CONTRIBUTI DI RICERCA CRENOS

MODELLING REALIZED COVARIANCE MATRICES: A CLASS OF HADAMARD EXPONENTIAL MODELS

Luc Bauwens
Edoardo Otranto

WORKING PAPERS

2020/07

arkadia

CENTRO RICERCHE ECONOMICHE NORD SUD
(CRENOS)
UNIVERSITÀ DI CAGLIARI
UNIVERSITÀ DI SASSARI

CRENOS was set up in 1993 with the purpose of organising the joint research effort of economists from the two Sardinian universities (Cagliari and Sassari) investigating dualism at the international and regional level. CRENoS' primary aim is to improve knowledge on the economic gap between areas and to provide useful information for policy intervention. Particular attention is paid to the role of institutions, technological progress and diffusion of innovation in the process of convergence or divergence between economic areas. To carry out its research, CRENoS collaborates with research centres and universities at both national and international level. The centre is also active in the field of scientific dissemination, organizing conferences and workshops along with other activities such as seminars and summer schools.
CRENoS creates and manages several databases of various socio-economic variables on Italy and Sardinia. At the local level, CRENoS promotes and participates to projects impacting on the most relevant issues in the Sardinian economy, such as tourism, environment, transports and macroeconomic forecasts.

www.crenos.unica.it
crenos@unica.it

CRENoS – CAGLIARI
VIA SAN GIORGIO 12, I-09124 CAGLIARI, ITALIA
TEL. +39-070-6756397; FAX +39-070- 6756402

CRENoS - SASSARI
VIA MURONI 25, I-07100 SASSARI, ITALIA
TEL. +39-079-213511

Title: MODELLING REALIZED COVARIANCE MATRICES: A CLASS OF HADAMARD EXPONENTIAL MODELS

Prima Edizione: Novembre 2020

ISBN: 978 88 68513 283

Viale Bonaria 98 - 09125 Cagliari
Tel. 070/6848663 - info@arkadiaeditore.it
www.arkadiaeditore.it

Modelling Realized Covariance Matrices: a Class of Hadamard Exponential Models

Luc Bauwens[+]
Université catholique de Louvain, CORE
Edoardo Otranto[±]
University of Messina & CRENoS

Abstract

Time series of realized covariance matrices can be modelled 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. An empirical study on twenty-nine assets reveals that the extended models have superior forecasting performances than their simpler versions.

Keywords: realized covariances, dynamic covariances and correlations, Hadamard exponential matrix.

Jel Classification: C32, C58.

[+] Université catholique de Louvain, CORE, Louvain-La-Neuve, Belgium. E-mail: luc.bauwens@uclouvain.be

[±] University of Messina, Dipartimento di Economia, and CRENoS. E-mail: eotranto@unime.it.

1 Introduction

The dynamic modelling of conditional covariance matrices is the topic of a large number of contributions in financial econometrics. The literature started by extending the univariate GARCH model (Bollerslev, 1996) to the multivariate case, developing progressively the family of MGARCH models (for a review, see, e.g., Bauwens et al., 2006). Due to the availability of intraday prices and the development of realized volatility measures, attention shifted to the dynamic modelling of realized covariances and correlations. This has resulted in new classes of models for positive definite matrices, such as the Conditional Autoregressive Wishart (CAW) models proposed by Golosnoy et al. (2012).

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 BEKK (for covariances) and DCC (for correlations) formulations of conditional processes: BEKK in MGARCH was introduced by Engle and Kroner (1995) and adapted to CAW by Golosnoy et al. (2012); DCC was introduced in MGARCH by Engle (2002) and extended to CAW by Bauwens et al. (2012b, 2016).

In this work, we evaluate empirically the merits of modelling realized covariance matrices through correlations and variances or through covariances and variances. We perform this analysis using realized covariance daily data for twenty-nine assets and introducing a set of new parameterizations of the CAW model family which extend the existing BEKK-type formulation of Golosnoy et al. (2012) and the DCC-type formulation of Bauwens et al. (2102b). 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-1 BEKK and DCC versions, while adding a single scalar parameter to these models, hence, they remain parsimonious.[1]

To illustrate the idea, we refer to equation (2.3), a diagonal BEKK-type CAW process for the conditional covariance matrix $\boldsymbol{S}_t$, and we consider the term $\boldsymbol{A} \odot \boldsymbol{C}_{t-1}$ where $\boldsymbol{C}_{t-1}$ is the lagged realized covariance matrix. One of the proposed new models parameterizes the (i,j)-element of the impact matrix $\boldsymbol{A}$ as $a \exp(\phi_A r_{ij,t-1}) / \exp(\phi_A)$, where a and ϕ_A are positive scalar parameters, and $r_{ij,t-1}$ is the lagged realized correlation for the asset pair (i,j). If ϕ_A is set to zero, a scalar model is obtained because each element of $\boldsymbol{A}$ is then equal to a. When ϕ_A is strictly positive, the off-diagonal elements of $\boldsymbol{A}$ differ because the lagged realized correlations differ between asset pairs, hence contrary to the simple scalar version, the dynamics of the covariances are asset-pair specific and the coefficients representing the impacts of the lagged conditional covariances on the next conditional covariances are time-varying.

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

[1] Alternatively, a conditional covariance or correlation process can be parsimoniously parameterized by assuming that a small number of factors drive its dynamics: see Engle et al. (1990) in the MGARCH case, and Sheppard and Xu (2019) for realized covariances.

or 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 (2020) in the MGARCH framework to develop DCC models where the conditional correlations also have asset pair-specific and time-varying dynamics. We adapt these models to the CAW framework. 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 nonscalar formulations. We adopt the approximate solution proposed by Hafner and Franses (2009), which uses an average of the unknown parameters in the targeting estimator. With our dataset of 29×29 covariance matrices, it turns out to be impossible to estimate the robust standard errors, but this evaluation is possible when we consider more parsimonious models. The latter are obtained by a testing procedure that eventually reduces the number of parameters of the variance processes, resulting in a small number of groups of assets that have 'identical' (ie, not significantly different) variance parameters.

An empirical exercise compares the forecasting performance of the CAW models in their BEKK- and DCC-type simple versions with their HE-CAW extended versions, using two statistical loss functions (QLIK, a quasi-likelihood loss, and FN, a Frobenius norm) and an economic one (GMVP, the variance of a global minimum variance portfolio). The results reveal that the DCC-type models have smaller in–sample and out–of–sample one-step ahead forecast losses than the BEKK-type models, whereas the opposite occurs when the forecasting horizon increases. In estimation results, the parameters of the HE-CAW models which imply time-varying and pair-specific impact coefficients are statistically significant at conventional levels. In terms of forecast comparisons, most of these flexible models produce smaller (or in some cases non-superior) losses than the simpler models, especially for the FN and GMVP loss functions.

The rest of the paper is organized as follows: the CAW modelling framework is defined in the next section. The HE-CAW models are defined in Section 3. The data set is described in Section 4. Section 5 presents the empirical results. Final remarks conclude the paper.

2 CAW Modelling 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:

$$\boldsymbol{C}_t|\mathcal{I}_{t-1} \sim W_n\left(\nu, \boldsymbol{S}_t/\nu\right), \tag{2.1}$$

where $\boldsymbol{S}_t$, of dimension $n \times n$, is the (positive semidefinite) 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 $\boldsymbol{\theta}$):

$$l(\boldsymbol{\theta}|\, C_1, \ldots, C_T) = -\frac{\nu}{2} \sum_{t=1}^{T} \left\{\log|S_t(\boldsymbol{\theta})| + \operatorname{trace}\left[S_t(\boldsymbol{\theta})^{-1}\, C_t\right]\right\}. \tag{2.2}$$

The parameter ν does not affect in (2.2) the estimation of $\boldsymbol{\theta}$, so that it can be set equal to 1.

2.1 BEKK-type Models

There is a large set of options to specify $\boldsymbol{S}_t$, inspired by the MGARCH literature; for example, Golosnoy et al. (2012) adopt a BEKK model. The diagonal BEKK-CAW model is the following process:

$$\boldsymbol{S}_t = \boldsymbol{C} + \boldsymbol{A} \odot \boldsymbol{C}_{t-1} + \boldsymbol{B} \odot \boldsymbol{S}_{t-1}, \tag{2.3}$$

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 (co)variance $\boldsymbol{S}_t$ depends only on the corresponding lagged conditional (co)variance and realized (co)variance, not on other (co)variances. The number of parameters in (2.3), equal to $3n(n+1)/2$, renders ML estimation unfeasible already for $n \geq 5$. The number of parameters is reduced by considering the covariance targeting version of (2.3), defined by setting $\boldsymbol{C} = (\boldsymbol{J}_n - \boldsymbol{A} - \boldsymbol{B}) \odot \bar{\boldsymbol{C}}$, where $\boldsymbol{J}_n$ is an $n \times n$ matrix of ones and $\bar{\boldsymbol{C}}$ 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 Frances (2009), replaces the possibly non-positive definite matrix $\boldsymbol{J}_n - \boldsymbol{A} - \boldsymbol{B}$ by the scalar $1 - \bar{a} - \bar{b} \in (0, 1)$, where $\bar{a}$ and $\bar{b}$ are the averages of the elements in matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively.

The scalar version of (2.3), 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-1 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 $\boldsymbol{b} = (b_1^{1/2}, b_2^{1/2} \ldots, b_n^{1/2})'$. The advantage of this parameterization is to have different dynamics for the elements of the conditional covariance matrix, even if they are derived from the parameters of the variances.

We propose another diagonal BEKK model, the EO (Equal Off-diagonal) one, where $\boldsymbol{A}$ (or $\boldsymbol{B}$) is a matrix with diagonal elements $a_1, \ldots, a_n$ (or $b_1, \ldots, b_n$), and equal off-diagonal elements, all equal to a_c (or b_c). The constraint $0 \leq a_c \leq a_i < 1$ (or $0 \leq b_c \leq b_i < 1$) for each $i = 1, ...n$ provides a sufficient condition for $\boldsymbol{A}$ (or $\boldsymbol{B}$) to be positive semidefinite, but it is very restrictive and can be relaxed. See the proof and discussion in Appendix A. For practical targeting, $\bar{a}$ and $\bar{b}$ are set at the averages of the elements of $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively, for example $\bar{a} = (a_1 + a_2 + \cdots + a_n + n(n-1)a_c)/n^2$.

2.2 DCC-type Models

Bauwens et al. (2012b, 2016) specify $\boldsymbol{S}_t$ in the CAW framework using the DCC formulation of MGARCH. They name the model "Re–cDCC". 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):

$$S_{ii,t} = (1 - \alpha_i - \beta_i)\bar{C}_{ii} + \alpha_i C_{ii,t-1} + \beta_i S_{ii,t-1}, \tag{2.4}$$

where $\bar{C}_{ii}$, $S_{ii,t}$ and $C_{ii,t}$ represent the $i-$th element of the diagonal of matrices $\bar{\boldsymbol{C}}$, $\boldsymbol{S}_t$ and $\boldsymbol{C}_t$ respectively.

