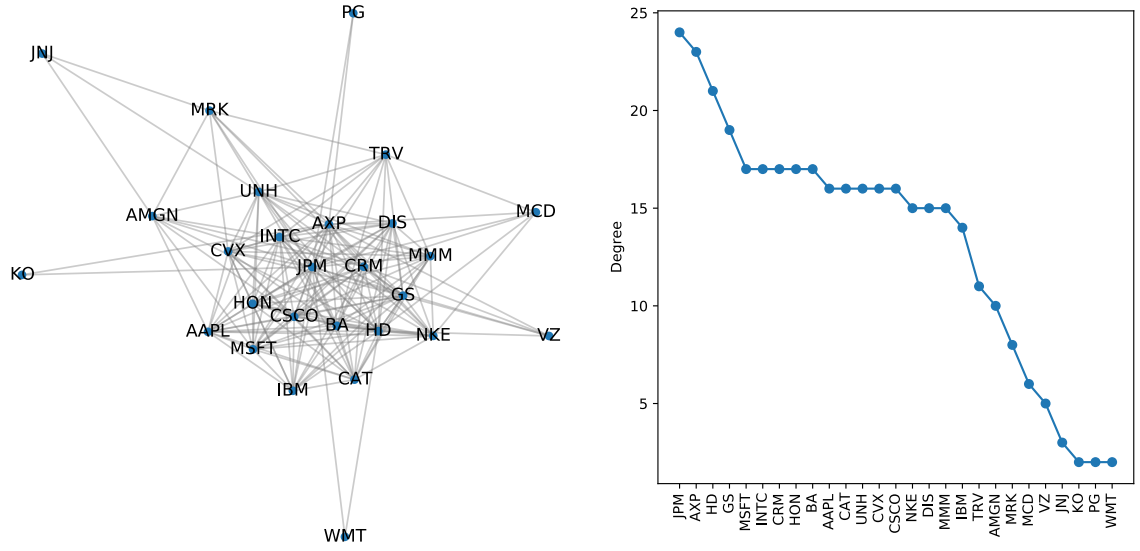
Figure 4 displays the adjacency matrices of various graphs for volatilities. The GICS sector graph in Figure 4(b) has a block-diagonal structure, with fewer edges than the GLASSO and complete graphs. The

neighborhood aggregation via the GICS sector graph in Eqn (8) is equivalent to calculating the average volatility within each sector, while ignoring the inter-sector links.



(a) Complete (K) (b) Sector (S) (c) GLASSO (GL)

Figure 4: Adjacency matrices for realized volatility. The figure shows the adjacency matrices from various methods, for forecasting realized volatility. The stocks are sorted by their industrial sectors, as listed in Table 2. The red unit represents 1, i.e. a connection between two stocks, while the white unit represents 0, i.e. no connection.



(a) Connected components (b) Degree-rank plot

Figure 5: Illustration of the graph corresponding to the adjacency matrix in 4(c) and degree-rank plot.

The adjacency matrix chosen by GLASSO (shown in Figure 4(c)) maintains a GICS sector structure to some extent, such as information technology, consumer discretionary, health care, financials, and industrials.

Meanwhile, we observe that there are substantial inter-sector connections, especially in financials and information technology, to capture the market information.

To better illustrate the interactions between stocks modeled by GLASSO, we construct a network based on the adjacency matrix of 4(c), shown in Figure 5(a). We also calculate the node degrees and sort stocks in descending order of degree. Figure 5(b) reveals that financial stocks, such as JPM, AXP, and GS, appear to be the center of the volatility network, followed by technology companies, including MSFT, INTC, etc.

4.4 Out-of-sample results

We now focus our discussion on the out-of-sample predictive performance of the competing models in a rolling window setup.

4.4.1 Model evaluation

Following Bollerslev et al. (2016, 2018); Symitsi et al. (2018); Pascalau and Poirier (2021), we estimate the parameters of each model using a fixed length rolling window of the most recent 1000 observations and obtain the rolling 1-day forecasts for each trading day in the next month. To this end, the out-of-sample period lasts from July 2011 to July 2021. All of the models are re-estimated every month.[7]

To ensure that the predicted correlations lie in $[-1, 1]$, we winsorize the values, meaning that we set the values below -1 to -1, and those above 1 to 1 in the following experiments.[8] Furthermore, to ensure the positiveness and definiteness of covariances, we replace the negative definite covariance matrices with the simple average realized covariance matrix over the relevant estimation sample period (see Bollerslev et al. (2018)).

Table 5 illustrates the average losses over the entire out-of-sample period, i.e. July 2011-July 2021. We first focus on the comparison between HAR-Cholesky and HAR-DRD. Consistent with the in-sample results, HAR-DRD significantly outperforms HAR-Cholesky across all losses. Furthermore, the results show that the GHAR(GL, $\widetilde{L}$) model yields the most accurate out-of-sample forecasts across all loss functions. Specifically, based on the $\mathcal{L}_E$ (resp. $\mathcal{L}_F$, $\mathcal{L}_Q$) loss function, the GHAR(GL, $\widetilde{L}$) model has about 2.5% (resp. 2.5%, 1.8%) lower average forecast error compared to the baseline HAR-DRD. To disentangle the prediction improvements in covariances, we compare the performance of HAR-DRD, GHAR(GL, -), GHAR(-, $\widetilde{L}$), and GHAR(GL, $\widetilde{L}$). We observe that the graph information helps improve the predictive performance in both volatilities and correlations forecasting, which subsequently leads to an improved prediction of the covariance matrix. Indeed,

[7] In Appendix C, we also report the performance of the models updated more frequently, such as daily and weekly, which leads to similar conclusions.

[8] Another common method is to use some transformation functions to convert volatilities and correlations, e.g. log transformation for volatilities (see Andersen et al. (2003); Bucci (2020b); Zhang et al. (2022)) and Fisher transformation for correlations (see Andersen et al. (2006)). We report the results in Appendix D.

GHAR(GL, $\widetilde{L}$), which utilizes both graphs in volatility and correlation simultaneously, clearly yields superior results compared to the models that only utilize one separate graph either in volatility or correlation. The MCS test indicates that in addition to GHAR(GL, $\widetilde{L}$), the GHAR(K, $\widetilde{L}$) model is also included in the subset of best models, across all loss functions.

In addition, GHAR(S, $\cdot$) has an inferior predictive performance, compared to GHAR(GL, $\cdot$). Except for the fact that there are no inter-sector links in sector graphs, another possible reason might be the lack of temporal dynamics in sector graphs.[9] This also motivates us to examine the dynamics of graphs captured by GLASSO in the following subsection.

Table 5: Out-of-sample losses over 1-day forecast horizon.

	$\mathcal{L}^E$			$\mathcal{L}^F$			$\mathcal{L}^Q$		
	Ratio	Rank	p-val	Ratio	Rank	p-val	Ratio	Rank	p-val
HAR-Cholesky	1.031	13	<0.001	1.032	13	<0.001	1.071	13	<0.001
HAR-DRD	1.000	12	0.002	1.000	12	0.003	1.000	9	<0.001
GHAR(-, $\widetilde{K}$)	0.993	11	0.002	0.991	11	0.003	0.986	7	<0.001
GHAR(-, $\widetilde{L}$)	0.991	10	0.004	0.989	9	0.006	0.984	3	0.048
GHAR(S, -)	0.990	9	0.002	0.990	10	0.003	1.002	12	<0.001
GHAR(S, $\widetilde{K}$)	0.983	8	0.003	0.983	7	0.005	0.987	8	<0.001
GHAR(S, $\widetilde{L}$)	0.982	7	0.004	0.981	5	0.006	0.985	6	<0.001
GHAR(K, -)	0.982	5	0.004	0.983	6	0.006	1.001	11	<0.001
GHAR(K, $\widetilde{K}$)	0.976	4	0.004	0.976	3	0.006	0.985	5	<0.001
GHAR(K, $\widetilde{L}$)	0.975	2	0.869	0.975	1	1.000	0.982	2	0.277
GHAR(GL, -)	0.982	6	0.004	0.983	8	0.006	1.001	10	<0.001
GHAR(GL, $\widetilde{K}$)	0.976	3	0.005	0.976	4	0.006	0.984	4	<0.001
GHAR(GL, $\widetilde{L}$)	0.975	1	1.000	0.975	2	0.768	0.982	1	1.000

Note: The table reports the out-of-sample losses of various models over the 1–day forecast horizon, averaged over the entire testing sample. $\mathcal{L}^E$ is the Euclidean distance, $\mathcal{L}^F$ is the Frobenius distance, and $\mathcal{L}^Q$ is the Quasi-Likelihood loss function.

