MODELING REALIZED COVARIANCE MATRICES: A CLASS OF HADAMARD EXPONENTIAL MODELS

Luc Bauwens, Edoardo Otranto

REPRINT | 3202

CORE

Voie du Roman Pays 34, L1.03.01

B-1348 Louvain-la-Neuve

Tel (32 10) 47 43 04

Email:lidam-library@uclouvain.be

https://uclouvain.be/en/research-institutes/lidam/core/core-reprints.html

Journal of Financial Econometrics, 2022, 1–26
https://doi.org/10.1093/jjfinec/nbac007
Article

OXFORD

Modeling Realized Covariance Matrices: A Class of Hadamard Exponential Models

Luc Bauwens [1] **and Edoardo Otranto**[2]

[1]CORE, Université catholique de Louvain, Belgium and [2]Dipartimento di Economia, Universitá di Messina, Italy

Address correspondence to Luc Bauwens, CORE, Université catholique de Louvain, Voie du Roman Pays 34, B-1348 Louvain-La-Neuve, Belgium, or e-mail: luc.bauwens@uclouvain.be.

Received November 26, 2020; revised February 19, 2022; editorial decision March 2, 2022; accepted March 2, 2022

Abstract

Time series of realized covariance matrices can be modeled in the conditional autoregressive Wishart model family via dynamic correlations or via dynamic covariances. Extended parameterizations of these models are proposed, which imply a specific and time-varying impact parameter of the lagged realized covariance (or correlation) on the next conditional covariance (or correlation) of each asset pair. The proposed extensions guarantee the positive definiteness of the conditional covariance or correlation matrix with simple parametric restrictions, while keeping the number of parameters fixed or linear with respect to the number of assets. Two empirical studies reveal that the extended models have superior forecasting performances than their simpler versions and benchmark models.

Key words: dynamic covariances and correlations, Hadamard exponential matrix, realized covariances

JEL classification: C32, C58

The dynamic modeling of return covariance matrices is the topic of a large number of contributions in financial econometrics. The literature started by extending the univariate generalized autoregressive conditional heteroskedasticity (GARCH) model of Bollerslev (1986) to the multivariate case, developing progressively the family of multivariate generalized conditional heteroskedasticity (MGARCH) models; for a review, see for example, Bauwens, Laurent, and Rombouts (2006). Due to the availability of intraday prices and the development of realized volatility measures, attention shifted to the dynamic modeling of realized covariances and correlations. This has resulted in new models for positive definite matrices, among which we focus our attention on the conditional autoregressive Wishart

(CAW) models proposed by Golosnoy, Gribisch, and Liesenfeld (2012), and the multivariate heterogeneous autoregressive (HAR)-type models[1] extending the univariate HAR model of Corsi (2009). Both MGARCH models and models for realized measures are popular tools for risk management and portfolio allocation. MGARCH models have the advantage of being applicable in much larger dimensions; models for realized measures are based on more precise measures of covariation.

MGARCH and CAW models require to specify a dynamic process for a conditional covariance matrix (i.e., the conditional expectation of a covariance matrix). In that respect, they use the same kind of Baba-Engle-Kent-Kroner (BEKK) (for covariances) and dynamic conditional correlations (DCC) (for correlations) formulations of conditional processes: BEKK in MGARCH was introduced by Engle and Kroner (1995) and adapted to CAW by Golosnoy, Gribisch, and Liesenfeld (2012); DCC was introduced in MGARCH by Engle (2002) and extended to CAW by Bauwens, Storti, and Violante (2012) and Bauwens, Braione, and Storti (2016).

This research provides empirical evaluations of the merits of modeling realized covariance matrices through correlations and variances or through covariances and variances. These evaluations use new parameterizations of the CAW model family, which extend the existing BEKK-type formulation of Golosnoy, Gribisch, and Liesenfeld (2012) and the DCC-type formulation of Bauwens, Storti, and Violante (2012). The proposed new parameterizations imply a specific impact parameter of the lagged realized covariance (or correlation) on the next conditional covariance (or correlation) of each asset pair; moreover, these impact parameters are time-varying. They nevertheless guarantee the positive definiteness of the conditional covariance (or correlation) matrix with simple parametric restrictions, while keeping the number of parameters fixed or at most linear in the number of assets. In brief, they are more flexible than existing scalar or rank-one BEKK and DCC versions, while adding a scalar parameter to these models, hence, they remain parsimonious.[2]

To illustrate the idea, in the simple scalar BEKK-type CAW process, each conditional covariance $S_{ij,t}$ is specified as a linear function of its own lag and the corresponding lagged realized covariance $C_{ij,t-1}$, so that $S_{ij,t} = c_{ij} + bS_{ij,t-1} + aC_{ij,t-1}$, where c_{ij}, a, and b are parameters. One of the proposed models replaces the impact coefficient a of $C_{ij,t-1}$ on $S_{ij,t}$ by $a\exp(\phi_A R_{ij,t-1})/\exp(\phi_A)$, where ϕ_A is a parameter and $R_{ij,t-1}$ is the lagged realized correlation for the asset pair (i, j). Hence, if ϕ_A differs from zero, the impact coefficient differs between asset pairs and is time-varying. This type of extension adds flexibility in the dynamics, by adding a single parameter. It can be used in non-scalar specifications as well.

These extended parameterizations use the element-by-element (Hadamard) exponential function of a matrix to define the impact parameter matrix of the lagged realized or

1 Notably, the vec-HAR model derived from Chiriac and Voev (2011) and the HAR-DRD model of Oh and Patton (2016). These models are used, with some variations in their scope and specification, in several papers, for example, Hautsch, Kyj, and Malec (2015); Callot, Kock, and Medeiros (2017); deBrito, Medeiros, and Ribeiro (2018); Bollerslev et al. (2021); and Vassallo, Buccheri, and Corsi (2021).

2 Alternatively, a conditional covariance or correlation process can be parsimoniously parameterized by assuming that a small number of factors drive its dynamics. See (among others) Engle, Ng, and Rothschild (1990) in the MGARCH case. Papers centered on or using factor models for realized covariances are mentioned in Section 6.

conditional covariances (or correlations) on the subsequent conditional covariances (or correlations). The Hadamard exponential matrix benefits from several mathematical properties, exploited by Bauwens and Otranto (2020b) in the MGARCH framework to develop DCC models where the conditional correlations also have asset pair-specific and time-varying dynamics. These models are extended to the CAW model family. CAW models that use a Hadamard exponential function in their parameterization, whether in the BEKK-type or in the DCC-type of specifications, will be named "Hadamard Exponential CAW" (HE-CAW). Every HE-CAW model can be simplified to a corresponding simpler CAW model by imposing a parameter restriction that can be tested ($\phi_A = 0$ in the previous paragraph).

BEKK-type CAW models are estimated by maximizing the Wishart log-likelihood function in one step, whereas DCC-type CAW models can be estimated in one step and also in two steps. One-step estimators are in principle efficient statistically, while two-step ones incur an efficiency loss; see Engle and Sheppard (2001) for the MGARCH DCC models. The efficiency issue is complicated by the targeting issue because it is difficult to define a practical targeting estimator in non-scalar formulations. The approximate targeting solution proposed by Hafner and Franses (2009) is adopted, which makes use of an average of the unknown parameters in the targeting estimator. Because the HE parameterization of any CAW model adds a single parameter to its simpler version, the estimation of the HE version is not more difficult than its simpler version. What can make the estimation difficult in practice[3] is the dimension of the model (the number of assets), in non-scalar versions. For example, if the dimension is 100, and a rank-one parameterization is used, the number of parameters to be estimated simultaneously is equal to 200. More parsimonious models can be obtained by a statistical clustering procedure that eventually reduces the number of parameters, resulting in a small number of asset groups that have identical parameters. Alternatively, the groups of assets can be based on an a priori (sectoral or other) classification.

Two empirical exercises, using a data set of 29 stocks and another of 100 stocks, serve to compare the forecasting performance of the BEKK- and DCC-type HE-CAW models with their simple versions, and with benchmark models, after including in the latter a HE extension. The out-of-sample forecasts of the different models are compared using two statistical loss functions (QLIK, a quasi-likelihood loss, and FN, a Frobenius norm) through the model confidence set (MCS) procedure of Hansen, Lunde, and Nason (2003). The forecasts are also compared in term of the standard deviation of the time series of the estimated global minimum variance portfolio (GMVP) returns.

In the estimation results for both data sets, the parameters of the HE-CAW models which imply time-varying and pair-specific impact coefficients are statistically significant at conventional levels and improve the model fit with respect to the simpler versions. In the forecast evaluations, the models with the HE term tend to perform better than the simpler models, and the CAW-type models than the benchmark models, but non-systematic exceptions and variations occur, depending on the data set and the forecast horizon.

The paper is organized as follows: the CAW modeling framework is defined in Section 1. Scalar HE-CAW models are introduced in Section 2. Using the data set of 29 stocks, the scalar models are compared empirically in Section 3 with benchmark models in term of

3 The difficulty we encountered is the computation of robust standard errors, rather than the convergence of the maximization procedure.

covariance matrix out-of-sample forecasting performance. Extended parameterizations of HE-CAW models, and ways to reduce their parameterization, are defined in Section 4. Section 5 presents detailed estimation results and covariance matrix out-of-sample forecast evaluations for 13 HE-CAW models. Section 6 presents empirical results for the data set of 100 stocks. Final remarks conclude the paper. An Online Supplementary Appendix (henceforth referred to as SA) presents additional empirical information and results.

1 Caw Modeling Framework

Let $\boldsymbol{C}_t$ denote the $(n \times n)$ realized covariance matrix of day t $(t = 1, \ldots, T)$, and $\mathcal{I}_t$ the information set at time t, consisting of the current and past values of $\boldsymbol{C}_t$. Several ways of defining $\boldsymbol{C}_t$ as a function of intraday returns are available in the literature. In the CAW framework, the conditional distribution of $\boldsymbol{C}_t$ is a n-dimensional central Wishart with ν $(> n)$ degrees of freedom; in symbols:

$$\boldsymbol{C}_t | \mathcal{I}_{t-1} \sim W_n(\nu, \boldsymbol{S}_t / \nu), \tag{1}$$

where $\boldsymbol{S}_t$, of dimension $n \times n$, is the (positive definite) expected value of the conditional distribution of $\boldsymbol{C}_t$. Under this distributional hypothesis, the marginal conditional distributions of the realized variances (on the diagonal of $\boldsymbol{C}_t$) are univariate Gamma distributions.