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 et al. (2012b), the Wishart log–likelihood can be split into two parts (excluding a constant part):

$$\begin{array}{l} l(\boldsymbol{\theta}|C_1, \ldots, C_T) = l_v(\boldsymbol{\theta}_v) + l_c(\boldsymbol{\theta}_c) \\ l_v(\boldsymbol{\theta}_v) = -\frac{\nu}{2}\left[\log|\boldsymbol{D}_t^2| + \text{trace}\left(\boldsymbol{D}_t^{-1}\boldsymbol{C}_t\boldsymbol{D}_t^{-1}\right)\right] = -\frac{\nu}{2}\left[\sum_{i=1}^n \log(S_{ii,t}) + \sum_{i=1}^n S_{ii,t}^{-1}C_{ii,t}\right] \\ l_c(\boldsymbol{\theta}_c|\boldsymbol{\theta}_v) = -\frac{\nu}{2}\left\{\log\left|\boldsymbol{D}_t^{-1}\boldsymbol{S}_t\boldsymbol{D}_t^{-1}\right| + \text{trace}\left[\left(\boldsymbol{D}_t\boldsymbol{S}_t^{-1}\boldsymbol{D}_t - \boldsymbol{I}_n\right)\boldsymbol{D}_t^{-1}\boldsymbol{C}_t\boldsymbol{D}_t^{-1}\right]\right\}, \end{array} \tag{2.5}$$

where $\boldsymbol{I}_n$ is the $(n \times n)$ identity matrix, $\boldsymbol{D}_t$ is a diagonal matrix with elements $S_{ii,t}^{1/2}$ and $S_{ii,t}$ are the diagonal elements of $\boldsymbol{S}_t$, and $C_{ii,t}$ are the diagonal elements of $\boldsymbol{C}_t$. 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 in a first step. The parameters relative to the correlation part, $\boldsymbol{\theta}_c$, can be estimated in a second step by maximizing $l_c(\boldsymbol{\theta}_c|\boldsymbol{\theta}_v)$, conditional on the estimator of $\boldsymbol{\theta}_v$ obtained in the first step. Like for (2.2), ν does not affect the estimation of θ and can be set to 1. Because the term $\text{trace}\left(\boldsymbol{D}_t^{-1}\boldsymbol{C}_t\boldsymbol{D}_t^{-1}\right)$ appearing in the expression of $l_c(\boldsymbol{\theta}_c|\boldsymbol{\theta}_v)$ does not depend on $\boldsymbol{\theta}_c$, 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 $\boldsymbol{D}_t^{-1}\boldsymbol{C}_t\boldsymbol{D}_t^{-1}$, with parameters ν and $\boldsymbol{D}_t^{-1}\boldsymbol{S}_t\boldsymbol{D}_t^{-1}$.

The Re–cDCC model of $\boldsymbol{R}_t$, in its diagonal version, is defined as the following set of equations:

$$\begin{array}{l} \boldsymbol{R}_t = \tilde{\boldsymbol{Q}}_t^{-1/2}\boldsymbol{Q}_t\tilde{\boldsymbol{Q}}_t^{-1/2}, \\ \boldsymbol{Q}_t = \boldsymbol{Q} + \boldsymbol{A} \odot \left(\tilde{\boldsymbol{Q}}_{t-1}^{1/2}\boldsymbol{D}_{t-1}^{-1}\boldsymbol{C}_{t-1}\boldsymbol{D}_{t-1}^{-1}\tilde{\boldsymbol{Q}}_{t-1}^{1/2}\right) + \boldsymbol{B} \odot \boldsymbol{Q}_{t-1}, \\ \tilde{\boldsymbol{Q}}_t = \text{diag}(\boldsymbol{Q}_t), \end{array} \tag{2.6}$$

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

It is interesting to notice that the EO parameterization for the BEKK formulation (2.3) is in the spirit of the scalar version of the Re-cDCC model, defined by (2.4) and (2.6) with $\boldsymbol{A} = a\boldsymbol{J}_n$ and $\boldsymbol{B} = b\boldsymbol{J}_n$: firstly, a univariate model with specific parameters a_i and b_i holds for each variance, like for the variances in (2.4); secondly, a common dynamic model holds for the covariances, with coefficient a_c for $C_{ij,t-1}$ and b_c for $S_{ij,t-1}$. This is a scalar structure for the covariances, similar to the scalar Re-cDCC structure for the correlations. Another difference is the scalar used for targeting, which for the EO model is common to all the variances and covariances and equal to $1-\bar{a}-\bar{b}$, whereas it is $1-\alpha_i-\beta_i$ for the variances and $1-a-b$ for the correlations in the scalar Re-DCC model. Differently from Re-cDCC, the estimation of the EO model must be performed in one step, but it is feasible for moderate values of n since the number of parameters in $(a_1,\ldots,a_n,b_1,\ldots,b_n,a_c,b_c)$ is equal to $2n+2$.

Model names: The following acronyms are used in the rest of the paper: CAW for all models falling in the framework defined in this section. The CAW model class contains two families: the first one uses BEKK-type processes for the dynamics of $\boldsymbol{S}_t$, such as (2.3) and (3.3): they are named COV models. The second family uses univariate processes for the conditional variances (the diagonal elements of $\boldsymbol{S}_t$) and DCC-type processes for the conditional correlation matrix $\boldsymbol{R}_t$, such as (2.6) and (3.4); they are named COR models.

3 Hadamard Exponential CAW Models

A clear disadvantage of the parameterizations of the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ in (2.3) and (2.6) is either that they are too heavy for large n, or that they lack flexibility when they are of scalar or rank-1 type: the scalar version imposes the same dynamics for all the variances and covariances, whereas the rank-1 version imposes that the covariances depend on the product of the corresponding parameters of the variances. Bauwens and Otranto (2020), in the framework of MGARCH conditional correlation models, provide extensions of the scalar DCC model of Engle (2002), where the elements of $\boldsymbol{A}$ 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 (2.3) and (2.6), while maintaining a parsimonious parameterization, can be obtained by extending and generalizing the Bauwens and Otranto (2020) 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 (2.3) and of (2.6) are introduced below. In our empirical experiments, we find that only $\boldsymbol{A}$ is time–varying for COV models and only $\boldsymbol{B}$ for COR models. This finding is in contrast with papers which

assume the scalar restriction $\boldsymbol{B} = b\boldsymbol{J}_n$, as Noureldin et al. (2014), Clements et al. (2018), Bauwens and Otranto (2020).

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

We define two parameterizations of $\boldsymbol{A}_t$, which can be applied similarly to $\boldsymbol{B}_t$ (replace a by b, ϕ_A by ϕ_b, and $\boldsymbol{a}$ by $\boldsymbol{b}$). They are:

$$\begin{array}{ll} Sc \text{ (Scalar):} & \boldsymbol{A}_t = a \exp^{\odot}(\phi_A \boldsymbol{M_t}) = a\boldsymbol{J}_n \odot \exp^{\odot}(\phi_A \boldsymbol{M}_t), \\ R1 \text{ (Rank-1):} & \boldsymbol{A}_t = \boldsymbol{a}\boldsymbol{a}' \odot \exp^{\odot}(\phi_A \boldsymbol{M}_t), \end{array} \tag{3.1}$$

where $\phi_A \geq 0$, $\boldsymbol{M}_t$ is a positive definite symmetric matrix known at date t, $a \in (0,1)$ is a scalar in the first parameterization, and $\boldsymbol{a}$ in the second one is the n-dimensional vector $(a_1^{1/2}, \ldots, a_n^{1/2})'$ in which each element is in $(0,1)$. Using the square root of a_i, the coefficient of the lagged variance of the $i-$th conditional GARCH-type variance equation is a_i, like in (2.4).

Notice that if $\phi_A = 0$, $\exp^{\odot}(\phi_A \boldsymbol{M}_t)$ is equal to $\boldsymbol{J}_n$, so that $\boldsymbol{A}_t$ is constant, being equal to $a\boldsymbol{J}_n$ (scalar model) or $\boldsymbol{a}\boldsymbol{a}'$ (rank-1 model). When ϕ_A is strictly positive, 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 $\boldsymbol{A}_t$ when $\phi_A > 0$:

$$\begin{array}{ll} Pt: & \boldsymbol{M}_t = \boldsymbol{P}_{t-1} - \boldsymbol{J}_n, \\ Rt: & \boldsymbol{M}_t = \boldsymbol{R}_{t-1} - \boldsymbol{J}_n, \end{array} \tag{3.2}$$

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 the first line of (2.6).

Each matrix $\boldsymbol{A}_t$ obtained by combining (3.1) and (3.2) is the Hadamard product of a positive definite (Sc case) or semidefinite matrix ($R1$ case) with strictly positive diagonal entries 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, 2020). 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, 2020). 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 $(0,1)$, where $r \in (-1,+1)$ is a correlation coefficient.

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 covari-

ances and correlations. In the COV models, an off-diagonal element of $\boldsymbol{A}_t$ represents the impact coefficient of the corresponding lagged realized covariance ($C_{ij,t-1}$) on the next conditional covariance ($S_{ij,t-1}$), being for example $a_{ij}\exp(\phi_A r_{ij,t-1})/\exp(\phi_A)$, in the $R1$ parameterization, where $a_{ij} = a_i^{1/2}a_j^{1/2}$ and $r_{ij,t-1}$ is the lagged conditional or realized correlation. 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 $r_{ij,t-1}$. Hence, when the lagged correlation increases (due to volatility clustering or an idiosyncratic factor), the next conditional covariance increases (for given values of $a_{ij}C_{ij,t-1}$ 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.[2] 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 (with $\boldsymbol{B}_t$), the impact coefficient $b_{ij}\exp(\phi_B r_{ij,t-1})/\exp(\phi_B)$ represents the impact of the lagged quasi-correlation $\boldsymbol{Q}_{t-1}$ on the next quasi-correlation $Q_{ij,t}$, whereas the impact of $C_{ij,t-1}/(S_{ii,t-1}S_{jj,t-1})^{1/2}$, which is a pseudo-correlation, multiplied by $(Q_{ii,t-1}Q_{jj,t-1})^{1/2}$ (the Aielli type of correction of the DCC model in this context), is constant. The time-varying impact is thus similar to what it is for COV models, but it operates through the quasi-correlation terms.

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 or the other kind of correlation is better adapted to fit the kind of impact embedded via the Hadamard exponential matrix. In our empirical experiments, we find that the lagged realized correlation provides a better fitting and out–of–sample performance.

COR and COV practical equations

The combination of the two parameterizations of $\boldsymbol{A}_t$ for COV (and $\boldsymbol{B}_t$ for COR) with the two choices of $\boldsymbol{M}_t$ when $\phi_A > 0$ ($\phi_B > 0$), and the two cases where $exp^{\odot}(\phi_A \boldsymbol{M}_t)$

[2]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 et al., 2006); in particular, the HE term does not change the impact of $C_{ii,t-1}$ on $S_{ii,t}$.

($exp^{\odot}(\phi_B \boldsymbol{M}_t)$) are equal to $\boldsymbol{J}_n$, provides a family of six COV models (and six COR models).

