DCC- AND DECO-HEAVY: MULTIVARIATE GARCH MODELS BASED ON REALIZED VARIANCES AND CORRELATIONS

Luc Bauwens, Yongdeng Xu

REPRINT | 3345

UCLouvain

CORE | LIDAM
Louvain Institute of Data Analysis and Modeling in economics and statistics

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

International Journal of Forecasting 39 (2023) 938–955

Contents lists available at ScienceDirect

International Journal of Forecasting

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

DCC- and DECO-HEAVY: Multivariate GARCH models based on realized variances and correlations

Luc Bauwens [a], Yongdeng Xu [b,*]

[a] LIDAM/CORE, UCLouvain, Voie du Roman Pays, 34, B-1348 Louvain-La-Neuve, Belgium
[b] Cardiff Business School, Cardiff University, CF103EU, UK

ARTICLE INFO

Keywords:
Correlation forecasting
Dynamic conditional correlation
Equicorrelation
High-frequency data
Multivariate volatility

ABSTRACT

This paper introduces the scalar DCC-HEAVY and DECO-HEAVY models for conditional variances and correlations of daily returns based on measures of realized variances and correlations built from intraday data. Formulas for multi-step forecasts of conditional variances and correlations are provided. Asymmetric versions of the models are developed. An empirical study shows that in terms of forecasts the scalar HEAVY models outperform the scalar BEKK-HEAVY model based on realized covariances and the scalar BEKK, DCC, and DECO multivariate GARCH models based exclusively on daily data.

© 2022 The Authors. Published by Elsevier B.V. on behalf of International Institute of Forecasters. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

The covariance matrix of daily asset returns is used in several applications of financial management. Therefore, modeling the temporal dependence in the elements of the covariance matrix, often with the ultimate aim of forecasting, is a key area of financial econometrics. The most widespread model class for this purpose is that of multivariate generalized autoregressive conditional heteroskedasticity (MGARCH) models, wherein the conditional covariance matrix of daily returns is specified as a deterministic function of past daily returns. A survey of MGARCH models is provided by Bauwens, Laurent, and Rombouts (2006).

The increasing availability of intraday data has led to the development of so-called High-frEquency-bAsed VolatilitY (HEAVY) models, initially in the univariate case (Engle & Gallo, 2006; Shephard & Sheppard, 2010). In the multivariate case, such models have been introduced in the stochastic volatility framework by Jin and Maheu (2013), and in the MGARCH one by Noureldin, Shephard, and Sheppard (2012). In the latter framework, the main difference is that the conditional covariance matrix of daily returns is specified in the HEAVY case as a function of lagged realized covariances, instead of lagged outer products of daily returns in the MGARCH case. Thus, multi-step forecasts of daily conditional covariance matrices by HEAVY models require as input forecasts of realized covariances, which are obtained from a dynamic model for the realized covariance matrix. HEAVY models are based on more accurate measurements of covariances than GARCH models, and they improve forecasts of the conditional covariance matrix of daily returns, as illustrated by Noureldin et al. (2012). A different class of models that relate the daily return volatility to a realized volatility measure is the realized GARCH model of Hansen, Huang, and Shek (2012); this type of model has been extended to the multivariate setup by Gorgi, Hansen, Janus, and Koopman (2019) and Hansen, Lunde, and Voev (2014).

A well-known difficulty with MGARCH and HEAVY models is that the number of parameters they require tends to be large, being at least a quadratic function of the dimension of the return vector. Therefore, when the dimension is more than a handful of assets, a scalar parameterization is adopted. For example, Noureldin et al. (2012), for 10 assets, and Opschoor, Janus, Lucas, and Van Dijk (2018), for 30 assets, adopt the scalar BEKK

* Corresponding author.
E-mail addresses: luc.bauwens@uclouvain.be (L. Bauwens), xuy16@cardiff.ac.uk (Y. Xu).

https://doi.org/10.1016/j.ijforecast.2022.03.005
0169-2070/© 2022 The Authors. Published by Elsevier B.V. on behalf of International Institute of Forecasters. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

parameterization of Engle and Kroner (1995). This parameterization, coupled with a targeting procedure, which is a way to estimate the constant parameter matrices of the model, facilitates the estimation of the remaining scalar parameters enormously, since their number is independent of the dimension. The scalar BEKK parameterization implies that the conditional variances and covariances have the same persistence and the same sensitivity to the past realized covariances, which may be unrealistic for a large number of assets. A factor HEAVY model, proposed by Sheppard and Xu (2019), avoids this drawback of the scalar BEKK while keeping the number of parameters manageable for estimation.