The CAW approach consists in specifying $\boldsymbol{S}_t$ as a function of the information set, $\boldsymbol{S}_t$ being indexed by unknown parameters $\boldsymbol{\theta}$, estimated by maximizing the log-likelihood function (excluding terms that do not depend on $\boldsymbol{\theta}$):

$$l(\boldsymbol{\theta} | \boldsymbol{C}_1, \ldots, \boldsymbol{C}_T) = -\frac{\nu}{2} \sum_{t=1}^{T} \left\{ \log |\boldsymbol{S}_t(\boldsymbol{\theta})| + \mathrm{trace}\left[\boldsymbol{S}_t(\boldsymbol{\theta})^{-1} \boldsymbol{C}_t \right] \right\}. \tag{2}$$

In the case of BEKK-type models, the function (2) is maximized with respect to $\boldsymbol{\theta}$ in one step, whereas in the case of DCC-type models, the function is maximized in two steps, as explained in Section 1.2. In both cases, the parameter ν does not affect the estimation of $\boldsymbol{\theta}$, so that it can be set equal to 1.

1.1 BEKK-Type Models

There is a large variety of ways to specify $\boldsymbol{S}_t$, inspired by the MGARCH literature[4]; for example, Golosnoy, Gribisch, and Liesenfeld (2012) adopt a BEKK model. The diagonal BEKK-CAW model is the following process:

$$\boldsymbol{S}_t = \boldsymbol{C} + \boldsymbol{A} \odot \boldsymbol{C}_{t-1} + \boldsymbol{B} \odot \boldsymbol{S}_{t-1}, \tag{3}$$

where $\boldsymbol{C}$, $\boldsymbol{A}$, and $\boldsymbol{B}$ are unknown symmetric matrices of parameters, the first positive definite, the other positive semidefinite, and $\odot$ represents the element-by-element (Hadamard) product. The model is said to be diagonal because each conditional covariance $\boldsymbol{S}_t$ depends only on the corresponding lagged conditional covariance and realized covariance, not on other covariances. The number of parameters in Equation (3), equal to $3n(n+1)/2$, renders maximum-likelihood (ML) estimation unfeasible already for $n \geq 5$. The number of

4 A related but different Wishart-type model by Gorgi et al. (2019) specifies the dynamics of $\boldsymbol{C}_t$ as a function of a vector of time-varying parameters that follow a GAS-type process.

parameters is reduced by considering the covariance targeting version of Equation (3), defined by setting $\boldsymbol{C} = (\boldsymbol{J}_n - \boldsymbol{A} - \boldsymbol{B}) \odot \overline{\boldsymbol{C}}$, where $\boldsymbol{J}_n$ is an $n \times n$ matrix of ones and $\overline{\boldsymbol{C}}$ is the sample mean of the realized covariances $\boldsymbol{C}_t$. In the above nonscalar framework, it is not possible to guarantee that the matrix used for targeting is positive definite; a practical solution, proposed by Hafner and Franses (2009), replaces the possibly non-positive definite matrix $\boldsymbol{J}_n - \boldsymbol{A} - \boldsymbol{B}$ by the scalar $1 - \overline{a} - \overline{b} \in (0, 1)$, where $\overline{a}$ and $\overline{b}$ are the averages of the elements of the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively.

The scalar version of Equation (3), where $\boldsymbol{A} = a\boldsymbol{J}_n$ and $\boldsymbol{B} = b\boldsymbol{J}_n$ with a and b unknown, reduces drastically the number of parameters in $\boldsymbol{A}$ and $\boldsymbol{B}$, so that the dynamics is common for all the variances and covariances. A less important reduction is obtained by adopting a rank-one parameterization, with $\boldsymbol{A} = \boldsymbol{a}\boldsymbol{a}'$ and $\boldsymbol{B} = \boldsymbol{b}\boldsymbol{b}'$, where $\boldsymbol{a} = (a_1^{1/2}, a_2^{1/2} \ldots, a_n^{1/2})'$ and $\boldsymbol{b} = (b_1^{1/2}, b_2^{1/2} \ldots, b_n^{1/2})'$.[5] The advantage of this parameterization is to have different dynamics for the elements of the conditional covariance matrix, even if the parameters for the covariances are derived from the parameters of the variances.

The parameter vector $\boldsymbol{\theta}$ in Equation (2) consists of a and b in the scalar version and of the elements of $\boldsymbol{a}$ and $\boldsymbol{b}$ in the rank-one parameterization.

1.2 DCC-Type Models

Bauwens, Storti, and Violante (2012) and Bauwens, Braione, and Storti (2016) specify $\boldsymbol{S}_t$ in the CAW framework using a formulation similar to that of the dynamic conditional correlation model of Engle (2002) for multivariate GARCH models. They name the model "Re-cDCC" (realized consistent DCC). It consists of n univariate models for the conditional variances and a scalar DCC model for the realized correlation matrix. The conditional variance models are specified as the GARCH-type model (for each asset i):

$$S_{ii,t} = (1 - \alpha_i - \beta_i)\overline{C}_{ii} + \alpha_i C_{ii,t-1} + \beta_i S_{ii,t-1}, \tag{4}$$

where $\overline{C}_{ii}$, $S_{ii,t}$, and $C_{ii,t}$ represent the i-th element of the diagonal of the matrices $\overline{\boldsymbol{C}}$, $\boldsymbol{S}_t$, and $\boldsymbol{C}_t$, respectively, while α_i and β_i are positive scalar parameters.

The conditional correlation model is a DCC model with a correction similar to the "consistent" correction proposed by Aielli (2013) for DCC-MGARCH. As shown by Bauwens, Storti, and Violante (2012), the Wishart log-likelihood (2) can be split into two parts (excluding a constant part):

$$\begin{aligned}
&l(\boldsymbol{\theta}|\boldsymbol{C}_1, \ldots, \boldsymbol{C}_T) = l_v(\boldsymbol{\theta}_v) + l_c(\boldsymbol{\theta}_c|\boldsymbol{\theta}_v) \\
&l_v(\boldsymbol{\theta}_v) = -\frac{\nu}{2}\left[\log|\boldsymbol{D}_t^2| + \mathrm{trace}\left(\boldsymbol{D}_t^{-1}\boldsymbol{C}_t\boldsymbol{D}_t^{-1}\right)\right] = -\frac{\nu}{2}\left[\sum_{i=1}^{n}\log(S_{ii,t}) + \sum_{i=1}^{n} S_{ii,t}^{-1} C_{ii,t}\right] \\
&l_c(\boldsymbol{\theta}_c|\boldsymbol{\theta}_v) = -\frac{\nu}{2}\left\{\log|\boldsymbol{D}_t^{-1}\boldsymbol{S}_t\boldsymbol{D}_t^{-1}| + \mathrm{trace}\left[\left(\boldsymbol{D}_t\boldsymbol{S}_t^{-1}\boldsymbol{D}_t - \boldsymbol{I}_n\right)\boldsymbol{D}_t^{-1}\boldsymbol{C}_t\boldsymbol{D}_t^{-1}\right]\right\},
\end{aligned} \tag{5}$$

where $\boldsymbol{I}_n$ is the $(n \times n)$ identity matrix and $\boldsymbol{D}_t$ is a diagonal matrix with diagonal elements $S_{ii,t}^{1/2}$. The estimation of $\boldsymbol{\theta}$ is split into two steps. The log-likelihood relative to the variance part, $l_v(\boldsymbol{\theta}_v)$, is the sum of the n univariate log-likelihood functions of the conditional variances, that can be maximized consistently in a first step. The parameters relative to the

5 Using the square root of a_i (b_i), the coefficient of the lagged (conditional) variance of the i-th GARCH-type variance equation is a_i (b_i), like a (b) in the scalar version.

correlation part, θ_c, can be estimated in a second step by maximizing $l_c(\theta_c|\theta_v)$, conditional on the estimator of θ_v obtained in the first step. Like for Equation (2), ν does not affect the estimation of θ and can be set to 1. Because the term trace$(D_t^{-1}C_tD_t^{-1})$ appearing in the expression of $l_c(\theta_c|\theta_v)$ does not depend on θ_c, it can be dropped from it in the maximization, so that the second step objective function is actually the log-likelihood of a Wishart density function for $D_t^{-1}C_tD_t^{-1}$, with parameters ν and $D_t^{-1}S_tD_t^{-1}$.

The Re-cDCC model of R_t, in its diagonal version, is defined as the following set of equations:

$$R_t = \tilde{Q}_t^{-1/2} Q_t \tilde{Q}_t^{-1/2}, \tag{6}$$

$$Q_t = Q + A \odot \left(\tilde{Q}_{t-1}^{1/2} D_{t-1}^{-1} C_{t-1} D_{t-1}^{-1} \tilde{Q}_{t-1}^{1/2}\right) + B \odot Q_{t-1}, \tag{7}$$

$$\tilde{Q}_t = \mathrm{diag}(Q_t), \tag{8}$$

where for any square matrix X, diag(X) is the diagonal matrix obtained by setting to zero all the off-diagonal elements of X. Like for the BEKK formulation, a parsimonious version of the constant matrix Q replaces it by $(1-\overline{a}-\overline{b})\overline{R}$, $\overline{R}$ being the sample correlation matrix, computed from $\overline{C}$. The scalar version is obtained by setting $A = aJ_n$ and $B = bJ_n$, and the rank-one version by setting $A = aa'$ and $B = bb'$.

The conditional correlation matrix R_t, defined by Equation (6), should not be confused with the realized correlation matrix that is obtained by transforming the realized covariance matrix C_t into a correlation matrix; the latter is denoted P_t in the sequel. Both S_t and P_t are observable, whereas R_t is not.

1.2.1 Model names

The following acronyms are used in the rest of the paper: CAW for all models falling in the model class defined in this section. The CAW model class contains two families: the first one uses BEKK-type processes for the dynamics of S_t, such as Equations (3) and (11): they are named COV (covariance) models. The second family uses univariate processes for the conditional variances (the diagonal elements of S_t) and DCC-type processes for the conditional correlation matrix R_t, such as Equations (7) and (12); they are named COR (correlation) models.

2 Hadamard Exponential Scalar CAW Models

A clear disadvantage of the parameterizations of the matrices A and B in Equations (3) and (7) is either that they are too heavy for large n, or that they lack flexibility when they are of scalar or rank-one type: the scalar version imposes the same dynamics for all the variances and covariances, whereas the rank-one version imposes that the covariances depend on the product of the corresponding parameters of the variances. Bauwens and Otranto (2020b), in the framework of MGARCH conditional correlation models, provide extensions of the scalar DCC model of Engle (2002), where the elements of A depend in a nonlinear way on the lagged conditional correlations. In particular, in their model, called NonLinear AutoRegressive Correlation (NLARC) model, the effect of the lagged conditional correlations enters through the element-by-element (Hadamard) exponential function.

The objective of adding flexibility in models (3) and (7), while maintaining a parsimonious parameterization, can be obtained by extending and generalizing the Bauwens and

Otranto (2020b) NLARC parameterization to the CAW model family. The matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ become time-varying and are denoted, respectively, by $\boldsymbol{A}_t$ and $\boldsymbol{B}_t$ in the sequel. Two parameterizations of the time-varying matrices $\boldsymbol{A}_t$ and $\boldsymbol{B}_t$ for Equations (3) and (7) are introduced below. In almost all estimations for our empirical experiments, we find that specifying both matrices to be time-varying does not improve the fit, and a better fit is achieved by a time-varying $\boldsymbol{A}$ than by a time-varying $\boldsymbol{B}$.