The COV version of CAW models in the empirical applications of Section 5 is specified as

$$\boldsymbol{S}_t = (1 - \bar{a}_t - \bar{b})\bar{\boldsymbol{C}} + \boldsymbol{A}_t \odot \boldsymbol{C}_{t-1} + \boldsymbol{B} \odot \boldsymbol{S}_{t-1}, \tag{3.3}$$

with $\boldsymbol{A}_t$ as defined in the Sc and $R1$ parameterizations shown in (3.1), or with constant $\boldsymbol{A}$ when $\phi_A = 0$. In the latter case, $\bar{a}_t$ is constant, being equal to a in the scalar model, and to the average of the elements of $\boldsymbol{A} = \boldsymbol{a}\boldsymbol{a}'$ in the rank-1 model. When $\boldsymbol{A}_t$ is time-varying, $\bar{a}_t$, defined the average of its elements, is also time-varying. The matrix $\boldsymbol{B}$ and the scalar $\bar{b}$ are constant, with $\boldsymbol{B} = b\boldsymbol{J}_n$ and $\bar{b} = b$ in the scalar model, or $\boldsymbol{B} = \boldsymbol{b}\boldsymbol{b}'$ and $\bar{b}$ the average of the elements of $\boldsymbol{b}\boldsymbol{b}'$ in the rank-1 model.

For COR models, (2.6) is changed to

$$\boldsymbol{Q}_t = (1 - \bar{a} - \bar{b}_t)\bar{\boldsymbol{R}} + \boldsymbol{A} \odot \left(\tilde{\boldsymbol{Q}}_{t-1}^{1/2} \boldsymbol{D}_{t-1}^{-1} \boldsymbol{C}_{t-1} \boldsymbol{D}_{t-1}^{-1} \tilde{\boldsymbol{Q}}_{t-1}^{1/2}\right) + \boldsymbol{B}_t \odot \boldsymbol{Q}_{t-1}, \tag{3.4}$$

where the two parameterizations defined in (3.1) are applied to $\boldsymbol{B}_t$, and where $\boldsymbol{A}$ is either the scalar matrix $a\boldsymbol{J}_n$ (with $\bar{a} = a$) or the rank-one matrix $\boldsymbol{a}\boldsymbol{a}'$ (with $\bar{a}$ equal to the average of the elements of this matrix). When $\phi_B > 0$, $\boldsymbol{B}_t$ is time-varying, and $\bar{b}_t$ is the average of its elements; when $\phi_B = 0$, $\boldsymbol{B}_t$ is constant, being either $b\boldsymbol{J}_n$ (and $\bar{b}_t = b$), or $\boldsymbol{b}\boldsymbol{b}'$ (and $\bar{b}_t$ is the average of the elements of this matrix).

The conditional variance dynamic equation of the first step of the COR model for each i is specified as (2.4).

Parsimonious parameterizations by asset grouping

To reduce the number of parameters in a given model, we 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 $COV - R1$ ($COR - R1$) model has 58 (116) parameters, but if four groups are formed, this is reduced to 8 (66). 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 estimated asymptotic covariance matrix of the estimator may be singular; secondly, the reduction typically reduces the estimator variance if the restrictions are valid.

To do this for the four *non-scalar* COV models and the three *non-scalar* COR models, we test the joint hypothesis $a_i = a_j$ and $b_i = b_j$ for each pair of assets (i, j) and then we put in 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 highest p-value. For example, if the pair of assets (i, j) is the one with highest p-value (greater than the nominal size of 0.05), and then we find that the hypothesis concerning the pair (i, k) is not rejected, we include the three assets in the same group only if the hypothesis concerning the equality of the coefficients of the pair (j, k) is not rejected. We continue this grouping for all the p-values greater than 0.05. Bauwens and Otranto (2020) use a

very similar algorithm for the DCC MGARCH model and report that it detects suitably the correct number of groups of several simulated data generating processes.

We can also form groups for the *scalar* COR model, by testing jointly $\alpha_i = \alpha_j$ and $\beta_i = \beta_j$, the parameters of the variance processes (2.4), and proceeding as described in the previous paragraph. These parameters can be estimated in the first step of the 2-step procedure, or simultaneously with the parameters of the correlation part in the 1-step procedure, hence the resulting number of groups and group compositions are not necessarily identical.

Stationarity conditions

Golosnoy et al. (2012) provide the covariance stationarity conditions (i.e., the conditions for the existence of the unconditional second-order moments) of the BEKK-type CAW (or COV) stochastic process as a function of the model parameters, for a more general BEKK(1,1) process than in (2.3), i.e., 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 (3.3) with the constant $\boldsymbol{A}$ parameterizations in (3.1), the stationarity condition is $a + b < 1$ in the Sc parameterization, $\max(\boldsymbol{a}\boldsymbol{a}' + \boldsymbol{b}\boldsymbol{b}') < 1$ in $R1$: moreover, in the EO representation of the COV model, we impose $\max(\boldsymbol{A} + \boldsymbol{B}) < 1$, where $\max$ applied to a matrix selects its largest entry. When the HE matrix depending on the lagged realized or conditional correlations is added in these parameterizations, 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 a constant $\boldsymbol{A}$ version, it holds at each t for the corresponding time-varying version. For example, in the Sc case, $a \exp(\phi_A r)/\exp(\phi_A) + b < 1$ 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 (2.4) is $\alpha_i + \beta_i < 1$. For the correlation process (3.4), the stationarity conditions are the same as for the COV parameterizations, and if the stationarity condition holds for a constant $\boldsymbol{B}$ parameterization, it holds for the corresponding time-varying one.

4 The Dataset

To investigate empirically the questions we are interested in, we use a time-series of daily realized covariance matrices computed from a high-frequency dataset for 29 stocks of the Dow Jones Industrial Average (DJIA) index; the 30th stock was dropped since it is not permanently in the index during the sample period. The data source is the TAQ database. The sample period is 3 January 2001–16 April 2018, resulting in 4319 observations. Each daily realized covariance matrix is computed as the sum of the outer products of the one minute (log-)returns of the day. The one minute returns are obtained from synchronized intra-day prices. The synchronization was done globally for the 29 stocks, using one

minute intervals, the price closest (from the left) to the respective sampling point was taken; the first and last 15 minutes of the day (9:30-16:00) were excluded. The data are annualized in percentage (multiplied by 25,200).

The stock names and tickers are: Apple Inc.(AAPL), American Express Company (AXP), The Boeing Company (BA), Caterpillar Inc. (CAT), Cisco Systems, Inc. (CSCO), Chevron Corporation (CVX), The Walt Disney Company (DIS), DowDuPont Inc. (DWDP), General Electric Company (GE), The Goldman Sachs Group, Inc. (GS), Home Depot Inc.(HD), International Business Machines Corporation (IBM), Intel Corporation(INTC), Johnson & Johnson (JNJ), JPMorgan Chase & Co. (JPM), The Coca-Cola Company (KO), McDonald's Corporation (MCD), 3M Company (MMM), Merck &Co., Inc. (MRK), Microsoft Corporation (MSFT), NIKE, Inc. (NKE), Pfizer Inc.(PFE), The Procter & Gamble Company (PG), The Travelers Companies, Inc.(TRV), United Health Group Incorporated (UNH), United Technologies Corporation (UTX), Verizon Communications Inc. (VZ), Walmart Inc. (WMT), Exxon Mobil Corporation (XOM).

Table 1 shows some summary statistics of realized variances, covariances and correlations. The variances have larger (time series) average levels and standard deviations than the covariances. The average variances are more heterogenous across assets than the covariances: the minimum and maximum averages are 2.44 and 9.56 for the variances: for covariances, without averaging over 28 values (i.e., for the 406 average covariances), they are 0.57 and 3.73. The average correlations of each stock with the other stocks fluctuate around 0.22 with little variation. The minimum and maximum average correlations (over the 406 average correlations) are 0.097 and 0.443.

5 Empirical Results

We estimate and evaluate the performance of fourteen models on the dataset for $n = 29$ stocks of the Dow Jones Industrial Average (DJIA) index described in section 4:

- Seven correlation models as in (2.6) modified by (3.4):
  -COR-S-$1s$ $(2n+2)$: the scalar version with $\boldsymbol{A} = a\boldsymbol{J}_n$, $\boldsymbol{B} = b\boldsymbol{J}_n$ and the variance coefficients estimated in one step (simultaneously);
  -COR-S $(2n+2)$: the same model as above, but estimated in two steps;
  -COR-S-Pt $(2n{+}3)$: the scalar version with $\boldsymbol{A} = a\boldsymbol{J}_n$ and $\boldsymbol{B}_t = b\exp^{\odot}[\phi_B(\boldsymbol{P_t} - \boldsymbol{J_n})]$;
  -COR-S-Rt $(2n{+}3)$: the scalar version with $\boldsymbol{A} = a\boldsymbol{J}_n$ and $\boldsymbol{B}_t = b\exp^{\odot}[\phi_B(\boldsymbol{R_t} - \boldsymbol{J_n})]$;
  -COR-$R1$ $(2n + 2n_g)$: the rank-1 version with $\boldsymbol{A} = \boldsymbol{a}\boldsymbol{a}'$ and $\boldsymbol{B} = \boldsymbol{b}\boldsymbol{b}'$;
  -COR-$R1$-Pt $(2n + 2n_g + 1)$: the rank-1 version with $\boldsymbol{A} = \boldsymbol{a}\boldsymbol{a}'$ and $\boldsymbol{B}_t = \boldsymbol{b}\boldsymbol{b}' \odot \exp^{\odot}[\phi_B(\boldsymbol{P_t} - \boldsymbol{J_n})]$;
  -COR-$R1$-Rt $(2n + 2n_g + 1)$: the rank-1 version with $\boldsymbol{A} = \boldsymbol{a}\boldsymbol{a}'$ and $\boldsymbol{B}_t = \boldsymbol{b}\boldsymbol{b}' \odot \exp^{\odot}[\phi_B(\boldsymbol{R_t} - \boldsymbol{J_n})]$.