We further investigate whether the coefficients in each model vary across time. Figure 6 displays the rolling 4-year betas (and gammas) and their respective 95% confidence intervals for modeling volatilities over the testing period. The plots of HAR first show that there are three substantial shocks in β_d, which occur at the end of 2012, the end of 2015, and March 2020. In all GHAR models, β_d rapidly drops at the end of 2012 and 2015 as well. However, β_d appears to be more consistent during the pandemic period (March 2020), while the graph coefficient γ_d displays a high jump during the COVID-19 pandemic. This may indicate that the spillover effect is a major driver of market volatility during the pandemic. Another surprising aspect of

[9] We use the up-to-date GICS sector to construct a static graph employed during the entire period.

the plots is that β_w has a similar trend across various models, and similarly for β_m. We also observe that γ_w is insignificant during 2016-2020 at the 5% confidence level. These findings indicate that mid- and long-term graph effects may have a minor influence on volatility dynamics.[10]

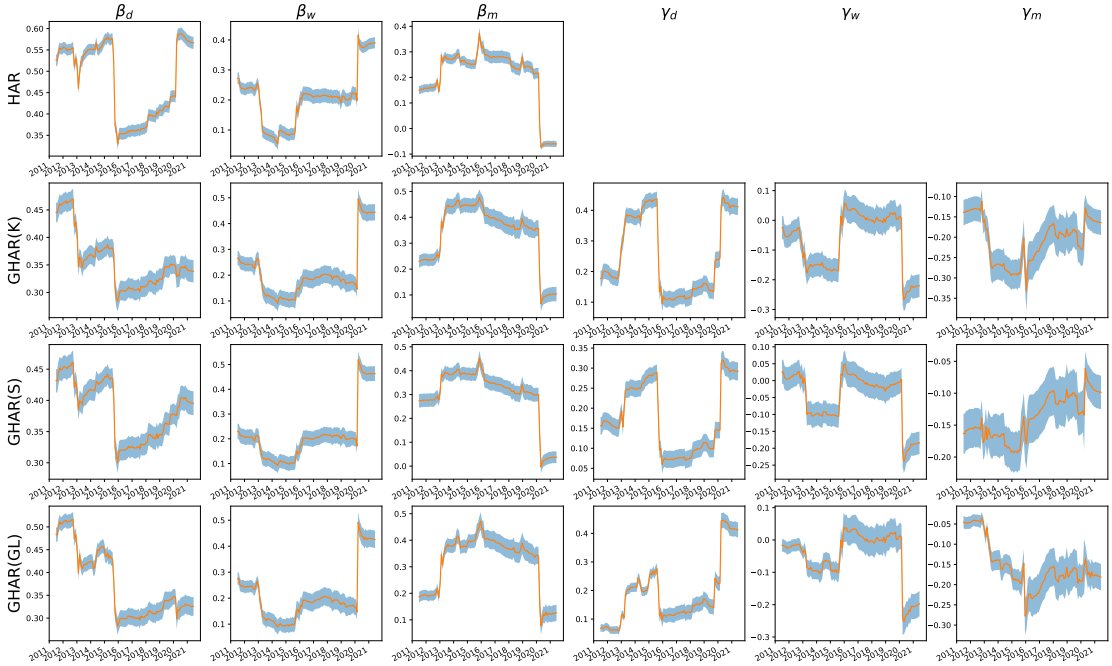


Figure 6: Rolling coefficients of various models for forecasting volatilities. The orange curve represents the estimated coefficients, with the blue area covering the 95% confidence interval.

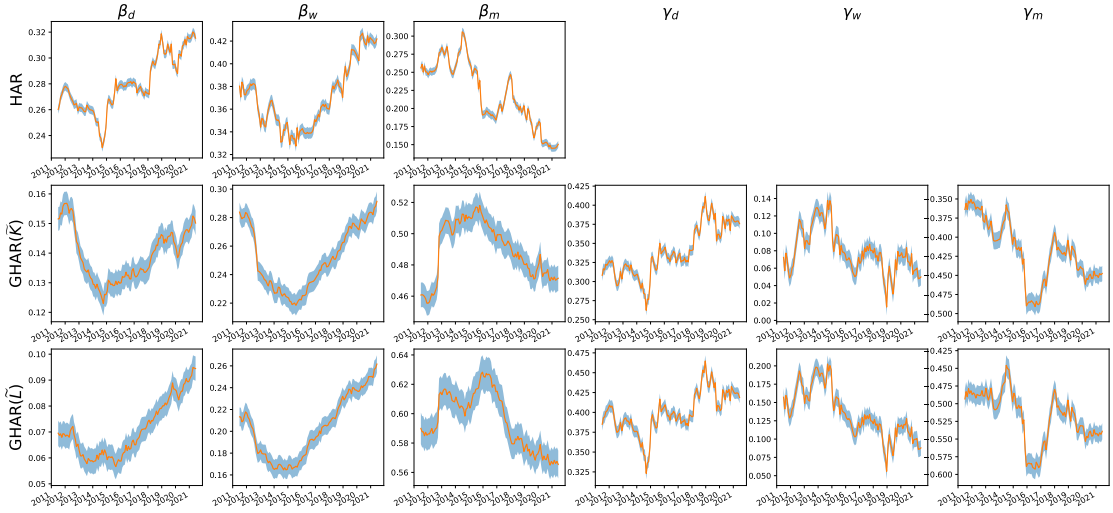


Figure 7: Rolling coefficients of various models for forecasting correlations. The orange curve represents the estimated coefficients, with the blue area covering the 95% confidence interval.

[10] We study the graph-based model with only short-term graph effect in Appendix B.

We provide a similar analysis of the models for correlations, illustrated in Figure 7. First, the rolling coefficients appear to be less volatile than their counterparts when modeling volatilities. As expected, the rolling coefficients look similar across graph-based methods. In addition, we observe that the coefficients β_d, β_w, and β_m are predominantly smoother for graph-based models compared oo their counterparts in the standard HAR. Interestingly, the short-term graph effect γ_d in graph-based models exhibits a similar trend as β_d in HAR.

4.4.2 Graph analysis

Since GHAR(GL, $\widetilde{L}$) delivers the best forecasting results, we further dive into the dynamics of the graphs uncovered by GLASSO. For each adjacency matrix at a timestamp, we calculate the fraction of non-sector stocks that are connected to sector members, a.k.a *group degree centrality* (see more in Everett and Borgatti (1999)). For ease of visualization, in Figure 8, we only plot group degree centrality of four GICS sectors with most constituents. We first observe that the centrality for each sector maintains at a higher level since 2015, implying that sector connectedness becomes increasingly important in the U.S. stock market, in line with the findings of Baruník et al. (2016); Costa et al. (2022). In addition, we observe that at the beginning of the COVID-19 pandemic, healthcare and information technology companies exhibit increasing influence on other stocks, which is not surprising as they are two critical sectors whose prices have been largely affected during the pandemic.