2.1 Parameterizations of $\boldsymbol{A}_t$ and $\boldsymbol{B}_t$

We define a scalar parameterization of $\boldsymbol{A}_t$, which can be applied similarly to $\boldsymbol{B}_t$ (replace a by b and ϕ_A by ϕ_b) as

$$Sc\ (\text{Scalar}): \boldsymbol{A}_t = a \exp^{\odot}(\phi_A \boldsymbol{M}_t) = a\boldsymbol{J}_n \odot \exp^{\odot}(\phi_A \boldsymbol{M}_t), \tag{9}$$

where $a \in [0, 1)$ and $\boldsymbol{M}_t$ is a positive definite symmetric matrix known at date t.

If ϕ_A is equal to zero, $\exp^{\odot}(\phi_A \boldsymbol{M}_t)$ is equal to $\boldsymbol{J}_n$, so that $\boldsymbol{A}_t$ is constant, being equal to $a\boldsymbol{J}_n$ (scalar model). When ϕ_A is not equal to zero, the elements of $\boldsymbol{A}$ differ because the elements of $\boldsymbol{M}_t$ differ; hence, the dynamics of the variances and covariances (in the BEKK version) or of the correlations (in the DCC version) are different and the coefficients representing the impact of the lagged conditional covariances (or correlations) on the next conditional covariances (or correlations) are time-varying since $\boldsymbol{M}_t$ is time-varying. Two time-varying versions of $\boldsymbol{M}_t$ are used in the Hadamard exponential function of $\boldsymbol{A}_t$:

$$\begin{aligned} Pt: &\quad \boldsymbol{M}_t = \boldsymbol{P}_{t-1} - \boldsymbol{J}_n, \\ Rt: &\quad \boldsymbol{M}_t = \boldsymbol{R}_{t-1} - \boldsymbol{J}_n, \end{aligned} \tag{10}$$

where $\boldsymbol{P}_{t-1}$ is the realized correlation matrix obtained by transforming the realized covariance matrix $\boldsymbol{C}_{t-1}$ into a correlation matrix and $\boldsymbol{R}_{t-1}$ is the conditional correlation matrix. In the COV models, the latter is obtained by transforming $\boldsymbol{S}_{t-1}$ into a correlation matrix and in the COR models, it is the matrix defined in Equation (6).

If $\phi_A \geq 0$, each matrix $\boldsymbol{A}_t$ obtained by combining Equations (9) and (10) is the Hadamard product of the positive definite matrix $a\boldsymbol{J}_n$ and a positive definite matrix $(\exp^{\odot}(\phi_A \boldsymbol{M}_t))$, so that it is a positive definite matrix (see Lemma 3 in Bauwens and Otranto 2020b). It can be directly checked that $\exp^{\odot}(\phi_A \boldsymbol{M}_t)$, for each $\boldsymbol{M}_t$ proposed above, is a positive definite matrix. For example, $\exp^{\odot}[\phi_A(\boldsymbol{R}_{t-1} - \boldsymbol{J}_n)] = \exp^{\odot}(\phi_A \boldsymbol{R}_{t-1})/\exp(\phi_A)$, and since $\boldsymbol{R}_{t-1}$ is positive definite and $\phi_A > 0$, the HE matrix $\exp^{\odot}(\phi_A \boldsymbol{R}_{t-1})$ is positive definite (see Lemma 1 in Bauwens and Otranto 2020b). Moreover, the diagonal elements of $\exp^{\odot}[\phi_A(\boldsymbol{R}_{t-1} - \boldsymbol{J}_n)]$ are equal to 1, since the diagonal elements of $\boldsymbol{R}_{t-1} - \boldsymbol{J}_n$ are equal to zero. Each off-diagonal element is of the type $\exp(\phi_A r)/\exp(\phi_A)$ and therefore in (0, 1) if $\phi_A > 0$, where $r \in (-1, +1)$ is a correlation coefficient. If $\phi_A < 0$, $\exp(\phi_A r)/\exp(\phi_A)$ is larger than 1, but this does not imply that the off-diagonal elements of $\boldsymbol{A}_t$ are larger than 1, because each element of the exponential matrix is multiplied by $a \in [0, 1)$. Hence, it is possible that $\boldsymbol{A}_t$ is positive definite. In other words, the condition $\phi_A \geq 0$ is sufficient, but not necessary, for positive definiteness of $\boldsymbol{A}_t$.

2.2 Interpretation of the HE Term

The question can be raised whether the proposed form of dependence of $\boldsymbol{A}_t$ on lagged (conditional or realized) correlations makes sense for the dynamics of conditional covariances and correlations. In the COV models, an off-diagonal element of $\boldsymbol{A}_t$, equal to $a\exp(\phi_A r_{ij,t-1})/\exp(\phi_A)$ where $r_{ij,t-1}$ is the lagged conditional or realized correlation between assets i and j, represents the impact coefficient of the corresponding lagged realized covariance ($C_{ij,t-1}$) on the next conditional covariance ($S_{ij,t-1}$). The use of a lagged correlation in the impact coefficient can be justified in relation with the phenomenon of volatility clustering. Clustering characterizes financial market volatility, which itself affects the correlations: when a cluster of high volatility occurs, correlations increase with a certain persistence, but the changes in correlations can differ between pairs of assets. Adding a dependence of the impact coefficient on the past correlation of each asset pair through the exponential function is a way to include the impact of the clustering effect on the next conditional covariance in a way that is specific for each asset pair and is time-varying. This time-varying impact element ($\exp(\phi_A r_{ij,t-1})$) is an increasing convex function of $r_{ij,t-1} \in (-1, +1)$ when ϕ_A is positive (as we always find in estimations of HE-CAW models). Hence, when the lagged correlation increases (due to volatility clustering or an idiosyncratic factor), the next conditional covariance increases (for given values of $aC_{ij,t-1}$ and of the other terms); said differently, the higher (lower) the lagged correlation, the higher (lower) the persistence of the lagged realized covariance on the current conditional covariance.[6] The effect on the next conditional correlation, defined as $S_{ij,t}/(S_{ii,t}S_{jj,t})^{1/2}$, is also positive for given values of the conditional variances; however, in case of increased market volatility in the past (resulting in the increased value of $r_{ij,t-1}$), these variances also increase, so that the positive effect in the numerator can be countered. Typically, however, according to empirical evidence, the correlations increase when a strong and persistent volatility clustering episode occurs.

In the COR models, the impact coefficient $a\exp(\phi_A r_{ij,t-1})/\exp(\phi_A)$ represents the impact of the pseudo-correlation $C_{ij,t-1}/(S_{ii,t-1}S_{jj,t-1})^{1/2}$ multiplied by $(Q_{ii,t-1}Q_{jj,t-1})^{1/2}$ (the Aielli type of correction of the DCC model in this context), on the next quasi-correlation $Q_{ij,t}$. The time-varying impact is thus similar to what it is for COV models, but it operates through the quasi-correlation term.

Regarding the difference between the two choices of correlations (realized or conditional), a conditional correlation is a moving average of the realized correlations of the past, including the most recent one. Using the conditional correlations implies thus a smoother dynamic reaction to the past than using the most recent realized correlation. It is an empirical question whether one kind of correlation or the other is better adapted to fit the kind of impact embedded via the Hadamard exponential matrix. In our empirical experiments, we find that the lagged conditional correlation often provides a better fit and out-of-sample forecast performance.

2.3 COR and COV Practical Equations

The COV version of CAW models in the empirical applications of this paper is specified as

$$\boldsymbol{S}_t = (1 - \bar{a}_t - \bar{b})\overline{\boldsymbol{C}} + \boldsymbol{A}_t \odot \boldsymbol{C}_{t-1} + \boldsymbol{B} \odot \boldsymbol{S}_{t-1}, \tag{11}$$

6 This is quite different from the asymmetric effects, whereby the impact of the lagged variance on the next conditional variance is stronger when the lagged return is negative, while the same holds for a covariance when both lagged returns are negative (Cappiello, Engle, and Sheppard 2006); in particular, the HE function does not change the impact of $C_{ii,t-1}$ on $S_{ii,t}$, which is equal to a and larger than the covariance impact when $\phi_A > 0$.

with $\boldsymbol{A}_t$ as defined in the scalar parameterization (9), or with constant $\boldsymbol{A}$ when $\phi_A = 0$. In the latter case, $\overline{a}_t$ is constant, being equal to a. When $\boldsymbol{A}_t$ is time-varying, $\overline{a}_t$ is the average of its elements. The matrix $\boldsymbol{B}$ and the scalar $\overline{b}$ are constant, with $\boldsymbol{B} = b\boldsymbol{J}_n$ and $\overline{b} = b$.

For COR models, Equation (7) is changed to

$$\boldsymbol{Q}_t = \left(1 - \overline{a}_t - \overline{b}\right)\overline{\boldsymbol{R}} + \boldsymbol{A}_t \odot \left(\tilde{\boldsymbol{Q}}_{t-1}^{1/2}\boldsymbol{D}_{t-1}^{-1}\boldsymbol{C}_{t-1}\boldsymbol{D}_{t-1}^{-1}\tilde{\boldsymbol{Q}}_{t-1}^{1/2}\right) + \boldsymbol{B} \odot \boldsymbol{Q}_{t-1}, \tag{12}$$

with the same definitions as above. The conditional variance dynamic equation of the first step of the COR model for each i is specified as Equation (4).

2.4 Stationarity Conditions

Golosnoy, Gribisch, and Liesenfeld (2012) provide the covariance stationarity conditions (i.e., the conditions for the existence of the unconditional second-order moments) of the BEKK-type (or COV) CAW stochastic process as a function of the model parameters, for a more general BEKK(1,1) process than in Equation (3), that is, not necessarily a diagonal process. They obtain the results by writing the vectorized process of $\boldsymbol{C}_t$ as a VARMA(1,1) process and using the stationarity conditions for such a process. Translating these results to the case of Equation (11) with the constant $\boldsymbol{A}$ parameterization in Equation (9), the stationarity condition is $a + b < 1$. When the HE matrix depending on the lagged realized or conditional correlations is included in the parameterization, the $\boldsymbol{C}_t$ process cannot be written as a VARMA process with fixed parameters. The process is nonlinear due to the exponential function; hence, the unconditional moments are not known. However, given that the entries of the HE matrix are all positive, equal to 1 on the diagonal, and smaller than 1 elsewhere, it is obvious that if the stationarity condition holds for the constant $\boldsymbol{A}$ version, it holds at each t for the corresponding time-varying version (since $a\exp(\phi_A r)/\exp(\phi_A) + b < 1$ holds if $a + b < 1$). Intuitively, these extended conditions (for each t) seem sufficient for covariance stationarity.