- Seven covariance models as in (3.3):
  -COV-S (2): the scalar version with $\boldsymbol{A} = a\boldsymbol{J}_n$ and $\boldsymbol{B} = b\boldsymbol{J}_n$;
  -COV-S-Pt (3): the scalar version with $\boldsymbol{A}_t = a\exp^{\odot}[\phi_A(\boldsymbol{P_t} - \boldsymbol{J_n})]$ and $\boldsymbol{B} = b\boldsymbol{J}_n$;
  -COV-S-Rt (3): the scalar version with $\boldsymbol{A}_t = a\exp^{\odot}[\phi_A(\boldsymbol{R_t} - \boldsymbol{J_n})]$ and $\boldsymbol{B} =$

Table 1: Descriptive statistics for the dataset of 29 DJIA Stocks

Stock	RVar		RCov		RCor	
Ticker	Mean	SD	$\overline{\text{Mean}}$	$\overline{\text{SD}}$	$\overline{\text{Mean}}$	$\overline{\text{SD}}$
AAPL	9.557	16.727	1.548	4.683	0.196	0.133
AXP	7.450	17.623	1.641	5.112	0.220	0.136
BA	5.221	8.002	1.332	3.967	0.208	0.136
CAT	5.867	10.893	1.578	5.043	0.229	0.140
CSCO	7.476	12.721	1.656	4.585	0.227	0.136
CVX	4.197	9.972	1.353	5.082	0.224	0.152
DIS	5.665	9.851	1.480	4.266	0.234	0.147
DWDP	4.869	8.259	1.516	4.577	0.245	0.145
GE	6.123	14.871	1.663	5.004	0.247	0.140
GS	8.040	30.538	1.797	6.283	0.242	0.139
HD	5.567	10.353	1.597	4.867	0.245	0.145
IBM	3.652	7.564	1.351	4.066	0.259	0.143
INTC	7.389	10.853	1.770	4.478	0.249	0.140
JNJ	2.405	4.340	0.938	2.865	0.231	0.146
JPM	8.828	21.905	1.978	6.037	0.261	0.145
KO	2.620	4.724	0.991	3.070	0.221	0.141
MCD	3.656	6.542	1.090	3.553	0.215	0.143
MMM	3.369	8.549	1.315	4.024	0.272	0.151
MRK	4.259	9.504	1.171	3.889	0.214	0.144
MSFT	4.909	7.290	1.474	3.864	0.243	0.143
NKE	4.355	6.543	1.189	3.454	0.210	0.142
PFE	4.276	5.930	1.087	3.213	0.195	0.136
PG	2.440	5.779	0.860	3.246	0.201	0.135
TRV	4.716	12.319	1.071	4.265	0.185	0.141
UNH	5.345	9.910	0.955	3.543	0.153	0.125
UTX	3.865	7.334	1.103	4.044	0.206	0.136
VZ	4.208	7.906	1.002	3.833	0.164	0.128
WMT	3.284	6.047	0.884	3.346	0.165	0.124
XOM	4.034	9.946	1.062	4.748	0.177	0.131
Column Mean	5.091	10.441	1.326	4.242	0.218	0.139
Column SD	1.875	5.566	0.310	0.835	0.030	0.007

Columns 2-3: time series mean and standard deviation (SD) of realized variances
Column 4-5: mean of 28 time series means and SD of realized covariances between stock in column 1 and the other 28 stocks
Column 6-7: like in columns 4-5, but for realized correlations

$b\boldsymbol{J}_n$;
-COV-$R1$ ($2n_g$): the rank-1 version with $\boldsymbol{A} = \boldsymbol{a}\boldsymbol{a}'$ and $\boldsymbol{B} = \boldsymbol{b}\boldsymbol{b}'$;
-COV-$R1$-Pt ($2n_g + 1$): the rank-1 version with $\boldsymbol{A}_t = \boldsymbol{a}\boldsymbol{a}' \odot \exp^{\odot}[\phi_A(\boldsymbol{P_t} - \boldsymbol{J_n})]$ and $\boldsymbol{B} = \boldsymbol{b}\boldsymbol{b}'$;
-COV-$R1$-Rt ($2n_g + 1$): the rank-1 version with $\boldsymbol{A}_t = \boldsymbol{a}\boldsymbol{a}' \odot \exp^{\odot}[\phi_A(\boldsymbol{R_t} - \boldsymbol{J_n})]$ and $\boldsymbol{B} = \boldsymbol{b}\boldsymbol{b}'$;
-COV-EO ($2n_g + 2$): The EO version (see Section 2) of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ with equal elements out of the diagonal.

After each model acronym, the number of parameters is indicated between parentheses; n_g is the number of groups with equal coefficients identified by the grouping procedure defined in Section 3. In the COR models we count also the $2n$ coefficients of the first step estimation, where n univariate models as (2.4) for variances are estimated.

5.1 Estimation Results

5.1.1 Groups and synthesis of estimation results for 14 models

Each non-scalar model is estimated in its full parameterization, then the grouping algorithm of the parameters described in Section 3 is applied. The obtained groups are shown in Table 2. The algorithm detects different numbers of groups: 4, 6 or 8 for the COR models, 4 for the COV models. The 'first step' column shows the groups obtained from the first step estimation of the COR models (or, equivalently, from univariate estimations of the variance equations): the number of groups is 5 and the grouping is very similar to the grouping (in 4 groups) obtained from the one-step estimation (COR-S-$1s$). Using the first column as a benchmark, it is interesting to notice that the groups obtained for the COV models are more similar to it than the groups obtained from the second step estimation of the non-scalar COR models; this is due to the fact that the grouping procedure of these COR models does not depend on the parameters of the variance processes, contrary to the procedure for the COV models.

We have estimated the 14 models 1) in their full parametrizations; 2) in their reduced ones, using the grouping shown in Table 2; 3) in reduced parameterizations using for each model the grouping in 5 groups shown in the first column of Table 2 (obtained in the first step estimation of COR models). The third set of estimations is interesting because the grouping is simple and quick to obtain since it does not require the estimation of a full parameterization, which, for large n, is time consuming and may not be feasible. Actually, we were able to get the parameter estimates of all models, but for the full parameterizations of COR-S-$1s$, COR-$R1$-Pt, COR-$R1$-Rt, COV-$R1$-Rt and COV-EO, we could not get a proper Hessian and hence the 'sandwich' (robust) covariance estimator; for this model we use the OPG (outer-product of gradient) covariance estimator. The gradient and Hessian are computed by numerical differentiation (using GAUSS).

In Table 3 we report a synthesis of the three sets of estimations. The table shows for each model the number of estimated parameters (including those of the first step for COR models), the maximized log-likelihood value (MaxLL), and the Akaike and Bayesian information criteria (AIC and BIC).

Table 2: Groups of parameters (identified by the stock ticker) for non-scalar models

		COR Models				COV Models			
Stock	first step	S-1s	$R1$	$R1$-Pt	$R1$-Rt	$R1$	$R1$-Pt	$R1$-Rt	EO
AAPL	1	1	1	1	1	1	1	1	1
AXP	2	2	2	1	2	2	2	2	2
BA	3	3	2	2	1	3	3	3	3
CAT	3	3	1	1	2	3	3	3	3
CSCO	4	4	2	2	1	4	4	4	4
CVX	1	1	1	1	2	1	1	1	1
DIS	1	1	2	1	1	1	1	1	1
DWDP	3	3	1	1	2	1	1	4	3
GE	2	4	2	1	1	2	4	1	4
GS	1	1	1	3	3	1	1	1	1
HD	1	1	2	1	1	4	1	4	1
IBM	1	3	2	2	1	1	1	1	1
INTC	1	1	2	2	1	3	1	4	3
JNJ	3	3	3	4	4	1	1	4	1
JPM	2	2	1	3	3	2	2	2	4
KO	3	3	2	4	4	3	3	3	3
MCD	5	1	2	4	4	3	3	3	4
MMM	1	1	3	4	4	1	1	4	3
MRK	3	3	4	5	5	3	3	3	3
MSFT	3	3	3	6	4	3	3	3	3
NKE	3	3	3	6	4	3	3	3	1
PFE	3	3	5	7	5	3	3	3	3
PG	3	3	3	6	4	1	1	4	3
TRV	4	4	5	7	5	4	4	4	2
UNH	3	3	6	8	6	3	3	3	1
UTX	3	3	3	6	4	3	1	4	3
VZ	5	1	6	8	6	3	3	3	3
WMT	3	3	6	8	6	3	3	3	3
XOM	1	1	3	6	4	1	1	4	3
number of groups	5	4	6	8	6	4	4	4	4

For each model, the group to which a stock belongs is identified by a value from 1 to the number indicated in the last row of the table.

Table 3: Number of estimated parameters (Np), maximized log–Likelihood (MaxLL), Akaike (AIC) and Bayesian (BIC) information criteria of fully parameterized models, reduced models with n_g groups and with the five groups shown in Table 2, column 2

	Full Parameterization				n_g groups					5 groups			
	Np	MaxLL	AIC	BIC	n_g	Np	MaxLL	AIC	BIC	Np	MaxLL	AIC	BIC
COR-S-1s	60	-116916	54.181	54.269	4	10	-116922	54.160	54.175	12	-116922	54.161	54.179
COR-S	60	-116945	54.194	54.282						12	-116949	54.174	54.192
COR-S-Pt	61	-116938	54.191	54.281						13	-116942	54.171	54.190
COR-S-Rt	61	-116940	54.192	54.282						13	-116945	54.172	54.191
COR-$R1$	116	-116935	54.215	54.387	6	70	-116937	54.195	54.298	20	-116948	54.177	54.206
COR-$R1$-Pt	117	-116930	54.214	54.386	8	75	-116932	54.194	54.302	21	-116941	54.174	54.205
COR-$R1$-Rt	117	-116932	54.215	54.387	6	71	-116934	54.194	54.299	21	-116942	54.174	54.205
COV-S	2	-117444	54.398	54.401									
COV-S-Pt	3	-117404	54.380	54.385									
COV-S-Rt	3	-117384	54.371	54.376									
COV-$R1$	58	-117409	54.408	54.494	4	8	-117418	54.389	54.401	10	-117420	54.391	54.406
COV-$R1$-Pt	59	-117385	54.397	54.484	4	9	-117389	54.376	54.389	11	-117390	54.378	54.394
COV-$R1$-Rt	59	-117372	54.391	54.478	4	9	-117372	54.368	54.381	11	-117373	54.370	54.386
COV-EO	60	-117177	54.302	54.390	4	10	-117183	54.281	54.296	12	-117187	54.284	54.302

It is clear that each full parameterization provides a MaxLL value that hardly improves that of the corresponding reduced model and, as a consequence, the reduced models have higher AIC and BIC values. For example, COR-$R1$-Pt has 117 parameters in the full parameterization, but its MaxLL is only 2 points greater than in the model with 8 groups and 75 parameters, and 11 points greater than the 5 groups model with 21 parameters. In general, for the COR models, each 5-group model identified from the variance parameters is a valid alternative to the same model with groups identified from the correlation parameters: the AIC and BIC are lower, except for COR-S-1s model (where the difference is 0.001 for AIC and 0.004 for BIC), which identifies the groups again from the variance parameters. The opposite is true for COV models: the models with 5 groups have higher AIC and BIC than the models with $n_g = 4$ groups, but the differences between the two reduced alternatives are rather small. We conclude that in practice, it is acceptable to use the reduced models obtained by the grouping based on the variance parameters, which is good news since this type of grouping is feasible even with a large number of assets.

In the set of COR models, the COR-S-1s models have much higher MaxLL than the other models. The two restricted COR-S-1s models are acceptable reductions of the unrestricted one according to likelihood ratio (LR) chi-square statistics (both equal to 12, with 56 or 48 degrees of freedom); this is confirmed by their lowest AIC and BIC. Among the models estimated in two steps, the unrestricted COR-$R1$-Pt has the highest MaxLL. LR tests do not reject each unrestricted S version against the corresponding $R1$ version, all p-values being larger than 0.99; considering the 5-group models, the same holds, except that $S - Rt$ is rejected against $R1$-Rt at the 7% level (p-value = 0.067). Interestingly, each model with constant $\boldsymbol{B}$ matrix is rejected against the corresponding HE model, the largest p-value being 0.047 for S versus S-Rt in the 5-group case, all the other being less than 0.014.