The first contribution of this research is the extension of the HEAVY class by the DCC specification of Engle (2002) and the DECO of Engle and Kelly (2012). Scalar DCC-HEAVY and DECO-HEAVY models are developed. Like it was done for the BEKK-HEAVY, these models use well-known formulations of the corresponding MGARCH models, modifying them by specifying the dynamics of the daily conditional correlation matrix as a function of past realized correlations. Likewise, the model we propose for the realized correlation matrix is of DCC type, instead of BEKK type. In both parts of the model, there is a decoupling of the parameters of the variance dynamics (daily or realized) from those of the model for the corresponding correlation matrix. The decoupling has been shown to be advantageous in several respects for MGARCH models: the scalar DCC is a more flexible model than the scalar BEKK and usually provides a better empirical fit, even after taking into account its heavier parameterization; each part of the model can be estimated by QML in two steps (one for the variance equations, and one for the correlation), which makes it practical for handling a dimension of more than a handful of assets; and it usually improves the forecast quality. The same advantages occur in the scalar HEAVY model context, and this is confirmed empirically.

The proposed DCC-HEAVY formulation is different from the model of Braione (2016), where the dynamics of the daily correlation matrix are driven by the outer product of the past degarched returns (which is not a correlation matrix), while the model for the realized covariances and variances is of BEKK type. An advantage of the new specifications is that they allow us to forecast directly the correlations several steps ahead, avoiding an approximation due to the fact that otherwise, correlations are obtained by normalizing quasi-correlations.

Stationarity conditions and formulas for multi-step forecasts are derived from a vector multiplicative error representation of the DCC-HEAVY model, where the dynamics are driven by both lagged realized measures and outer products of past degarched returns, thus encompassing both DCC-HEAVY and DCC-GARCH. Furthermore, asymmetric impact and HAR-type terms (see Corsi (2009)) are added to the DCC- and DECO-HEAVY models.

The second contribution of this research is a detailed empirical comparison of the BEKK-, DCC-, and DECO-HEAVY and GARCH models. All models are applied to the stocks in the Dow Jones Industrial Average (DJIA) index. Like in Shephard and Sheppard (2010), the effects of the lagged squared returns are insignificant when lagged realized variances are included in the conditional variance equations. Likewise, the effect of the lagged outer product of degarched returns is insignificant when the lagged realized correlation matrix is included in the conditional correlation equation. Moreover, when applying the model confidence set approach based on statistical and economic loss functions, the empirical results show that the DCC- and DECO-HEAVY models provide better out-of-sample forecasts than the DCC-GARCH, DECO-GARCH, BEKK-GARCH, and BEKK-HEAVY models. Including asymmetric and HAR terms further improves the DCC- and DECO-HEAVY model forecasts.

The remainder of the paper is organized as follows. Section 2 introduces the DCC- and DECO-HEAVY models. Section 3 provides the multiplicative error representation, the multi-step forecast formulas, and model extensions. Section 4 presents the estimation procedure. Section 5 provides the empirical results. Section 6 concludes. A supplementary appendix includes additional theoretical and empirical results, in particular for a dataset used by Noureldin et al. (2012).

2. DCC-HEAVY and DCC-DECO models

In the first subsection, the multivariate HEAVY framework of Noureldin et al. (2012) is recalled, in particular the scalar BEKK-HEAVY model. The scalar DCC-HEAVY and DCC-DECO models are defined in the next subsections.