For the COR models, the stationarity condition for each variance process (4) is $\alpha_i + \beta_i < 1$. For the correlation process (12), the stationarity condition is the same as for the COV parameterizations, and if the stationarity condition holds for a constant $\boldsymbol{A}$ parameterization, it holds for the corresponding time-varying one.

3 Empirical Comparison of Scalar Models

3.1 Data Set of 29 Stocks

To investigate empirically the questions we are interested in, we use a time-series of daily realized covariance matrices computed from a high-frequency data set for 29 stocks of the DOW Jones Industrial Average (DJIA) index; the 30th stock was dropped since it was not permanently in the index during the sample period. The data source is the TAQ database. The sample period is January 3, 2001–April 16, 2018, resulting in 4319 observations. Each daily realized covariance matrix is computed as the sum of the outer products of the 5-min log-returns of the day. The 5-min returns are obtained from synchronized intra-day prices. The synchronization was done globally for the 29 stocks, using 5-min intervals, the price closest (from the left) to the respective sampling point was taken; the first and last 15 min of the day (9:30–16:00) was excluded. The data are annualized in percentage (multiplied by

25,200). The stock names and tickers are listed in Online SA I, where a table shows some summary statistics of realized variances, covariances, and correlations.

3.2 Out-of-Sample Forecast Comparison

As first empirical implementation, we compare the proposed COV and COR scalar models with three types of benchmark scalar models. The first one is the Exponential Weighted Moving Average (EWMA) recursion, with smoothing coefficient 0.94, also used by Lunde, Shephard, and Sheppard (2016)[7]:

$$\text{EWMA}: \; \boldsymbol{S}_t = 0.06\boldsymbol{C}_{t-1} + 0.94\boldsymbol{S}_{t-1}. \tag{13}$$

The other models are the scalar vech-HAR (vHAR) model of Chiriac and Voev (2011) and the scalar HAR–DRD model of Oh and Patton (2016). In the latter, the scalar aspect is in the correlation model for the realized correlation matrix, whereas the variance models are univariate HAR (Corsi 2009). Moreover, we also introduce the same HE term as in the scalar COV and COR models in the vHAR and HAR–DRD models. The HE version of vHAR is

$$\begin{aligned}&\text{HE–vHAR}:\\ &\text{vech}(\boldsymbol{C}_t) = (1-\overline{\alpha}_{D,t}-\alpha_W-\alpha_M)\overline{\boldsymbol{C}} + \alpha_D \exp\big(\phi_A \text{vech}(\boldsymbol{P}_{t-1}-\boldsymbol{J}_n)\big) \odot \text{vech}(\boldsymbol{C}_{t-1}) + \\ &\alpha_W \text{vech}\big(\overline{\boldsymbol{C}}_{t-2:t-5}\big) + \alpha_M \text{vech}\big(\overline{\boldsymbol{C}}_{t-6:t-22}\big),\end{aligned} \tag{14}$$

where vech$(\cdot)$ is the operator that stacks the lower triangular part of its matrix argument as a vector, $\overline{\boldsymbol{C}}_{r:s}$ is the average of the realized covariance matrices from time r to time s, and $\overline{\alpha}_{D,t}$ is the average of the entries of $\alpha_D \exp\big(\phi_A \text{vech}(\boldsymbol{P}_{t-1}-\boldsymbol{J}_n)\big)$.

Similarly, the HE version of HAR–DRD is

$$\begin{aligned}&\text{HE–DRD}:\\ &\text{vech}(\boldsymbol{P}_t) = (1-\overline{\alpha}_{D,t}-\alpha_W-\alpha_M)\overline{\boldsymbol{R}} + \alpha_D \exp\big(\phi_A \text{vech}(\boldsymbol{P}_{t-1}-\boldsymbol{J}_n)\big) \odot \text{vech}(\boldsymbol{P}_{t-1}) + \\ &\alpha_W \text{vech}\big(\overline{\boldsymbol{P}}_{t-2:t-5}\big) + \alpha_M \text{vech}\big(\overline{\boldsymbol{P}}_{t-6:t-22}\big),\end{aligned} \tag{15}$$

with similar interpretation of the symbols. Both vHAR and HAR–DRD, which correspond to $\phi_A = 0$, are estimated by ordinary least squares, whereas HE-vHAR and HE-DRD are estimated by nonlinear least squares (see results in Online SA III).

To compare the model performances in out-of-sample forecasts of the covariance matrix, the 11 models (three COR-S, three COV-S, EWMA, and the four HAR-type models) have been estimated using an expanding window scheme, adding 25 observations at a time. The initial estimation is on the period from January 2, 2001 to December 31, 2015 ($T = 3744$). Using these estimates, 25 forecasts (1-step, 5-step, and 22-step ahead) are computed after the last date of the estimation window. The next estimation is on the initial period plus 25 observations, from which 25 forecasts are computed after the end of this extended estimation sample. This procedure is repeated until the end of the data set which contains 575 realized covariance matrices after December 31, 2015. These observed matrices are

7 deBrito, Medeiros, and Ribeiro (2018) and Gorgi et al. (2019) also use the EWMA equation for forecasting realized covariance matrices, but with coefficient 0.96.

compared with the h-step ahead forecasted covariance matrices using loss functions that allow us to compare the forecasted and realized covariance matrices, for $h = 1$, 5, and 22 (so that $T_1 = 575$, $T_5 = 571$, and $T_{22} = 564$ are the lengths of the forecast samples). The statistical loss functions adopted are the forecast sample means of the Quasi-Likelihood function (QLIK) and of the squared Frobenius norm (FN); both functions are consistent in the sense of Patton (2011), Patton and Sheppard (2009), and Laurent, Rombouts, and Violante (2013). They are defined as

$$\mathrm{QLIK}_h = \frac{1}{T_h}\sum_{t=T+h}^{T+T_h}\left\{\ln|\widehat{S}_{t,h}| + \mathrm{trace}\left(\widehat{S}_{t,h}^{-1}C_t\right)\right\}, \tag{16}$$

$$\mathrm{FN}_h = \frac{1}{T_h}\sum_{t=T+h}^{T+T_h}\mathrm{trace}\left[\left(\widehat{S}_{t,h} - C_t\right)'\left(\widehat{S}_{t,h} - C_t\right)\right] = \frac{1}{T_h}\sum_{t=T+h}^{T+T_h}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\widehat{s}_{ij,t,h} - c_{ij,t}\right)^2, \tag{17}$$

where $\widehat{S}_{t,h}$ is the h-step ahead forecasted conditional covariance of day t.[8]

For each loss function, the MCS procedure of Hansen, Lunde, and Nason (2003) and Hansen, Lunde, and Nason (2011) is used to identify the best models with a chosen level of confidence. To compute it, we adopt the semi-quadratic test statistic $\sum_{i\neq j\in\mathcal{M}}[\bar{l}_{ij}^2/\widehat{Var}(\bar{l}_{ij})]$, where $\bar{l}_{ij}$ is the mean of the loss differences between model i and model j belonging to the set of models $\mathcal{M}$; the variance of $\bar{l}_{ij}$ is obtained by the bootstrap procedure of Hansen, Lunde, and Nason (2003), with 10,000 replications. In brief, the first test is for the full set of candidate models and checks the null hypothesis that $\bar{l}_{ij} = 0$; the model presenting the highest differences (on average) with respect to all other models is eliminated from the set of candidate models and the test is re-applied to the set of remaining models until we obtain just one. In Table 1 we show, for each forecast horizon and each loss function, which models belong to the MCS at the 95% level of confidence.

The MCS results for the statistical loss functions are mainly favoring the models using the HE extension, and among them, especially COV and COR models. Which models of these families are included in the 95% MCS set depends on the forecast horizon and the loss function. Except for QLIK_5 and QLIK_{22}, each MCS includes COR or COV models using the HE specifications. In the two cases where a vHAR or HAR-DRD model is in the MCS, it is the version with the HE extension which is in the MCS.

In addition to the statistical loss functions, we consider an economic loss function, which compares the forecasts in term of the GMVP risk. For a given model and forecast horizon h, at each time period t of the forecast sample the weight vector $(\boldsymbol{w}_{t,h})$ of the GMVP is computed by minimizing the forecasted portfolio variance $(\boldsymbol{w}_{t,h}'\widehat{S}_{t,h}\boldsymbol{w}_{t,h})$ under the constraint that the weights add-up to unity ($\boldsymbol{j}_n'\boldsymbol{w}_{t,h} = 1$, where $\boldsymbol{j}_n$ is a vector of ones). The solution is $\widehat{\boldsymbol{w}}_{t,h} = \widehat{S}_{t,h}{}^{-1}\boldsymbol{j}_n/(\boldsymbol{j}_n'\widehat{S}_{t,h}{}^{-1}\boldsymbol{j}_n)$. The weights are applied to the observed returns of the forecast period, which results in 575 portfolio returns (for $h = 1$) and the standard deviation of this time series of forecasted portfolio returns is computed. The best model corresponds

8 The QLIK_1 loss function (16) is equal to the estimation objective function (2) (setting $S_t(\boldsymbol{\theta}) = \widehat{S}_{t,1}$), multiplied by $-2/\nu$, to make it a loss and remove the nuisance parameter ν. Another possible estimation criterion consists in minimizing the FN loss function (17) where $\widehat{S}_{t,h}$ is replaced by the specified $S_t(\boldsymbol{\theta})$.

Table 1 Out-of-sample forecast analysis of 11 scalar models for 29 stocks: 95% MCSs for two loss functions and three forecast horizons h

Model	QLIK			FN		
COR-S	1			1		
COR-S-Pt	1			1		
COR-S-Rt	1			1		
COV-S		5	22		5	22
COV-S-Pt					5	22
COV-S-Rt					5	22
EWMA						
vHAR						
HE-vHAR				1		
HAR-DRD						
HE-HAR-DRD						22

Notes: The symbol 1 identifies the models belonging to the MCS for $h=1$; the symbol 5 for $h=5$; and the symbol 22 for $h=22$. The definitions of the COR-S- and COV-S- models are summarized in Table 3 and the other models are defined in Equations (13)–(15).

to the smallest standard deviation. The MCS procedure is applied to compare the variances of the different models.[9] The results are shown in Table 2.

At horizon 1, the MCS consists of EWMA (the lowest loss), the three COV models (with losses <1% higher), and the COR-S-Pt model (loss 2.2% higher). At horizon 5, it consists of COV-S-Rt, and at horizon 22, the MCS includes COV-S-Pt and COV-S-Rt (lowest losses), COV-S (+0.4%), EWMA (+3.6%), vHAR (+3.7%), and HE-vHAR (+4.3%). The COV-S-Rt is in the three MCS and has the smallest loss for two horizons.