Among the COV models, the three COV-EO models have much higher MaxLL and lower AIC and BIC than the other models. COV-EO can be considered the COV version of COR-S-1s in term of specification and estimation, since both have specific parameters for each variance process, and equal parameters for the covariances in the case of COV and the correlations in the case of COR; moreover, they are estimated in

one step. Nevertheless, COV-EO is dominated by COR-S-1s in term of AIC and BIC.

Concerning the comparison of COV-$R1$ and COV-S models, the result is quite different from what it is for COR models: each S model is *rejected* against the corresponding $R1$ and against EO, both for unrestricted models as for 5-group ones (all p-values being less the 0.001, except one that is equal to 0.10 for unrestricted S against unrestricted $R1$). Like for COR models, each model with constant $\boldsymbol{A}$ matrix is rejected against the corresponding HE model, all p-values being less than 0.001.

Globally, COR models have lower AIC and BIC than COV models. The MaxLL values of COV and COR models are not comparable because the models have different parameterizations, in particular of the constant terms of the variance processes – compare (2.4) and (3.3). Other model comparisons, in terms of forecasting performances, are presented in subsections 5.2 and 5.3.

5.1.2 Estimation results for COR and COV models with 5 groups

COR models

The estimation results for the scalar COR models and the rank-one ($R1$) models with 5 groups are reported in Table 4. The upper part of the Table shows the parameters of the variance equations estimated with the COR-S-1s model and in the first step of 2-step estimation for the other models. Each lagged realized variance impact coefficient (α_i) is smaller in joint estimation than in univariate estimation (with differences between 0.08 and 0.04), whereas each lagged conditional variance coefficient (β_i) is larger (with differences between 0.09 and 0.04), hence the estimates of the persistence effect ($\alpha_i + \beta_i$) is almost the same in both estimations. One can also notice the smaller standard errors of the 1-step estimation.

Considering the correlation parameters (bottom part of the Table), the differences are small within the scalar models and within the $R1$ models, and they are slightly stronger between them. The small but significant[3] estimates of ϕ_B imply a time-varying impact coefficient for the lagged $\boldsymbol{Q}_t$ matrix in (3.4); in particular the impact coefficient varies between 0.864 and 0.876 in the COR-S-Pt model and between 0.866 and 0.875 in the COR-S-Rt model. For the $R1$ models, we can notice slightly different coefficients for the different groups; in particular for COR-$R1$-Pt (COR-$R1$-Rt) the b_i minimum coefficient is 0.867 (0.866) for group 2, and the maximum is 0.886 (0.885) in group 4. Actually, the scalar model (estimated in two steps) is not rejected against any $R1$ version: the largest likelihood ratio statistic between $R1$ and S occurs for $R1$-Pt and is about 6, for 9 degrees of freedom (see last column of Table 3 for MaxLL values).

[3]The z-ratios are 3.5, 2.5, 3 and 1.66. LR tests also point to the same conclusion (see subsection 5.1.1). However, the test is not standard, the null ($\phi_B = 0$) being at the boundary of the parameter admissible values ($\phi_B \geq 0$). Bauwens and Otranto (2020) 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 COR model and $R1$ versions with 5 groups (robust standard errors in parentheses)

	Variance Part						
Parameters	S-1s	first step					
α_1	0.390	0.469					
	(0.024)	(0.031)					
α_2	0.425	0.508					
	(0.022)	(0.029)					
α_3	0.372	0.434					
	(0.021)	(0.024)					
α_4	0.377	0.446					
	(0.021)	(0.029)					
α_5	0.339	0.376					
	(0.026)	(0.027)					
β_1	0.591	0.508					
	(0.025)	(0.033)					
β_2	0.562	0.476					
	(0.023)	(0.031)					
β_3	0.603	0.537					
	(0.023)	(0.027)					
β_4	0.611	0.539					
	(0.023)	(0.031)					
β_5	0.645	0.608					
	(0.028)	(0.029)					
	Correlation Part						
	S-1s	S	S-Pt	S-Rt	$R1$	$R1$-Pt	$R1$-Rt
a_1	0.096	0.099	0.101	0.102	0.108	0.107	0.108
	(0.006)	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)
a_2					0.108	0.110	0.110
					(0.009)	(0.008)	(0.008)
a_3					0.095	0.097	0.098
					(0.008)	(0.007)	(0.007)
a_4					0.090	0.096	0.096
					(0.010)	(0.009)	(0.009)
a_5					0.092	0.093	0.093
					(0.006)	(0.006)	(0.006)
b_1	0.881	0.875	0.876	0.875	0.866	0.870	0.868
	(0.007)	(0.009)	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)
b_2					0.865	0.867	0.866
					(0.012)	(0.011)	(0.011)
b_3					0.880	0.880	0.879
					(0.011)	(0.009)	(0.010)
b_4					0.891	0.885	0.886
					(0.013)	(0.012)	(0.012)
b_5					0.878	0.881	0.880
					(0.008)	(0.008)	(0.008)
ϕ_B			0.007	0.005		0.006	0.005
			(0.002)	(0.002)		(0.002)	(0.003)

The models are defined in the beginning of Section 5. All models except COR-S-1s are estimated in two steps, the first step results being the same (given in the third column of the upper part). For COR-S-1s, all parameters are estimated simultaneously.

COV models

The estimation results for the scalar COV models and the non-scalar ones with 5 groups are shown in Table 5. The positive and significant[4] estimated ϕ_A imply time–varying impact coefficients for the lagged realized covariances through the $\boldsymbol{A}_t$ matrix in (3.3); in particular the coefficients vary between 0.250 and 0.274 for COV-S-Pt, and between 0.253 and 0.279 for the COV-S-Rt.

In the $R1$ models, the (a_i, b_i) parameters vary between groups, for example between (0.225, 0.759) for group 5 and (0.296,0.683) for group 2 in the $R1$ model of the fifth column of the table. One can notice that when $a_i > a_j$ is higher, $b_j > b_i$, so that persistence $(a_i + b_i)$ is more stable across the groups than each of the two parameters.

The EO estimates vary less between groups than in the $R1$ models. It can be noticed that $a_c < a_i$ but that $b_c > b_i$ for each i; nevertheless, $c_a c_b < a_i b_i$, so that as discussed in Appendix A, the conditional covariance matrices $\boldsymbol{S}_t$ are positive definite (for all t) at the QML estimates.

Graphical illustration of time-varying impacts

The pattern of the time–varying coefficients is illustrated in the example of Figure 1, where the time-varying coefficient $b_{ij,t}$ (for $i = AXP$, $j = JPM$) for the COR-$R1$-Pt and COR-$R1$-Rt models (upper part of the figure) is shown, whereas in the bottom part $a_{ij,t}$ is shown for COV-$R1$-Pt and COV-$R1$-Rt. Obviously, the dynamics is smoother when the driving variable is the conditional correlation (Rt models) than the realized correlation (Pt models).

Comparison of estimated $\boldsymbol{A}$ and $\boldsymbol{B}$ matrices

The comparison of the estimated $\boldsymbol{B}$ matrices (of size 29) of the different COR models and the estimated $\boldsymbol{A}$ matrices of the different COV models is not easy by inspection of Tables 4 and 5. For each model the implied matrix of size 29 is computed using the estimated parameters. Table 6 shows the squared Frobenius distances for all pairs of $\boldsymbol{B}$ of the seven COR models and for all pairs of $\boldsymbol{A}$ of the seven COV models. For the time–varying models (S-Pt, S-Rt, $R1$-Pt, $R1$-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 in some cases practically null, as between the two time–varying models S-Pt and S-Rt and between $R1$-Pt and $R1$-Rt; the largest (but also contained) differences occur when one of the terms of the comparison is S-$1s$ (in which all the parameters are estimated simultaneously). Higher differences are present between the $\boldsymbol{A}$ matrices in the COV models, where there are no common estimations (whereas the first step estimation is common in six COR models). The $\boldsymbol{A}$ matrix which seems really different from the other is that of the COV-EO model, where the covariance part has proper parameters, not dependent on the parameters of the variance part as in the $R1$ models.

[4]Three z-ratios are larger than 4.45, the fourth one being equal to 2.56. LR tests also point to the same conclusion (see subsection 5.1.1). The caveat discussed in footnote 3 applies.

Table 5: Estimation results of scalar COV model and non-scalar versions with 5 groups (robust standard errors in parentheses)

Parameters	S	S-Pt	S-Rt	$R1$	$R1$-Pt	$R1$-Rt	EO
a_1	0.251	0.274	0.279	0.277	0.290	0.291	0.274
	(0.010)	(0.014)	(0.016)	(0.016)	(0.020)	(0.019)	(0.014)
a_2				0.296	0.300	0.303	0.287
				(0.015)	(0.017)	(0.016)	(0.016)
a_3				0.235	0.262	0.267	0.282
				(0.009)	(0.019)	(0.016)	(0.016)
a_4				0.249	0.266	0.286	0.302
				(0.013)	(0.015)	(0.013)	(0.021)
a_5				0.225	0.255	0.264	0.283
				(0.011)	(0.018)	(0.013)	(0.019)
b_1	0.736	0.725	0.721	0.712	0.710	0.709	0.721
	(0.011)	(0.013)	(0.014)	(0.016)	(0.016)	(0.016)	(0.012)
b_2				0.683	0.691	0.689	0.704
				(0.016)	(0.015)	(0.015)	(0.015)
b_3				0.754	0.737	0.733	0.716
				(0.010)	(0.016)	(0.014)	(0.015)
b_4				0.731	0.726	0.713	0.691
				(0.013)	(0.014)	(0.012)	(0.020)
b_5				0.759	0.742	0.736	0.712
				(0.011)	(0.015)	(0.012)	(0.017)
a_c							0.201
							(0.009)
b_c							0.786
							(0.010)
ϕ_A		0.045	0.049		0.041	0.046	
		(0.009)	(0.011)		(0.016)	(0.009)	

The models are defined in the beginning of Section 5.

