ARTICLE IN PRESS

International Journal of Forecasting xxx (xxxx) xxx

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Outlier-robust methods for forecasting realized covariance matrices

Dan Li[a,c,*], Christopher Drovandi[b,c], Adam Clements[a,c]

[a] School of Economics and Finance, Queensland University of Technology, Australia
[b] School of Mathematical Sciences, Queensland University of Technology, Australia
[c] QUT Centre for Data Science, Queensland University of Technology, Australia

ARTICLE INFO

Keywords:
Multivariate volatility
HAR
Portfolio allocation
Minimum covariance determinant
Multivariate regression
Least-trimmed squares estimator

ABSTRACT

This paper proposes two new approaches to improve the estimation of the coefficients of the multivariate HAR (MHAR) model with the primary purpose of improving forecast performance. A robust estimator of the covariance matrix is adopted to replace the realized covariance matrix while estimating the MHAR model. The robustness to outliers of the new estimator makes the OLS estimation scheme for the MHAR model more reliable. In addition, a robust estimation scheme is developed for the MHAR model, which is based on the multivariate least-trimmed squares method. Both approaches provide significant improvements in forecasting performance based on both statistical loss and portfolio outcomes. The forecast performance of the multivariate HARQ model can also be improved with the proposed approaches, as evidenced by robustness checks.

1. Introduction

Forecasting multivariate volatility has attracted much attention in the literature, as forecasts of the covariance matrix are important inputs in applications such as portfolio allocation. A number of competing approaches have been proposed for modeling the dynamics of multivariate financial volatility, including the widely used multivariate GARCH model and its various extensions (reviewed in Bauwens et al., 2006), and multivariate stochastic volatility models (see, for example, Harvey et al., 1994). However, the parametric multivariate GARCH-type and stochastic volatility models traditionally incorporated low-frequency data and treated multivariate volatilities as latent. In contrast, the emergence of high-frequency intraday data has stimulated the construction of realized measures, which makes volatilities observable and motivates methods that directly employ the realized measures in forecasting models.

Realized volatility (RV) is computed by summing intraday squared returns sampled at short intervals for a particular trading day. Andersen et al. (2001) revealed that RV is an unbiased ex post estimator of daily return volatility. Consequently, there has been an increasing interest in modeling and forecasting RV. One of the widely used models for forecasting RV is the heterogeneous autoregressive (HAR) model of Corsi (2009). The simplicity of the HAR model, and its superior forecasting performance, means it has become the workhorse in the literature (see, for example, Bubák et al., 2011; Busch et al., 2011; Souček & Todorova, 2013). In the context of multivariate volatility, a realized covariance (RCov) matrix of financial asset returns is constructed. Chiriac and Voev (2011) extended the univariate HAR to the multivariate case, namely multivariate HAR (MHAR), to model and forecast RCov. As with the univariate HAR model, the MHAR model has a simple and easy-to-implement structure and is often estimated using the method of ordinary least squares (OLS). Due to the stylized features of RV and RCov, the OLS estimator can become inefficient and may lead to bias in the estimates of parameters in the models. It is well known that RV exhibits large spikes and

* Corresponding author at: School of Economics and Finance, Queensland University of Technology, Australia.
E-mail address: d31.li@hdr.qut.edu.au (D. Li).

https://doi.org/10.1016/j.ijforecast.2023.04.004
0169-2070/

Please cite this article as: D. Li, C. Drovandi and A. Clements, Outlier-robust methods for forecasting realized covariance matrices. International Journal of Forecasting (2023), https://doi.org/10.1016/j.ijforecast.2023.04.004.

outliers and is non-Gaussian, which can make the OLS estimator inefficient, as OLS is highly sensitive to (i.e. not robust against) outliers and suboptimal in the presence of non-normality. In this paper, we propose two approaches to avoiding this issue. One is to adjust the underlying covariance matrix estimate to mitigate the outliers. The other is to employ an estimation scheme that is robust to the existence of outliers.

Clements and Preve (2021) used the method of weighted least squares (WLS) for estimating the univariate HAR model. They showed that the predictive accuracy of the HAR model improved by using the estimator of WLS. The idea of using data transformation, such as the square root and log transformations, has been widely raised in the literature in the context of modeling RV (see, for example, Corsi, 2009; Taylor, 2017a), which was motivated by the non-Gaussian feature of RV. Data transformations appear useful for reducing skewness, and hence the possible effect of outliers and potential heteroskedasticity in RV (Clements & Preve, 2021). Taylor (2017a) showed that the forecast accuracy of RV can be improved by applying transformations in the HAR model to deal with these issues. In the context of the multivariate HAR model, the forecasting method proposed by Bauer and Vorkink (2011) used matrix logarithm transformations of the covariance matrix to improve the accuracy of the RCov forecast. Furthermore, apart from improving the estimation scheme or transforming RV or RCov, there has been a focus on dealing with the issue of measurement error to get more accurate forecasting results. The HARQ model of Bollerslev et al. (2016) and the multivariate HARQ (MHARQ) model of Bollerslev et al. (2018) account for the measurement errors in the original HAR and multivariate HAR models, respectively, and hence provide better forecasting performance. The HARQ-type models allow for less (resp. more) weight to be placed on historical values of RV or RCov when the measurement error is higher (resp. lower). Cipollini et al. (2021) investigated the correlation between the square root of realized quarticity and realized variance. Based on this relationship, they proposed an approach to capture the attenuation bias, which reduces the need for the quarticity term in the HARQ-type models. Recently, Quaedvlieg (2020) used pre-shrunk realized quantities for modeling and forecasting volatilities, which provides an alternative way to deal with the measurement error in the realized measures.

In this paper, instead of directly dealing with the measurement errors or transforming RCov when applying the MHAR model, we propose two approaches that make the estimation scheme more robust to outliers, resulting in more accurate forecasts accordingly. Firstly, a new estimate of the covariance matrix, which is more robust to outliers, is proposed to replace the raw RCov while estimating the parameters in the MHAR model. By employing the robust estimate of RCov, the OLS estimator becomes more reliable, improving parameter estimation and hence forecast performance. The robust estimate of RCov is an alternative approach for reducing the skewness in the individual elements of the RCov matrix. As discussed above, reducing the skewness in RV leads to improvements in forecast accuracy in the univariate setting. This may be due to a range of reasons. Reducing the skewness in RV, whether by transformations or otherwise, reduces the effects of the peaks in the series and increases its persistence and hence predictability (see Ding & Granger, 1996; Ding et al., 1993; Proietti & Lütkepohl, 2013). As pointed out by Taylor (2017b), some models may be better suited to the properties of altered or transformed data. Given that the HAR model is specifically designed to parsimoniously capture the persistence in RV, the HAR model may be better suited to a version of RV that is more persistent. Opschoor and Lucas (2022) find that modeling the skewness (along with its volatility) in RV is important for forecasting the one-day-ahead density of RV and its associated tail risk. As opposed to the full density of RV, here we are following a long line of research which focuses on forecasting the mean level of RV or RCov within the context of regression-type models. Within this context, reducing the skewness (via a range of methods) in the estimate of volatility has been found to be beneficial for forecasting. Secondly, we employ the multivariate least-trimmed squares (MLTS) method of Agulló et al. (2008) to develop a more robust estimation scheme for the MHAR model. With the presence of extreme values in RCov, MLTS is more robust than the OLS estimation when using standard RCov, leading to more accurate forecasts of RCov.

The merits of our proposed approaches are assessed by undertaking a simulation study and an empirical analysis of the performance of out-of-sample forecasts. In the empirical analysis, we consider portfolios with five different sizes. Based on the simulation and empirical studies, we find significant gains, both statistically and economically, using our proposed approaches. In addition, the forecasting improvements of the MLTS approach increase with the portfolio dimension in terms of the quasi-likelihood (QLIKE) loss function. Moreover, robustness checks are provided by applying the proposed approaches with other existing approaches that produce more accurate forecasts than the MHAR model with the OLS estimator. By utilizing our proposed approaches to the MHARQ model, more accurate forecasts can be generated compared to the ones using the OLS estimator. Furthermore, price jumps are taken into consideration, as the effect of jumps on the estimation of volatility is nontrivial (Bollerslev et al., 2020). A jump-robust covariance estimate and truncated intraday returns with the removal of jumps are considered. The benefits of the proposed approaches are confirmed even after mitigating the effect of jumps.

The paper is organized as follows. Section 2 introduces the concept of RCov and the MHAR model. Section 3 describes the robust estimator of the covariance matrix and the MLTS framework for the MHAR model. This section also explains how the resulting coefficient estimates are used for making RCov forecasts, along with the approaches for evaluating forecasting performance. Section 4 describes the simulation study. Section 5 reports the empirical results. Robustness checks are presented in Section 6. Concluding remarks are provided in Section 7.

2. Realized covariance and the multivariate heterogeneous autoregressive (MHAR) model

Realized covariance (RCov) matrices are generated from the high-frequency intraday prices of a set of financial assets. Suppose the N-dimensional log-price process $P(t)$ follows a continuous time diffusion process:

$$P(t) = \int_0^t \mu(u)du + \int_0^t \sigma(u)dW(u), \tag{1}$$

where $\mu(u)$ is a vector of drift components; $\sigma(u)$ is the instantaneous volatility matrix, with the corresponding spot covariance matrix $\Sigma(u) \equiv \sigma(u)\sigma(u)'$; and $W(u)$ is an N-dimensional vector of independent Brownian motion. The quantity of interest here is

$$\boldsymbol{\Sigma_t} = \int_{t-1}^{t} \Sigma(u)du, \tag{2}$$

known as the integrated covariance of $P(t)$ over day t. It is a measure of the daily ex post covariation of $P(t)$. The latent integrated covariance $\boldsymbol{\Sigma_t}$ can be approximated by the realized covariance matrix S_t, which is computed from intraday returns sampled at equally spaced intervals.

Suppose $r_{t,i}$ is the ith period intraday return of these N assets, which is computed as

$$r_{t,i} = P_{t-1+i\Delta} - P_{t-1+(i-1)\Delta}, \; i = 1, 2, \ldots, M, \tag{3}$$

for day t with M equidistant intervals of length Δ. Assuming intraday returns have a mean of zero, then the RCov S_t is given by

$$S_t = \sum_{i=1}^{M} r_{t,i} r'_{t,i}, \tag{4}$$

which converges to the true covariance $\boldsymbol{\Sigma_t}$ as M approaches infinity (Andersen et al., 2003).

Chiriac and Voev (2011) extended the widely used heterogeneous autoregressive (HAR) model of Corsi (2009), originally designed for forecasting univariate RV, to a simple-to-implement multivariate version. Let $s_t \equiv \text{vech}\, S_t$ denote the $N^* = N(N+1)/2$ dimensional vectorized version of the realized covariance matrix. The scalar version of the multivariate HAR (MHAR) model can be expressed as

$$s_t = \theta_0 + \theta_1 s_{t-1} + \theta_2 s_{t-1|t-5} + \theta_3 s_{t-1|t-22} + u_t, \tag{5}$$

where $s_{t-1|t-k} = \frac{1}{k}\sum_{i=1}^{k} s_{t-i}$ denotes the vectorized version of the k-day realized covariance matrix, u_t is an error term, θ_0 is an $N^* \times 1$ parameter vector, and θ_1, θ_2, and θ_3 are scalar parameters.

3. Methodology

Since outliers are often present in realized measures of volatility, standard OLS estimation may lead to biases in the estimates of parameters of the MHAR model. We aim to address these issues from two perspectives. First, a more robust covariance matrix estimator is proposed to replace the sample covariance matrix RCov while using OLS to estimate the MHAR coefficients. The minimum covariance determinant (MCD) method of Rousseeuw and Driessen (1999) is employed to obtain a robust estimator of the covariance matrix, the so-called MCD covariance. Here, the OLS estimation scheme remains unchanged, but the raw RCov is replaced by MCD covariance. Second, a more robust estimation scheme relative to OLS is considered. The multivariate least-trimmed squares (MLTS) method of Agulló et al. (2008) is adopted to develop a more robust estimator for the MHAR model.

In this section, we first consider a replacement of the MCD covariance matrix for the raw RCov to estimate the coefficients of the MHAR model via the OLS estimation scheme. Next, instead of replacing the RCov, we change the OLS estimator to the MLTS estimation scheme, which is more robust to outliers. Moreover, we combine these two approaches to estimating the MHAR model with the MCD covariance matrix via MLTS.