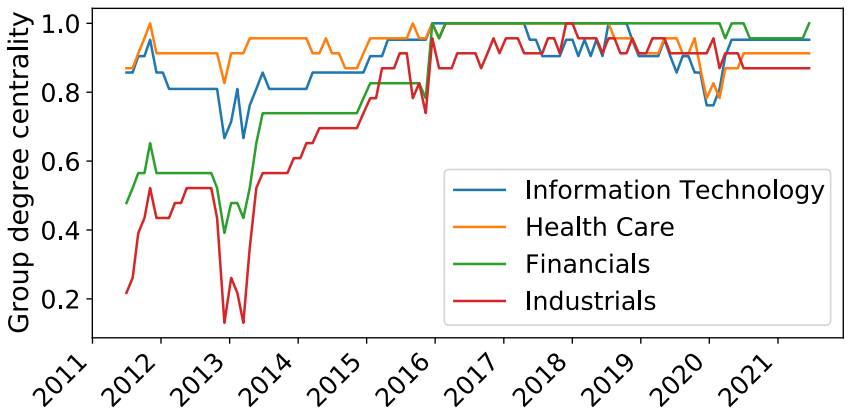


Figure 8: Group degree centrality of 4 sectors in the graphs from GLASSO.

To study the stability and temporal dynamics of graphs learned by GLASSO, we also calculate the similarity of two consecutive graphs at monthly basis in terms of their edge sets. The Jaccard index of the edge set is defined as the ratio between the number of intersection edges and the number of union edges of two consecutive graphs. Specifically, at time t, the edge set of the GLASSO estimate is denoted as $\mathcal{E}^{(t)}$, with

the Jaccard index computed as $\frac{|\mathcal{E}^{(t)} \cap \mathcal{E}^{(t-1)}|}{|\mathcal{E}^{(t)} \cup \mathcal{E}^{(t-1)}|}$.

We plot the Jaccard index between consecutive GLASSO estimates against the timespan in Figure 9. The graphs are stable in general, with only a few exceptions: Aug 2011 - Dec 2011, Nov 2012 - June 2013, Dec 2015 - Feb 2016, and March 2020 - April 2020. This suggests that the equity market might be experiencing significant changes during these periods. In fact, some periods correspond to stock market turmoils, as highlighted in Figure 9. To the best of our knowledge, there was no major financial event from the end of 2012 until mid 2013 documented in the literature, yet very recent methods in the change-point detection machine learning literature have also picked up this regime (see Sulem et al. (2022)).

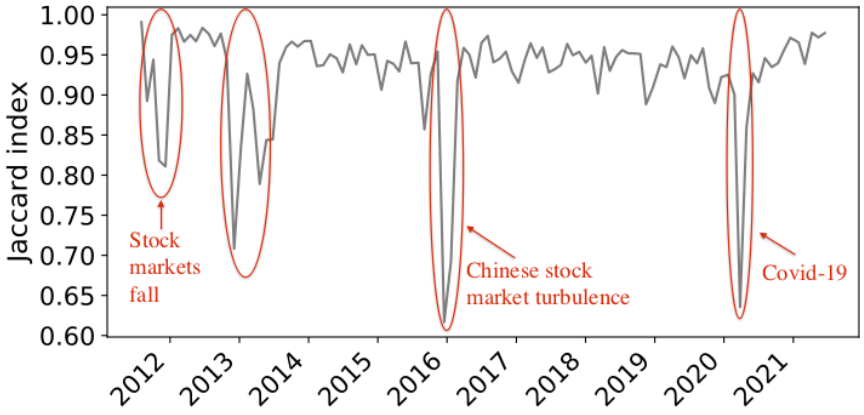


Figure 9: Jaccard index of similarity between consecutive graphs.

4.5 Portfolio performance

To further quantify the benefit of our proposed GHAR models, we examine the portfolio performance based on the covariance forecasts (see Symitsi et al. (2018); Bollerslev et al. (2018)). We construct the global minimum variance portfolio (GMVP) to bypass the problem of forecasting future returns, as the scope of this paper is on covariance forecasting. We calculate the weights of GMVP as

$$\min_{\boldsymbol{w}_t} \boldsymbol{w}_t' \widehat{\boldsymbol{\Sigma}}_t \boldsymbol{w}_t \quad \text{s.t.} \quad \boldsymbol{w}_t' \mathbf{1} = 1 \tag{12}$$

where $\mathbf{1} = (1, \ldots, 1)' \in \mathbb{R}^N$. The optimal weights of the GMVP are given by $\hat{\boldsymbol{w}}_t = \frac{\widehat{\boldsymbol{\Sigma}}_t^{-1}\mathbf{1}}{\mathbf{1}'\widehat{\boldsymbol{\Sigma}}_t^{-1}\mathbf{1}}$.

In addition, we consider another portfolio imposed with the **long-only** constraint, denoted as GMVP$^+$, which is a more realistic setting in certain markets. In this case, we are not allowed to short-sell stocks, i.e. $w_{i,t} \geq 0, \ \forall i, t$. Since there is no closed-form solution for GMVP$^+$, we use the quadratic programming solver in the built-in Python package CVXPY[11] to compute the optimized portfolio weights. Finally, we compute

[11] https://www.cvxpy.org

the portfolio returns as $r_t^{(p)} = \hat{\boldsymbol{w}}_t' \boldsymbol{r}_t$.

Based on the covariance forecasts from the competing models, we obtain multiple corresponding portfolios. We then use the following metrics to evaluate the out-of-sample performance of each portfolio.

• (1) Annualized portfolio standard deviation:

$$\sigma^{(p)} = \sqrt{252 \times \frac{1}{T} \sum_{t=1}^{T} \left(r_t^{(p)} - \bar{r}_t^{(p)} \right)^2}, \tag{13}$$

where T is the length of the out-of-sample period, and $\bar{r}_t^{(p)}$ is the average return of a particular portfolio over the out-of-sample period.

• (2) Average portfolio turnover:

$$\begin{aligned} TO_t &= \sum_{i=1}^{N} \left| w_{i,t+1} - w_{i,t} \frac{1 + r_{t,i}}{1 + \boldsymbol{w}_t' \boldsymbol{r}_t} \right|, \\ \tau^{(p)} &= \frac{1}{T-1} \sum_{t=1}^{T-1} TO_t, \end{aligned} \tag{14}$$

where TO_t represents the portfolio turnover from day t to day $t+1$. Recall that $w_{i,t}$ (resp. $r_{i,t}$) is the i-th element in $\boldsymbol{w}_t$ (resp. $\boldsymbol{r}_t$).

To further gauge the economic value of covariance forecasts, we also compare the aforementioned portfolios with the naive equally-weighted portfolio (denoted as 1/N), a common benchmark in the literature.[12] Table 6 illustrates the annualized standard deviation and average turnover for GMVP and GMVP$^+$ over the out-of-sample period, assuming no transaction cost. The results show that using high-frequency data for forecasting realized covariance leads to superior portfolios compared to the naive 1/N portfolio in terms of standard deviation. Portfolios with graph information, especially GHAR(GL, $\widetilde{L}$), are able to generate lower out-of-sample variance compared to the one without any graphs, i.e. HAR-DRD, further proving evidence for the existence of network/spillover effects.

[12] DeMiguel et al. (2009) demonstrated that of the 14 models they evaluate across seven empirical datasets, none is consistently better than the 1/N portfolio in terms of Sharpe ratio, certainty-equivalent return, or turnover.

Table 6: Out-of-sample portfolio performance over 1-day forecast horizon.

	GMVP		GMVP$^+$	
	$\sigma^{(p)}$	$\tau^{(p)}$	$\sigma^{(p)}$	$\tau^{(p)}$
1/N	11.890	0.007	11.890	0.007
HAR-Cholesky	9.948	0.664	9.917	0.411
HAR-DRD	9.918	0.771	9.936	0.511
GHAR(-, $\widetilde{K}$)	9.967	0.730	9.935	0.482
GHAR(-, $\widetilde{L}$)	9.970	0.761	9.951	0.510
GHAR(S, -)	9.906	0.738	10.005	0.503
GHAR(S, $\widetilde{K}$)	9.933	0.675	9.983	0.468
GHAR(S, $\widetilde{L}$)	9.930	0.717	9.993	0.499
GHAR(K, -)	9.857	0.653	9.947	0.432
GHAR(K, $\widetilde{K}$)	9.896	0.590	9.930	0.389
GHAR(K, $\widetilde{L}$)	9.893	0.634	9.945	0.426
GHAR(GL, -)	9.831	0.695	9.911	0.475
GHAR(GL, $\widetilde{K}$)	9.825	0.630	**9.883**	0.435
GHAR(GL, $\widetilde{L}$)	**9.799**	0.675	9.901	0.470

Note: The table reports the out-of-sample performance of GMVP and GMVP$^+$ constructed using the 1-day-ahead covariance forecasts of various models. $\sigma^{(p)}$ is the annualized portfolio standard deviation, and $\tau^{(p)}$ is the average portfolio turnover. The lowest $\sigma^{(p)}$ in each column is indicated in bold.