2.1. Multivariate HEAVY framework

Let $r_t = (r_{1t}, r_{2t} \dots r_{kt})'$ denote the $k \times 1$ daily return vector of day t corresponding to k assets, and $r_{(j)t} = (r_{(j)1t}, r_{(j)2t} \dots r_{(j)kt})'$ the corresponding jth intra-daily return vector at time j on day t, where $j = 1, 2, \dots, m$. Assuming, for instance, six and a half hours of trading per day and five-minute returns, m is equal to 78. The outer product of daily returns is the $k \times k$ matrix $r_t r_t'$. The most simple realized covariance measure for the k assets on day t is the $k \times k$ matrix, defined as

$$RC_t = \sum_{j=1}^{m} r_{(j)t} r_{(j)t}'. \tag{1}$$

Assuming that $m > k$, RC_t is positive definite (assuming that no asset is a linear combination of other assets). Denote by v_t the $k \times 1$ realized variance vector of day t, consisting of the diagonal elements of RC_t, and by RL_t the realized correlation matrix of day t, defined as

$$RL_t = \{\text{diag}(RC_t)\}^{-1/2} RC_t \{\text{diag}(RC_t)\}^{-1/2}, \tag{2}$$

where $\text{diag}(RC_t)$ is the diagonal matrix obtained by setting the off-diagonal elements of RC_t equal to zero, and the exponent $-1/2$ transforms each diagonal element into the inverse of its square root. Thus, the off-diagonal elements of RL_t are the realized correlation coefficients for the asset pairs, and its diagonal elements are equal to unity.

A multivariate HEAVY model specifies a dynamic process for the conditional covariance matrix H_t of the daily return and another one for the conditional mean M_t of the realized covariance matrix of day t:

$$E(r_t r_t' | \mathcal{F}_{t-1}) := H_t, \tag{3}$$

$$E(RC_t|\mathcal{F}_{t-1}) := M_t, \tag{4}$$

where $\mathcal{F}_{t-1}$ is the information set generated by the daily and intra-daily observations, and $E(r_t|\mathcal{F}_{t-1}) = 0$ is assumed for simplicity (otherwise, r_t denotes the demeaned return vector). The link between both expectations comes from the dependence of H_t on past values of functions of RC_t.

For the specification of the dynamics of H_t and M_t, Noureldin et al. (2012) adopt the BEKK-type model to ensure that the conditional covariance matrix is positive semidefinite. The scalar version of the model, with the "targeting" parameterization of the constant terms, is

$$H_t = (1-\beta_H)\bar{H} - \alpha_H\bar{M} + \alpha_H RC_{t-1} + \beta_H H_{t-1}, \tag{5}$$

$$M_t = (1-\alpha_M-\beta_M)\bar{M} + \alpha_M RC_{t-1} + \beta_M M_{t-1}, \tag{6}$$

where the $k \times k$ matrices $\bar{H} = E(r_t r_t')$ and $\bar{M} = E(RC_t)$ are assumed to be positive definite (PD), and targeting means that these matrices are replaced by their empirical counterparts. For (6), sufficient restrictions for the positivity of M_t are that $\alpha_M \geq 0$, $\beta_M \geq 0$, $\beta_M = 0$ if $\alpha_M = 0$, and $\alpha_M + \beta_M < 1$. For (5), the restrictions $\alpha_H \geq 0$, $\beta_H \geq 0$, $\beta_H = 0$ if $\alpha_H = 0$, and $\beta_H < 1$ are not sufficient to ensure that $(1-\beta_H)\bar{H} - \alpha_H\bar{M}$, and hence H_t, be positive definite. The issue is explained in detail in Section A of the supplemental appendix, where the equivalence between (5) and one of the covariance targeting parameterizations of Noureldin et al. (2012) is shown. For estimation, in practice we check that H_t, rather than $(1-\beta_H)\bar{H} - \alpha_H\bar{M}$, is positive definite in the sample period during the numerical maximizing of the log-likelihood function, but for both datasets we used, no lack of PDness occurred.

The scalar BEKK-GARCH model corresponds to Eq. (5) for H_t above, where RC_{t-1} is replaced by $r_{t-1}r_{t-1}'$ and $\bar{M}$ by $\bar{H}$, thus using only the daily information.

2.2. DCC-HEAVY model

An alternative to the BEKK specification for H_t and M_t is the DCC model of Engle (2002). Any covariance matrix can be written as the product DRD, where D is the corresponding diagonal matrix of standard deviations, and R is the correlation matrix.