4 Extended Parameterizations

4.1 Rank-One Parameterizations

In Section 1, we reminded that the scalar specifications of the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ of COV and COR models can be extended by parameterizing them as $\boldsymbol{A}=\boldsymbol{a}\boldsymbol{a}'$ and $\boldsymbol{B}=\boldsymbol{b}\boldsymbol{b}'$, where $\boldsymbol{a}=(a_1^{1/2},a_2^{1/2}\ldots,a_n^{1/2})'$ and $\boldsymbol{b}=(b_1^{1/2},b_2^{1/2}\ldots,b_n^{1/2})'$. The advantage of this parameterization is to have different impact coefficients for the elements of the conditional covariance or quasi-correlation matrix, instead of coefficients that do not vary with asset pairs (i,j) as in the scalar models. Nevertheless, these impact coefficients remain constant through time for each pair (i,j). By combining the rank-one parameterizations with the Hadamard exponential matrix, we obtain time-varying coefficients. The rank-one parameterization of $\boldsymbol{A}_t$ in Equations (11) and (12) is defined as

$$R1\ (\text{Rank}-1){:}\ \boldsymbol{A}_t=\boldsymbol{a}\boldsymbol{a}'\odot\exp^{\odot}(\phi_A\boldsymbol{M}_t), \tag{18}$$

9 This is done like for the statistical loss functions, with $\bar{l}_{ij}$ in the semi-quadratic test statistic is the sample average of $(y_{ti}-\bar{y}_i)^2-(y_{tj}-\bar{y}_j)^2$, where y_{ti} is the GMVP return at time t for model i, and $\bar{y}_i$ the average of these returns over the forecast sample.

Table 2 Annualized standard deviations and MCS of out-of-sample *h*-step ahead forecasts of GMVP returns of 11 scalar models for 29 stocks

Model	$h=1$	$h=5$	$h=22$
COR-S	8.24	8.15	8.04
COR-S-Pt	**8.21**	8.14	8.01
COR-S-Rt	8.23	8.15	8.02
COV-S	**8.10**	7.53	**7.63**
COV-S-Pt	**8.08**	7.51	**7.60**
COV-S-Rt	**8.07**	**7.49**	**7.60**
EWMA	**8.03**	8.05	**7.87**
vHAR	8.33	8.01	**7.88**
HE-vHAR	8.36	8.04	**7.93**
HAR-DRD	8.53	7.94	8.18
HE-HAR-DRD	8.53	7.94	8.16

Note: Bold values identify the models forming the 95% MCS for the GMVP variance loss function defined in Section 3.2, see footnote 9.

with $\phi_A \geq 0$, $\boldsymbol{M}_t$ as in Equation (10) and $\boldsymbol{a} = (a_1^{1/2}, \ldots, a_n^{1/2})'$ in which each element is in (0, 1). Each matrix $\boldsymbol{A}_t$ obtained by combining Equations (18) and (10) is positive definite by application of Lemma 3 in Bauwens and Otranto (2020b). The stationarity condition for the rank-one model when $\phi_A = 0$ is $\max(\boldsymbol{aa}' + \boldsymbol{bb}') < 1$, where max applied to a matrix selects its largest entry. In the rank-one version of Equations (11) and (12), $\overline{a}_t$ is the average of the elements of $\boldsymbol{A}_t$; when $\phi_A = 0$, $\overline{a}_t$ is constant, being the average of the elements of $\boldsymbol{aa}'$.

4.2 Parsimonious Parameterizations by Asset Grouping

To reduce the number of parameters in a rank-one model, we can form groups of assets having similar parameters by applying Wald tests using the estimates of the model parameters (and an estimated covariance matrix of the estimator). If more than one group is formed, the model can subsequently be estimated under the restrictions that the parameters of the assets belonging to a group are equal. This enables us to reduce the number of estimated parameters; for example, for 29 assets, the initial COV-rank-one (COR-R1) model has 58 (116) parameters, but if three groups are formed, this is reduced to 6 (64). The reduction can be useful for two reasons: firstly, for a large number of assets, the estimation algorithm may fail due to the large number of parameters, or even if it converges, the computation of the estimated asymptotic covariance matrix of the estimator may fail to yield a positive definite matrix; and secondly, the reduction typically reduces the estimator variance if the restrictions are valid.

In principle, we could use an algorithm similar to the one used in Bauwens and Otranto (2020b), testing for each rank-one COR and COV models the joint hypothesis $a_i = a_j$ and $b_i = b_j$ for each pair of assets (i, j) and then assigning to the same group all the assets for which the hypothesis is not rejected consistently with the other assets belonging to the same group, starting from the pair with the highest p-value. Alternatively, we can follow the same procedure by testing jointly $\alpha_i = \alpha_j$ and $\beta_i = \beta_j$, the parameters of the variance processes (4) estimated in the first step estimation procedure of COR models, and adopting this grouping for all the rank-one COR and COV models.[10]

This type of procedure becomes unfeasible when the number of assets is large. An alternative method is to use an a priori grouping, for example, based on an industrial classification of firms, as in Fan, Furger, and Xiu (2016) and deBrito, Medeiros, and Ribeiro (2018).[11] We have applied our models using an industry-based classification of the 29 stocks in seven groups; notice that there is no unique way to specify the number of groups and their composition, since several firms are multi-industry. We find that the fit is slightly less good using this grouping than the grouping based on the algorithm described below.

Another statistical clustering algorithm, related to the grouping algorithm of Bauwens and Otranto (2020b) described above, is based on a classical agglomerative clustering procedure (Anderberg 1973) using a dissimilarity measure given by 1 minus the p-value of the test statistic of the equality of the parameters of pairs of univariate models in the first step estimation of the COR models. Furthermore, the number of groups can be decided automatically by imposing a stopping rule in the agglomerative steps; for example, when the dissimilarity measure at which the clusters are glued together is greater than 0.95 (corresponding to a test at a 5% significance level in our algorithm). Using an average linkage criterion, we obtain a classification of the 29 assets in three groups very similar to those of the algorithm of Bauwens and Otranto (2020b) using the tests on the parameters of the variance processes. The latter identifies four groups; with respect to the three groups, only one series is classified differently, and three series of the first group form a fourth group; Table 13 in Online SA II shows the group composition of the two algorithms, and of the a priori grouping in seven industries. The agglomerative clustering procedure is feasible with a large number of assets and is also used in detecting the number of groups in the application to 100 assets reported in Section 6.1.

5 Empirical Comparisons of COV and COR Models

We estimate and evaluate the forecasting performance of 13 models on the dataset of $n = 29$ stocks of the DJIA index described in Section 3.1. Table 3 provides the model list with references to the equations that define them and their specific parameterizations.

5.1 Estimation Results for COR and COV Models

Like for the scalar models (see Section 3.2), the rank-one versions have been estimated on the period from January 2, 2001 to December 31, 2015 ($T = 3744$). The models have been estimated with the parametric restrictions implied by the three groups obtained by the agglomerative algorithm described in the previous section.

10 We adopted this procedure in a previous version of this paper (Bauwens and Otranto 2020a), obtaining a grouping similar to the results when the procedure is performed specifically for each model.

11 Another possible extension for COR models is to consider a block structure for Q_t, based on the canonical representation proposed in Archakov and Hansen (2020), in which the matrix can be represented by a smaller $K \times K$ matrix, where K is the number of blocks, which could be modeled in the same manner as Q_t in Equation (7).

Table 3 List of models

Model name	Number of parameters	Specific parameterizations of $\boldsymbol{A}$ and $\boldsymbol{B}$
COR models defined in Equations (4) and (12), (6) and (8)		
COR-S-1s	$2n_g + 2$	$\boldsymbol{A} = a\boldsymbol{J}_n$, $\boldsymbol{B} = b\boldsymbol{J}_n$; estimated in one step
COR-S	$2n + 2$	$\boldsymbol{A} = a\boldsymbol{J}_n$, $\boldsymbol{B} = b\boldsymbol{J}_n$
COR-S-Pt	$2n + 3$	$\boldsymbol{A}_t = a \exp^{\odot}[\phi_A(\boldsymbol{P}_t - \boldsymbol{J}_n)]$, $\boldsymbol{B} = b\boldsymbol{J}_n$
COR-S-Rt	$2n + 3$	$\boldsymbol{A}_t = a \exp^{\odot}[\phi_A(\boldsymbol{R}_t - \boldsymbol{J}_n)]$, $\boldsymbol{B} = b\boldsymbol{J}_n$
COR-R1	$2n + 2n_g$	$\boldsymbol{A} = \boldsymbol{a}\boldsymbol{a}'$, $\boldsymbol{B} = \boldsymbol{b}\boldsymbol{b}'$
COR-R1-Pt	$2n + 2n_g + 1$	$\boldsymbol{A}_t = \boldsymbol{a}\boldsymbol{a}' \odot \exp^{\odot}[\phi_A(\boldsymbol{P}_t - \boldsymbol{J}_n)]$; $\boldsymbol{B} = \boldsymbol{b}\boldsymbol{b}'$
COR-R1-Rt	$2n + 2n_g + 1$	$\boldsymbol{A}_t = \boldsymbol{a}\boldsymbol{a}' \odot \exp^{\odot}[\phi_A(\boldsymbol{R}_t - \boldsymbol{J}_n)]$; $\boldsymbol{B} = \boldsymbol{b}\boldsymbol{b}'$
COV models defined in Equation (11)		
COV-S	2	$\boldsymbol{A} = a\boldsymbol{J}_n$, $\boldsymbol{B} = b\boldsymbol{J}_n$
COV-S-Pt	3	$\boldsymbol{A}_t = a \exp^{\odot}[\phi_A(\boldsymbol{P}_t - \boldsymbol{J}_n)]$, $\boldsymbol{B} = b\boldsymbol{J}_n$
COV-S-Rt	3	$\boldsymbol{A}_t = a \exp^{\odot}[\phi_A(\boldsymbol{R}_t - \boldsymbol{J}_n)]$, $\boldsymbol{B} = b\boldsymbol{J}_n$
COV-R1	$2n_g$	$\boldsymbol{A} = \boldsymbol{a}\boldsymbol{a}'$, $\boldsymbol{B} = \boldsymbol{b}\boldsymbol{b}'$
COV-R1-Pt	$2n_g + 1$	$\boldsymbol{A}_t = \boldsymbol{a}\boldsymbol{a}' \odot \exp^{\odot}[\phi_A(\boldsymbol{P}_t - \boldsymbol{J}_n)]$; $\boldsymbol{B} = \boldsymbol{b}\boldsymbol{b}'$
COV-R1-Rt	$2n_g + 1$	$\boldsymbol{A}_t = \boldsymbol{a}\boldsymbol{a}' \odot \exp^{\odot}[\phi_A(\boldsymbol{R}_t - \boldsymbol{J}_n)]$; $\boldsymbol{B} = \boldsymbol{b}\boldsymbol{b}'$

Notes: n is the number of assets; n_g is the number of groups with equal coefficients identified by the grouping procedure defined in Section 4.2. COR models are estimated in two steps, except COR-S-1s.