4.6 Longer future horizons

One-day forecasting is not the only time horizon of interest to practitioners. In this section, we consider various prediction horizons, which allows us to examine how the impact of graph information varies on future covariances. To obtain longer horizon forecasts, we adopt the direct approach (in line with Andersen et al. (2003); Bollerslev et al. (2018)), rather than the iterated approach.[13] Specifically, for the GHAR formulation of volatility in Eqn (8), we implement the following model over longer forecast horizons. We apply the same idea to the modeling of correlations, or the vector of Cholesky decomposition of covariances.

$$\begin{aligned}\textbf{GHAR}(\boldsymbol{A}): \quad \boldsymbol{RV}_{t:t+h} = \alpha^{(D)} + \beta_d^{(D)}\boldsymbol{RV}_{t-1} + \beta_w^{(D)}\boldsymbol{RV}_{t-5:t-2} + \beta_m^{(D)}\boldsymbol{RV}_{t-22:t-6} \\ + \gamma_d^{(D)}\boldsymbol{W}\cdot\boldsymbol{RV}_{t-1} + \gamma_w^{(D)}\boldsymbol{W}\cdot\boldsymbol{RV}_{t-5:t-2} + \gamma_m^{(D)}\boldsymbol{W}\cdot\boldsymbol{RV}_{t-22:t-6},\end{aligned} \tag{15}$$

where $\boldsymbol{RV}_{t:t+h} = \sum_{k=0}^{h}\boldsymbol{RV}_{t+k}$, and $h = 4$ and 21 for the weekly and monthly forecasts, respectively.

Table 7 illustrates the statistical performance of GHAR models over longer prediction horizons.[14] We

[13] Bollerslev et al. (2018) demonstrated that even minor model mis-specifications will be amplified in the iterated one-day-ahead forecasts, and result in worse estimations than the direct forecast procedure in practice.

[14] We report the portfolio performance results in the Appendix A.

observe that GHAR(GL, $\widetilde{L}$) continuously outperforms the baseline HAR-DRD, with only one exception: the $\mathcal{L}^Q$ loss over the monthly forecasting horizon. Furthermore, the ratios converge to one as the prediction horizon gets longer, as expected. Longer time horizon forecasting is less sensitive to graph information. A possible explanation for this might be that unexpected information (e.g. shocks) of a specific stock is diffused to the broad equity market in the short term (such as several days), rendering less necessary the need to explicitly incorporate graph effects into the model, at longer horizons.

Table 7: Out-of-sample losses over longer forecast horizons.

	$\mathcal{L}^E$			$\mathcal{L}^F$			$\mathcal{L}^Q$		
	Ratio	Rank	p-val	Ratio	Rank	p-val	Ratio	Rank	p-val
Panel A: 1-Week									
HAR-Cholesky	1.051	13	<0.001	1.058	13	<0.001	1.028	13	<0.001
HAR-DRD	1.000	12	0.006	1.000	12	0.013	1.000	7	0.002
GHAR(-, $\widetilde{K}$)	0.993	11	0.008	0.992	11	0.016	0.997	4	0.492
GHAR(-, $\widetilde{L}$)	0.992	10	0.013	0.991	9	0.025	0.997	3	0.492
GHAR(S, -)	0.991	9	0.008	0.992	10	0.011	1.005	12	<0.001
GHAR(S, $\widetilde{K}$)	0.985	8	0.008	0.985	8	0.016	1.003	10	0.002
GHAR(S, $\widetilde{L}$)	0.984	7	0.015	0.983	7	0.025	1.003	11	0.054
GHAR(K, -)	0.981	6	0.028	0.982	6	0.027	1.002	9	<0.001
GHAR(K, $\widetilde{K}$)	0.976	4	0.030	0.977	4	0.027	0.998	5	0.062
GHAR(K, $\widetilde{L}$)	0.975	3	0.064	0.976	2	0.449	0.998	6	0.104
GHAR(GL, -)	0.980	5	0.030	0.982	5	0.027	1.000	8	<0.001
GHAR(GL, $\widetilde{K}$)	0.975	2	0.064	0.976	3	0.055	0.997	2	0.492
GHAR(GL, $\widetilde{L}$)	0.974	1	1.000	0.975	1	1.000	0.997	1	1.000
Panel B: 1-Month									
HAR-Cholesky	1.108	13	<0.001	1.102	13	<0.001	1.023	13	<0.001
HAR-DRD	1.000	11	0.105	1.000	11	0.114	1.000	3	0.006
GHAR(-, $\widetilde{K}$)	0.994	4	0.480	0.993	4	0.581	0.999	1	1.000
GHAR(-, $\widetilde{L}$)	0.993	3	0.697	0.992	3	0.837	0.999	2	0.006
GHAR(S, -)	0.999	9	0.053	0.999	9	0.055	1.010	11	0.006
GHAR(S, $\widetilde{K}$)	0.995	6	0.314	0.994	6	0.501	1.010	10	0.006
GHAR(S, $\widetilde{L}$)	0.994	5	0.374	0.993	5	0.542	1.011	12	0.006
GHAR(K, -)	1.000	12	0.036	1.000	12	0.036	1.010	9	0.005
GHAR(K, $\widetilde{K}$)	1.000	10	0.065	1.000	10	0.064	1.008	7	0.006
GHAR(K, $\widetilde{L}$)	0.999	8	0.146	0.999	8	0.160	1.008	8	0.006
GHAR(GL, -)	0.996	7	0.322	0.996	7	0.393	1.003	6	0.006
GHAR(GL, $\widetilde{K}$)	0.992	2	0.697	0.991	2	0.837	1.001	4	0.006
GHAR(GL, $\widetilde{L}$)	0.991	1	1.000	0.991	1	1.000	1.001	5	0.006

Note: The table reports the out-of-sample losses of forecasting 1-week-ahead (Panel A) and 1-month-ahead (Panel B) realized covariances.

5 Robustness analysis

In this section, we address potential concerns regarding the robustness of our results. In particular, we aim to answer two important questions related to (i) the stability across different market regimes, and (ii) the measurement errors of volatilities. Due to space considerations and to improve the readability, discussions about other relevant concerns, such as the short-term graph effect, alternative model update frequencies, and transformations for volatilities and correlations, are provided in Appendices B, C, and D.

5.1 Stability across market regimes

The first essential question is whether the model performance varies across different market regimes. To this end, we perform a stratified out-of-sample analysis over two sub-samples: a relatively calm period when the realized volatility of S&P500 ETF index is below the 90% quantile of its entire sample distribution, and a turbulent period when this realized volatility is above its 90% quantile (Pascalau and Poirier (2021)).

Table 8 indicates the relative losses, ranks and p-values of the MCS test for the bottom 90% (low and moderate volatility regimes) and top 10% (high volatility regimes) of market RVs. In general, it appears that GHAR(GL, $\widetilde{L}$) is superior across market regimes over all loss functions in both situations. Interestingly, the larger percentage improvements in terms of $\mathcal{L}^E$ and $\mathcal{L}^F$ stem from the turbulent period, while the opposite holds true for $\mathcal{L}^Q$.

Additionally, it is worth noting that HAR-Cholesky achieves worse predictive performance relative to HAR-DRD during the turbulent period. This finding provides robust evidence in favor of the DRD decomposition for modeling covariance matrices, especially during periods of high volatility, in line with findings reported in Andersen et al. (2006).[15]

5.2 Measurement errors of volatilities

Bollerslev et al. (2016) revealed that the beta coefficients in the standard HAR model may be affected by measurement errors in the realized volatilities. By exploiting the asymptotic theory for high-frequency realized volatility estimation, the authors proposed an easy-to-implement model, termed as HARQ, shown in Eqn (17). The realized quarticity (RQ) is estimated according to Eqn (16), aiming to correct the measurement

[15] Andersen et al. (2006) proposed to characterize the dependencies in volatilities and correlations by regime-switching models.