3.1. Minimum covariance determinant estimator

For a large-dimensional multivariate dataset, it is challenging to detect outliers by visual inspection. The MCD method of Rousseeuw (1985) employed in this work provides a robust estimator of the multivariate location and covariance matrix. The MCD method is based on the computation of the ellipsoid with the smallest covariance determinant that would contain at least half of the data points (see Rousseeuw, 1985; Rousseeuw & Driessen, 1999 for details). Consider a set of intraday returns of N assets $\mathbf{r}_M = \{r_i; i = 1, \ldots, M\}$ for any trading day with M equidistant intervals. Let $H \subset \{1, \ldots, M\}$ where the size of H is $h \geq M/2$, and $\mathbf{r}_H = \{r_i; i \in H\}$ is a subset of $\mathbf{r}_M$. The objective of MCD is to find an optimal subset with h observations (out of M), $\hat{\mathbf{r}}_H$, whose covariance matrix has the smallest determinant. Then the MCD estimate of location is the mean of these h observations:

$$\boldsymbol{l} = \frac{1}{h}\sum_i r_i, \quad r_i \in \hat{\mathbf{r}}_H. \tag{6}$$

The covariance estimate is given by

$$\boldsymbol{C} = a_1 a_2 \frac{1}{h}\sum_i (r_i - \boldsymbol{l}) \times (r_i - \boldsymbol{l})^{\intercal}, \quad r_i \in \hat{\mathbf{r}}_H, \tag{7}$$

where a_1 is a consistency factor, and a_2 is a small-sample correction factor (Rousseeuw et al., 2004). Since the RCov defined in Eq. (4) is an approximation of the integrated covariance, $\boldsymbol{C}$ in Eq. (7) needs to be scaled by the sample size h to generate an integrated covariance estimate $\text{RbsC} = h \times \boldsymbol{C}$.

According to Rousseeuw and Driessen (1999), the MCD method is based on the fact that starting from any approximation to the MCD, another approximation with a lower determinant can be computed until the newer approximation is the MCD matrix, which has been proven by Rousseeuw and Driessen (1999). The step for finding the matrix with the lowest determinant is called the C-step. Consider an initial h-subset $H_1 \subset \{1, \ldots, M\}$ that contains intraday returns $\{r_i; i \in H_1\}$ and the mean μ_1 calculated by (6). Compute $\Sigma_1 = \frac{1}{h}\sum_{i \in H_1}(r_i - \mu_1)(r_i - \mu_1)'$ and the relative distance $d_1(i) = \sqrt{(r_i - \mu_1)'\Sigma^{-1}(r_i - \mu_1)}$ for $i = 1, \ldots, M$. Then the C-step works as follows in algorithmic terms:

- Sort the distances $d_1(i)$ for $i = 1, \ldots, M$, which yields a permutation π for which $d_1(\pi(1)) \leq d_1(\pi(2)) \leq \cdots \leq d_1(\pi(M))$

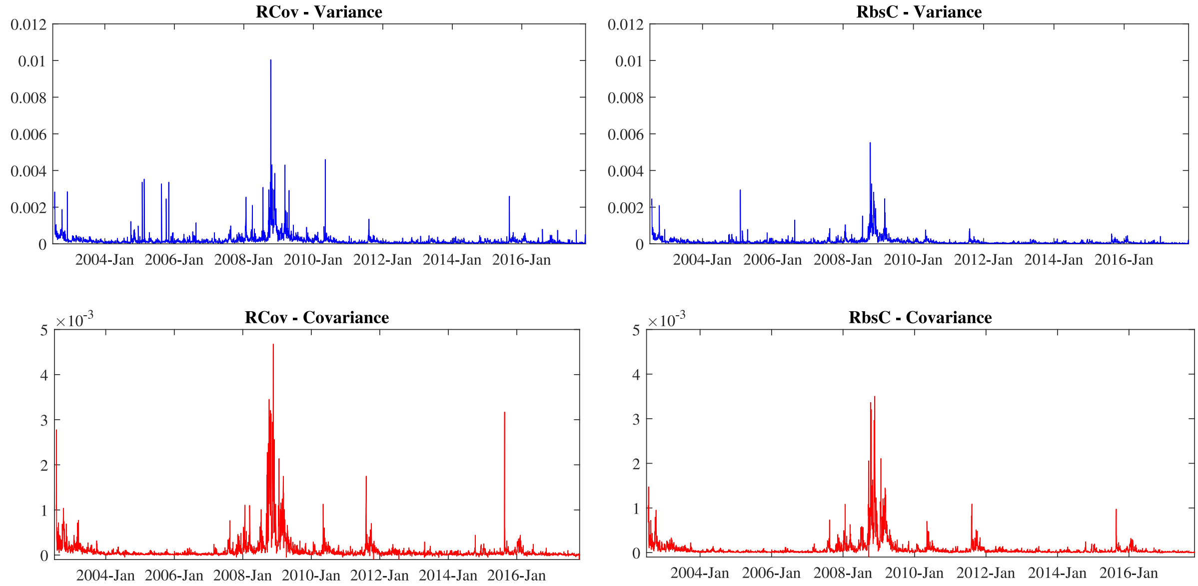

Fig. 1. Fifteen-asset portfolio: The series of variance of one stock and the series of covariance of two assets.

- Set $H_2 = \{\pi(1), \pi(2), \ldots, \pi(h)\}$
- Compute μ_2 and Σ_2 based on H_2

Then iterate the above C-steps until $\det(\Sigma_{\text{new}}) = 0$ or $\det(\Sigma_{\text{new}}) = \det(\Sigma_{\text{old}})$. The C-step is also a crucial step for the MLTS approach, and the details for implementing the C-step in MLTS are provided in Section 3.2.

An exact MCD algorithm can be computationally expensive, especially when the dataset is of high dimension (i.e. when the number of assets N is large). Rousseeuw and Driessen (1999) proposed a fast algorithm for computing the MCD estimator (FAST-MCD), which is available via the Matlab function "robustcov". We choose $h = 0.75M$, which provides a good balance between breakdown value and statistical efficiency as suggested by Rousseeuw and Driessen (1999). Fig. 1 plots a series of daily realized variances in the first row and a series of realized covariances of two assets in the second row. The assets are selected from the 15-asset portfolio in Table 22. The realized variance is for stock MRK, and the realized covariance is for JNJ and MSFT. The left panel shows elements of the raw Rcov, and the right panel provides elements of the RbsC. It is evident that the elements from RbsC contain smaller spikes relative to RCov, with the dynamics of the RbsC series remaining consistent with the overall movement level in RCov. The skewnesses of the 120 unique elements of the RCov and MCD covariance matrices for the 15-asset portfolio are provided in Fig. 2. Lower skewness exhibited by most elements in RbsC reflects much fewer outliers in the robust estimator.

As RbsC can mitigate the effect of extreme intraday returns, it is used to generate forecasts from the MHAR model in two approaches. Firstly, the coefficients of the MHAR model in Eq. (5) are obtained from

$$s_t^{rbs} = \theta_0^{rbs} + \theta_1^{rbs} s_{t-1}^{rbs} + \theta_2^{rbs} s_{t-1|t-5}^{rbs} + \theta_3^{rbs} s_{t-1|t-22}^{rbs} + u_t, \quad (8)$$

where $s_t^{rbs} \equiv \text{vech}\,\text{RbsC}_t$ denotes the $N^* = N(N+1)/2$ dimensional vectorized version of the MCD covariance matrix, and $s_{t-1|t-k}^{rbs} = \frac{1}{k}\sum_{i=1}^{k} s_{t-i}^{rbs}$ denotes the vectorized version of the k-day MCD covariance matrix.

After obtaining the MHAR coefficient estimates from Eq. (8) that are robust to the outliers in the raw covariance matrix estimates, the forecasts of RCov can be generated in two ways as follows.

$$\widehat{s}_t = \widehat{\theta}_0^{rbs} + \widehat{\theta}_1^{rbs} s_{t-1} + \widehat{\theta}_2^{rbs} s_{t-1|t-5} + \widehat{\theta}_3^{rbs} s_{t-1|t-22}, \quad (9)$$

where the lags of the vectorized version of RCov are used together with the coefficient estimates of Eq. (8). The other approach includes the MCD covariance in the MHAR model but still targets the RCov forecasts:

$$\widehat{s}_t = \widehat{\theta}_0^{rbs} + \widehat{\theta}_1^{rbs} s_{t-1}^{rbs} + \widehat{\theta}_2^{rbs} s_{t-1|t-5}^{rbs} + \widehat{\theta}_3^{rbs} s_{t-1|t-22}^{rbs}. \quad (10)$$

Both the approaches of Eqs. (9) and (10) are implemented in the empirical analysis, and the resulting forecasts are compared.

3.2. Multivariate least-trimmed squares estimator

Under the MCD approach, the variables are trimmed before implementing the OLS regression. An alternative is to leave the covariance matrix unchanged but trim the covariance matrix of the residuals to ensure that the parameter estimation is robust to outliers. We employ the MLTS approach of Agulló et al. (2008), which aims to find an optimal subset of h observations yielding the minimum determinant of the covariance matrix of its residuals from the OLS fitting. The interrelationship among different components of the error term is considered by taking the determinant. Agulló et al. (2008) proved that the MLTS estimator is Fisher-consistent.

Let the collection of all subsets of size h be denoted as $\mathcal{H} = \{H \subset \{1, \ldots, \mathcal{M}\} | \#H = h\}$, where $\mathcal{M}$ is the size of the RCov observations. Let $\widehat{\boldsymbol{\theta}}_{\text{OLS}}(H)$ be the OLS coefficient estimates for the model in Eq. (5) based on the observations $s_H = \{s_t; t \in H\}$ with the dimension

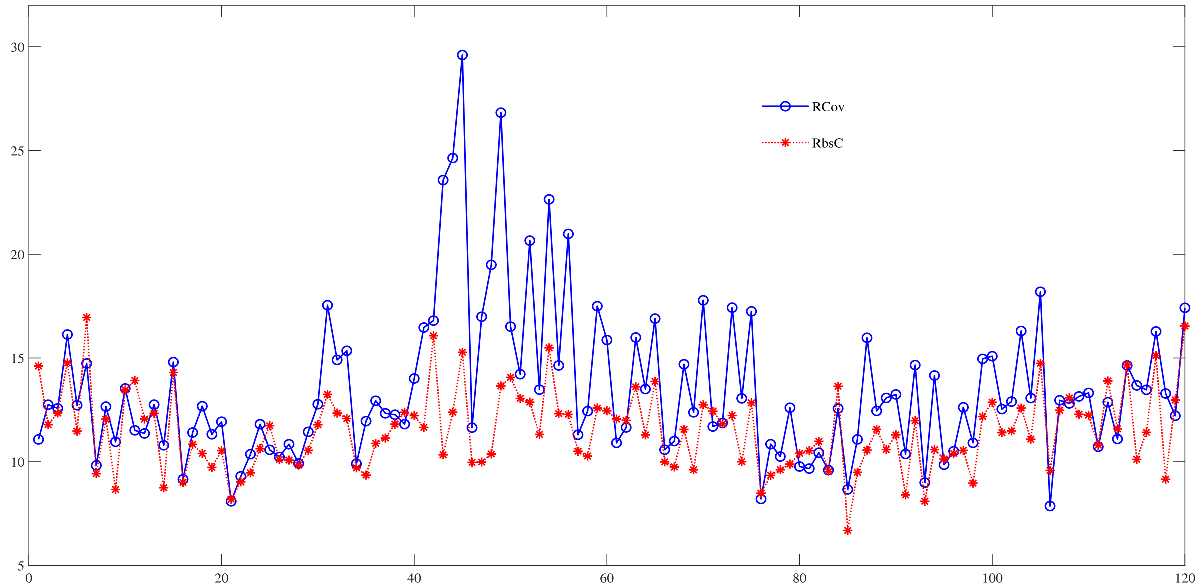

Fig. 2. Fifteen-asset portfolio: The skewness of 120 unique elements of the RCov and MCD covariance matrices (RbsC).