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

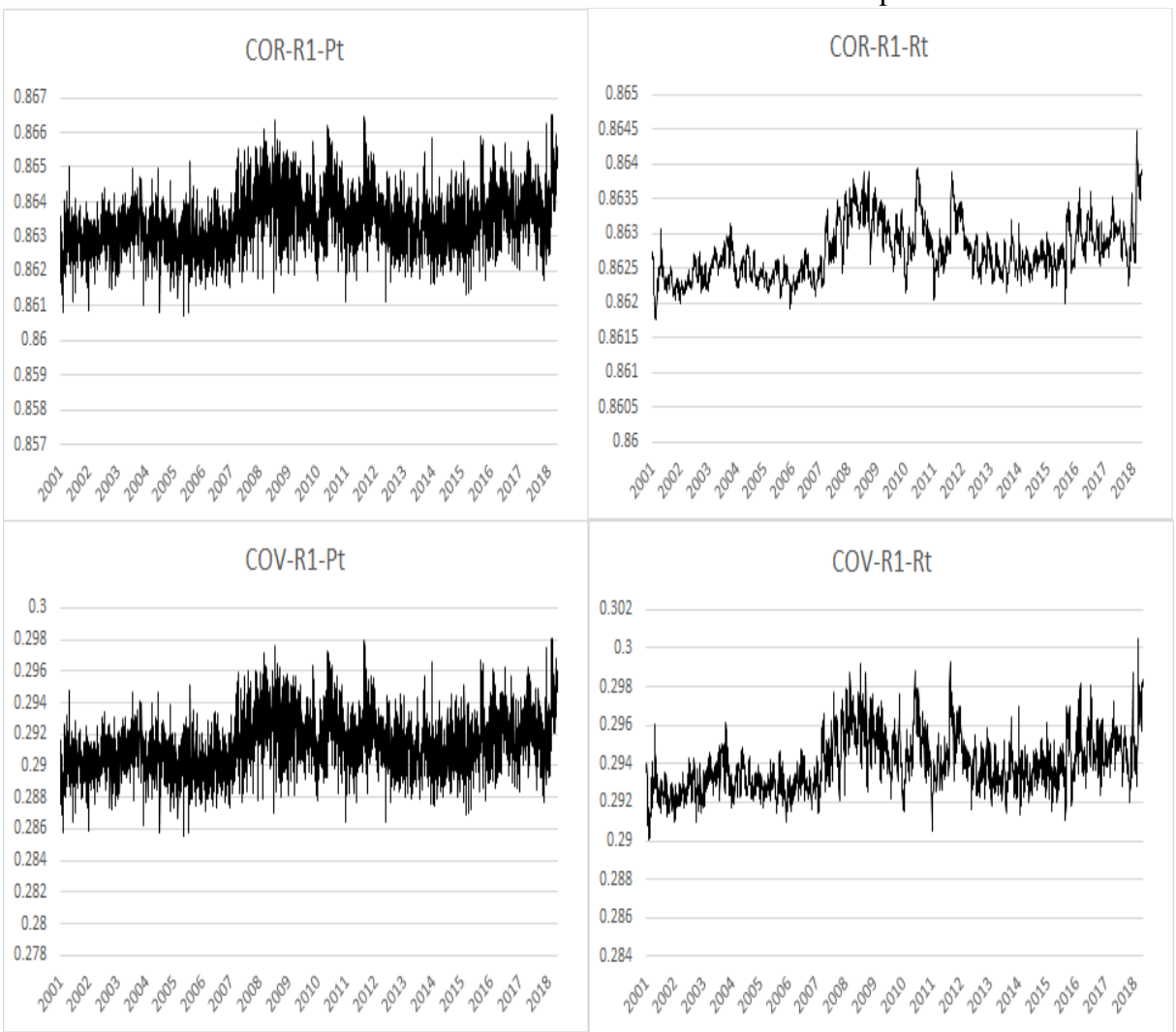


Table 6: Distances between the $\boldsymbol{B}$ matrices of the different COR models and the $\boldsymbol{A}$ matrices of the different COV models

Distance between $\boldsymbol{B}$ matrices in COR models						
	S	$S-Pt$	$S-Rt$	$R1$	$R1-Pt$	$R1-Rt$
$S-1s$	0.500	1.119	1.200	0.948	1.302	1.399
S		0.129	0.155	0.404	0.323	0.377
$S-Pt$			0.002	0.525	0.211	0.248
$S-Rt$				0.545	0.211	0.245
$R1$					0.149	0.149
$R1-Pt$						0.003
Distance between $\boldsymbol{A}$ matrices in COV models						
	$S-Pt$	$S-Rt$	$R1$	$R1-Pt$	$R1-Rt$	EO
S	2.470	3.915	3.434	4.055	5.285	30.394
$S-Pt$		0.166	4.819	1.346	1.295	50.044
$S-Rt$			5.995	1.458	1.072	55.982
$R1$				2.161	3.635	37.307
$R1-Pt$					0.300	52.272
$R1-Rt$						58.040

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

In practice, it seems that when the correlation and covariance coefficients do not depend on the variance coefficients, their estimation is a bit different with respect to the models containing this constraint.

5.2 Covariance Matrix In–sample Fit Evaluation

An additional 'in-sample' comparison of the estimated models is performed using loss functions that allow us to compare the fitted and realized covariance matrices. The statistical loss functions adopted are the 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 et al. (2013). They are defined as

$$\text{QLIK} = \frac{1}{T}\sum_{t=1}^{T}\left\{\ln|\hat{\boldsymbol{S}}_t| + \text{trace}\left(\hat{\boldsymbol{S}}_t^{-1}\boldsymbol{C}_t\right)\right\}, \tag{5.1}$$

$$\text{FN} = \frac{1}{T}\sum_{t=1}^{T}\text{trace}\left[\left(\hat{\boldsymbol{S}}_t-\boldsymbol{C}_t\right)'\left(\hat{\boldsymbol{S}}_t-\boldsymbol{C}_t\right)\right] = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{n}(\hat{s}_{ij,t}-c_{ij,t})^2, \tag{5.2}$$

where $\hat{\boldsymbol{S}}_t$ is the estimated conditional covariance at time t and T is the sample size.[5] Moreover, we consider an economic loss function, based on theoretical portfolio performances, following the global minimum variance portfolio approach of Engle and Colacito (2006). In practice, the loss consists in evaluating the variance of a portfolio with weights

[5]The QLIK loss function (5.1) is equal to the estimation objective function (2.2) (setting $\boldsymbol{S}_t(\boldsymbol{\theta}) = \hat{\boldsymbol{S}}_t$), multiplied by $-2/\nu$, to make it a loss and remove the nuisance parameter ν. Another estimation criterion consists in minimizing the FN loss function (5.2) where $\hat{\boldsymbol{S}}_t$ is replaced by the specified $\boldsymbol{S}_t(\boldsymbol{\theta})$.

obtained as in the classical portfolio problem combining expected returns and the covariance matrix, setting all the expected returns to $1/\sqrt{n}$ (so that the corresponding vector has length equal to one). In this way the portfolios are compared only in terms of covariance matrix and not of return dimension. The GMVP loss function is defined as

$$\text{GMVP} = \frac{1}{T}\sum_{t=1}^{T} \hat{\boldsymbol{w}}_{\boldsymbol{t}}{}' \hat{\boldsymbol{S}}_t \hat{\boldsymbol{w}}_{\boldsymbol{t}} \tag{5.3}$$

where $\hat{\boldsymbol{w}}_{\boldsymbol{t}} = \sqrt{n}\hat{\boldsymbol{S}}_t^{-1}\boldsymbol{j}_n/(\boldsymbol{j}_n'\hat{\boldsymbol{S}}_t^{-1}\boldsymbol{j}_n)$ (with $\boldsymbol{j}_n$ a vector of ones) is the global minimum variance weight vector associated to $\hat{\boldsymbol{S}}_t$ when the expected return vector is equal to $\boldsymbol{j}_n/\sqrt{n}$.[6]

For each loss function, the Model Confidence Set (MCS) procedure of Hansen et al. (2003, 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 et al. (2003), with 10,000 replications.

Figure 2 shows the values of the three losses for the fourteen models listed in the beginning of this section, using their 5-group versions (ie, with the grouping resulting from the algorithm applied to the univariate models for the variances). The results for the models using their full parameterizations are not shown because they are almost identical. The COR models clearly provide lower QLIK losses than the COV models, with the lowest value for COR-S-1s. The 95% (confidence level) MCS consists of COR-S-1s, COR-S-Pt and COR-R1-Pt (the black squares in the graph identify the models with non significant differences of the losses at the 5% nominal size of the tests). The COR-S-1s model and the non time–varying COV models have the lowest FN losses, but all models belong to the 95% MCS, meaning they cannot be discriminated in terms of in–sample fit performance of this kind. For the GMVP loss, the COR models have lower values than the COV models, and COR-R1-Pt has the lowest one and is alone in the 95% MCS.

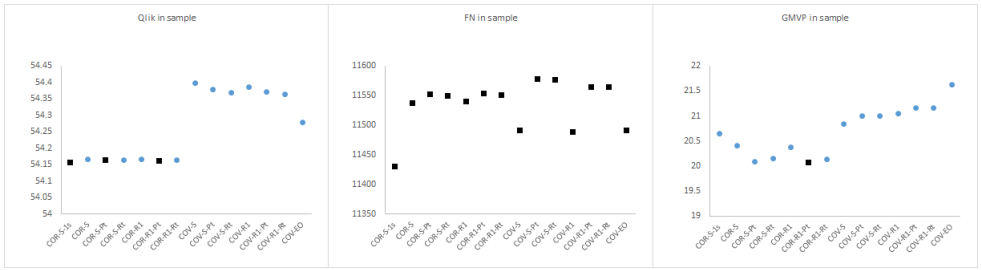
5.3 Covariance Matrix Out–of–sample Forecast Evaluation

To compare the model performances in out–of–sample forecasts of the covariance matrix, the models (in their 5-group versions) have been estimated once for all on the reduced period from January 2, 2001 to December 31, 2015 ($T = 3744$). The remaining 575 realized covariance matrices serve for the out–of–sample evaluations. They are compared to the 1-, 5- and 22-step forecasted covariance matrices. The QLIK, FN and GMVP losses are computed for the 575 forecasts (a few less at horizons 5 and 22). Figure 3 shows the values of the losses and the 95% MCS for the three forecast horizons.

Regarding QLIK, we notice an improved performance, compared to the in-sample results, of the simplest models. This is in line with the empirical observation that more sophisticated models have a better in–sample performance, whereas simpler models tend

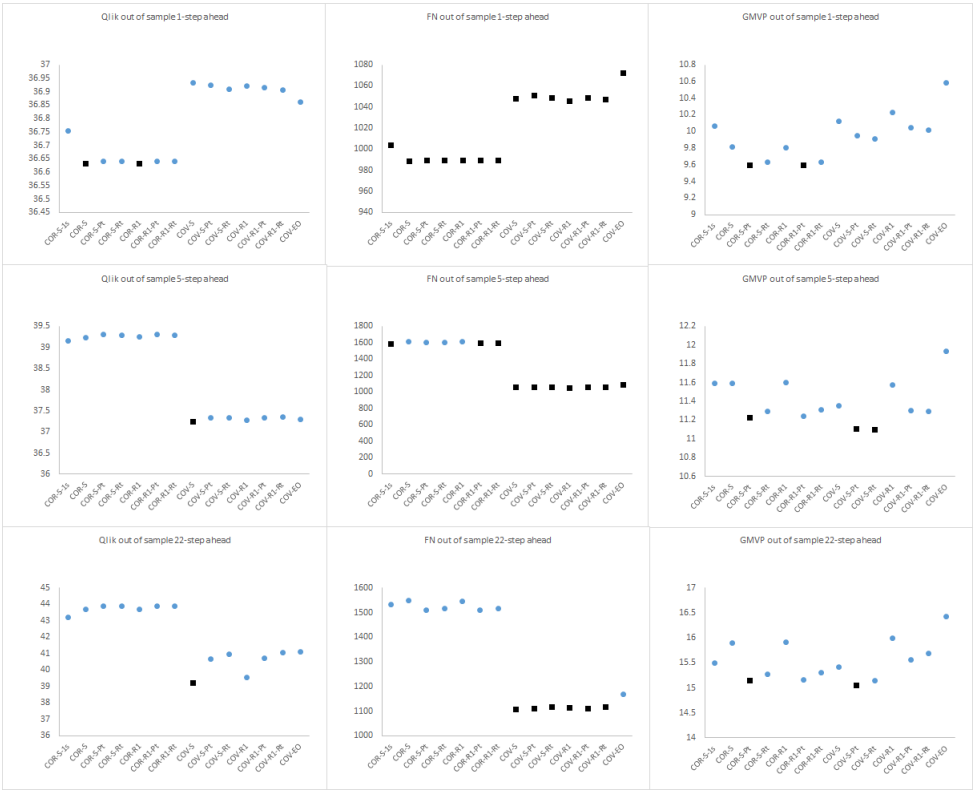
[6] GMVP as defined by (5.3) does not seem to be a loss function: actually it should be defined as $\frac{1}{T}\sum_{t=1}^{T}(\hat{\boldsymbol{w}}_{\boldsymbol{t}}{}'\hat{\boldsymbol{S}}_t\hat{\boldsymbol{w}}_{\boldsymbol{t}} - \boldsymbol{w}_{\boldsymbol{t}}'\boldsymbol{\Sigma}_t\boldsymbol{w}_{\boldsymbol{t}})$ where Σ_t is the true covariance matrix and $\boldsymbol{w}_{\boldsymbol{t}}$ is defined like $\hat{\boldsymbol{w}}_{\boldsymbol{t}}$ but using $\boldsymbol{\Sigma}_t$. Because $\hat{\boldsymbol{w}}_{\boldsymbol{t}}{}'\hat{\boldsymbol{S}}_t\hat{\boldsymbol{w}}_{\boldsymbol{t}}{}' \geq \boldsymbol{w}_{\boldsymbol{t}}'\boldsymbol{\Sigma}_t\boldsymbol{w}_{\boldsymbol{t}}$ (see Theorem 1 of Engle and Colacito, 2006) and $\boldsymbol{w}_{\boldsymbol{t}}'\boldsymbol{\Sigma}_t\boldsymbol{w}_{\boldsymbol{t}}$ does not depend on a model, neglecting these terms does not change the ranking of different models.