Table 8: Stratified out-of-sample losses.

	$\mathcal{L}^E$			$\mathcal{L}^F$			$\mathcal{L}^Q$		
	Ratio	Rank	p-val	Ratio	Rank	p-val	Ratio	Rank	p-val
Panel A: Bottom 90%									
HAR-Cholesky	0.987	7	<0.001	0.988	7	<0.001	1.067	13	<0.001
HAR-DRD	1.000	13	<0.001	1.000	13	<0.001	1.000	9	<0.001
GHAR(-, $\widetilde{K}$)	0.999	12	<0.001	0.998	12	<0.001	0.983	7	<0.001
GHAR(-, $\widetilde{L}$)	0.997	11	<0.001	0.996	11	<0.001	0.980	3	0.729
GHAR(S, -)	0.992	10	<0.001	0.993	10	<0.001	1.003	12	<0.001
GHAR(S, $\widetilde{K}$)	0.992	9	<0.001	0.993	9	<0.001	0.984	8	<0.001
GHAR(S, $\widetilde{L}$)	0.990	8	<0.001	0.990	8	<0.001	0.981	4	0.001
GHAR(K, -)	0.986	3	0.676	0.987	3	0.503	1.002	11	<0.001
GHAR(K, $\widetilde{K}$)	0.986	6	<0.001	0.988	5	<0.001	0.982	5	<0.001
GHAR(K, $\widetilde{L}$)	0.985	2	0.780	0.986	2	0.664	0.979	2	0.965
GHAR(GL, -)	0.986	4	0.651	0.988	6	0.367	1.002	10	<0.001
GHAR(GL, $\widetilde{K}$)	0.986	5	<0.001	0.988	4	<0.001	0.982	6	<0.001
GHAR(GL, $\widetilde{L}$)	0.984	1	1.000	0.986	1	1.000	0.979	1	1.000
Panel B: Top 10%									
HAR-Cholesky	1.119	13	<0.001	1.115	13	<0.001	1.091	13	<0.001
HAR-DRD	1.000	12	0.034	1.000	12	0.035	1.000	12	0.011
GHAR(-, $\widetilde{K}$)	0.983	10	0.071	0.981	10	0.117	0.999	10	0.015
GHAR(-, $\widetilde{L}$)	0.982	9	0.147	0.981	9	0.136	0.998	9	0.015
GHAR(S, -)	0.986	11	0.028	0.986	11	0.029	1.000	11	0.006
GHAR(S, $\widetilde{K}$)	0.970	6	0.116	0.968	5	0.210	0.996	7	0.015
GHAR(S, $\widetilde{L}$)	0.970	5	0.292	0.968	6	0.343	0.996	6	0.015
GHAR(K, -)	0.976	7	0.055	0.976	7	0.057	0.998	8	0.015
GHAR(K, $\widetilde{K}$)	0.961	2	0.991	0.959	1	1.000	0.993	4	0.024
GHAR(K, $\widetilde{L}$)	0.961	4	0.991	0.960	2	0.985	0.992	3	0.050
GHAR(GL, -)	0.976	8	0.071	0.976	8	0.076	0.996	5	0.015
GHAR(GL, $\widetilde{K}$)	0.961	3	0.991	0.960	3	0.985	0.991	2	0.188
GHAR(GL, $\widetilde{L}$)	0.961	1	1.000	0.960	4	0.980	0.991	1	1.000

Note: Stratified losses during trading days with the bottom 90% (Panel A) and top 10% (Panel B) realized volatility of S&P500 ETF index.

errors.

$$RQ_{i,t} = \frac{M}{3}\sum_{l=1}^{M} r_{i,t(l)}^4, \quad \boldsymbol{RQ}_{t-1} = (RQ_{1,t}, \ldots, RQ_{N,t})', \tag{16}$$

$$\boldsymbol{RV}_t = \alpha_0^{(Q)} + \beta_d^{(Q)}\boldsymbol{RV}_{t-1} + \phi_d^{(Q)}\boldsymbol{RQ}_{t-1} \circ \boldsymbol{RV}_{t-1} + \beta_w^{(Q)}\boldsymbol{RV}_{t-5:t-2} + \beta_m^{(Q)}\boldsymbol{RV}_{t-22:t-6} + \boldsymbol{u}_t^{(Q)}, \tag{17}$$

where $\circ$ denotes the Hadamard (element-wise) product. The authors show that the HARQ model outperforms other commonly used models, including GARCH and HAR, when applied to the S&P500 ETF index and individual stocks.

To examine if our results are sensitive to measurement errors of volatilities, we also implement the following models with graph information, denoted as **GHARQ** (Eqn (18)). As demonstrated by Bollerslev et al. (2018), the measurement errors for the correlations are fairly small; therefore, we follow their setting and ignore the attenuation for the correlations. Finally, as summarized in Table 1, we also consider various graphs for modeling volatilities and correlations.

$$\begin{aligned}\boldsymbol{RV}_t = \alpha^{(Q)} &+ \beta_d^{(Q)}\boldsymbol{RV}_{t-1} + \phi_d^{(Q)}\boldsymbol{RQ}_{t-1} \circ \boldsymbol{RV}_{t-1} + \beta_w^{(Q)}\boldsymbol{RV}_{t-5:t-2} + \beta_m^{(Q)}\boldsymbol{RV}_{t-22:t-6} \\ &+ \gamma_d^{(Q)}\boldsymbol{W}\cdot\boldsymbol{RV}_{t-1} + \gamma_w^{(Q)}\boldsymbol{W}\cdot\boldsymbol{RV}_{t-5:t-2} + \gamma_m^{(Q)}\boldsymbol{W}\cdot\boldsymbol{RV}_{t-22:t-6} + \boldsymbol{u}_t^{(Q)}.\end{aligned} \tag{18}$$

Table 9: Out-of-sample losses of HARQ and GHARQ.

	$\mathcal{L}^E$			$\mathcal{L}^F$			$\mathcal{L}^Q$		
	Ratio	Rank	p-val	Ratio	Rank	p-val	Ratio	Rank	p-val
HARQ-DRD	1.000	12	<0.001	1.000	12	0.003	1.000	8	<0.001
GHARQ(-, $\widetilde{K}$)	0.995	10	<0.001	0.994	10	0.003	0.992	6	<0.001
GHARQ(-, $\widetilde{L}$)	0.993	9	0.006	0.992	8	0.016	0.989	4	<0.001
GHARQ(S, -)	0.996	11	0.003	0.997	11	0.003	1.010	11	<0.001
GHARQ(S, $\widetilde{K}$)	0.992	8	0.003	0.993	9	0.003	0.991	5	<0.001
GHARQ(S, $\widetilde{L}$)	0.991	7	0.008	0.991	7	0.016	0.989	3	<0.001
GHARQ(K, -)	0.989	6	0.008	0.990	6	0.016	1.023	12	<0.001
GHARQ(K, $\widetilde{K}$)	0.988	5	0.008	0.989	5	0.016	1.002	10	<0.001
GHARQ(K, $\widetilde{L}$)	0.986	3	0.008	0.987	3	0.016	0.999	7	<0.001
GHARQ(GL, -)	0.987	4	0.008	0.988	4	0.016	1.001	9	<0.001
GHARQ(GL, $\widetilde{K}$)	0.985	2	0.008	0.986	2	0.016	0.984	2	0.006
GHARQ(GL, $\widetilde{L}$)	0.983	1	1.000	0.985	1	1.000	0.982	1	1.000

Note: The table reports the out-of-sample losses of forecasting 1-day-ahead realized covariances of HARQ and GHARQ with different graphs. The baseline here is the HARQ-DRD model.

Table 9 presents the results obtained from **GHARQ**. We conclude from this table that GHARQ(GL,

$\widetilde{L}$) remains the best performing model. This indicates that volatility and correlation graphs indeed provide additional information for covariance forecasting. We also observe (unreported) that HARQ-DRD provides significant better forecasts than HAR-DRD, consistent with Bollerslev et al. (2018).