of $N^* \times h$. The covariance matrix of the residuals of this subset is given by

$$\widehat{\Sigma}_{\mathrm{OLS}}(H) = \frac{1}{h-p}\left(s_H - \widehat{s}_H\right)\left(s_H - \widehat{s}_H\right)^{\intercal}, \tag{11}$$

where p is the number of parameters, and $\widehat{s}_H$ is the forecasts of s_H based on $\widehat{\boldsymbol{\theta}}_{\mathrm{OLS}}(H)$. Then the MLTS estimator is given by

$$\widehat{\boldsymbol{\theta}}_{\mathrm{MLTS}}(s_{\mathcal{M}}) = \widehat{\boldsymbol{\theta}}_{\mathrm{OLS}}(\hat{H}) \quad \text{where } \hat{H} \in \underset{H \in \mathcal{H}}{\arg\min} \det\ \widehat{\Sigma}_{\mathrm{OLS}}(H), \tag{12}$$

and the covariance matrix of the errors is estimated by

$$\widehat{\Sigma}_{\mathrm{MLTS}}(s_{\mathcal{M}}) = a_0 \widehat{\Sigma}_{\mathrm{OLS}}(\hat{H}), \tag{13}$$

where a_0 is a consistent factor (Agulló et al., 2008). The MLTS estimator has an equivalent characterization with the estimator based on the squared Mahalanobis distance of the residuals. For parameters $\boldsymbol{\theta} \in \mathbb{R}^{p\times q}$ and a positive definite and symmetric matrix of size q, Σ, the squared distance is given by

$$d^2 = \left(s - \hat{s}\right)^{\intercal} \Sigma^{-1} \left(s - \hat{s}\right). \tag{14}$$

The MLTS estimator algorithm for the MHAR model is as follows:

1. Draw m random $(p+q)$-subsets J_m from the vectorized version of $\mathcal{M}$ RCov observations.
2. For each subset, compute the corresponding OLS estimates $\widehat{\boldsymbol{\theta}}_{\mathrm{OLS}}(J_m)$ and sample covariance matrix $\widehat{\Sigma}_{\mathrm{OLS}}(J_m)$ of its residuals. If the determinant of $\widehat{\Sigma}_{\mathrm{OLS}}(J_m)$ is zero, then draw additional points until the determinant is greater than zero or $\#J_m = h$.
3. Compute the squared residual distances defined in Eq. (14) for all $\mathcal{M}$ observations based on the OLS estimates and denote H_1 as the subset corresponding to the h observations with the smallest squared residual distances.
4. Construct a new subset H_2 that is more concentrated than H_1 where $\det(\widehat{\Sigma}_{H_2}) < \det(\widehat{\Sigma}_{H_1})$ using the approach described next. Compute OLS estimates $\widehat{\boldsymbol{\theta}}_{\mathrm{OLS}}(H_1)$ and sample covariance matrix $\widehat{\Sigma}_{\mathrm{OLS}}(H_1)$ of the residuals based on the subset H_1. Then repeat Step 3 based on the OLS estimates from H_1 to get a new subset H_2 with h smallest residual distances. By doing this, Agulló et al. (2008) argued that $\det(\widehat{\Sigma}_{H_2}) \leq \det(\widehat{\Sigma}_{H_1})$ *with equality if and only if* $\widehat{\boldsymbol{\theta}}_{\mathrm{OLS}}(H_2) = \widehat{\boldsymbol{\theta}}_{\mathrm{OLS}}(H_1)$ *and* $\widehat{\Sigma}_{\mathrm{OLS}}(H_2) = \widehat{\Sigma}_{\mathrm{OLS}}(H_1)$. This step is called the C-step. Two or three C-steps are applied to reduce the value of the objective function.
5. After completing the above steps for all m subsets, 10 subsets with the lowest determinant are selected, and further C-steps are carried out until convergence. Then the final solution is the subsample with the lowest determinant among these 10.

The number of the initial random subsets h should be large enough, and Agulló et al. (2008) suggested that using $m = 1000$ is usually sufficient.

The MLTS estimator is originally designed for a full-sized multivariate model, where the dimension of parameters is the multiplication of dimensions of responses and regressors $p \times q$. It is worth noting that even though the number of scale parameters is three in the MHAR model, the parameter matrix can be transformed to a $p \times q$ matrix similar to the one in the full version of the

MHAR model when implementing the MLTS estimator. However, unlike the one in the full MHAR model, the transformed coefficient parameters matrix contains some zero elements, as shown below:

$$\underset{(p\times q)}{\boldsymbol{\theta}} = \begin{pmatrix} \theta_0^1 & \theta_0^2 & \dots & \theta_0^q \\ \theta_1 & 0 & \dots & 0 \\ 0 & \theta_1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \theta_1 \\ \theta_2 & 0 & \dots & 0 \\ 0 & \theta_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \theta_2 \\ \theta_3 & 0 & \dots & 0 \\ 0 & \theta_3 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \theta_3 \end{pmatrix}, \tag{15}$$

where θ_1, θ_2, and θ_3 are scalar parameters, $q = N^*$, and p is the number of regressors.

After obtaining the MLTS estimates, the RCov forecasts can be generated from the MHAR model as follows:

$$\hat{s}_t = \hat{\theta}_0^{\mathrm{MLTS}} + \hat{\theta}_1^{\mathrm{MLTS}} s_{t-1} + \hat{\theta}_2^{\mathrm{MLTS}} s_{t-1|t-5} + \hat{\theta}_3^{\mathrm{MLTS}} s_{t-1|t-22}. \tag{16}$$

3.3. Combined method

We also consider the combination of the two approaches above by using MLTS to estimate the MHAR model with RbsC data. The coefficients are produced from

$$s_t^{rbs} = \theta_0^{\mathrm{CMB}} + \theta_1^{\mathrm{CMB}} s_{t-1}^{rbs} + \theta_2^{\mathrm{CMB}} s_{t-1|t-5}^{rbs} + \theta_3^{\mathrm{CMB}} s_{t-1|t-22}^{rbs} + u_t. \tag{17}$$

After generating the estimated coefficients for Eq. (17) via MLTS, the forecasts of RCov can be generated using either raw realized covariances or MCD covariances as covariates, as follows:

$$\hat{s}_t = \hat{\theta}_0^{\mathrm{CMB}} + \hat{\theta}_1^{\mathrm{CMB}} s_{t-1} + \hat{\theta}_2^{\mathrm{CMB}} s_{t-1|t-5} + \hat{\theta}_3^{\mathrm{CMB}} s_{t-1|t-22}, \tag{18}$$

or

$$\hat{s}_t = \hat{\theta}_0^{\mathrm{CMB}} + \hat{\theta}_1^{\mathrm{CMB}} s_{t-1}^{rbs} + \hat{\theta}_2^{\mathrm{CMB}} s_{t-1|t-5}^{rbs} + \hat{\theta}_3^{\mathrm{CMB}} s_{t-1|t-22}^{rbs}. \tag{19}$$

3.4. Forecast evaluation

Two statistical loss functions, the Frobenius distance and the empirical quasi-likelihood (QLIKE) loss are used for forecast evaluation. Let F_t denote the one-step-ahead forecast of the ex post realized covariance RCov_t. The Frobenius norm that measures the distance between two matrices is given by

$$L_t^F = \sqrt{Tr[(F_t - \mathrm{RCov}_t)(F_t - \mathrm{RCov}_t)^\intercal]}. \tag{20}$$

The QLIKE measure used in this work is the so-called James–Stein loss, which is usually referred to as the multivariate quasi-likelihood loss function (see, for example, Luo & Chen, 2020; Quaedvlieg, 2020). The QLIKE measure is expressed as

$$L_t^Q = Tr(F_t^{-1}\mathrm{RCov}_t) - \log|F_t^{-1}\mathrm{RCov}_t| - N. \tag{21}$$

The sums of L_t^F and L_t^Q over the out-of-sample period T will be reported, which are $L^F = \sum_{t=1}^T L_t^F$ and $L^Q = \sum_{t=1}^T L_t^Q$, respectively.

In order to detect whether the forecasts from different approaches differ significantly, the model confidence set (MCS) of Hansen et al. (2011) is employed. The MCS approach identifies the best forecasting model or models among all candidates with $(1-\alpha)$ confidence based on the loss functions. The reported results in this work consider both the range and squared statistics proposed in Hansen et al. (2003).

Following Bollerslev et al. (2018), in addition to the statistical measures, we consider the economic evaluations based on the construction of global minimum variance (GMV) portfolios. Suppose N risky assets comprise a portfolio and the assets are allocated based on the forecasts of daily covariance matrix F_{t+1}. The optimal GMV portfolio allocation weights minimizing the conditional volatility can be obtained by solving

$$\hat{w}_{t+1} = \arg\min w_{t+1}^\intercal F_{t+1} w_{t+1}, \; s.t. \; w_{t+1}^\intercal \iota = 1, \tag{22}$$

where ι is an $N \times 1$ vector of ones, and the solution is

$$\hat{w}_{t+1} = \frac{F_{t+1}^{-1}\iota}{\iota^\intercal F_{t+1}^{-1}\iota}. \tag{23}$$

Lower portfolio turnover is preferred, as this reduces the transaction costs of implementing trading strategies based on a forecast. According to Bollerslev et al. (2018), the total portfolio turnover from day t to $t+1$ can be measured by

$$TO_t = \sum_{n=1}^{N} \left| w_{t+1}^{(n)} - w_t^{(n)} \frac{1 + r_t^{d(n)}}{1 + w_t^\intercal r_t^{d(n)}} \right|, \tag{24}$$

where $w_t^{(n)}$ is the weight of the nth asset, and $r_t^{d(n)}$ is the daily return on the nth asset. In addition, we report the portfolio variance based on the raw RCov to assess the stability of a portfolio. The portfolio variance is computed as

$$V_p^t = w_t \; \mathrm{RCov}_t \; w_t^\intercal. \tag{25}$$

Moreover, we report the portfolio concentration that measures the extremeness in the portfolio allocation weights:

$$CO_t = \left(\sum_{n=1}^{N} w_t^{(n)2} \right)^{1/2}. \tag{26}$$

Finally, the total amount of the following short positions is reported:

$$SP_t = \sum_{n=1}^{N} w_t^{(n)} \mathbb{I}_{w_t^{(n)}<0}. \tag{27}$$

Since implementing short positions is usually more expensive than long positions, fewer short positions are generally preferred.

4. Simulation study

In the simulation study, we simulate daily covariances for a portfolio with 15 assets to evaluate the forecasting

Table 1
Summary of forecasting approaches.

	Variables		Estimator
	Estimation	Forecast	
$F_{\mathrm{RCov;RCov}}$	RCov	RCov	OLS
$F_{\mathrm{MCD;RCov}}$	RbsC	RCov	OLS
$F_{\mathrm{MCD;MCD}}$	RbsC	RbsC	OLS
F_{MLTS}	RCov	RCov	MLTS
$F_{\mathrm{CMB;RCov}}$	RbsC	RCov	MLTS
$F_{\mathrm{CMB;MCD}}$	RbsC	RbsC	MLTS

Table 2
Out-of-sample forecast results: Ratios of loss scores (15-asset portfolio simulation).

	L^F	L^Q	V_p	*TO*	*SP*	*CO*
$F_{\mathrm{RCov;RCov}}$	1.000	1.000	1.000	1.000	1.000	1.000
$F_{\mathrm{MCD;RCov}}$	1.004	**0.997***	0.997	0.916	0.966	0.993
$F_{\mathrm{MCD;MCD}}$	1.000	1.000	0.999	0.930	0.984	0.997
F_{MLTS}	**0.947***	**0.996***	**0.962***	0.559	0.452	0.855
$F_{\mathrm{CMB;RCov}}$	**0.949***	0.999	0.962	0.531	0.428	0.849
$F_{\mathrm{CMB;MCD}}$	**0.950***	1.002	0.963	0.538	0.433	0.850

Notes: Bold indicates member(s) of the MCS with the range statistics at 5%. * indicates member(s) of the MCS with the squared statistics at 5%. Shaded cells indicate the lowest value in each column.

performance of the proposed approaches. Following the simulation design of Bollerslev et al. (2018), 4193 daily multivariate kernel (MK) covariance (Barndorff-Nielsen et al., 2011) estimates are obtained for the 15 U.S. stocks in Table 22. This 15-asset portfolio is also analyzed empirically in Section 5, where the data are discussed in more detail. We then treat the MK estimate as the integrated covariance matrix for each day and simulate one-second returns based on the MK covariance estimate for that day. A standard Euler scheme is adopted to generate simulations from the following process:

$$\begin{aligned} dP(u) &= S(u)^{1/2} dW(u), \\ S(u) &= \sigma_d(u)\Sigma_{\mathrm{MK}}, \\ \sigma_d(u) &= C + Ae^{-au} + Be^{-b(1-u)}, \end{aligned} \tag{28}$$

where $\sigma_d(u)$ is a diurnal "U"-shaped function. The parameter values of $\sigma_d(u)$ are set to $A = 0.75$, $B = 0.25$, $C = 0.88929198$, and $a = b = 10$, following Andersen et al. (2012). We then generate five-minute returns ($M = 78$) based on 23,400 one-second returns simulated from the process in (28). A rolling window scheme with a window size of 2000 is adopted to estimate models, and one-step-ahead forecasts are generated accordingly. By testing on an Intel(R) Core(TM) i7-8665U CPU with the Windows 10 Enterprise operating system and MATLAB R2021b, using the OLS estimator and one core, the computation time for the 15-asset portfolio with a 2000 sample size is 3 seconds. The computation time for computing RbsC for 2000 days is 4 minutes, and the time for running MLTS is about 1.5 hours.