2.2.1. DCC-HEAVY specification of H_t

Let us denote by h_t the $k \times 1$ vector of conditional variances (that is, the diagonal elements of H_t), by $h_t^{1/2}$ the vector of conditional standard deviations (obtained by taking the square root of each entry of h_t), and by R_t the corresponding conditional correlation matrix. Then, the conditional covariance matrix H_t can be written as

$$H_t = \mathrm{Diag}(h_t^{1/2})R_t\,\mathrm{Diag}(h_t^{1/2}), \tag{7}$$

where, for x being a $k \times 1$ vector, Diag(x) is the $k \times k$ diagonal matrix with the entries of x as diagonal elements.

Assumption (3) for the HEAVY-BEKK is replaced by

$$E(\mathrm{diag}(r_t r_t')|\mathcal{F}_{t-1}) := \mathrm{Diag}(h_t), \tag{8}$$

$$E(u_t u_t'|\mathcal{F}_{t-1}) := R_t, \text{ where } u_t = r_t \odot h_t^{-1/2}. \tag{9}$$

Thus, instead of specifying altogether the dynamics of the conditional variances and covariances of the returns, as for instance in (5), the DCC-HEAVY model specifies the dynamics of the conditional variances of the returns and of the conditional correlation matrix (R_t) of the degarched returns (i.e., the observed returns divided by their conditional standard deviations). Notice indeed that

$$\begin{aligned} E(u_t u_t'|\mathcal{F}_{t-1}) &= E(r_t r_t'|\mathcal{F}_{t-1}) \odot \left(h_t^{-1/2}(h_t^{-1/2})'\right) \\ &= H_t \odot \left(h_t^{-1/2}(h_t^{-1/2})'\right) \end{aligned}$$

is a matrix with unit diagonal elements and off-diagonal elements that are the conditional correlation coefficients, that is, the conditional covariances divided by the corresponding conditional standard deviations. This is the same setting as in the DCC-GARCH model of Engle (2002), and since a covariance is the product of two standard deviations and a correlation, the expectation of the covariance is not the corresponding function of the expected standard deviations and expected correlation.

The dynamics of the conditional variance vector are specified as

$$h_t = \omega_h + A_h v_{t-1} + B_h h_{t-1}, \tag{10}$$

where v_{t-1} is defined below (1), ω_h is a $k \times 1$ positive vector, and A_h and B_h are $k \times k$ matrices, such that each entry of h_t is positive. To ease the restrictions necessary for this and to avoid parameter proliferation, A_h and B_h are restricted to be diagonal matrices with positive entries on the diagonal, and the elements on the diagonal of B_h to be smaller than unity. The diagonality restrictions imply that each conditional variance depends on its own lag and the corresponding previous realized variance, as in the HEAVY-r model of Shephard and Sheppard (2010). More generally, A_h can be non-diagonal to allow spillover effects. If B_h is restricted to be diagonal, the model of the k variances can be estimated in k separate parts (see Section 4). The DCC-GARCH model of Engle (2002) for the conditional variances is similar to (10), with r_{t-1}^2 (the squared elements of r_{t-1}) replacing v_{t-1}.

The conditional correlation matrix is specified through a scalar dynamic equation:

$$R_t = \tilde{R} + \alpha_r RL_{t-1} + \beta_r R_{t-1}, \tag{11}$$

$$\tilde{R} = (1-\beta_r)\bar{R} - \alpha_r\bar{P}, \tag{12}$$

where $\alpha_r \geq 0$, $\beta_r \geq 0$, $\beta_r = 0$ if $\alpha_r = 0$, $\beta_r < 1$, $\bar{R}$ is the $k \times k$ unconditional correlation matrix of u_t, and $\bar{P}$ is the $k \times k$ unconditional expectation of RL_t. The elements of $\bar{R}$ and $\bar{P}$ can be set to their empirical counterpart to simplify the estimation, so that only two parameters (α_r and β_r) remain to be estimated. By substituting (12) in (11), R_t is equal to $\bar{R} + \alpha_r(RL_{t-1} - \bar{P}) + \beta_r(R_{t-1} - \bar{R})$, and by taking the unconditional expectation on both sides, $E(R_t) = \bar{R}$ if $E(RL_t) = \bar{P}$. The specification of R_t is similar in spirit to the specification of H_t in (5), by taking into account that $E(R_t)$ is not equal to $E(RL_t)$, like $E(RC_t)$ is not equal to $E(H_t)$.