6 Conclusion

This study investigates whether the graph information can provide additional predictive power when forecasting realized variances and covariances in the U.S. equity market. We propose a new approach that augments the HAR-DRD model by considering graph effects from connected assets. We evaluate the in-sample and out-of-sample contributions of such graph effects in realized covariance matrices forecasting, and observe that the proposed model GHAR(GL, $\widetilde{L}$) achieves the best predictive accuracy consistently across various evaluation metrics. Furthermore, based on the covariance predictions, we construct a global minimum variance portfolio. The portfolio analysis shows that portfolios considering graph effects are able to attain significantly lower out-of-sample variance compared to the traditional models without any graph information or the naive equally-weighted portfolio. An assessment of the forecast performance over time shows that the forecast improvements over longer horizons (1-week, 1-month) remain significant, but start to decay as the prediction horizon increases. Furthermore, the robustness tests demonstrate the forecast improvements are observed consistently over the different out-of-sample sub-periods and are insensitive to measurement errors of volatilities. Overall, our results show that graph structures in volatilities and correlations are informative for forecasting realized covariance matrices.

One interesting finding from our in-sample analysis is that the short-term graph component has a positive effect, while the long-term graph effect is negative, both for volatility and correlation. To the best of our knowledge, this result has not been discussed in the literature, along with any economic forces that may drive this, particularly with regard to the long-term graph effect. Further studies will need to be undertaken.

We are aware that there could be other types of graphs, e.g. supply/demand chains (see Tokman et al. (2007); Herskovic et al. (2020)), shared analyst coverage (Ali and Hirshleifer (2020)), and news co-mentions (Sidorov et al. (2016)) that may contain rich structural information about equities. Our framework can be directly applied to new sources of graphs. It would be very interesting to compare (and potentially ensemble) the predictive power of different types of graphs.

Another interesting direction pertains to performing a similar analysis as in the present paper, but across different markets. For example, according to Rapach et al. (2013); Buncic and Gisler (2016); Choi et al. (2021), U.S.-based equity market information can be used to improve returns/volatility forecasts in a large cross-section of international equity markets. However, there seems to be less information about the impact

of international equity markets on the U.S. market (Wilms et al. (2021)). It would be an interesting study to explore the two-way interplay between U.S. equity market and international equity markets, in a bipartite (and potentially multipartite) graph setting.

Yet another potential interesting direction to explore concerns the integration into the models of any available node-level feature information, such as price and volume-based technical indicators, news sentiment, or microstructure-based features, which modern graph-based neural networks can naturally accommodate, while also accounting for any potential nonlinearities inherent in the data (e.g. Zhang et al. (2022)).

References

Ali, U. and D. Hirshleifer (2020). Shared analyst coverage: Unifying momentum spillover effects. *Journal of Financial Economics 136*(3), 649–675.

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

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens (2001). The distribution of realized stock return volatility. *Journal of Financial Economics 61*(1), 43–76.

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

Andersen, T. G., T. Bollerslev, and N. Meddahi (2011). Realized volatility forecasting and market microstructure noise. *Journal of Econometrics 160*(1), 220–234.

Andrada-Félix, J., F. Fernández-Rodríguez, and A.-M. Fuertes (2016). Combining nearest neighbor predictions and model-based predictions of realized variance: Does it pay? *International Journal of Forecasting 32*(3), 695–715.

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

Baele, L. (2005). Volatility spillover effects in european equity markets. *Journal of Financial and Quantitative Analysis 40*(2), 373–401.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (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.

Baruník, J., E. Kočenda, and L. Vácha (2016). Asymmetric connectedness on the us stock market: Bad and good volatility spillovers. *Journal of Financial Markets 27*, 55–78.

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

Benzaquen, M., I. Mastromatteo, Z. Eisler, and J.-P. Bouchaud (2017). Dissecting cross-impact on stock markets: An empirical analysis. *Journal of Statistical Mechanics: Theory and Experiment 2017*(2), 023406.

Berben, R.-P. and W. J. Jansen (2005). Comovement in international equity markets: A sectoral view. *Journal of International Money and Finance 24*(5), 832–857.

Bollerslev, T., B. Hood, J. Huss, and L. H. Pedersen (2018). Risk everywhere: Modeling and managing volatility. *Review of Financial Studies 31*(7), 2729–2773.

Bollerslev, T., J. Marrone, L. Xu, and H. Zhou (2014). Stock return predictability and variance risk premia: Statistical inference and international evidence. *Journal of Financial and Quantitative Analysis 49*(3), 633–661.

Bollerslev, T., M. C. Medeiros, A. J. Patton, and R. Quaedvlieg (2021). From zero to hero: Realized partial (co) variances. *Journal of Econometrics*.

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

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

Boyer, B. H., M. S. Gibson, and M. Loretan (1999). Pitfalls in tests for changes in correlations.

Bucci, A. (2020a). Cholesky–ann models for predicting multivariate realized volatility. *Journal of Forecasting 39*(6), 865–876.

Bucci, A. (2020b). Realized volatility forecasting with neural networks. *Journal of Financial Econometrics 18*(3), 502–531.

Buncic, D. and K. I. Gisler (2016). Global equity market volatility spillovers: A broader role for the united states. *International Journal of Forecasting 32*(4), 1317–1339.

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

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

Choi, D., W. Jiang, and C. Zhang (2021). Alpha go everywhere: Machine learning and international stock returns. *Available at SSRN 3489679*.

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

Costa, A., P. Matos, and C. da Silva (2022). Sectoral connectedness: New evidence from us stock market during covid-19 pandemics. *Finance Research Letters 45*, 102124.

Cubadda, G., B. Guardabascio, and A. Hecq (2017). A vector heterogeneous autoregressive index model for realized volatility measures. *International Journal of Forecasting 33*(2), 337–344.

Degiannakis, S. and G. Filis (2017). Forecasting oil price realized volatility using information channels from other asset classes. *Journal of International Money and Finance 76*, 28–49.

DeMiguel, V., L. Garlappi, and R. Uppal (2009). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *Review of Financial Studies 22*(5), 1915–1953.

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. and K. F. Kroner (1995). Multivariate simultaneous generalized arch. *Econometric Theory 11*(1), 122–150.

Everett, M. G. and S. P. Borgatti (1999). The centrality of groups and classes. *The Journal of Mathematical Sociology 23*(3), 181–201.

Fan, J., A. Furger, and D. Xiu (2016). Incorporating global industrial classification standard into portfolio allocation: A simple factor-based large covariance matrix estimator with high-frequency data. *Journal of Business & Economic Statistics 34*(4), 489–503.

Fiszeder, P. and W. Orzeszko (2021). Covariance matrix forecasting using support vector regression. *Applied Intelligence 51*(10), 7029–7042.

Forbes, K. J. and R. Rigobon (2002). No contagion, only interdependence: measuring stock market comovements. *Journal of Finance 57*(5), 2223–2261.

Friedman, J., T. Hastie, and R. Tibshirani (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics 9*(3), 432–441.

Gilmer, J., S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl (2017). Neural message passing for quantum chemistry. In *International Conference on Machine Learning*, pp. 1263–1272. PMLR.

Glasserman, P. and H. P. Young (2015). How likely is contagion in financial networks? *Journal of Banking & Finance 50*, 383–399.

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

Hallac, D., Y. Park, S. Boyd, and J. Leskovec (2017). Network inference via the time-varying graphical lasso. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 205–213.

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

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

Hautsch, N., L. M. Kyj, and P. Malec (2015). Do high-frequency data improve high-dimensional portfolio allocations? *Journal of Applied Econometrics 30*(2), 263–290.

Herskovic, B., B. Kelly, H. Lustig, and S. Van Nieuwerburgh (2016). The common factor in idiosyncratic volatility: Quantitative asset pricing implications. *Journal of Financial Economics 119*(2), 249–283.

Herskovic, B., B. Kelly, H. Lustig, and S. Van Nieuwerburgh (2020). Firm volatility in granular networks. *Journal of Political Economy 128*(11), 4097–4162.