5.1.1 COR models

The estimation results for the scalar models and the R1 models with three groups are reported in Table 4. The upper part of the table, under the heading 'S-1s', shows the estimates of the parameters of the variance equations of the three groups. These parameters have been estimated jointly with those of the scalar correlation model, reported in the bottom part of the table (heading 'S-1s'). For the six models estimated in two steps (S, S-Pt, S-Rt, R1, R1-Pt, R1-Rt), the first step consists in estimating the 29 variance Equation (4), but instead of reporting the 29 estimates of α_i and of β_i, we report the average, minimum and maximum of these estimates in each group ('Variance Part' of the table, under the heading 'first step'); group 1 has 20 stocks, group 2 has 4, and group 3 has 5 (see column 3 of Table 13 in SA II).

Concerning the variance parameter estimates, the lagged realized variance impact coefficients (α_i) are smaller (on average) in joint estimation than in univariate estimation, whereas the lagged conditional variance coefficients (β_i) are larger, but the estimates of the persistence effect ($\alpha_i + \beta_i$) are close in both estimations.

Considering the correlation parameters a_i and b_i (bottom part of the table), the differences are small within the scalar models and within the $R1$ models, and even between them. Actually, each scalar model (estimated in two steps) is not rejected against the corresponding $R1$ version: the largest likelihood ratio (LR) statistic is equal to 6.2 for 4 restrictions. On the contrary, the HE parameter (ϕ_A) is positive and significant in each model.[12] This

12 The z-ratios are 2.31, 1.86, 2.47, and 2.36. However, the test is not standard, the null being at the boundary of the parameter admissible values. Bauwens and Otranto (2020b) show by a Monte Carlo study that in the DCC MGARCH model (with the HE extension), the distribution of the z-ratio is close to $N(0,1)$ if the sample size is "large enough" and the true value is not "too close" to zero.

Table 4 Estimation results of scalar and R1 (with three groups) COR models, for 29 stocks (robust standard errors in parentheses)

Parameters	S-1s	Variance part First step		
		Average	Min	Max
α_1	0.284	0.361	0.276	0.411
	(0.018)			
α_2	0.310	0.395	0.369	0.455
	(0.022)			
α_3	0.251	0.325	0.270	0.436
	(0.030)			
β_1	0.697	0.616	0.567	0.712
	(0.024)			
β_2	0.679	0.591	0.530	0.618
	(0.024)			
β_3	0.725	0.645	0.510	0.710
	(0.033)			

	Correlation part						
	S-1s	S	S-Pt	S-Rt	R1	R1-Pt	R1-Rt
a_1	0.051	0.052	0.057	0.054	0.052	0.056	0.054
	(0.003)	(0.004)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
a_2					0.059	0.062	0.061
					(0.004)	(0.004)	(0.004)
a_3					0.048	0.053	0.050
					(0.007)	(0.007)	(0.007)
b_1	0.933	0.930	0.926	0.928	0.931	0.928	0.929
	(0.005)	(0.006)	(0.007)	(0.006)	(0.006)	(0.006)	(0.005)
b_2					0.921	0.919	0.929
					(0.006)	(0.006)	(0.006)
b_3					0.928	0.925	0.927
					(0.015)	(0.014)	(0.015)
ϕ_A			0.071	0.030		0.067	0.026
			(0.031)	(0.016)		(0.027)	(0.011)
	Variance and correlation parts						
Log-lik	−92657.3	−92698.6	−92690.5	−92695.8	−92695.5	−92688.2	−92694.0
AIC	49.51	49.56	49.56	49.56	49.56	49.56	49.56
BIC	49.53	49.66	49.66	49.66	49.67	49.67	49.67

Notes: The models are defined in Table 3. All models except COR-S-1s are estimated in two steps, the first step results being the same and obtained from 29 univariate variance models. In the upper part, we show the average, the minimum, and the maximum in each group. In the lower part, we show the log-likelihood, AIC, and BIC values for the full model.

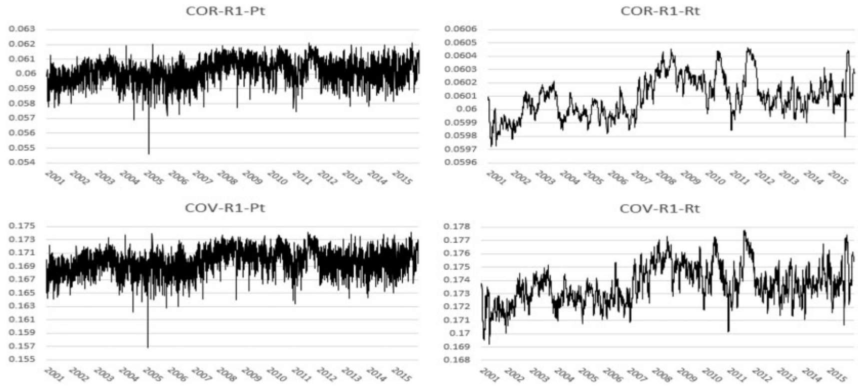

Figure 1 Time series of $a_{ij,t}$ coefficients estimated with the COR-R1-Pt and the COR-R1-Rt models (upper graphs) and with the COV-R1-Pt and the COV-R1-Rt models (lower graphs). The coefficients are for the asset pair AXP–JPM.

implies time–varying elements in the $\boldsymbol{A}_t$ matrices of (12); an illustration is shown in the upper part of Figure 1, where the time-varying coefficients $a_{ij,t}$ (for $i = AXP$, $j = JPM$) for the COR-R1-Pt and COR-R1-Rt models are shown over the estimation period. Obviously, the dynamics is smoother when the driving variable is the conditional correlation (Rt model) than the realized correlation (Pt model), because R_t is by construction a smoothed correlation matrix (being based on all past realized correlations), whereas P_t is based on the last realized correlation.

5.1.2 COV models

The estimation results for the scalar models and the R1 models with three groups are shown in Table 5. The positive and significant[13] estimated ϕ_A imply time-varying elements in the $\boldsymbol{A}_t$ matrix of Equation (11). An illustration is shown in the lower part of Figure 1.

In the R1 models, the (a_i, b_i) parameters vary between groups, for example, between (0.116, 0.875) for group 3 and (0.168, 0.818) for group 2 in the R1 model. One can notice that when $a_i > a_j$, then $b_j > b_i$, so that persistence $(a_i + b_i)$ is more stable across the groups than each of the two parameters.

5.1.3 Comparison of estimated *A* matrices

The comparison of the estimated $\boldsymbol{A}$ matrices of the different COR and COV models is not easy by inspection of Tables 4 and 5. For each model, the implied matrix (of size 29×29) is computed using the estimated parameters. Table 6 shows the squared Frobenius distances for all pairs of $\boldsymbol{A}$ of the six COR models and the same for the five COV models. For the models with time-varying $\boldsymbol{A}$ (S-Pt, S-Rt, R1-Pt, and R1-Rt), the value used in the distance computations is the average of the time-varying matrices.

For the COR models, the distances are very small and practically null between S and S-1s, S-Pt and S-Rt, S-Pt and R1, and between R1-Rt and R1-Pt. The distance between S-Pt and S-Rt is much smaller than the distances between each of them and S. The same

13 The four z-ratios are larger than 3.94.

Table 5 Estimation results of scalar and R1 (with three groups) COV models, for 29 stocks (robust standard errors in parentheses)

Parameters	S	S-Pt	S-Rt	R1	R1-Pt	R1-Rt
a_1	0.135	0.153	0.160	0.138	0.153	0.162
	(0.006)	(0.010)	(0.010)	(0.006)	(0.010)	(0.010)
a_2				0.168	0.175	0.180
				(0.010)	(0.011)	(0.011)
a_3				0.116	0.138	0.144
				(0.008)	(0.011)	(0.012)
b_1	0.855	0.845	0.840	0.853	0.845	0.838
	(0.007)	(0.009)	(0.009)	(0.007)	(0.009)	(0.009)
b_2				0.818	0.819	0.815
				(0.011)	(0.010)	(0.010)
b_3				0.875	0.861	0.856
				(0.008)	(0.010)	(0.011)
ϕ_A		0.061	0.072		0.055	0.070
		(0.015)	(0.012)		(0.014)	(0.013)
Log-lik	−93522.1	−93464.4	−93428.6	−93485.7	−93442.2	−93409.4
AIC	49.97	49.94	49.92	49.96	49.93	49.92
BIC	49.98	49.95	49.93	49.97	49.94	49.93

Note: The models are defined in Table 3.

Table 6 Distances between the different A matrices of the COR models of Table 4 and the COV models of Table 5

Distance between $\boldsymbol{A}$ matrices in COR models

	S	S-Pt	S-Rt	R1	R1-Pt	R1-Rt
S-1s	0.055	0.706	0.396	1.060	1.032	0.781
S		0.372	0.156	0.639	0.684	0.529
S-Pt			0.048	0.058	0.261	0.369
S-Rt				0.175	0.330	0.343
R1					0.286	0.460
R1-Pt						0.053

Distance between $\boldsymbol{A}$ matrices in COV models

	S-Pt	S-Rt	R1	R1-Pt	R1-Rt
S	12.105	27.710	9.169	17.881	34.948
S-Pt		3.192	15.346	4.031	8.255
S-Rt			27.961	6.361	3.985
R1				9.139	23.276
R1-Pt					3.521

Note: Each value is 100 times the squared FN (= trace$\boldsymbol{X}'\boldsymbol{X}$) of the difference ($\boldsymbol{X}$) between the estimated matrices of the models in the row and column headers.

comment applies to R1-Pt, R1-Rt, and R1. The distances between S-1s and the other models are larger than the distances between S and the other models, reflecting that the other models share a common estimation of the variance equations.

Higher differences are present between the A matrices of the COV models, where there are no common estimations (whereas the first step is common in five COR models). The distance between S-Pt and S-Rt is much smaller than the distances between each of them and S. The same comment applies to R1-Pt, R1-Rt, and R1. The four distances between the pairs formed by combining each element of {S-Pt, S-Rt} with each element of {R1-Pt, R1-Rt} are small in comparison with the distances between S and each element of {R1-Pt, R1-Rt} and between R1 and {S-Pt, S-Rt}. The distance between S and R1 is larger than between S-Pt and R1-Pt and between S-Rt and R1-Rt.

5.2 Covariance Matrix Out-of-Sample Forecast Evaluation

To compare the model performances in out-of-sample forecasts, we proceed exactly as explained in Section 3.2, considering this time only the scalar and R1 COR and COV models, as reported in the previous subsection. Table 7 shows which models belong to the 95% MCS for each statistical loss function and forecast horizon.