Since $\bar{R}$, $\bar{P}$, and RL_{t-1} have unit diagonal elements, and assuming that the initial matrix R_0 is a correlation matrix, it is obvious that R_t has unit diagonal elements, but to be a well-defined correlation matrix it must be positive definite (PD). This is not necessarily the case for the set of values of (α_r, β_r) stated above. The issue is illustrated in Section B of the supplemental appendix. For

estimation, we proceed (like for the BEKK-HEAVY model) by checking that R_t (rather than $\tilde{R}$) is positive definite in the sample period during the numerical maximizing of the log-likelihood function. Like for the BEKK-HEAVY model, for the datasets we used, no issue of a lack of PDness occurred.

An advantage of the proposed specifications of R_t, based on the realized correlation matrix, is that there is no need to transform the covariance matrix of the degarched returns into a correlation one, as in the DCC-GARCH model of Engle (2002), which is specified by

$$R_t = \{\text{diag}(Q_t)\}^{-1/2} Q_t \{\text{diag}(Q_t)\}^{-1/2}, \tag{13}$$

with the $k \times k$ symmetric PD matrix

$$Q_t = (1 - \alpha_q - \beta_q)\bar{Q} + \alpha_q u_{t-1} u'_{t-1} + \beta_q Q_{t-1}, \tag{14}$$

where $\alpha_q \geq 0$, $\beta_q \geq 0$, $\beta_q = 0$ if $\alpha_q = 0$, $\alpha_q + \beta_q < 1$, and $\bar{Q}$ is a $k \times k$ PD matrix. Q_t is actually the conditional covariance matrix of u_t, with non-unit diagonal elements, and by (13), it is transformed into a correlation matrix. However, this parameterization raises two issues, one about estimation, the other about forecasting.

First, because $E(u_t u'_t | \mathcal{F}_{t-1}) = R_t \neq Q_t$, $\bar{Q}$ is not equal to $E(Q_t)$ and therefore is not consistently estimated by $\sum_t u_t u'_t / T$. Thus, using this average to estimate $\bar{Q}$ as proposed by Engle (2002), so that only α_q and β_q remain to be estimated by QML, introduces an asymptotic bias in their estimator. See Aielli (2013) for details and an alternative formulation of (14) that avoids this problem.

Second, at date t, the $(s+1)$-step-ahead forecast of R_{t+s+1} requires $E_t(u_{t+s} u'_{t+s}) = E_t(R_{t+s})$, which is not available in closed form due to the nonlinear relation (13) between R_{t+s} and Q_{t+s}. By assuming $E_t(R_{t+s}) \approx E_t(Q_{t+s})$, Engle and Sheppard (2001) obtain the closed-form forecast recurrence relation

$$E_t(R_{t+s+1}) = \bar{Q} + (\alpha_q + \beta_q) E_t(R_{t+s} - \bar{Q}) \tag{15}$$

that starts with $E_t(R_{t+1}) = R_t$. The correlation forecasts are thus approximate and biased.

The DCC-HEAVY model differs from the DCC-GARCH model in three ways: 1) the dynamics of conditional variances h_t are driven by the lagged realized variances v_{t-1}; 2) the conditional correlation R_t is modeled directly rather than parameterized in a sandwich form as in (13); and 3) the dynamics of the conditional correlation matrix R_t are driven by the lagged realized correlation matrix. The last two features allow us to obtain exact closed forms for s-step-ahead correlation forecasts, as explained in Section 3.

2.2.2. DCC-HEAVY specification of M_t

The specification of the DCC-HEAVY model requires defining the dynamics of

$$M_t = \text{Diag}(m_t^{1/2}) P_t \, \text{Diag}(m_t^{1/2}), \tag{16}$$

where m_t is the vector containing the main diagonal of M_t, that is, the conditional means of the realized variances, and P_t is the corresponding conditional mean of the realized correlation matrix.