Izzeldin, M., M. K. Hassan, V. Pappas, and M. Tsionas (2019). Forecasting realised volatility using arfima and har models. *Quantitative Finance 19*(10), 1627–1638.

King, M. A. and S. Wadhwani (1990). Transmission of volatility between stock markets. *Review of Financial Studies 3*(1), 5–33.

Lee, T.-H. and X. Long (2009). Copula-based multivariate garch model with uncorrelated dependent errors. *Journal of Econometrics 150*(2), 207–218.

Ling, S. and M. McAleer (2003). Asymptotic theory for a vector arma-garch model. *Econometric Theory 19*(2), 280–310.

Liu, L. Y., A. J. Patton, and K. Sheppard (2015). Does anything beat 5-minute rv? a comparison of realized measures across multiple asset classes. *Journal of Econometrics 187*(1), 293–311.

Meinshausen, N. and P. Bühlmann (2006). High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics 34*(3), 1436–1462.

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

Pascalau, R. and R. Poirier (2021). Increasing the information content of realized volatility forecasts. *Journal of Financial Econometrics*.

Rapach, D. E., J. K. Strauss, and G. Zhou (2013). International stock return predictability: what is the role of the united states? *Journal of Finance 68*(4), 1633–1662.

Schwert, G. W. (1990). Stock volatility and the crash of'87. *Review of Financial Studies 3*(1), 77–102.

Sheppard, K. (2010). Financial econometrics notes. *University of Oxford*, 333–426.

Sidorov, S., A. Faizliev, V. Balash, A. Gudkov, A. Chekmareva, and P. Anikin (2016). Company co-mention network analysis. In *International Conference on Network Analysis*, pp. 341–354. Springer.

Sulem, D., H. Kenlay, M. Cucuringu, and X. Dong (2022). Graph similarity learning for change-point detection in dynamic networks. *arXiv preprint arXiv:2203.15470*.

Symitsi, E., L. Symeonidis, A. Kourtis, and R. Markellos (2018). Covariance forecasting in equity markets. *Journal of Banking & Finance 96*, 153–168.

Tokman, M., R. G. Richey, L. D. Marino, and K. M. Weaver (2007). Exploration, exploitation and satisfaction in supply chain portfolio strategy. *Journal of Business Logistics 28*(1), 25–56.

Varneskov, R. and V. Voev (2013). The role of realized ex-post covariance measures and dynamic model choice on the quality of covariance forecasts. *Journal of Empirical Finance 20*, 83–95.

Vassallo, D., G. Buccheri, and F. Corsi (2021). A dcc-type approach for realized covariance modeling with score-driven dynamics. *International Journal of Forecasting 37*(2), 569–586.

Wang, Y., Z. Pan, and C. Wu (2018). Volatility spillover from the us to international stock markets: A heterogeneous volatility spillover garch model. *Journal of Forecasting 37*(3), 385–400.

Wilms, I., J. Rombouts, and C. Croux (2021). Multivariate volatility forecasts for stock market indices. *International Journal of Forecasting 37*(2), 484–499.

Zhang, C., Y. Zhang, M. Cucuringu, and Z. Qian (2022). Volatility forecasting with machine learning and intraday commonality. *Available at SSRN 4022147*.

Zhu, X., R. Pan, G. Li, Y. Liu, and H. Wang (2017). Network vector autoregression. *The Annals of Statistics 45*(3), 1096–1123.

A Additional results of Section 4.6

Following Bollerslev et al. (2018), the results for GMVP and GMVP$^+$ based on longer horizon forecasting models are averaged across all possible weekly (i.e., Monday-to-Monday, Tuesday-to-Tuesday, etc.) and monthly (i.e., 1st June-to-1st July, 2nd June-to-2nd July, etc.) horizons. Table A.1 details the findings of portfolios re-balanced weekly (Panel A) and monthly (Panel B). To begin, note that modeling volatilities and correlations separately (HAR-DRD) manifests larger economic benefits than the one based on Cholesky decomposition. Compared to the HAR-DRD portfolio, the GHAR portfolios (especially GHAR(GL, $\widetilde{K}$)) achieve better ex-post standard deviations and reduce the turnover over the 1-week horizon. In contrast, the performances of most GHAR models are worse over 1-month horizon. This again provides evidence that the graph effect in the equity market mainly manifests itself at short-term horizons and decays rapidly in time.

Table A.1: Out-of-sample portfolio performance over longer forecast horizons.

	GMVP		GMVP$^+$	
	$\sigma^{(p)}$	$\tau^{(p)}$	$\sigma^{(p)}$	$\tau^{(p)}$
Panel A: 1-Week				
1/N	10.679	0.016	10.679	0.016
HAR-Cholesky	10.562	0.808	9.641	0.455
HAR-DRD	10.015	0.787	9.536	0.530
GHAR(-, $\widetilde{K}$)	10.000	0.740	9.483	0.488
GHAR(-, $\widetilde{L}$)	10.014	0.768	9.520	0.517
GHAR(S, -)	10.200	0.804	9.703	0.558
GHAR(S, $\widetilde{K}$)	10.177	0.731	9.624	0.510
GHAR(S, $\widetilde{L}$)	10.183	0.770	9.658	0.540
GHAR(K, -)	10.040	0.702	9.700	0.473
GHAR(K, $\widetilde{K}$)	10.033	0.620	9.624	0.414
GHAR(K, $\widetilde{L}$)	10.040	0.664	9.660	0.450
GHAR(GL, -)	9.924	0.754	9.516	0.519
GHAR(GL, $\widetilde{K}$)	**9.884**	0.671	**9.434**	0.463
GHAR(GL, $\widetilde{L}$)	9.892	0.717	9.479	0.498
Panel B: 1-Month				
1/N	10.286	0.035	10.286	0.035
HAR-Cholesky	10.721	1.107	10.130	0.630
HAR-DRD	9.990	0.778	**9.598**	0.558
GHAR(-, $\widetilde{K}$)	9.903	0.814	9.633	0.566
GHAR(-, $\widetilde{L}$)	**9.869**	0.812	9.601	0.578
GHAR(S, -)	10.085	0.840	9.684	0.605
GHAR(S, $\widetilde{K}$)	10.019	0.864	9.714	0.608
GHAR(S, $\widetilde{L}$)	9.989	0.870	9.704	0.622
GHAR(K, -)	10.085	0.795	9.717	0.575
GHAR(K, $\widetilde{K}$)	10.000	0.813	9.727	0.572
GHAR(K, $\widetilde{L}$)	9.969	0.821	9.705	0.588
GHAR(GL, -)	10.258	0.844	9.814	0.615
GHAR(GL, $\widetilde{K}$)	10.194	0.865	9.832	0.615
GHAR(GL, $\widetilde{L}$)	10.178	0.875	9.828	0.633

Note: The table reports the out-of-sample performance of GMVP and GMVP$^+$ constructed using the 1-week-ahead (Panel A) or 1-month-ahead (Panel B) covariance forecasts of various models. $\sigma^{(p)}$ is the annualized portfolio standard deviation, and $\tau^{(p)}$ is the average portfolio turnover. The lowest $\sigma^{(p)}$ attained in each column is indicated in bold.

B Short-term graph effect

The main aim of this section is to assess the performance of models with only the short-term graph effect. In other words, we consider models similar to **GHAR**($\boldsymbol{A}$) (Eqn (8)) and **GHAR**($\widetilde{\boldsymbol{A}}$) (Eqn (9)), but without the weekly and monthly graph effects. We refer to these models as **GHAR-S**. The results in Table B.1 show that forecast improvements relative to the benchmark HAR-DRD model remain significant even when we only incorporate the short-term graph effect.

Table B.1: Out-of-sample losses with the short-term graph effect.