Table 1 provides a summary of the approaches to forecasting RCov. $F_{\mathrm{RCov;RCov}}$ denotes that both the estimation and forecasting are based on RCov. $F_{\mathrm{MCD;RCov}}$ denotes that the estimation is based on the MCD covariance matrices and the forecasts are based on RCov. Similarly, $F_{\mathrm{MCD;MCD}}$ indicates that both the estimation and forecasting use MCD covariance matrices. All of these approaches are based on estimation under OLS. F_{MLTS} relates to the estimation using the MLTS estimator with forecasts from Eq. (16). $F_{\mathrm{CMB;RCov}}$ denotes that the estimations are based on the combined method and the forecasts are based on RCov. $F_{\mathrm{CMB;MCD}}$ denotes that the estimations are based on the combined method and the forecasts are based on MCD covariance matrices.

Given the nature of the simulation design, the true values of the parameters of the MHAR model are unknown. Therefore, the accuracy of coefficient estimates in isolation cannot be assessed. Given the true integrated covariance matrix provided for simulation, and the primary focus of producing more accurate forecasts, we compare the forecasting performance of each candidate approach.

Table 2 reports the ratios of statistical and economic loss scores of forecasts generated from the proposed approaches to that of the existing OLS approach. The approaches of F_{MLTS}, $F_{\mathrm{CMB;RCov}}$, and $F_{\mathrm{CMB;MCD}}$ are selected by MCS based on L^F, and the forecasts of $F_{\mathrm{MCD;RCov}}$ and F_{MLTS} are included in the MCS based on the loss function L^Q. All of the proposed approaches produce lower portfolio variance than $F_{\mathrm{RCov;RCov}}$, with F_{MLTS} being the only member of the MCS. We can see that the approach based solely on MLTS is always included in the MCS based on the loss measures of L^F, L^Q, and V_p. In terms of the economic measures *TO*, *SP*, and *CO*, the combined approach $F_{\mathrm{CMB;RCov}}$ is preferred, where the corresponding portfolio produces the lowest portfolio turnover, the lowest extremeness, and the smallest short position. Overall, the simulation study demonstrates that using MCD and MLTS provides benefits to portfolios in terms of forecasting performance under both statistical and economic loss functions. The superiority of the proposed approaches is empirically explored in the next section.

In addition to the above simulation setup that is commonly adopted in the literature, another simulation exercise is provided, which allows the comparison of the true parameters of the data generating process (DGP) and parameter estimates from the OLS and MLTS estimation schemes. In this simulation, the OLS parameter estimates of the MHAR model based on 3022 daily realized covariance matrices of the 15 U.S. stocks in Table 22 starting from July 22, 2002 are taking as the true parameters for the simulation exercise. The first 22 realized covariance matrices, along with the 3000 daily residuals from the OLS estimation, are used to construct the simulated RCov series. The simulation process involves three steps. First, each simulation uses the first 22 realized covariance matrices as starting values for the MHAR model. Next, the MHAR OLS coefficients are used to generate a one-day-ahead conditional mean. Finally, a residual u_t from the OLS estimation of the MHAR model in Eq. (5) is randomly selected, and then combined with the conditional mean to generate a simulated one-day-ahead value of RCov. This process is repeated to generate a simulated series of RCov of length 3000.

We generate 500 simulated time series of RCov and employ the OLS and MLTS approaches for estimation. The parameter estimates are compared with true parameters by the following mean squared error (MSE) loss:

$$L^{MSE} = \frac{1}{N_S}\sum_{j=1}^{N_S}\sum_{i=1}^{N^*+3}(\theta^i - \hat{\theta}_j^i)^2, \quad (29)$$

where $N_S = 500$ is the number of simulated samples, $N^* + 3 = 123$ is the number of total parameters in the DGP of a 15-asset portfolio, θ^i is the true ith parameter, and $\hat{\theta}_j^i$ is the ith parameter estimate for simulated dataset j. The MSE for parameters estimates obtained by MLTS is $L^{MSE}_{\text{MLTS}} = 0.00004$, while the MSE obtained by the OLS approach is $L^{MSE}_{\text{OLS}} = 0.00133$, which indicates that the estimates obtained by the MLTS approach are closer to the true values.

5. Empirical analysis

5.1. Data

The empirical analysis is based on 26 large U.S. stocks. The names and ticker symbols for each of the stocks are provided in Table 19 in Appendix A. Portfolios of five different sizes are constructed, with 5, 10, 15, 20, and 26 stocks. The stocks included in each portfolio are provided in Tables 19 to 23. The high-frequency data used to construct the covariance matrices are five-minute returns over the period from July 22, 2002 to June 27, 2019, representing a total of 4193 daily observations. Since the MHAR model of Eq. (5) requires initial conditions of s_{t-1}, $s_{t-1|t-5}$, and $s_{t-1|t-22}$, the first response variable is set to be the realized covariance on day $t = 23$. Ultimately, the full sample size is 4171.

Three estimation and forecasting schemes are selected to compare the forecasting performance of the proposed approaches. Firstly, we adopt a fixed window scheme where we split the sample into an in-sample part for estimation only and an out-of-sample part for prediction and forecast evaluation, with an in-sample period of 2000. In addition, an expanding window approach is employed. With a starting in-sample size of 2000, the next 250 out-of-sample forecasts $F_{2001:2250}$ are produced based on the in-sample estimation. Then a further 250 forecasts $F_{2251:2500}$ are generated based on the first 2250 observations, and so on. A rolling window scheme with a fixed window size of 2000 is also used to estimate the MHAR model. The forecasts at day $t + 1$ are based on observations from $t + 1 - W$ to t, for $W = 2000$ and $W \leq t \leq 4170$.

5.2. Forecasting results

This section presents the out-of-sample forecasts for the five different dimensions from the multivariate HAR models. The forecast comparisons are based on the statistical and economic measures introduced in Section 3.4. Tables 3, 5 and 7 present the forecasting results based on the statistical loss functions of L^F and L^Q for the three different estimation and forecasting schemes, together with the corresponding MCS results. Forecasting results for the portfolio-based economic loss measures for the three estimation and forecasting schemes are provided in Tables 4, 6 and 8. The MCS results based on the portfolio variance V_p are also presented. The economic evaluations are based on the GMV portfolios where the portfolio weights only depend on the covariance matrix of returns.

Table 3
Out-of-sample forecast results: Ratios of statistical measures (fixed window scheme).

	95% MCS (Range & Squared)				
	5 Assets L^F	10 Assets L^F	15 Assets L^F	20 Assets L^F	26 Assets L^F
$F_{\text{RCov;RCov}}$	1.000	1.000	1.000	1.000	1.000
$F_{\text{MCD;RCov}}$	0.981	0.972	0.987	0.984	0.987
$F_{\text{MCD;MCD}}$	**0.917***	**0.919***	0.941	**0.932***	**0.933***
F_{MLTS}	0.936	0.939	0.944	0.951	0.956
$F_{\text{CMB;RCov}}$	**0.904***	**0.918***	**0.927***	**0.927***	**0.932***
$F_{\text{CMB;MCD}}$	**0.932**	0.974	0.968	**0.946***	**0.924***
	L^Q	L^Q	L^Q	L^Q	L^Q
$F_{\text{RCov;RCov}}$	**1.000***	**1.000***	**1.000***	1.000	1.000
$F_{\text{MCD;RCov}}$	**0.994***	**1.001***	1.021	1.016	1.019
$F_{\text{MCD;MCD}}$	1.336	1.401	1.381	1.272	1.163
F_{MLTS}	**0.999***	**1.007***	**0.995***	**0.990***	**0.992***
$F_{\text{CMB;RCov}}$	**0.998***	1.026	1.021	1.008	1.002
$F_{\text{CMB;MCD}}$	1.369	1.419	1.316	1.215	1.111

Notes: Bold indicates member(s) of the MCS with the range statistics at 5%. * indicates member(s) of the MCS with the squared statistics at 5%. Shaded rows indicate that the approach is favored by all five portfolios.

Here, we focus on a daily investment horizon, and the loss ratios of proposed approaches relative to the benchmark of $F_{\text{RCov;RCov}}$ are provided. In all the cases of the fixed window scheme and expanding window scheme, $F_{\text{CMB;RCov}}$ is always a member of the MCS based on the loss function L^F, and F_{MLTS} is always included in the MCS based on the loss function L^Q, as shown in Tables 3 and 5. In terms of economic loss, TO, SP, and CO, the approach solely based on MLTS exhibits the best performance for all instances, as shown in Tables 4 and 6. Based on the forecasts of F_{MLTS}, all portfolios produce the lowest portfolio turnover, which implies more stable covariance matrix forecasts, less extreme portfolio positions, and the smallest short positions. For the expanding window scheme, the portfolio variance from $F_{\text{CMB;RCov}}$ is significantly lower for the portfolios with a lower dimension, as indicated by the MCS results in Table 6, along with the resulting loss ratios always being lower than 1. The portfolio variance based on F_{MLTS} is contained in the MCS and provides the lowest variance in many cases. In particular, when the portfolio dimension is high, only F_{MLTS} is selected by MCS, as shown in the cases of the 20-asset and 26-asset portfolios in Table 6.

It is worth mentioning that although F_{MLTS} is not always selected by MCS based on L^F, the MLTS approach still outperforms the vanilla OLS approach, given that the ratios of L^F are always lower than 1. Such improvements are enhanced further by incorporating the MCD covariance matrices into the MLTS approach, which is shown by the fact that the forecast $F_{\text{CMB;RCov}}$ is always chosen by MCS, as indicated by the shaded rows in the upper panel of Tables 3 and 5.

For the rolling window scheme, both $F_{\text{MCD;MCD}}$ and $F_{\text{CMB;RCov}}$ are selected by MCS based on L^F in all cases,

Table 4
Out-of-sample forecast results: Ratios of economic measures (fixed window scheme).

	5 Assets				10 Assets				15 Assets			
	TO	*SP*	*CO*	V_p	*TO*	*SP*	*CO*	V_p	*TO*	*SP*	*CO*	V_p
$F_{RCov;RCov}$	1.000	1.000	1.000	**1.000***	1.000	1.000	1.000	**1.000***	1.000	1.000	1.000	1.000
$F_{MCD;RCov}$	0.946	1.162	1.018	**1.001***	1.125	1.155	1.018	1.002	0.973	1.146	1.027	1.005
$F_{MCD;MCD}$	0.926	1.737	1.043	1.006	1.159	1.508	1.089	1.013	0.969	1.624	1.110	1.025
F_{MLTS}	0.870	0.652	0.996	1.004	0.748	0.649	0.933	1.001	0.646	0.516	0.884	0.995
$F_{CMB;RCov}$	1.061	0.771	0.996	**1.001***	0.891	0.765	0.950	**0.999***	0.867	0.655	0.913	**0.993***
$F_{CMB;MCD}$	0.958	0.957	1.011	1.006	0.819	0.877	0.978	1.005	0.788	0.854	0.953	1.003
	20 Assets				26 Assets							
	TO	*SP*	*CO*	V_p	*TO*	*SP*	*CO*	V_p				
$F_{RCov;RCov}$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000				
$F_{MCD;RCov}$	1.018	1.089	1.022	1.004	1.074	1.102	1.029	1.006				
$F_{MCD;MCD}$	1.037	1.414	1.111	1.025	1.086	1.325	1.111	1.028				
F_{MLTS}	0.658	0.678	0.889	**0.993***	0.883	0.757	0.903	**0.996***				
$F_{CMB;RCov}$	0.922	0.800	0.924	0.994	1.043	0.894	0.945	0.998				
$F_{CMB;MCD}$	0.869	0.983	0.973	1.004	1.007	1.061	1.000	1.013				

Notes: Bold indicates member(s) of the MCS with the range statistics at 5%. * indicates member(s) of the MCS with the squared statistics at 5%. Shaded cells indicate the lowest value in each column.

Table 5
Out-of-sample forecast results: Ratios of statistical measures (expanding window scheme).