The composition of the MCS depends mainly on the forecast horizon. For horizon 22, only COV models are in the MCS of the two loss functions; for horizon 5, the same result holds for QLIK, while for FN, the rank-one COR models also belong to the MCS. For horizon 1, only COR models belong to the MCS for QLIK; for FN, all models (except COV-R1) are in the MCS.

Table 8 reports the results for the GMVP loss. At horizons 1 and 22, the COV models are in the MCS, with standard deviations smaller than COR models; at horizon 5, the COV models are also in the MCS, together with the R1 COR models. For each horizon, the COV models provide smaller standard deviations than the COR models.

5.3 Decomposing FN Loss between Variance and Covariance Contributions

The FN loss function (17) can be decomposed as the sum of the variance contribution and the covariance one:

$$\mathrm{FN}_h = \frac{1}{T_h} \sum_{t=T+h}^{T+T_h} \left[\sum_i \left(\widehat{s}_{ii,t,h} - c_{ii,t} \right)^2 + 2 \sum_{i<j} \left(\widehat{s}_{ij,t,h} - c_{ij,t} \right)^2 \right]. \tag{19}$$

The other loss functions function cannot be broken down into a part that depends only on the covariances and a part that depends only on the variances.

Table 9 shows the variance and covariance contributions to FN for COV and COR models. The values are reported as averages with respect to each model class, because there is little variation within each class. It is clear that the contributions are similar between COR and COV models. The respective contributions are stable with respect to the forecast horizon.

6 Dealing with Large Datasets

Portfolio managers very often deal with a large number of assets to mitigate price impacts when trading large sums of money. Several authors have contributed to adapt existing

Table 7 Out-of-sample forecast analysis of 13 COR and COV models for 29 stocks: 95% MCSs for two loss functions and three forecast horizons *h*

Model	QLIK			FN		
COR-S-1s				1		
COR-S	1			1		
COR-S-Pt	1			1		
COR-S-Rt				1		
COR-R1	1			1	5	
COR-R1-Pt	1			1	5	
COR-R1-Rt	1			1	5	
COV-S			22	1	5	22
COV-S-Pt				1	5	22
COV-S-Rt		5		1	5	22
COV-R1					5	22
COV-R1-Pt				1	5	22
COV-R1-Rt				1	5	22

Notes: The symbol 1 identifies the models belonging to the best set for $h = 1$; the symbol 5 for $h = 5$; and the symbol 22 for $h = 22$. The models are defined in Table 3.

Table 8 Annualized standard deviations and MCS of out-of-sample *h*-step ahead forecasts of GMVP returns of 13 COV and COR models for 29 stocks

Model	$h = 1$	$h = 5$	$h = 22$
COR-S-1step	8.40	8.34	8.16
COR-S	8.24	8.15	8.04
COR-S-Pt	8.21	8.14	8.01
COR-S-Rt	8.23	8.15	8.02
COR-R1	8.25	**7.65**	8.03
COR-R1-Pt	8.22	**7.66**	8.01
COR-R1-Rt	8.23	**7.65**	8.02
COV-S	8.10	**7.53**	**7.63**
COV-S-Pt	**8.08**	**7.51**	**7.60**
COV-S-Rt	**8.07**	**7.49**	**7.60**
COV-R1	**8.05**	**7.53**	**7.66**
COV-R1-Pt	**8.04**	**7.51**	**7.62**
COV-R1-Rt	**8.03**	**7.49**	**7.62**

Note: Bold values identify the models forming the 95% MCS for the GMVP variance loss function defined in Section 3.2, see footnote 9.

methods or develop new models to deal with a large number of assets. In the multivariate GARCH framework, Engle, Ledoit, and Wolf (2019) and De Nard et al. (2021b) apply in different ways the DCC model of Engle (2002), which uses daily return data, from which the daily conditional covariance matrix of returns is modeled. An impressive aspect of these papers is the ability to estimate the DCC model for a genuinely large dimension (up to 1500 in the first paper and to 1000 in the second one). De Nard et al. (2021a) improve the DCC

Table 9 Variance and covariance contributions (in %) to FN loss for the 3-group COR and COV models for 29 stocks: average contribution in all models of each class

	1 step		5 steps		22 steps	
Models	Var	Cov	Var	Cov	Var	Cov
COR	11.6	88.4	11.5	88.5	13.8	86.2
COV	11.4	88.6	11.5	88.5	12.8	87.2

model of Engle, Ledoit, and Wolf (2019), maintaining the possibility to work with very large datasets, using an intraday counterpart of the simple daily return, based on open/high/low/close price data, in conjunction with a smoothed sign of the return, called the regularized return, used in the second-step estimation.

Dealing with realized covariance matrices, the most important contributions extend existing models (e.g., HAR, GAS, and HEAVY) or adopt factor models. In the first group, it is worth noting the work of Oh and Patton (2016), who propose the HAR-DRD model, where the (log of) realized variances are specified as univariate HAR processes estimated by OLS, whereas the vech of the realized correlation matrix is specified as a scalar HAR process of the vech of the lagged correlation matrices; they apply the models to data for 104 stocks of the NYSE TAQ database. Vassallo, Buccheri, and Corsi (2021) propose a model in the spirit of Gorgi et al. (2019) and in the spirit of DCC, but using the GAS idea for variance and correlation, so that estimation is in two steps; the dimension is 100 at most.

Examples of large datasets of realized covariances analyzed with factor models are[14] Hautsch, Kyj, and Malec (2015) (dimension 400) and deBrito, Medeiros, and Ribeiro (2018) (dimension 430). Other examples are given by Fan, Furger, and Xiu (2016), and Aït-Sahalia and Xiu (2017), forecasting covariances in dimension 500 and 491, respectively, using the random walk recursion, which does not require an econometric estimation. Ke, Lian, and Zhang (2022) propose a new dynamic structure for high-dimensional covariance matrices, combining common risk factors embedded with a homogeneous structure, showing the possibility of using it with ultra-high dimensions with a simulation study. Also, interesting the work of Sheppard and Xu (2019), introducing factor HEAVY models, but applying them to a reduced dimension (40 firms). The empirical implementation of a factor model in the context of intra-day data requires a larger information set than just realized covariance matrices; for example, deBrito, Medeiros, and Ribeiro (2018) use (in addition to the realized matrices) the 5-min distant returns inside the trading period and accounting and market information to construct the weights of the factors.

Considering the proposed class of HE models, a nice characteristic is that, depending on a small number of parameters and providing positive definite covariance matrices, they can be adopted, in principle, also for large dimensions. This is of course most easy with the scalar versions of our models, where the number of coefficients is very small, but also possible with the R1 parameterization coupled with the clustering algorithm. So, the advantage is that HE models can work with small and large dimensions, without the necessity to

14 See (among others) Engle Ng, and Rothschild (1990) in the MGARCH case.

introduce new parameterizations or factors to reduce the dimension. In this framework, a problem that arises is the creation of vast realized correlations with the property of positive definiteness. In general, for n stocks, we need substantially more than n intraday returns to get a well-behaved positive-definite realized covariance matrix; for example, for 6.5 h of trading, n must be sufficiently less than 78 if we use 5-min returns. As the number of intraday observations tends to n, the realized covariance matrix tends to be singular and it should be modeled as a singular matrix (see Alfelt et al. 2021), or it must be regularized to become positive definite (see e.g., Lunde, Shephard, and Sheppard [2016] and Hautsch, Kyj, and Malec [2015] for different ways). Increasing considerably the intra-day sampling frequency, so that the number of intra-day returns is sufficiently larger than the number of assets, is not a recommended solution, because as n gets large it is difficult to have enough synchronized observations to avoid the Epps effect (a bias toward zero in covariation statistics). Moreover, if the sampling frequency is too high, the realized covariance matrices become dominated by microstructure noise.

6.1 Empirical Results with 100 Assets

In this section, we apply the proposed models to the data set of 100 stocks traded on NYSE created and analyzed by Vassallo, Buccheri, and Corsi (2021). The realized covariance matrices are computed using the multivariate realized kernel estimator of Barndorff-Nielsen et al. (2011), which is positive definite, robust to microstructure noise, and deals with unsynchronized data by the "Refresh Time" approach. The time span ranges from January 3, 2006 to December 31, 2014 (2265 observations). Table 15 in Online SA IV reports some global summary statistics of realized variances, covariances, and correlations.

We use the first 2013 observations to estimate the models, leaving the last 252 observations (the last year of the data set) as out-of-sample set. In the estimation of COR models, we obtain insignificant, very close to zero, ϕ_A coefficients, so that, within this family, we do not consider the Pt and Rt specifications. Moreover, COR-R1 does not outperform COR-S in terms of AIC and BIC, so only the COR-S model is kept for the COR models. Hence, we compare the out-of-sample performance of the following 12 models: COR-S, COV-S, COV-S-Pt, COV-S-Rt, COV-R1, COV-R1-Pt, COV-R1-Rt, EWMA, vHAR, HE-vHAR, HAR-DRD, and HE-HAR-DRD. In all models with HE functions, we detect only $\boldsymbol{A}_t$ as time-varying matrix, both for Pt and Rt specifications. For the R1 versions, we have grouped the assets using the agglomerative clustering method described in Section 4.2. We obtain six groups containing 21, 16, 2, 59, 1, and 1 assets. The estimation results are reported in Tables 16 and 17 in Online SA IV. The estimated ϕ_A coefficients are positive and significant in the COV models. The R1 versions improve the fit with respect to the corresponding scalar models. In the benchmark models, ϕ_A is estimated to be negative, being insignificant in the HE-vHAR model but significant in the HE-HAR-DRD model.

In Table 10, we show the results of the forecast comparisons of the 12 models for the statistical loss functions. Estimates are refreshed every 25-th observation (as explained in Section 3.2). At horizon 1, the MCS for the QLIK loss function consists of the R1 COV models that use the HE extension, and for the FN loss, of all COR and COV models, plus the HE-vHAR model. At horizon 5, the MCS for QLIK includes only COV-S, while for the FN loss, the scalar COV models using the HE extension are also in the MCS, together with COV-R1-Rt and EWMA. At horizon 22, vHAR defines the MCS for QLIK and EWMA for FN.

Table 10 Out-of-sample forecast analysis of 12 models for 100 stocks: 95% MCSs for two loss functions and three forecast horizons h

Model	QLIK			FN		
COR-S				1		
COV-S		5		1	5	
COV-S-Pt				1	5	
COV-S-Rt				1	5	
COV-R1				1		
COV-R1-Pt	1			1		
COV-R1-Rt	1			1	5	
EWMA					5	22
vHAR			22			
HE-vHAR				1		
HAR-DRD						
HE-HAR-DRD						

Notes: The symbol 1 identifies the models belonging to the best set for $h=1$; the symbol 5 for $h=5$; and the symbol 22 for $h=22$. The models are defined in Table 3.