The conditional expectation of the realized variances is specified as

$$m_t = \omega_m + A_m v_{t-1} + B_m m_{t-1}, \tag{17}$$

where ω_m is a positive $k \times 1$ vector, and A_m and B_m are $k \times k$ matrices that are restricted to be diagonal matrices with positive entries, as discussed after (10).

The dynamic process for the conditional expectation of the realized correlation matrix is defined in the following way:

$$P_t = (1 - \alpha_p - \beta_p)\bar{P} + \alpha_p RL_{t-1} + \beta_p P_{t-1}, \tag{18}$$

where $\alpha_p \geq 0$, $\beta_p \geq 0$, $\beta_p = 0$ if $\alpha_p = 0$, $\alpha_p + \beta_p < 1$, and $\bar{P}$ is a correlation matrix that is the unconditional mean of RL_t. The elements of $\bar{P}$ can be set to their empirical counterpart to render the estimation simpler. $E(RL_t)$ is not equal to the unconditional correlation matrix $E(P_t)$, due to the nonlinearity of the transformation from covariance to correlation. However, Bauwens, Storti, and Violante (2012) show that if RC_t is computed from a large enough number of high-frequency returns, $\bar{P}$ should be almost equal to $E(RL_t)$.

Eqs. (10)–(12) and (17)–(18) define the DCC-HEAVY model. By setting $\alpha_r = \beta_r = \alpha_p = \beta_p = 0$, the model simplifies into a constant conditional correlation HEAVY model. Estimation is discussed in Section 4.

2.3. DECO-HEAVY model

The DECO-HEAVY model differs from the DCC-HEAVY in the specification of the conditional correlation matrix corresponding to H_t and of the conditional mean of the realized correlation matrix corresponding to M_t.

The specification of the conditional correlation matrix corresponding to H_t, denoted by R_t^E, is based on the assumption that all the conditional correlations are the same time-varying correlation $\rho_t \in (-1/(k-1), 1)$, chosen to be the average of the correlation coefficients of $R_t = (r_{t,ij})$, defined by (11)–(12):

$$R_t^E = (1 - \rho_t^E) I_k + \rho_t^E J_k, \tag{19}$$

$$\rho_t^E = \frac{2}{k(k-1)} \sum_{i>j} r_{t,ij}. \tag{20}$$

where J_k is a $k \times k$ matrix of ones. The DECO-GARCH model of Engle and Kelly (2012) uses as the dynamic equicorrelation coefficient the average of the DCC-GARCH correlations defined by (13) together with the modification of (14) proposed by Aielli (2013).

Likewise, the conditional mean of the realized correlation matrix corresponding to M_t, denoted by P_t^E, is specified as:

$$P_t^E = (1 - \rho_t^P) I_k + \rho_t^P J_k, \tag{21}$$

$$\rho_t^P = \frac{2}{k(k-1)} \sum_{i>j} p_{t,ij}. \tag{22}$$

The main advantage of DECO with respect to DCC, especially when k is very large, is the availability of analytical expressions of the inverse and determinant of the equicorrelated matrices, which are used in the computation of the likelihood function for estimation, and of economic loss functions for forecast evaluations. For more details, see Engle and Kelly (2012).

3. Representation, forecasting, and extensions

In this section, the DCC-HEAVY and the closely related DCC-GARCH and DCCX-GARCHX models are represented as multiplicative error models (MEMs), from which stationarity conditions and closed-form formulas for multi-step forecasts follow directly. An extension of the DCC-HEAVY model is proposed by adding asymmetric impact and HAR terms.

3.1. Multiplicative error representation

An MEM for a positive variable x_t specifies it as the product of a positive conditional mean and a positive error that follows some distribution with expectation equal to one. If x_t is a squared centered return, the conditional mean is the conditional variance. This can be extended to the elements of a vector. In this subsection, the focus is on the conditional expectation formulation, and not on the distribution.