	95% MCS (Range & Squared)				
	5 Assets L^F	10 Assets L^F	15 Assets L^F	20 Assets L^F	26 Assets L^F
$F_{RCov;RCov}$	1.000	1.000	1.000	1.000	1.000
$F_{MCD;RCov}$	0.989	0.980	0.992	0.989	0.992
$F_{MCD;MCD}$	0.938	**0.940***	0.953	**0.944***	**0.942***
F_{MLTS}	0.933	**0.942***	0.949	0.952	0.950
$F_{CMB;RCov}$	**0.915***	**0.935***	**0.941***	**0.940***	**0.940***
$F_{CMB;MCD}$	0.970	1.011	0.995	0.978	**0.949***
	L^Q	L^Q	L^Q	L^Q	L^Q
$F_{RCov;RCov}$	**1.000***	**1.000***	**1.000***	1.000	1.000
$F_{MCD;RCov}$	**1.009***	1.014	1.030	1.024	1.026
$F_{MCD;MCD}$	1.465	1.518	1.461	1.336	1.210
F_{MLTS}	**1.007***	**1.005***	**0.999***	**0.989***	**0.984***
$F_{CMB;RCov}$	1.050	1.077	1.081	1.049	1.020
$F_{CMB;MCD}$	1.598	1.633	1.520	1.340	1.177

Notes: Bold indicates member(s) of the MCS with the range statistics at 5%. * indicates member(s) of the MCS with the squared statistics at 5%. Shaded rows indicate that the approach is favored by all five portfolios.

as shown in Table 7. As with the forecasting results under the fixed and expanding window schemes, the L^F ratios between F_{MLTS} and $F_{RCov;RCov}$ are lower than 1 in all instances under the rolling window scheme. From the lower panel of Table 7, we can see that none of the candidates are included in the MCS in every case, though MLTS is most frequently included. The forecasting performance of the MLTS approach improves relative to the competing methods as the portfolio size grows. For the portfolio with 26 assets, the standard OLS estimation approach is rejected by the MCS, but the forecast of F_{MLTS} is still a member of MCS. Furthermore, the approach of F_{MLTS} is preferred under the economics measures in all cases except the *TO* measure of the five-asset portfolio, as reported in Table 8. The portfolio variance based on the forecasts of F_{MLTS} is significantly low for all the portfolios, and as with the expanding scheme, F_{MLTS} is the only one included in MCS when the portfolio increases.

The results show that the MLTS robust estimation scheme and the robust covariance estimates can lead to significant improvements in forecasting performance. Overall, the pure MLTS estimator seems to be the best performer. The results show that F_{MLTS} leads to portfolios with the most appealing characteristics in terms of low turnover, lower concentration, smaller short positions, and low portfolio variance, especially when the number of assets in the portfolio is large. For the 26-asset portfolio, F_{MLTS} is the only member of MCS based on L^Q. Although F_{MLTS} is not always selected by MCS based on the loss function L^F, its loss ratios are always smaller than 1.

We argue that the combined approach $F_{CMB;RCov}$ provides the second-best performance. It is a member of the MCS based on L^F for all five portfolios, regardless of which estimation or forecasting scheme is used. It also provides reasonable improvements in some of the portfolio characteristics, such as lower short positions, less extreme weights, and lower portfolio variance compared with the original OLS approach. Overall, the empirical results are consistent with the simulation results, where the proposed approaches provide some benefits over the original OLS approach in terms of forecasting performance.

Furthermore, we investigate the statistical gains or losses based on the root mean square error (RMSE) loss adopted in Luo and Chen (2020) for realized variances and realized covariances of RCov, separately. The loss function of RMSE for a $t+1,\ldots,T$ out-of-sample forecast is computed by:

$$L^{RMSE} = \frac{1}{T-t}\sum_{n=t+1}^{T}\sqrt{\sum_{i}|e_{n_i}|^2}, \tag{30}$$

where e_{n_i} is the difference between the ith realized variance (or realized covariance) at time n, an element of RCov_n, and the corresponding forecast.

Table 9 reports the L^{RMSE} for realized variance and realized covariance. The gains in forecast performance for the full RCov discussed above can be attributed to improvements in both variances and covariances under most approaches. The only exception is the $F_{CMB;MCD}$, where

Table 6
Out-of-sample forecast results: Ratios of economic measures (expanding window scheme).

	5 Assets				10 Assets				15 Assets			
	TO	*SP*	*CO*	V_p	*TO*	*SP*	*CO*	V_p	*TO*	*SP*	*CO*	V_p
$F_{RCov;RCov}$	1.000	1.000	1.000	**1.000***	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$F_{MCD;RCov}$	0.944	1.148	1.017	1.001	1.111	1.134	1.019	1.002	0.990	1.144	1.028	1.005
$F_{MCD;MCD}$	0.938	1.758	1.044	1.005	1.165	1.492	1.088	1.012	1.008	1.617	1.113	1.025
F_{MLTS}	0.893	0.651	0.976	1.002	0.712	0.637	0.929	**0.998***	0.667	0.500	0.884	**0.994***
$F_{CMB;RCov}$	1.002	0.774	0.985	**1.000***	0.858	0.756	0.947	**0.997***	0.953	0.731	0.938	**0.995***
$F_{CMB;MCD}$	0.918	0.986	0.998	1.003	0.802	0.864	0.975	1.002	0.887	0.948	0.982	1.006
	20 Assets				26 Assets							
	TO	*SP*	*CO*	V_p	*TO*	*SP*	*CO*	V_p				
$F_{RCov;RCov}$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000				
$F_{MCD;RCov}$	1.019	1.092	1.024	1.004	1.069	1.102	1.030	1.006				
$F_{MCD;MCD}$	1.060	1.425	1.115	1.025	1.100	1.328	1.114	1.027				
F_{MLTS}	0.644	0.615	0.880	**0.991***	0.749	0.686	0.886	**0.991***				
$F_{CMB;RCov}$	0.880	0.759	0.927	0.994	0.958	0.819	0.934	0.995				
$F_{CMB;MCD}$	0.825	0.928	0.976	1.004	0.917	0.963	0.988	1.008				

Notes: Bold indicates member(s) of the MCS with the range statistics at 5%. * indicates member(s) of the MCS with the squared statistics at 5%. Shaded cells indicate the lowest value in each column.

Table 7
Out-of-sample forecast results: Ratios of statistical measures (rolling window scheme; window size: 2000).

	95% MCS (Range & Squared)				
	5 Assets L^F	10 Assets L^F	15 Assets L^F	20 Assets L^F	26 Assets L^F
$F_{RCov;RCov}$	1.000	1.000	1.000	1.000	1.000
$F_{MCD;RCov}$	0.990	0.981	0.993	0.989	0.992
$F_{MCD;MCD}$	**0.943***	**0.943***	**0.955***	**0.945***	**0.943***
F_{MLTS}	**0.929***	**0.942***	**0.951***	0.953	0.953
$F_{CMB;RCov}$	**0.925***	**0.945***	**0.946***	**0.944***	**0.942***
$F_{CMB;MCD}$	0.992	1.030	1.002	0.983	0.955
	L^Q	L^Q	L^Q	L^Q	L^Q
$F_{RCov;RCov}$	**1.000***	**1.000***	**1.000***	**1.000***	1.000
$F_{MCD;RCov}$	1.012	1.016	1.031	1.024	1.027
$F_{MCD;MCD}$	1.495	1.535	1.470	1.343	1.215
F_{MLTS}	1.023	**1.016***	**1.010***	**0.995***	**0.988***
$F_{CMB;RCov}$	1.126	1.147	1.127	1.081	1.042
$F_{CMB;MCD}$	1.807	1.838	1.651	1.429	1.236

Notes: Bold indicates member(s) of the MCS with the range statistics at 5%. * indicates member(s) of the MCS with the squared statistics at 5%. Shaded rows indicate that the approach is favored by all five portfolios.

the improvement for RCov is mainly from realized covariances. From the upper panel of Table 9, we can see that the combined approach $F_{CMB;RCov}$ provides the lowest RMSE for realized variances in most cases, whereas the lowest loss for realized covariances comes from $F_{MCD;MCD}$ for most portfolios, as shown in the lower panel of Table 9. Practitioners interested in specific elements of RCov can be guided by analyzing the decomposed effects of different estimation and forecast approaches.

5.3. Longer forecast horizons

In addition to the daily investment horizon, we consider longer weekly and monthly horizons where portfolios are constructed or re-balanced less frequently. For the longer-horizon forecasts, a direct forecasting approach is used, where we replace the one-day-ahead vectorized realized covariance matrix, the left-hand side of the MHAR model in Eq. (5), with an average over the desired horizon, as shown below:

$$\bar{s}_{t,t+h|t} = \theta_0 + \theta_1 s_{t-1} + \theta_2 s_{t-1|t-5} + \theta_3 s_{t-1|t-22} + u_t, \quad (31)$$

where $\bar{s}_{t,t+h|t} = \frac{1}{h}\sum_{i=0}^{h} s_{t+i}$ and $h = 4$ and 21 for the weekly and monthly horizons. We compare the performance of forecasts resulting from the vanilla OLS approach with proposed approaches, to investigate the robustness of the outperformance of these approaches for longer investment periods.

Table 10 reports the statistical loss ratios for both weekly and monthly forecasts. The forecasting evaluation results of economic measures reported in Tables 11 and 12 are based on GMV portfolios constructed with weekly $F_{t+4|t}$ and monthly $F_{t+21|t}$ covariance forecasts, respectively. The loss ratios based on statistical measures in Table 10 show strong evidence to support the superior forecasting performance of F_{MLTS} and $F_{CMB;RCov}$. In both the weekly and monthly cases, the combined approach of $F_{CMB;RCov}$ is always included in the MCS based on the loss function L^F, except for the monthly forecast of the 26-asset portfolio. The MLTS approach is contained in the MCS in all cases based on the L^Q loss function. Moreover, compared to the results obtained over the daily horizon, the improvements of forecasting accuracy based on $F_{CMB;RCov}$ and F_{MLTS} become larger over the long horizons, as indicated by the lower loss ratios. At the longer horizon, the robust approaches perform even better, as evidenced by the lower loss ratios in Table 10 for the monthly forecasts relative to the weekly ones.

The results for the economic loss functions *TO*, *SP*, and *CO*, reported in Tables 11 and 12, show that the MLTS approach outperforms the others in all cases, with the only exception being the monthly forecasts of the five-asset portfolio. As with the one-day-ahead case, the volatility of the GMV portfolio resulting from F_{MLTS} is significantly lower in many cases, and the MLTS approach is the only one included in the MCS when the portfolio's dimension is high. In addition, the combined approach $F_{CMB;RCov}$ provides improvements in the portfolio characteristics in many cases relative to the OLS.

Table 8
Out-of-sample forecast results: Ratios of economic measures (rolling window scheme; window size: 2000).

	5 Assets				10 Assets				15 Assets			
	TO	*SP*	*CO*	V_p	*TO*	*SP*	*CO*	V_p	*TO*	*SP*	*CO*	V_p
$F_{\text{RCov;RCov}}$	1.000	1.000	1.000	**1.000***	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$F_{\text{MCD;RCov}}$	0.939	1.142	1.016	1.001*	1.109	1.127	1.018	1.002	0.990	1.143	1.029	1.005
$F_{\text{MCD;MCD}}$	0.938	1.768	1.045	1.005	1.167	1.490	1.088	1.013	1.011	1.617	1.113	1.025
F_{MLTS}	0.970	0.690	0.968	**1.000***	0.751	0.665	0.929	**0.997***	0.742	0.556	0.913	**0.995***
$F_{\text{CMB;RCov}}$	1.143	0.827	0.981	**1.000***	0.912	0.786	0.952	**0.997***	1.042	0.820	0.964	0.998
$F_{\text{CMB;MCD}}$	1.058	1.068	0.989	**1.001***	0.860	0.897	0.980	1.001	0.980	1.063	1.014	1.011

	20 Assets				26 Assets			
	TO	*SP*	*CO*	V_p	*TO*	*SP*	*CO*	V_p
$F_{\text{RCov;RCov}}$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$F_{\text{MCD;RCov}}$	1.018	1.090	1.024	1.004	1.067	1.099	1.030	1.006
$F_{\text{MCD;MCD}}$	1.059	1.414	1.114	1.025	1.101	1.326	1.114	1.028
F_{MLTS}	0.744	0.657	0.912	**0.993***	0.842	0.713	0.916	**0.993***
$F_{\text{CMB;RCov}}$	0.957	0.824	0.954	0.996	1.001	0.851	0.956	0.997
$F_{\text{CMB;MCD}}$	0.910	0.993	1.009	1.007	0.970	0.991	1.015	1.012

Notes: Bold indicates member(s) of the MCS with the range statistics at 5%. * indicates member(s) of the MCS with the squared statistics at 5%. Shaded cells indicate the lowest value in each column.