Table 11 reveals that the MCS includes the EWMA and three scalar COV models for the three forecast horizons. The three R1-COV models are also in the MCS at horizon 1 and the COR-S at horizon 22. The models that are not in the MCS provide higher portfolio standard deviations than the included models.

7 Concluding Remarks

In the family of CAW models for realized covariance matrices, we have proposed a new class of models, with the characteristics of a greater flexibility than previous models, providing the possibility to estimate different and changing dynamics for each element of the realized covariance matrix. The new models just add one parameter with respect to their classical versions. This is obtained thanks to parameterizations based on the Hadamard exponential matrix function, which possesses the useful property to guarantee the positive definiteness of the matrix of parameters. The HE-CAW models show, in most cases, a better in-sample fit and out-of-sample forecasting performance than the classical CAW models, both in the COV (BEKK-type) version and the COR (DCC-type) one. The COR models have the advantage to be estimable in two steps, which is useful since they are heavily parameterized for a large number of assets. This heaviness occurs because the dynamic variance processes have asset-specific parameters, which is clearly an advantage in terms of fitting quality.

We have proposed a parameterization of the HE term of the new models based on lagged realized or conditional correlations, but in principle any positive definite matrix $\boldsymbol{M}_t$ in Equation (9) can be used provided it can be justified by an economic argument. The models can be extended to include so-called asymmetric effects, whereby the impact of the lagged variance on the next conditional variance is stronger when the lagged return is negative, while

Table 11 Annualized standard deviations and MCS of out-of-sample h-step ahead forecasts of GMVP returns of 12 models for 100 stocks

Model	$h=1$	$h=5$	$h=22$
COR-S	5.35	5.41	**5.10**
COV-S	**5.10**	**5.23**	**5.32**
COV-S-Pt	**5.09**	**5.18**	**5.21**
COV-S-Rt	**5.09**	**5.19**	**5.21**
COV-R1	**5.15**	5.45	5.60
COV-R1-Pt	**5.13**	5.35	5.53
COV-R1-Rt	**5.13**	5.34	5.50
EWMA	**5.14**	**5.14**	**4.96**
vHAR	5.32	5.51	5.46
HE-vHAR	5.31	5.50	5.46
HAR-DRD	5.60	5.54	5.43
HE-HAR-DRD	5.63	5.55	5.43

Note: Bold values identify the models forming the 95% MCS for the GMVP variance loss function defined in Section 3.2, see footnote 9.

the same holds for a covariance when both lagged returns are negative. For COV models, this asymmetric effect is captured by adding the term $\boldsymbol{G} \odot \boldsymbol{d}_{t-1}\boldsymbol{d}'_{t-1} \odot \boldsymbol{C}_{t-1}$ to Equation (11), where $d_{i,t-1} = 1$ if the daily return $r_{i,t-1}$ is negative and $d_{i,t-1} = 0$ if it is positive.

The HE-CAW models could be useful as alternatives to the vech-HAR models used by Hautsch, Kyj, and Malec (2015) and deBrito, Medeiros, and Ribeiro (2018) as building blocks of their factor models. One advantage of HE-CAW models is that they yield positive-definite forecasts by construction. Likewise, the HE-CAW models could be used to forecast the h-step ahead (for $h > 1$) realized covariance matrix needed in HEAVY models (see Noureldin, Shephard, and Sheppard 2012).

The HE parameterization can also be used in other models than CAW. We have done this for benchmark HAR-type models, and the same could perhaps be done in other models, such as score-driven models developed by Gorgi et al. (2019) and Vassallo, Buccheri, and Corsi (2021).

Supplemental Data

Supplemental data are available at https://www.datahostingsite.com.

Acknowledgments

We thank the associate editor and referees for their constructive reports on the manuscript. We have also benefited from the insightful comments of Alessandro Palandri (University of Firenze) when we presented the paper at the Ph.D. Seminar “Economics, Management and Statistics” of the University of Messina (October 25, 2020). We also thank Lyudmila Grigoryeva (University of Konstanz) and Juan-Pablo Ortega (University of St.Gallen) for providing the data sets of realized covariance matrices for the 29 DJIA companies and Oleksandra Kukharenko (University of Konstanz) for handling the data and computing the realized covariance matrices. Finally, we

thank the authors of Vassallo, Buccheri, and Corsi (2021) for providing their data set of realized covariance matrices for the 100 stocks.

References

Aielli, G. P. 2013. Dynamic Conditional Correlation: On Properties and Estimation. *Journal of Business & Economic Statistics* 31: 282–299.

Aït-Sahalia, Y., and D. Xiu. 2017. Using Principal Component Analysis to Estimate a High Dimensional Factor Model with High-Frequency Data. *Journal of Econometrics* 201: 384–399.

Alfelt, G., T. Bodnar, F. Javed, and J. Tyrcha. 2021. "Singular Conditional Autoregressive Wishart Model for Realized Covariance Matrices." Working Paper 1/2021, Örebro University, School of Business.

Anderberg, M. R. 1973. *Cluster Analysis for Applications*. New York: Academic Press.

Archakov, I., and P. R. Hansen. 2020. "A Canonical Representation of Block Matrices with Applications to Covariance and Correlation Matrices." *arXiv* 2012.02698v2 [econ.EM].

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: 149–169.

Bauwens, L., M. Braione, and G. Storti. 2016. Forecasting Comparison of Long Term Component Dynamic Models for Realized Covariance Matrices. *Annals of Economics and Statistics* 123/124: 103–134.

Bauwens, L., S. Laurent, and J. Rombouts. 2006. Multivariate GARCH Models: A Survey. *Journal of Applied Econometrics* 21: 79–109.

Bauwens, L., and E. Otranto. 2020a. "Modelling Realized Covariance Matrices: A Class of Hadamard Exponential Models." CORE Discussion Paper 2020/34, Université catholique de Louvain.

Bauwens, L., and E. Otranto. 2020b. Nonlinearities and Regimes in Conditional Correlations with Different Dynamics. *Journal of Econometrics* 217: 496–522.

Bauwens, L., G. Storti, and F. Violante. 2012. "Dynamic Conditional Correlation Models for Realized Covariance Matrices." CORE Discussion Paper 2012-60, Université catholique de Louvain.

Bollerslev, T. 1986. Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* 31: 307–327.

Bollerslev, T., M. Medeiros, A. Patton, and R. Quaedvlieg. 2021. From Zero to Hero: Realized Partial (Co)variances. *Journal of Econometrics* (https://doi.org/10.1016/j.jeconom.2021.04.013).

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: 140–158.

Cappiello, L., R. F. Engle, and K. Sheppard. 2006. Asymmetric Dynamics in the Correlations of Global Equity and Bond Returns. *Journal of Financial Econometrics* 4: 537–572.

Chiriac, R., and V. Voev. 2011. Modelling and Forecasting Multivariate Realized Volatility. *Journal of Applied Econometrics* 26: 922–947.

Corsi, F. 2009. A Simple Approximate Long-Memory Model of Realized Volatility. *Journal of Financial Econometrics* 7: 174–196.

De Nard, G., R. Engle, O. Ledoit, and M. Wolf. 2021a. Large Dynamic Covariance Matrices: Enhancements Based on Intraday Data. Available at SSRN:https://ssrn.com/abstract=3662143.

De Nard, G., O. Ledoit, and M. Wolf. 2021b. Factor Models for Portfolio Selection in Large Dimensions: The Good, the Better and the Ugly. *Journal of Financial Econometrics* 19: 236–257.

deBrito, D. S., M. C. Medeiros, and R. M. Ribeiro. 2018. Forecasting Large Realized Covariance Matrices: The Benefits of Factor Models and Shrinkage. Available at SSRN:https://ssrn.com/abstract=3163668.

Engle, R. 2002. Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models. *Journal of Business & Economic Statistics* 20: 339–350.

Engle, R., and F. Kroner. 1995. Multivariate Simultaneous Generalized ARCH. *Econometric Theory* 11: 122–150.

Engle, R., O. Ledoit, and M. Wolf. 2019. Large Dynamic Covariance Matrices. *Journal of Business & Economic Statistics* 37: 363–375.

Engle, R., V. Ng, and M. Rothschild. 1990. Asset Pricing with a Factor ARCH Covariance Structure: Empirical Estimates for Treasury Bills. *Journal of Econometrics* 45: 213–238.

Engle, R., and K. Sheppard. 2001. "Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH." Mimeo, UCSD.

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: 489–503.

Golosnoy, V., B. Gribisch, and R. Liesenfeld. 2012. The Conditional Autoregressive Wishart Model for Multivariate Stock Market Volatility. *Journal of Econometrics* 167: 211–223.

Gorgi, P., P. Hansen, P. Janus, and S. Koopman. 2019. Realized Wishart-GARCH: A Score-Driven Multi-asset Volatility Model. *Journal of Financial Econometrics* 17: 1–32.

Hafner, C., and P. Franses. 2009. "A Generalized Dynamic Conditional Correlation Model: Simulation and Application to Many Assets." DP 0904, Institut de Statistique, Université catholique de Louvain.

Hansen, P., A. Lunde, and J. 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: 453–497.

Hautsch, N., L. Kyj, and P. Malec. 2015. Do High-Frequency Data Improve High-Dimensional Portfolio Allocations? *Journal of Applied Econometrics* 30: 263–290.

Ke, Y., H. Lian, and W. Zhang. 2022. High Dimensional Dynamic Covariance Matrices with Homogeneous Structure. *Journal of Business & Economic Statistics* 1: 96–110.

Laurent, S., J. V. Rombouts, and F. Violante. 2013. On Loss Functions and Ranking Forecasting Performances of Multivariate Volatility Models. *Journal of Econometrics* 173: 1–10.

Lunde, A., N. Shephard, and K. Sheppard. 2016. Econometric Analysis of Vast Covariance Matrices Using Composite Realized Kernels and Their Application to Portfolio Choice. *Journal of Business & Economic Statistics* 34: 504–518.

Noureldin, D., N. Shephard, and K. Sheppard. 2012. Multivariate High-Frequency-Based Volatility (HEAVY) Models. *Journal of Applied Econometrics* 27: 907–933.

Oh, D., and A. Patton. 2016. High-Dimensional Copula-Based Distributions with Mixed Frequency Data. *Journal of Econometrics* 193: 349–366.

Patton, A., and K. Sheppard. 2009. "Evaluating Volatility and Correlation Forecasts." In T. Andersen, R. Davis, J.-P. Kreiss, and T. Mikosch (eds.), *Handbook of Financial Time Series*. Springer Verlag.

Patton, A. J. 2011. Volatility Forecast Comparison Using Imperfect Volatility Proxies. *Journal of Econometrics* 160: 246–256.

Sheppard, K., and W. Xu. 2019. Factor High-Frequency-Based Volatility (HEAVY) Models. *Journal of Financial Econometrics* 17: 33–65.

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: 569–586.