	$\mathcal{L}^E$			$\mathcal{L}^F$			$\mathcal{L}^Q$		
	Ratio	Rank	p-val	Ratio	Rank	p-val	Ratio	Rank	p-val
HAR-DRD	1.000	12	0.010	1.000	12	0.019	1.000	10	<0.001
GHAR-S(-, $\widetilde{K}$)	0.997	11	0.008	0.997	11	0.013	0.992	8	<0.001
GHAR-S(-, $\widetilde{L}$)	0.996	10	0.010	0.995	9	0.019	0.990	4	0.005
GHAR-S(S, -)	0.994	9	0.010	0.995	10	0.019	1.000	12	<0.001
GHAR-S(S, $\widetilde{K}$)	0.992	8	0.010	0.993	8	0.019	0.992	7	<0.001
GHAR-S(S, $\widetilde{L}$)	0.991	7	0.010	0.991	6	0.019	0.990	3	0.001
GHAR-S(K, -)	0.990	6	0.010	0.991	7	0.019	1.000	9	<0.001
GHAR-S(K, $\widetilde{K}$)	0.989	5	0.010	0.990	5	0.019	0.991	5	<0.001
GHAR-S(K, $\widetilde{L}$)	0.988	3	0.010	0.989	3	0.019	0.989	2	0.387
GHAR-S(GL, -)	0.988	4	0.028	0.990	4	0.026	1.000	11	<0.001
GHAR-S(GL, $\widetilde{K}$)	0.987	2	0.028	0.988	2	0.026	0.991	6	<0.001
GHAR-S(GL, $\widetilde{L}$)	0.986	1	1.000	0.987	1	1.000	0.988	1	1.000

Note: The table reports the out-of-sample losses of the models with the short-term graph effect over the 1–day forecast horizon, averaged over the entire testing sample. $\mathcal{L}^E$ is the Euclidean distance, $\mathcal{L}^F$ is the Frobenius distance, and $\mathcal{L}^Q$ is the Quasi-Likelihood loss function.

C Alternative model update frequencies

A potential concern one could raise is that the superior forecasting ability of our GHAR models could be driven to some extent by the low update frequency (every month in our previous analysis) of models. To investigate this possibility, we re-run our analysis and update all models at higher frequencies, specifically daily and weekly. The results in Table C.1 are generally consistent with those from our main analysis.

Table C.1: Out-of-sample losses under alternative model update frequencies.

	$\mathcal{L}^E$			$\mathcal{L}^F$			$\mathcal{L}^Q$		
	Ratio	Rank	p-val	Ratio	Rank	p-val	Ratio	Rank	p-val
Panel A: Daily									
HAR-Cholesky	1.026	13	<0.001	1.027	13	<0.001	1.115	13	<0.001
HAR-DRD	1.000	12	0.015	1.000	12	0.023	1.000	9	<0.001
GHAR(-, $\widetilde{K}$)	0.993	10	0.022	0.992	8	0.035	0.986	3	<0.001
GHAR(-, $\widetilde{L}$)	0.991	7	0.035	0.990	7	0.067	0.984	1	1.000
GHAR(S, -)	0.995	11	0.019	0.996	11	0.020	1.007	11	<0.001
GHAR(S, $\widetilde{K}$)	0.989	6	0.030	0.989	6	0.035	0.991	6	<0.001
GHAR(S, $\widetilde{L}$)	0.988	5	0.062	0.988	4	0.110	0.988	5	<0.001
GHAR(K, -)	0.992	9	0.022	0.994	10	0.032	1.015	12	<0.001
GHAR(K, $\widetilde{K}$)	0.987	4	0.035	0.988	5	0.067	0.997	8	<0.001
GHAR(K, $\widetilde{L}$)	0.986	3	0.062	0.987	2	0.140	0.995	7	<0.001
GHAR(GL, -)	0.992	8	0.030	0.993	9	0.033	1.003	10	<0.001
GHAR(GL, $\widetilde{K}$)	0.986	2	0.062	0.987	3	0.110	0.987	4	<0.001
GHAR(GL, $\widetilde{L}$)	0.986	1	1.000	0.986	1	1.000	0.984	2	0.713
Panel B: Weekly									
HAR-Cholesky	1.029	13	<0.001	1.031	13	<0.001	1.104	13	<0.001
HAR-DRD	1.000	12	0.003	1.000	12	0.006	1.000	9	<0.001
GHAR(-, $\widetilde{K}$)	0.991	10	0.008	0.990	8	0.021	0.986	4	<0.001
GHAR(-, $\widetilde{L}$)	0.990	8	0.014	0.988	7	0.037	0.984	2	0.542
GHAR(S, -)	0.994	11	0.006	0.994	11	0.008	1.006	11	<0.001
GHAR(S, $\widetilde{K}$)	0.986	6	0.014	0.986	6	0.033	0.991	8	<0.001
GHAR(S, $\widetilde{L}$)	0.985	5	0.039	0.985	4	0.094	0.988	5	<0.001
GHAR(K, -)	0.990	9	0.010	0.992	10	0.012	1.008	12	<0.001
GHAR(K, $\widetilde{K}$)	0.984	4	0.014	0.985	5	0.037	0.991	7	<0.001
GHAR(K, $\widetilde{L}$)	0.983	3	0.039	0.983	3	0.094	0.988	6	<0.001
GHAR(GL, -)	0.989	7	0.012	0.991	9	0.016	1.002	10	<0.001
GHAR(GL, $\widetilde{K}$)	0.983	2	0.039	0.983	2	0.094	0.986	3	<0.001
GHAR(GL, $\widetilde{L}$)	0.982	1	1.000	0.982	1	1.000	0.983	1	1.000

Note: The table reports the out-of-sample losses of the competitive models when updated every day (Panel A) and every week (Panel B).

D Transformations for volatilities and correlations

In line with Andersen et al. (2006), we now assess the ability of our proposed models under the transformation for volatilities and correlations. Specifically, we apply the log transformation to volatilities (e.g. Andersen et al. (2003); Herskovic et al. (2016); Bucci (2020b); Zhang et al. (2022)) and the Fisher transformation to correlations (see Andersen et al. (2006)). Table D.1 presents the results for modeling transformed volatilities and correlations. As shown, GHAR(K, $\widetilde{L}$) is the best performing model, followed by GHAR(GL, $\widetilde{L}$). The

baseline model HAR-DRD is never included in the MCS, demonstrating the importance of graph structural information.

Table D.1: Out-of-sample losses for forecasting transformed volatilities and correlations.

	$\mathcal{L}^E$			$\mathcal{L}^F$			$\mathcal{L}^Q$		
	Ratio	Rank	p-val	Ratio	Rank	p-val	Ratio	Rank	p-val
HAR-DRD	1.000	12	0.003	1.000	12	0.002	1.000	12	<0.001
GHAR(-, $\widetilde{K}$)	0.992	11	0.004	0.991	11	0.003	0.979	8	<0.001
GHAR(-, $\widetilde{L}$)	0.991	10	0.004	0.990	10	0.004	0.975	6	<0.001
GHAR(S, -)	0.987	9	0.004	0.986	9	0.002	0.998	11	<0.001
GHAR(S, $\widetilde{K}$)	0.980	8	0.004	0.978	8	0.004	0.976	7	<0.001
GHAR(S, $\widetilde{L}$)	0.979	7	0.004	0.977	7	0.004	0.972	3	<0.001
GHAR(K, -)	0.974	5	0.004	0.973	5	0.004	0.996	9	<0.001
GHAR(K, $\widetilde{K}$)	0.969	2	0.004	0.967	2	0.004	0.973	4	<0.001
GHAR(K, $\widetilde{L}$)	0.968	1	1.000	0.966	1	1.000	0.968	1	1.000
GHAR(GL, -)	0.978	6	0.004	0.977	6	0.004	0.998	10	<0.001
GHAR(GL, $\widetilde{K}$)	0.972	4	0.004	0.970	4	0.004	0.975	5	<0.001
GHAR(GL, $\widetilde{L}$)	0.972	3	0.004	0.970	3	0.004	0.969	2	0.523

Note: The table reports the out-of-sample losses of the models for forecasting transformed volatilities and correlations over the 1-day forecast horizon, averaged over the entire testing sample. $\mathcal{L}^E$ is the Euclidean distance, $\mathcal{L}^F$ is the Frobenius distance, and $\mathcal{L}^Q$ is the Quasi-Likelihood loss function.