Table 9
Out-of-sample forecast results: Decomposition of ratios of statistical measures for one-day-ahead RCov forecasts (rolling window scheme; window size: 2000).

	95% MCS (Loss Measure: RMSE)				
	5 Assets	10 Assets	15 Assets	20 Assets	26 Assets
	Realized Variances (diagonal elements of RCov)				
$F_{\text{RCov;RCov}}$	1.000	1.000	1.000	1.000	1.000
$F_{\text{MCD;RCov}}$	0.985	0.973	0.985	0.982	0.986
$F_{\text{MCD;MCD}}$	0.944	0.975	0.986	0.972	0.950
F_{MLTS}	0.919	**0.939**	0.948	0.951	0.954
$F_{\text{CMB;RCov}}$	**0.917**	0.950	**0.947**	**0.944**	**0.938**
$F_{\text{CMB;MCD}}$	1.030	1.113	1.089	1.063	1.008
	Realized Covariances (off-diagonal elements of RCov)				
$F_{\text{RCov;RCov}}$	1.000	1.000	1.000	1.000	1.000
$F_{\text{MCD;RCov}}$	0.995	0.983	0.994	0.990	0.993
$F_{\text{MCD;MCD}}$	0.933	**0.918**	**0.935**	**0.931**	**0.936**
F_{MLTS}	0.937	0.941	0.950	0.952	0.952
$F_{\text{CMB;RCov}}$	**0.931**	0.940	0.942	0.941	0.941
$F_{\text{CMB;MCD}}$	0.948	0.983	0.961	0.954	0.939

Notes: Bold indicates the lowest value in each column.

Overall, the results in this section highlight that the benefits of adopting the proposed approaches still apply in the case of longer horizons, especially for the MLTS estimation scheme and the combined method $F_{\text{CMB;RCov}}$ based on robust covariance matrices. In fact, the benefit may be somewhat larger under some of the loss functions at these longer horizons.

6. Robustness checks

We examine the robustness of the benefits of outlier-robust methods along a number of dimensions. First, the section considers the relative performance during different market conditions over time. Next, other benchmark HAR models are considered. Finally, the robustness of the benefits in relation to the effect of price jumps is considered.

6.1. Relative performance over time

Fig. 3 presents the difference in the cumulative loss scores between the $F_{\text{RCov;RCov}}$ and outlier-robust methods for the 26-asset portfolio. The top panel presents the difference between the $F_{\text{RCov;RCov}}$ and $F_{\text{CMB;RCov}}$, given that the combined approach provides the best forecast performance based on L^F. The second panel shows the difference between the $F_{\text{RCov;RCov}}$ and F_{MLTS} for L^Q. The final panel shows the realized variance of AXP, an asset within the 26-asset portfolio, treated as an indicator of the general market conditions.

The most obvious pattern is that the outlier-robust methods generate broadly consistent gains through the entire sample under both loss functions. Even during periods of consistently high volatility, such as late 2011, late 2015, early 2016, and early 2018, the outlier-robust methods on average provide superior performance, though sometimes associated with variability. Overall, the proposed approaches provide benefits under the vast majority of market conditions.

6.2. MHARQ model

In addition to the standard MHAR model, we apply the proposed outlier-robust approaches to the MHARQ model, which takes into account the measurement errors in RCov and can outperform the MHAR model in terms of forecasting. The MHARQ model of Bollerslev et al. (2018) takes the following form:

$$s_t = \theta_0 + \theta_{1,t} \circ s_{t-1} + \theta_2 s_{t-1|t-5} + \theta_3 s_{t-1|t-22} + u_t, \quad (32)$$
$$\theta_{1,t} = \theta_1 \iota + \theta_{1Q}\sqrt{\Pi_{t-1}},$$

where $\circ$ denotes the Hadamard product, θ_1 and θ_{1Q} are scalar parameters, ι is an $N(N+1)/2$ dimensional vector of ones, and Π_t is the estimate of measurement errors in the $s_t \equiv \text{vech}S_t$ vector, which can be estimated by using realized quarticity (Bollerslev et al., 2018).

In addition to the six approaches shown in Table 1, we use the raw RCov as the dependent variable in Eq. (8)

Table 10
Long-horizon forecasts: Ratios of statistical measures (expanding window scheme).

	Weekly Forecasts (95% MCS)					Monthly Forecasts (95% MCS)				
	5 Assets	10 Assets	15 Assets	20 Assets	26 Assets	5 Assets	10 Assets	15 Assets	20 Assets	26 Assets
	L^F	L^F	L^F	L^F	L^F	L^F	L^F	L^F	L^F	L^F
$F_{RCov;RCov}$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$F_{MCD;RCov}$	0.957	0.947	0.960	0.959	0.968	0.868	0.850	0.873	0.886	0.902
$F_{MCD;MCD}$	0.880	**0.872***	**0.885**	**0.875***	**0.885***	0.752	0.730	0.773	0.776	0.805
F_{MLTS}	0.888	**0.876***	**0.877***	0.881	0.887	0.758	0.735	0.765	0.771	0.803
$F_{CMB;RCov}$	**0.855***	**0.861***	**0.867***	**0.863***	**0.867***	**0.713***	**0.700***	**0.738***	**0.736***	0.750
$F_{Mix;MCD}$	0.924	0.960	0.941	0.908	**0.873***	0.772	0.771	0.790	**0.756***	**0.726***
	L^Q	L^Q	L^Q	L^Q	L^Q	L^Q	L^Q	L^Q	L^Q	L^Q
$F_{RCov;RCov}$	**1.000***	**1.000***	**1.000***	1.000	1.000	1.000*	1.000	1.000	1.000	1.000
$F_{MCD;RCov}$	**0.991***	**1.008***	1.060	1.048	1.049	**0.854***	**0.873***	**0.968***	0.969	**0.953***
$F_{MCD;MCD}$	1.646	1.807	1.867	1.639	1.402	1.203	1.450	1.730	1.541	1.238
F_{MLTS}	**0.993***	**0.986***	**0.981***	**0.973***	**0.966***	**0.846***	**0.861***	**0.912***	**0.905***	**0.907***
$F_{CMB;RCov}$	**1.044***	1.104	1.127	1.082	1.029	**0.932***	1.026	1.125	1.066	**0.970***
$F_{Mix;MCD}$	1.894	2.133	2.003	1.716	1.392	1.805	2.134	2.098	1.832	1.440

Notes: Bold indicates member(s) of the MCS with the range statistics at 5%. * indicates member(s) of the MCS with the squared statistics at 5%. Shaded cells indicate the lowest value in each column.

Table 11
Weekly forecasts: Ratios of economic measures (expanding window scheme).

	5 Assets				10 Assets				15 Assets			
	TO	*SP*	*CO*	V_p	*TO*	*SP*	*CO*	V_p	*TO*	*SP*	*CO*	V_p
$F_{RCov;RCov}$	1.000	1.000	1.000	**1.000***	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$F_{MCD;RCov}$	1.117	1.252	1.022	1.001*	1.140	1.184	1.019	1.002	1.322	1.226	1.036	1.006
$F_{MCD;MCD}$	1.081	1.819	1.046	1.007	1.141	1.552	1.090	1.013	1.326	1.773	1.124	1.027
F_{MLTS}	0.710	0.629	0.987	1.002*	0.697	0.661	0.935	**0.998***	0.623	0.497	0.890	**0.994***
$F_{CMB;RCov}$	0.882	0.841	1.001	**1.000***	0.830	0.833	0.958	**0.997***	0.896	0.757	0.944	**0.995***
$F_{Mix;MCD}$	0.791	1.020	1.019	1.005	0.762	0.974	0.991	1.002	0.814	1.036	0.993	1.005

	20 Assets				26 Assets			
	TO	*SP*	*CO*	V_p	*TO*	*SP*	*CO*	V_p
$F_{RCov;RCov}$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$F_{MCD;RCov}$	1.184	1.130	1.029	1.005	1.155	1.123	1.031	1.006
$F_{MCD;MCD}$	1.192	1.510	1.125	1.027	1.161	1.378	1.121	1.027
F_{MLTS}	0.635	0.630	0.887	**0.992***	0.689	0.700	0.891	**0.991***
$F_{CMB;RCov}$	0.867	0.797	0.935	0.994	0.869	0.837	0.935	0.994
$F_{Mix;MCD}$	0.801	1.013	0.991	1.005	0.819	1.018	0.997	1.006

Notes: Bold indicates member(s) of the MCS with the range statistics at 5%. * indicates member(s) of the MCS with the squared statistics at 5%. Shaded cells indicate the lowest value in each column.

Table 12
Monthly forecasts: Ratios of economic measures (expanding window scheme).

	5 Assets				10 Assets				15 Assets			
	TO	*SP*	*CO*	V_p	*TO*	*SP*	*CO*	V_p	*TO*	*SP*	*CO*	V_p
$F_{RCov;RCov}$	1.000	1.000	1.000	**1.000**	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$F_{MCD;RCov}$	1.139	1.381	1.021	**0.999***	1.138	1.331	1.019	1.001	1.284	1.234	1.029	1.003
$F_{MCD;MCD}$	1.023	1.681	1.044	1.009	1.042	1.708	1.093	1.014	1.170	1.812	1.120	1.024
F_{MLTS}	0.659	1.026	1.018	**0.995***	0.699	0.756	0.934	**0.994***	0.690	0.489	0.895	**0.992***
$F_{CMB;RCov}$	0.822	1.169	1.030	**0.995***	0.875	0.966	0.957	**0.994***	1.007	0.763	0.947	0.994
$F_{Mix;MCD}$	0.711	1.287	1.051	1.002	0.764	1.125	0.990	1.001	0.843	1.056	0.996	1.004

	20 Assets				26 Assets			
	TO	*SP*	*CO*	V_p	*TO*	*SP*	*CO*	V_p
$F_{RCov;RCov}$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$F_{MCD;RCov}$	1.169	1.141	1.028	1.004	1.131	1.139	1.029	1.004
$F_{MCD;MCD}$	1.069	1.545	1.124	1.028	1.041	1.412	1.119	1.027
F_{MLTS}	0.735	0.644	0.895	**0.991***	0.893	0.784	0.916	**0.991***
$F_{CMB;RCov}$	1.064	0.845	0.951	0.995	1.128	0.939	0.963	0.995
$F_{Mix;MCD}$	0.922	1.094	1.012	1.008	1.025	1.165	1.033	1.012

Notes: Bold indicates member(s) of the MCS with the range statistics at 5%. * indicates member(s) of the MCS with the squared statistics at 5%. Shaded cells indicate the lowest value in each column.

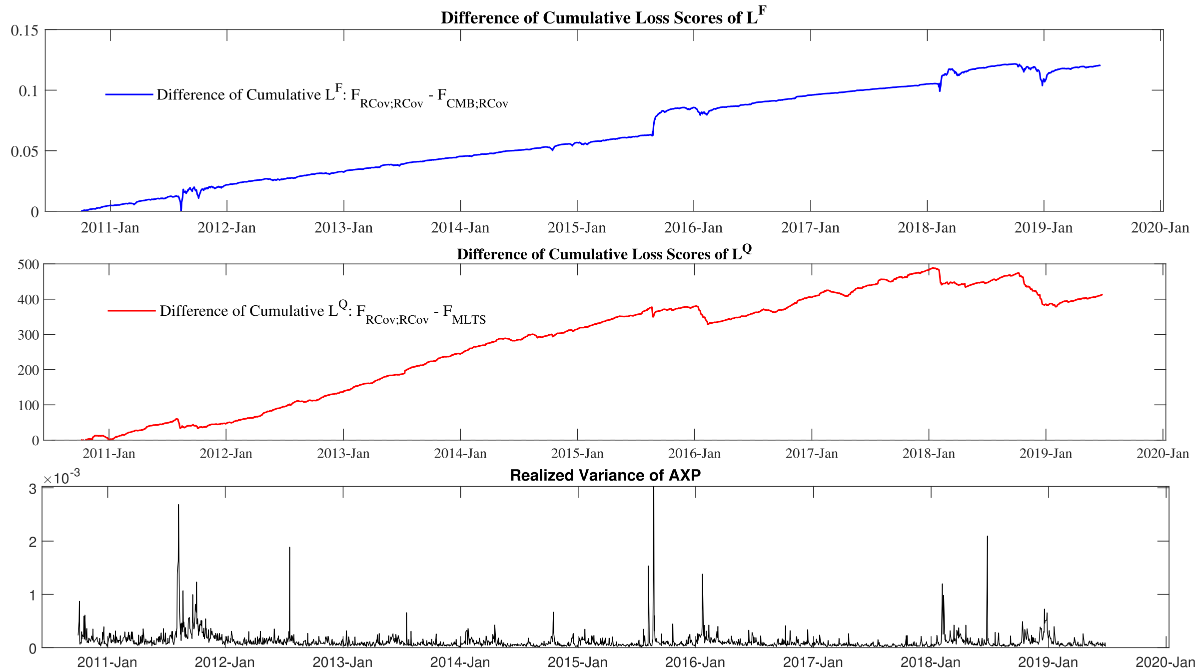

Fig. 3. Twenty-six-asset portfolio: The difference of cumulative L^F between $F_{\mathrm{RCov;RCov}}$ and $F_{\mathrm{CMB;RCov}}$.