Figure 2: In–sample evaluation of 5-group models: QLIK, FN and GMVP losses and 95% MCS (shown by the black squares) for the covariance matrices



The models are defined in the beginning of Section 5.

Figure 3: Out–of–sample evaluation of 5 group models for 1-, 5- and 22-step forecasts: QLIK, FN and GMVP losses and 95% MCS (shown by the black squares) for the covariance matrices



The models are defined in the beginning of Section 5.

to perform better in out–of–sample forecasts (see, for example, Hansen, 2010). In particular COR-S and COR-$R1$ are in the 95% MCS of the 1–step forecasts for covariance matrices. Increasing the forecast horizon, all COV models reverse their performance with respect to COR models. The model with the smallest number of parameters (COV-S) is alone in the 95% MCS.

The FN loss values are clearly the lowest for the COR models in the 1–step forecasts, but as for the in–sample results, all the models belong to the 95 % MCS. Increasing the forecast horizon, the COV models reverse their performance with respect to the COR models. COR models are completely ejected from the 95% MCS at horizon 22 (and partly at horizon 5), which contain all COV models (except one at horizon 22).

The GMVP loss provides the most stable results across the three horizons: the COR-S-Pt is included in the 95% MCS at the three horizons. COR-$R1$-Pt model is the best model in the in–sample results, and it is again the best one in the 1–step–ahead covariance forecasts, together with the simpler COR-S-Pt. The latter is in the 95% MCS also in the 5- and 22-step cases, with the analogous COV-S-Pt (and also COV-S-Rt in the 5-step case).

In short, the simplest models are favoured in out–of–sample evaluations, in particular the more parsimonious COV models when the forecast horizon increases. Interestingly, the HE parameterizations provide added value significantly (see the MCS), in particular using the realized correlations as driving variables of the time–varying coefficients.

5.4 Decomposing FN and GMVP between variance and covariance contributions

For given weight vectors, the GMVP loss function (5.3) can be decomposed as the sum of the variance contribution and the covariance one:

$$\text{GMVP} = \frac{1}{T}\sum_{t=1}^{T} \hat{\boldsymbol{w}}_{\boldsymbol{t}}{}'\hat{\boldsymbol{V}}_t\hat{\boldsymbol{w}}_{\boldsymbol{t}} + \frac{1}{T}\sum_{t=1}^{T} \hat{\boldsymbol{w}}_{\boldsymbol{t}}{}'\hat{\boldsymbol{W}}_t\hat{\boldsymbol{w}}_{\boldsymbol{t}}, \tag{5.4}$$

since $\hat{\boldsymbol{S}}_t = \hat{\boldsymbol{V}}_t + \hat{\boldsymbol{W}}_t$, where the first matrix is the diagonal matrix containing the variances and the second one has zero values on its main diagonal and the covariances elsewhere. Likewise, the FN loss function (5.2) can be decomposed as

$$\text{FN} = \frac{1}{T}\sum_{t=1}^{T}\Big(\sum_{i}(\hat{s}_{ii,t} - c_{ii,t})^2 + 2\sum_{i<j}(\hat{s}_{ij,t} - c_{ij,t})^2\Big). \tag{5.5}$$

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

Table 7 shows the variance and covariance contributions to FN and GMVP values for each model. Considering the 5- and 22-step results, two observations emerge about the FN results: 1) the covariance contributions of the COV models are much lower (by about 30%) than those of the COR models; 2) the models in the MCS sets always correspond to lower covariance contributions compared to the excluded models; notice that in the 5-step case this happens also for the COR models that are in the MCS, but much less than for

the COV models. Hence, it seems that at the larger forecasting horizons, the covariances determine more than the variances which models belong to the MCS sets.

For GMVP, the COR models have systematically a lower variance loss and a higher covariance loss than the COV models; the differences are of the order of 10% in-sample and out-of-sample at horizon 1, and 10 to 30% at the larger horizons. Since the total losses of the COR and COV models are close (they differ by 4% in-sample and out-of-sample at horizon 1, and by less than 1% at horizons 5 and 22), there is a kind of substitution effect between the two parts. Moreover, the shares of covariance contributions increase with the forecast horizon: they are on average (across models) 44% for COV and 51% for COR at horizon 1, 51 and 68 at horizon 5, and 62 and 74 at horizon 22.

6 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 nice property to guarantee the positive definiteness of the matrix of parameters. The HE-CAW models show, in most cases, a better performance, both in–sample and out–of–sample for different forecasting horizons, with respect to 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. The HE COR parameterizations provide empirically a better out–of–sample forecasting performance than the scalar COR model estimated in one step (called COR-S-1s) which can be considered as a benchmark model but has the disadvantage of being difficult to estimate for a large number of assets.

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 (3.1) 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 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 (3.3), 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 parameterization can also be used in other models such as of HAR-type (Chiriac and Voev, 2010; Oh and Patton, 2016). An extensive horse-race between CAW, HAR, and other models was not the goal of this paper and would be of interest in a different paper.

Table 7: Variance and covariance contributions to the FN and GVMP loss functions for the 5-group models

FN

	in-sample		out-of-sample forecasts					
			1 step		5 steps		22 steps	
Model	var	cov	var	cov	var	cov	var	cov
COR-S-1s	2081.0	9350.5	160.6	843.2	231.4	1355.0	214.6	1318.5
COR-S	2062.8	9475.5	160.8	828.5	231.3	1384.9	214.4	1334.2
COR-S-Pt	2062.8	9489.7	160.8	829.0	231.3	1369.6	214.4	1295.3
COR-S-Rt	2062.8	9487.0	160.8	828.7	231.3	1373.9	214.4	1302.6
COR-R1	2062.8	9477.9	160.8	828.6	231.3	1382.6	214.4	1331.1
COR-R1-Pt	2062.8	9491.7	160.8	829.0	231.3	1368.2	214.4	1294.5
COR-R1-Rt	2062.8	9489.0	160.8	828.7	231.3	1372.3	214.4	1301.6
COV-S	2065.2	9427.0	164.3	883.8	167.8	891.3	192.1	914.8
COV-S-Pt	2083.9	9494.1	165.7	885.2	171.4	893.5	215.9	895.4
COV-S-Rt	2085.3	9491.3	165.6	883.5	171.6	892.3	221.3	895.5
COV-R1	2060.2	9428.5	164.1	881.8	167.9	889.2	195.9	917.7
COV-R1-Pt	2077.8	9487.0	165.4	883.7	171.1	891.5	216.2	896.6
COV-R1-Rt	2078.9	9485.5	165.4	882.1	171.3	890.4	221.1	895.8
COV-EO	2072.9	9418.6	165.3	907.1	171.4	917.8	222.8	945.9

GMVP

	in-sample		out-of-sample forecasts					
			1 step		5 steps		22 steps	
Model	var	cov	var	cov	var	cov	var	cov
COR-S-1s	13.00	7.66	4.98	5.09	3.79	7.81	4.10	11.40
COR-S	13.00	7.41	4.84	4.98	3.74	7.85	4.16	11.74
COR-S-Pt	12.69	7.41	4.71	4.88	3.59	7.64	3.93	11.22
COR-S-Rt	12.75	7.41	4.74	4.90	3.62	7.67	3.97	11.31
COR-R1	12.93	7.46	4.83	4.98	3.72	7.88	4.13	11.78
COR-R1-Pt	12.64	7.44	4.71	4.88	3.58	7.66	3.91	11.25
COR-R1-Rt	12.70	7.44	4.74	4.90	3.61	7.70	3.96	11.35
COV-S	14.63	6.21	5.60	4.52	5.66	5.70	6.03	9.38
COV-S-Pt	14.26	6.75	5.61	4.34	5.61	5.49	5.82	9.23
COV-S-Rt	14.12	6.90	5.55	4.37	5.56	5.54	5.81	9.33
COV-R1	14.45	6.60	5.61	4.62	5.62	5.95	6.05	9.95
COV-R1-Pt	14.20	6.96	5.60	4.45	5.59	5.72	5.83	9.74
COV-R1-Rt	14.06	7.11	5.55	4.47	5.54	5.76	5.78	9.91
COV-EO	13.45	8.20	5.96	4.63	5.47	6.47	5.89	10.54

The gray cells identify the models belonging to the 95% MCS (see Figures 2 and 3).

Appendix: Positive semidefiniteness of the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ of the COV-EO model

We discuss the issue when the matrix $\boldsymbol{A}$, and therefore also the scalar $\bar{a}$, of (3.3) are constant, but the discussion is valid also for $\boldsymbol{B}$. The equation is

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

In the EO parameterization, $\boldsymbol{A}$ is a matrix with diagonal elements $a_1, \ldots, a_n$, and equal off-diagonal elements a_c, with $0 \leq a_c \leq a_i < 1$ for each $i = 1, ...n$. By expressing $\boldsymbol{A}$ as $\boldsymbol{D}_A + a_c\boldsymbol{J}_n$, where $\boldsymbol{D}_A = \mathrm{diag}(\boldsymbol{A}) - a_c\boldsymbol{I}_n$, an application of the Sherman and Morrison formula (see, for example, Hager, 1989) gives

$$\det(\boldsymbol{A}) = \det(\boldsymbol{D}_A) + a_c\boldsymbol{j}_n'\mathrm{adj}(\boldsymbol{D_A})\boldsymbol{j}_n = \prod_{i=1}^{n}(a_i - a_c) + a_c\sum_{i=1}^{n}\prod_{j\neq i}(a_j - a_c),$$

where $\boldsymbol{j}_n$ is a column of ones. It is straightforward to check that $0 \leq a_c \leq a_i$ for each $i = 1, ...n$, is a sufficient condition to obtain a non-negative determinant. Moreover, the nonnegativity of a_c is a logical requirement: considering a generic off-diagonal element of $\boldsymbol{S}_t$ (one of the conditional covariances), a negative a_c implies that an increase (decrease) in the corresponding lagged realized covariance decreases (increases) the conditional covariance, which does not make sense. The constraint $a_c \leq a_i$ for each i is a bit more restrictive, but in line with the constraints of a rank-1 parameterization, where the coefficient of the covariance in position (i, j) is given by the product of a_i and a_j, both less than 1, so that this covariance coefficient is smaller than a_i and a_j.