Table 13
Out-of-sample forecast results: Ratios of statistical measures (MHARQ model; rolling window scheme).

	95% MCS (Range & Squared)				
	5 Assets L^F	10 Assets L^F	15 Assets L^F	20 Assets L^F	26 Assets L^F
$F_{\mathrm{RCov;RCov}}$	1.000	1.000	1.000	1.000	1.000
$F_{\mathrm{MCD;RCov}}$	0.997	0.986	1.002	0.992	0.993
$F_{\mathrm{MCD;MCD}}$	**0.951***	**0.947***	0.955	**0.946***	**0.943***
F_{MLTS}	**0.942***	**0.944***	0.951	0.957	0.974
$F_{\mathrm{CMB;RCov}}$	**0.937***	**0.948***	**0.943***	**0.947***	**0.954***
$F_{\mathrm{CMB;MCD}}$	1.004	1.031	0.997	0.982	**0.942***
$F_{\mathrm{MCD-RCov;RCov}}$	1.339	1.332	1.288	1.22	1.149
	L^Q	L^Q	L^Q	L^Q	L^Q
$F_{\mathrm{RCov;RCov}}$	**1.000***	**1.000***	**1.000***	1.000	1.000
$F_{\mathrm{MCD;RCov}}$	1.016*	1.029	1.056	1.040	1.029
$F_{\mathrm{MCD;MCD}}$	1.563	1.610	1.551	1.397	1.247
F_{MLTS}	**1.009***	**1.003***	**0.991***	**0.979***	**0.976***
$F_{\mathrm{CMB;RCov}}$	1.118	1.127	1.094	1.056	1.009
$F_{\mathrm{CMB;MCD}}$	1.792	1.796	1.591	1.392	1.185
$F_{\mathrm{MCD-RCov;RCov}}$	1.355	1.310	1.198	1.144	1.083

Notes: Bold indicates member(s) of the MCS with the range statistics at 5%. * indicates member(s) of the MCS with the squared statistics at 5%. Shaded rows indicate that the approach is favored by all five portfolios.

for estimation and Eq. (9) for forecasting. We denote the forecasts from this approach as $F_{MCD-RCov;RCov}$. By doing this, we reveal why the approach of MCD only works well with RbsC when the estimation scheme is OLS. The results of loss ratios relative to the MHARQ model presented in Tables 13 and 14 are consistent with the ones of the MHAR model in Tables 7 and 8. The combined approach $F_{\mathrm{CMB;RCov}}$ is always included in the MCS based on L^F, as shown in the upper panel of Table 13. From the lower panel of Table 13, we can see that the MLTS approach F_{MLTS} is selected by the MCS in every case, and it is the only member of MCS for portfolios with higher dimensions. As reported in Table 14, the portfolio variance based on forecasts of F_{MLTS} is significantly lower for portfolios with more than 15 assets. Furthermore, the approach of F_{MLTS} is preferred by economic measures of *TO*, *SP*, and *CO* in all cases. Overall, the superiority of the proposed outlier-robust approaches is still observed in the context of the MHARQ model. The forecasting performance of $F_{MCD-RCov;RCov}$ is worse than the vanilla OLS approach in all instances in terms of higher loss ratios, as shown in Table 13. With the raw RCov that contains outliers as the dependent variable, it is difficult to get the best out of the OLS estimator even if the independent variables are the robust covariance estimate RbsC.

6.3. Jump-robust approaches

6.3.1. Jump-robust covariance estimate

We further investigate the robustness of the proposed approaches by taking the effect of price jumps into consideration. With the question of whether the benefits of the proposed approaches are still retained after containing the effect of jumps, we adopt a jump-robust estimator of the integrated covariance matrix, the realized bipower covariation (RBPCov) proposed by Barndorff-Nielsen and Shephard (2004a), to estimate the MHAR model. In this case, we limit the application to the pure OLS and MLTS estimation schemes, as the MCD approach cannot be applied to RBPCov since the MCD covariance is estimated based on intraday returns, but the RBPCov is already a

Table 14
Out-of-sample forecast results: Ratios of economic measures (MHARQ model; rolling window scheme).

	5 Assets				10 Assets				15 Assets			
	V_p	TO	SP	CO	V_p	TO	SP	CO	V_p	TO	SP	CO
$F_{\mathrm{RCov;RCov}}$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$F_{\mathrm{MCD;RCov}}$	0.999	0.876	1.066	1.006	1.006	1.067	1.133	1.021	1.012	0.985	1.162	1.032
$F_{\mathrm{MCD;MCD}}$	1.012	0.922	1.742	1.046	1.022	1.143	1.509	1.095	1.036	1.041	1.621	1.123
F_{MLTS}	0.991	0.719	0.624	0.948	0.989	0.697	0.681	0.930	**0.982***	0.650	0.551	0.900
$F_{\mathrm{CMB;RCov}}$	**0.989***	0.785	0.644	0.956	**0.986***	0.751	0.731	0.942	0.986	0.824	0.745	0.940
$F_{\mathrm{CMB;MCD}}$	0.997	0.779	0.974	0.970	0.991	0.710	0.840	0.972	1.000	0.781	0.969	0.991

	20 Assets				26 Assets			
	V_p	TO	SP	CO	V_p	TO	SP	CO
$F_{\mathrm{RCov;RCov}}$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$F_{\mathrm{MCD;RCov}}$	1.011	1.017	1.121	1.031	1.008	0.987	1.092	1.026
$F_{\mathrm{MCD;MCD}}$	1.037	1.077	1.451	1.128	1.033	1.044	1.320	1.116
F_{MLTS}	**0.980***	0.690	0.663	0.904	**0.984***	0.882	0.804	0.941
$F_{\mathrm{CMB;RCov}}$	0.985	0.808	0.797	0.938	0.992	0.944	0.930	0.978
$F_{\mathrm{CMB;MCD}}$	0.999	0.772	0.975	0.996	1.014	0.957	1.098	1.049

Notes: Bold indicates member(s) of the MCS with the range statistics at 5%. * indicates member(s) of the MCS with the squared statistics at 5%. Shaded cells indicate the lowest value in each column.

Table 15
Out-of-sample forecast results: Ratios of statistical measures (RBPCov; rolling window scheme).

	5 Assets	10 Assets	15 Assets	20 Assets	26 Assets
	L^F	L^F	L^F	L^F	L^F
$F_{\mathrm{RBPCov;RBPCov}}$	1.000	1.000	1.000	1.000	1.000
$F_{\mathrm{MLTS;RBPCov}}$	0.962	0.971	0.978	0.979	**0.968**
$F_{\mathrm{RBPCov;RCov}}$	1.027	1.022	1.019	1.020	1.018
$F_{\mathrm{MLTS;RCov}}$	**0.960**	**0.966**	**0.973**	**0.977**	**0.968**
	L^Q	L^Q	L^Q	L^Q	L^Q
$F_{\mathrm{RBPCov;RBPCov}}$	1.000	1.000	1.000	1.000	1.000
$F_{\mathrm{MLTS;RBPCov}}$	1.077	1.022	1.014	0.979	0.961
$F_{\mathrm{RBPCov;RCov}}$	**0.984**	**0.956**	**0.961**	0.948	0.935
$F_{\mathrm{MLTS;RCov}}$	1.021	0.980	0.977	**0.946**	**0.923**

Notes: Bold indicates the lowest value in each column.

matrix. The k, q-th element of RBPCov is equal to

$$\mathrm{RBPCov}[k,q]_t = \frac{\pi}{8}\left(\sum_{i=2}^{M} |r_{t,i}^k + r_{t,i}^q|\,|r_{t,i-1}^k + r_{t,i-1}^q| - |r_{t,i}^k - r_{t,i}^q|\,|r_{t,i-1}^k - r_{t,i-1}^q|\right), \tag{33}$$

where M is the number of intraday returns per trading day and $r_{t,i}^k$ is the ith period intraday return of asset k.

Tables 15 and 16 provide the loss ratios of statistical and economic measures, respectively. The notation for forecasts based on RBPCov follows that for the approaches in Table 1. $F_{\mathrm{RBPCov;RBPCov}}$ denotes that both the estimation and forecasting are based on RBPCov. $F_{\mathrm{RBPCov;RCov}}$ denotes that the estimation is based on the jump-robust covariance matrix RBPCov and the forecasts are based on RCov. Both approaches are based on estimation under OLS. $F_{\mathrm{MLTS;RBPCov}}$ and $F_{\mathrm{MLTS;RCov}}$ relate to the estimation using the MLTS estimator based on RBPCov, with forecasts based on RBPCov and RCov, respectively. The benchmark is the approach of $F_{\mathrm{RBPCov;RBPCov}}$. For the loss function L^F, the approaches based on MLTS always outperform the OLS approaches, as shown in the upper panel of Table 15. In terms of the loss function L^Q, $F_{\mathrm{MLTS;RCov}}$ produces lower loss scores for portfolios with 20 and 26 assets, while $F_{\mathrm{RBPCov;RCov}}$ results in lower loss scores for the portfolio with lower dimensions. In Table 16, the portfolio variance V_p yielding from $F_{\mathrm{MLTS;RCov}}$ is the lowest for all the portfolios. Moreover, in terms of the economic measures TO, SP, and CO, the approach of $F_{\mathrm{MLTS;RCov}}$ provides the lowest loss measures in all instances. Overall, we argue that the proposed MLTS approach still provides extra benefits beyond adopting a jump-robust covariance estimate.

6.3.2. Truncated intraday returns

In addition to directly applying MCD and MLTS to the raw RCov, we consider a realized covariance estimate, using truncated intraday returns with the effect of jumps removed. We employ the time-of-day (TOD) estimator of Bollerslev et al. (2013) to separate continuous price movements from jumps. To compute the threshold for separation, we first compute the realized variance (RV) and bipower variation (BV) by

$$RV_t^j = \sum_{i=1}^{M} |r_{t,i}^j|^2, \tag{34}$$

$$BV_t^j = \frac{\pi}{2}\left(\frac{M}{M-1}\right)\sum_{i=2}^{M} |r_{t,i}^j||r_{t,i-1}^j|, \tag{35}$$

where $r_{t,i}^j$ is the ith period intraday return of asset j. BV, proposed by Barndorff-Nielsen and Shephard (2004b), is a jump-robust estimator of the integrated variance. Based on RV and BV, we then estimate the TOD pattern for each asset:

$$TOD_i^j = \frac{M\sum_{t=1}^{T}|r_{t,i}^j|^2 \mathbb{1}(|r_{t,i}^j| \le \tau M^{-\omega}\sqrt{BV_t^j \wedge RV_t^j})}{\sum_{t=1}^{T}\sum_{i=1}^{M}|r_{t,i}^j|^2 \mathbb{1}(|r_{t,i}^j| \le \tau M^{-\omega}\sqrt{BV_t^j \wedge RV_t^j})}, \tag{36}$$

where $\mathbb{1}$ denotes an indicator function. The constants τ and ω are set to $\tau = 3$, and $\omega = 0.49$ by following Pelger (2019). The truncation level $\alpha_{t,i}^j$ for separating the realized

Table 16
Out-of-sample forecast results: Ratios of economic measures (RBPCov model; rolling window scheme).

	5 Assets				10 Assets				15 Assets			
	V_p	*TO*	*SP*	*CO*	V_p	*TO*	*SP*	*CO*	V_p	*TO*	*SP*	*CO*
$F_{\text{RBPCov;RBPCov}}$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$F_{\text{MLTS;RBPCov}}$	0.993	0.914	0.652	0.957	0.938	0.500	0.434	0.852	0.972	0.617	0.478	0.867
$F_{\text{RBPCov;RCov}}$	0.996	0.938	0.805	0.989	0.941	0.591	0.572	0.903	0.979	0.789	0.747	0.944
$F_{\text{MLTS;RCov}}$	**0.990**	**0.870**	**0.522**	**0.952**	**0.931**	**0.436**	**0.367**	**0.836**	**0.966**	**0.552**	**0.397**	**0.848**

	20 Assets				26 Assets			
	V_p	*TO*	*SP*	*CO*	V_p	*TO*	*SP*	*CO*
$F_{\text{RBPCov;RBPCov}}$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$F_{\text{MLTS;RBPCov}}$	0.968	0.641	0.593	0.868	0.943	0.678	0.616	0.849
$F_{\text{RBPCov;RCov}}$	0.973	0.770	0.774	0.934	0.942	0.654	0.716	0.894
$F_{\text{MLTS;RCov}}$	**0.958**	**0.533**	**0.488**	**0.837**	**0.923**	**0.527**	**0.495**	**0.804**

Notes: Bold indicates the lowest value in each column.

Table 17
Summary of forecasting approaches based on truncated returns.

	Variables		Estimator
	Estimation	Forecast	
$F_{\text{RJCov;RJCov}}$	RJCov	RJCov	OLS
$F_{\text{MCD-Tr;RJCov}}$	RbsTrC	RJCov	OLS
$F_{\text{MCD-Tr;MCD-Tr}}$	RbsTrC	RbsTrC	OLS
$F_{\text{MLTS-RJCov}}$	RJCov	RJCov	MLTS
$F_{\text{CMB-Tr;RJCov}}$	RbsTrC	RJCov	MLTS
$F_{\text{CMB-Tr;MCD-Tr}}$	RbsTrC	RbsTrC	MLTS

Note: RbsTrC represents the RbsC matrix generated based on truncated intraday returns.

jumps from the continuous price moves is given by

$$\alpha^j_{t,i} = \tau M^{-\omega}\sqrt{(BV^j_t \wedge RV^j_t) * TOD^j_i}. \quad (37)$$

Then the high-frequency returns caused by continuous movements are identified by the following condition:

$$|r^j_{t,i}| \leq \alpha^j_{t,i}. \quad (38)$$

We denote the realized covariance matrix constructed by using the intraday returns meeting the condition in Eq. (38) as RJCov. The procedure of utilizing the proposed approaches to estimate the MHAR model and forecast RCov based on the truncated intraday returns is the same as the case of using raw intraday returns.

Table 17 explains the notation for the forecasting approaches based on truncated high-frequency returns.

Table 18 provides ratios of the statistical loss functions relative to the benchmark $F_{\text{RJCov;RJCov}}$. The results show that while the outlier-robust methods do not provide extra gains for L^Q, they do offer gains of 1%–2% in performance relative to the benchmark for L^F. When the underlying intraday returns are available to construct realized covariance matrices, the outlier-robust estimates continue offering small benefits over truncating intraday returns. However, if one is working with publicly available realized covariance matrix data (without access to the underlying intraday data), the MLTS would be the only avenue for improving the forecast performance.

7. Conclusion

In this paper, we proposed two outlier-robust approaches for improving parameter estimation within the MHAR model, so that more reliable forecasts of RCov may be generated. We first replaced the raw RCov with the outlier-robust MCD covariance matrix, with OLS applied to estimating the parameters of the MHAR model. We demonstrated that the elements of the proposed robust covariance matrix contain smaller spikes when compared to raw RCov, making the MCD matrix more amenable to using OLS. We also employed the MLTS approach, as a robust estimation scheme for the MHAR model, where the covariance matrices of the OLS residuals are trimmed. Furthermore, we considered an approach that combines both MLTS estimation and the MCD covariance estimates.

The combined approach outperformed both MCD and MLTS according to the Frobenius distance measure. The pure MLTS approach was preferred under the QLIKE loss function as the portfolio dimension increased. Most importantly, in the context of portfolio allocation, the MLTS approach provided improved portfolio outcomes based on economic loss measures.

The approaches proposed here provide improvements in the forecasting performance relative to both the MHAR and MHARQ models. In addition, there are benefits relative to controlling the effect of jump activity on covariance matrices. The results point to some similarities in the effects of using MCD, MLTS, and truncating intraday returns when estimating covariance matrices. Gaining a deeper understanding of the links between these approaches may represent an interesting avenue for future research. Given the benefits of the outlier-robust covariance matrices identified here, extending these approaches to higher-dimensional portfolios, where the number of stocks is greater than the number of intraday observations, with the use of regularized methods, is also an interesting area for future work.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank the reviewers, the editor, and the associate editor for helpful comments to

Table 18
Out-of-sample forecast results: Ratios of statistical measures (truncated high-frequency returns; rolling window scheme).

	5 Assets	10 Assets	15 Assets	20 Assets	26 Assets
	L^F	L^F	L^F	L^F	L^F
$F_{RJCov;RJCov}$	1.000	1.000	1.000	1.000	1.000
$F_{MCD-Tr;RJCov}$	1.001	0.995	1.003	0.997	0.999
$F_{MCD-Tr;MCD-Tr}$	1.004	0.997	0.997	0.984	0.980
$F_{MLTS-RJCov}$	0.991	1.003	0.999	0.998	0.992
$F_{CMB-Tr;RJCov}$	1.005	1.021	1.007	1.003	**0.991**
$F_{CMB-Tr;MCD-Tr}$	1.066	1.097	1.057	1.034	1.005
	L^Q	L^Q	L^Q	L^Q	L^Q
$F_{RJCov;RJCov}$	1.000	1.000	1.000	1.000	1.000
$F_{MCD-Tr;RJCov}$	1.056	1.059	1.082	1.062	1.050
$F_{MCD-Tr;MCD-Tr}$	1.580	1.635	1.546	1.381	1.255
$F_{MLTS-RJCov}$	1.085	1.052	1.036	1.009	0.990
$F_{CMB-Tr;RJCov}$	1.229	1.220	1.172	1.111	1.056
$F_{CMB-Tr;MCD-Tr}$	1.839	1.855	1.640	1.412	1.237

Notes: Shaded cells indicate the lowest value in each column.

Table 19
Stock information.

Ticker	Company name
AAPL	Apple Inc.
AXP	American Express
BA	Boeing Co.
CAT	Caterpillar Inc.
CSCO	Cisco Systems Inc.
CVX	Chevron Corp.
DIS	Walt Disney Co.
GS	Goldman Sachs Group Inc.
HD	Home Depot Inc.
IBM	IBM
INTC	Intel Corp.
JNJ	Johnson & Johnson
JPM	JP Morgan Chase & Co.
KO	Coca-Cola Co.
MCD	McDonald's Corp.
MMM	3M Co.
MRK	Merck & Co. Inc.
MSFT	Microsoft Corporation
NKE	Nike Inc.
PFE	Pfizer Inc.
PG	Procter & Gamble Co.
UNH	UnitedHealth Group Inc.
UTX	United Technologies Corp.
VZ	Verizon Communications Inc.
WMT	Walmart Inc.
XOM	Exxon Mobil Corporation

improve the paper. The first author was supported by a scholarship from the QUT Centre for Data Science and a supervisor top-up scholarship from the School of Mathematical Sciences at QUT. Computational resources and services used in this work were provided by the HPC and Research Support Group, Queensland University of Technology, Brisbane, Australia.

Appendix A

See Table 19.

Appendix B. Investigated portfolios

See Tables 20–23.

Table 20
Five-asset portfolio.

Ticker	Company name
AAPL	Apple Inc.
AXP	American Express
BA	Boeing Co.
CAT	Caterpillar Inc.
CSCO	Cisco Systems Inc.

Table 21
Ten-asset portfolio.

Ticker	Company name
AAPL	Apple Inc.
AXP	American Express
BA	Boeing Co.
CAT	Caterpillar Inc.
CSCO	Cisco Systems Inc.
CVX	Chevron Corp.
DIS	Walt Disney Co.
GS	Goldman Sachs Group Inc.
HD	Home Depot Inc.
IBM	IBM

Table 22
Fifteen-asset portfolio.

Ticker	Company name
JNJ	Johnson & Johnson
JPM	JP Morgan Chase & Co.
KO	Coca-Cola Co.
MCD	McDonald's Corp.
MMM	3M Co.
MRK	Merck & Co. Inc.
MSFT	Microsoft Corporation
NKE	Nike Inc.
PFE	Pfizer Inc.
PG	Procter & Gamble Co.
UNH	UnitedHealth Group Inc.
UTX	United Technologies Corp.
VZ	Verizon Communications Inc.
WMT	Walmart Inc.
XOM	Exxon Mobil Corporation

Table 23
Twenty-asset portfolio.

Ticker	Company name
DIS	Walt Disney Co.
GS	Goldman Sachs Group Inc.
HD	Home Depot Inc.
IBM	IBM
INTC	Intel Corp.
JNJ	Johnson & Johnson
JPM	JP Morgan Chase & Co.
KO	Coca-Cola Co.
MCD	McDonald's Corp.
MMM	3M Co.
MRK	Merck & Co. Inc.
MSFT	Microsoft Corporation
NKE	Nike Inc.
PFE	Pfizer Inc.
PG	Procter & Gamble Co.
UNH	UnitedHealth Group Inc.
UTX	United Technologies Corp.
VZ	Verizon Communications Inc.
WMT	Walmart Inc.
XOM	Exxon Mobil Corporation

References

Agulló, J., Croux, C., & Van Aelst, S. (2008). The multivariate least-trimmed squares estimator. *Journal of Multivariate Analysis*, *99*(3), 311–338.

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

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

Andersen, T. G., Dobrev, D., & Schaumburg, E. (2012). Jump-robust volatility estimation using nearest neighbor truncation. *Journal of Econometrics*, *169*(1), 75–93.

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

Barndorff-Nielsen, O. E., & Shephard, N. (2004a). *Measuring the impact of jumps in multivariate price processes using bipower covariation*: *Technical report*, Discussion Paper, Nuffield College, Oxford University.

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

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

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

Bollerslev, T., Li, J., Patton, A. J., & Quaedvlieg, R. (2020). Realized semicovariances. *Econometrica*, *88*(4), 1515–1551.

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

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

Bollerslev, T., Todorov, V., & Li, S. Z. (2013). Jump tails, extreme dependencies, and the distribution of stock returns. *Journal of Econometrics*, *172*(2), 307–324.

Bubák, V., Kočenda, E., & Žikeš, F. (2011). Volatility transmission in emerging European foreign exchange markets. *Journal of Banking & Finance*, *35*(11), 2829–2841.

Busch, T., Christensen, B. J., & Nielsen, M. Ø. (2011). The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets. *Journal of Econometrics*, *160*(1), 48–57.

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

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

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

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

Ding, Z., & Granger, C. (1996). Modellling volatility persistence of speculative returns: A new approach. *Journal of Econometrics*, *73*, 185–215.

Ding, Z., Granger, C., & Engle, R. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, *1*, 83–106.

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

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

Harvey, A., Ruiz, E., & Shephard, N. (1994). Multivariate stochastic variance models. *The Review of Economic Studies*, *61*(2), 247–264.

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

Opschoor, A., & Lucas, A. (2022). Time-varying variance and skewness in realized volatility measures. *International Journal of Forecasting*.

Pelger, M. (2019). Large-dimensional factor modeling based on high-frequency observations. *Journal of Econometrics*, *208*(1), 23–42.

Proietti, T., & Lütkepohl, H. (2013). Does the Box–Cox transformation help in forecasting macroeconomic time series? *International Journal of Forecasting*, *29*(1), 88–99.

Quaedvlieg, R. (2020). Pre-shrinkage: Improved volatility forecasting using biased time-series. Available at SSRN 3716425.

Rousseeuw, P. J. (1985). Multivariate estimation with high breakdown point. *Mathematical Statistics and Applications*, *8*(283–297), 37.

Rousseeuw, P. J., & Driessen, K. V. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, *41*(3), 212–223.

Rousseeuw, P. J., Van Aelst, S., Van Driessen, K., & Gulló, J. A. (2004). Robust multivariate regression. *Technometrics*, *46*(3), 293–305.

Souček, M., & Todorova, N. (2013). Realized volatility transmission between crude oil and equity futures markets: A multivariate HAR approach. *Energy Economics*, *40*, 586–597.

Taylor, N. (2017a). Realised variance forecasting under Box-Cox transformations. *International Journal of Forecasting*, *33*(4), 770–785.

Taylor, N. (2017b). Realised variance forecasting under Box-Cox transformations. *International Journal of Forecasting*, *33*, 770–785.