In our estimations, we noticed that the positive semidefinitess of the matrix $\boldsymbol{A}$ can be obtained with weaker constraints. If we substitute $\boldsymbol{S}_{t-1}$ with its expression we obtain

$$\boldsymbol{S}_t = (1-\bar{a}-\bar{b})\bar{\boldsymbol{C}}+\boldsymbol{A}\odot\boldsymbol{C}_{t-1}+(1-\bar{a}-\bar{b})(\bar{\boldsymbol{C}}\odot\boldsymbol{B})+(\boldsymbol{A}\odot\boldsymbol{B})\odot\boldsymbol{C}_{t-2}.+(\boldsymbol{B}\odot\boldsymbol{B})\odot\boldsymbol{S}_{t-2}.$$

We always obtain in estimations that $c_a c_b < a_i b_i$ for each i; this implies that $\boldsymbol{A} \odot \boldsymbol{B}$ is positive definite because its diagonal elements are greater than its off–diagonal elements. Moreover $\bar{\boldsymbol{C}} \odot \boldsymbol{B}$ is also positive definite because $\bar{\boldsymbol{C}}$ and $\boldsymbol{B}$ are positive definite. Continuing the substitution of the lag of $\boldsymbol{S}_t$ in the previous equation, we obtain:

$$\boldsymbol{S}_t = (1 - \bar{a} - \bar{b})\sum_{k=1}^{t-1}(\bar{\boldsymbol{C}} \odot \boldsymbol{B}^{\odot k}) + \boldsymbol{A} \odot (\sum_{k=1}^{t-1}\boldsymbol{B}^{\odot k} \odot \boldsymbol{C}_{t-i-1}) + \sum_{k=1}^{t}\boldsymbol{B}^{\odot k} \odot \boldsymbol{S}_0$$

where $\boldsymbol{B}^{\odot k} = \boldsymbol{B} \odot \boldsymbol{B} \odot \cdots \odot \boldsymbol{B}$ (k times). The last term is negligible for large t; using a starting matrix as $\boldsymbol{S}_0 = \bar{\boldsymbol{C}}$, the previous comments apply as well.

In brief, the previous developments explain that imposing $c_a c_b < a_i b_i$ for each i did not result in non-positive definite $\boldsymbol{S}_t$ matrices during our estimations.

Acknowledgments

We thank Lyudmila Grigoryeva (University of Konstanz) and Juan-Pablo Ortega (University of St.Gallen) for providing the datasets for the DJIA companies and Oleksandra

Kukharenko (University of Konstanz) for handling the data and computing the realized covariance matrices from the TAQ data. 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).

References

[1] Aielli, G. (2013). Dynamic Conditional Correlations: On Properties and Estimation. *Journal of Business and Economic Statistics* 31, 282–299.

[2] Bauwens, L., Braione, M. and Storti, G. (2016). Forecasting Comparison of Long Term Component Dynamic Models for Realized Covariance Matrices. *Annals of Economics and Statistics* 123/124, 103-134.

[3] Bauwens, L., Hafner, C., and Laurent, S. (2012a): Volatility models. in *Handbook of Volatility Models and their Applications*, Bauwens, L., Hafner, C., and Laurent, S. (Eds.), Ch. 1. John Wiley and Sons.

[4] Bauwens, L., Laurent, S. and Rombouts J.V.K. (2006). Multivariate GARCH Models: A Survey. *Journal of Applied Econometrics* 21 79–109.

[5] Bauwens, L. and Otranto, E. (2016). Modeling the dependence of conditional correlations on market volatility. *Journal of Business & Economic Statistics* 34, 254–268.

[6] Bauwens, L. and Otranto, E. (2020). Nonlinearities and Regimes in Conditional Correlations with Different Dynamics. *Journal of Econometrics* 217, 496-522.

[7] Bauwens, L., Storti, G. and Violante, F. (2012b): Dynamic conditional correlation models for realized covariance matrices. *CORE Discussion Paper* 2012–60.

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

[9] Cappiello, L., Engle, R.F. and Sheppard, K. (2006). Asymmetric dynamics in the correlations of global equity and bond returns. *Journal of Financial Econometrics* 4, 537-572

[10] Chiriac, R. and Voev, V. (2010). Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics* 26, 922–947.

[11] Clements, A., Scott, A., and Silvennoinen, A. (2018). Volatility-dependent correlations: Further evidence of where, when and how. *Empirical Economics* 57, 505-540.

[12] Engle, R.F. (2002). Dynamic conditional correlation: A simple class of multivariate GARCH models. *Journal of Business and Economic Statistics* 20, 339–350.

[13] Engle, R.F. and Colacito, R. (2006). Testing and valuing dynamic correlations for asset allocation. *Journal of Business & Economic Statistics* 24, 238–253.

[14] Engle, R.F. and Kroner, K.F. (1995). Multivariate Simultaneous Generalized ARCH. *Econometric Theory* 11, 122–150.

[15] Engle, R.F., Ng, V.K. and Rothschild, M. (1990). Asset pricing with a factor ARCH covariance structure: empirical estimates for treasury bills. *Journal of Econometrics* 45, 213–238.

[16] Engle R. F. and Sheppard, K. (2001). Theoretical and Empirical Properties of Dynamic Conditional Correlation MVGARCH. *Working Paper* No. 2001–15, University of California, San Diego.

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

[18] Hafner, C.M. and Franses, P.H. (2009). A generalized dynamic conditional correlation model: simulation and application to many assets. *Econometric Reviews* 28, 612–631.

[19] Hager, W.W. (1989). Updating the inverse of a matrix. *SIAM Review* 31, 221–239.

[20] Hansen, P. R. (2010). A winner's curse for econometric models: on the joint distribution of in–sample fit and out–of–sample fit and its implications for model selection. *Technical report*, Stanford University.

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

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

[23] Laurent, S., Rombouts, J.V.K., and Violante, F. (2013). On loss functions and ranking forecasting performances of multivariate volatility models. *Journal of Econometrics* 173, 1–10.

[24] Noureldin, D., Shephard, N., and Sheppard, K. (2014). Multivariate Rotated ARCH, *Journal of Econometrics* 179, 16-30.

[25] Oh, D.H. and Patton, A.J. (2016). High-dimensional copula-based distributions with mixed frequency data. *Journal of Econometrics* 193, 34–366.

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

[27] Patton, A. and Sheppard, K. (2009). Evaluating volatility and correlation forecasts. In: Andersen, T., Davis, R., Kreiss, J., Mikosch, T. (Eds.), *Handbook of Financial Time Series*, 801–838. Springer

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

Ultimi Contributi di Ricerca CRENoS

I Paper sono disponibili in: http://www.crenos.unica.it

20/06 *Demetrio Lacava, Giampiero M. Gallo, Edoardo Otranto,* "Measuring the Effects of Unconventional Policies on Stock Market Volatility"

20/05 *Gianfranco Atzeni, Luca G. Deidda, Marco Delogu, Dimitri Paolini,* "Drop-out decisions in a cohort of Italian university students"

20/04 *Emanuela Marrocu, Raffele Paci, David Rigby, Stefano Usai,* "Smart Specialization Strategy: any relatedness between theory and practice?

20/03 *Giorgio Garau, Stefano Deriu,* "Total Factor Productivity and Relative Prices: the case of Italy"

20/02 *Fabio Cerina, Alessio Moro, Michelle Rendall,* "A Note on Employment and Wage Polarization in the U.S."

20/01 *Elias Carroni, Dimitri Paolini,* "Business models for streaming platforms: content acquisition, advertising and users"

19/16 *Daniela Sonedda*, "Regional variation in apprenticeship and permanent employment rates: which causes?"

19/15 *Daniela Sonedda,* "Regional disparities in the functioning of the labour markets"

19/14 *Bianca Biagi, Barbara Dettori, Raffaele Paci, Stefano Usai,* "Economic development in Sardinia: overcoming the insularity gap"

19/13 *Miguel Casares, Luca Deidda, Jose E. Galdon-Sanchez,* "On financial frictions and firm market power"

19/12 *Massimiliano Bratti, Maurizio Conti, Giovanni Sulis,* "Employment Protection and Firm-provided Training: Quasi-experimental Evidence from a Labour Market Reform"

19/11 *Jessica Goldberg, Mario Macis, Pradeep Chintagunta*, "Incentivized Peer Referrals for Tuberculosis Screening: Evidence from India"

19/10 *Julio J. Elías, Nicola Lacetera, Mario Macis,* "Paying for Kidneys? A Randomized Survey and Choice Experiment"

19/09 *Fabio Cerina, Elisa Dienesch, Alessio Moro, Michelle Rendall,* "Spatial Polarization"

19/08 *Michele Battisti, Massimo Del Gatto, Christopher F. Parmeter,* "Skill Biased Technical Change and Misallocation: a Unified Framework"

19/07 *Fabrizio Fusillo, Francesco Quatraro, Stefano Usai,* "Going Green: Environmental Regulation, eco-innovation and technological alliances"

19/06 *Oliviero A. Carboni, Giuseppe Medda,* "External R&D Acquisition and Product Innovation"

19/05 *José J. Cao-Alvira, Luca G. Deidda,* "Development of Bank Microcredit"

19/04 *Vania Licio,* "When history leaves a mark: a new measure of Roman roads"

19/03 *Simone Franceschini, Gerardo Marletto,* "Reorganization of supply chains as a key for the envision of socio- technical transitions. The case of tourism"

19/02 *Elias Carroni, Dimitri Paolini,* "The business model of a streaming platform"

19/01 *Stefania Capecchi, Marta Meleddu, Manuela Pulina, Giuliana Solinas,* "Mixture models for consumers' preferences in healthcare"

18/13 *Adelaide Baronchelli, Teodora Erika Uberti,* "Exports and FDI: Comparing Networks in the New Millennium"

18/12 *Gabriele Cardullo, Maurizio Conti, Giovanni Sulis,* "Unions, Two-Tier Bargaining and Physical Capital Investment: Theory and Firm-Level Evidence from Italy"

18/11 *Jing Guan, J.D. Tena,* "Estimating the Effect of Physical Exercise on Juveniles' Health Status and Subjective Well-Being in China"

www.crenos.unica.it

ISBN 9788868513283