For the conditional and realized variance equations, define the vectors of $2k \times 1$ elements $x_t = [(r_t^2)', v_t']'$ and $\mu_t = [h_t', m_t']'$. The conditional expectation formulation of the conditional and realized variance Eqs. (10) and (17) is

$$\begin{aligned} E(x_t|\mathcal{F}_{t-1}) &:= \mu_t, \\ \mu_t &= \omega + Ax_{t-1} + B\mu_{t-1}, \end{aligned} \tag{23}$$

where

$$\omega = \begin{bmatrix} \omega_h \\ \omega_m \end{bmatrix}, A = \begin{bmatrix} 0 & A_h \\ 0 & A_m \end{bmatrix}, B = \begin{bmatrix} B_h & 0 \\ 0 & B_m \end{bmatrix} \tag{24}$$

and the 0 symbol stands for a $k \times k$ matrix of zeros. If

$$A = \begin{bmatrix} A_h & 0 \\ 0 & A_m \end{bmatrix}, \tag{25}$$

the top part becomes the DCC-GARCH model, and the realized variance model is kept in the bottom part. If

$$A = \begin{bmatrix} A_h & A_{hm} \\ 0 & A_m \end{bmatrix}, \tag{26}$$

the model (referred to as DCC-GARCHX) includes in the top part both the lagged squared return and the lagged realized variance, encompassing the two previous models.

Defining the $2k \times k$ matrices $Y_t = [u_t u_t', RL_t]'$ and $\Phi_t = [R_t, P_t]'$, the conditional expectation formulation of the conditional and realized correlation matrix Eqs. (11) and (18) is

$$\begin{aligned} E(Y_t|\mathcal{F}_{t-1}) &:= \Phi_t, \\ \Phi_t &= \Omega + (\boldsymbol{\alpha} \otimes J_k)Y_{t-1} + (\boldsymbol{\beta} \otimes J_k)\Phi_{t-1}, \end{aligned} \tag{27}$$

where

$$\Omega = \begin{bmatrix} \tilde{R} \\ \tilde{P} \end{bmatrix}, \boldsymbol{\alpha} = \begin{bmatrix} 0 & \alpha_r \\ 0 & \alpha_p \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_r & 0 \\ 0 & \beta_p \end{bmatrix},$$

and the 0 symbol is scalar in this case. If $\boldsymbol{\alpha} = \begin{bmatrix} \alpha_q & 0 \\ 0 & \alpha_p \end{bmatrix}$, $\boldsymbol{\beta} = \begin{bmatrix} \beta_q & 0 \\ 0 & \beta_p \end{bmatrix}$, Q_t replaces R_t in the definition of Φ_t, and the constant term is adapted, then the top part becomes the "quasi-correlation" Eq. (14) of the DCC-GARCH model, the realized correlation equation being kept in the bottom part. If

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_q & \alpha_r \\ 0 & \alpha_p \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_r & 0 \\ 0 & \beta_p \end{bmatrix}, \tag{28}$$

both $u_{t-1}u_{t-1}'$ and RL_{t-1}' are included in the top part (referred to as DCCX-GARCH).

Processes such as defined by (23) and (27) can be written as VARMA(1,1) by defining appropriate error terms. From this representation, it follows that the covariance stationarity condition is that the largest eigenvalue of $A + B$ is smaller than unity for (23), and likewise for the largest eigenvalue of $\boldsymbol{\alpha} + \boldsymbol{\beta}$ for (27). For the matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ defined after (27), the previous condition is equivalent to $\beta_r < 1$ and $\alpha_p + \beta_p < 1$, as written after (11) and (18). The unconditional first moment is then obtained by applying standard results for the VARMA representations. For example, $E(x_t) = (I_{2k} - A - B)^{-1}\omega$, which gives $E(v_t) = (I_k - A_m - B_m)^{-1}\omega_m$ and $E(r_t^2) = (I_k - B_h)^{-1}[\omega_h + A_h E(v_t)]$ for the variances of the DCC-HEAVY model. For the correlations, $E(RL_t) = \bar{P}$ and $E(u_t u_t') = [\bar{R} + \alpha_r E(RL_t)]/(1 - \beta_r) = \bar{R}$, the last equality resulting from (12).

3.2. Multiple-step-ahead forecasting

Forecasts of the conditional covariance matrices of daily returns are used in several financial applications. The s-step-ahead forecast of H_{t+s}, computed at date t, is defined in the case of DCC-type models as

$$H_{t+s|t} = \text{Diag}\left\{[E_t(h_{t+s})]^{1/2}\right\} E_t(R_{t+s}) \text{Diag}\left\{[E_t(h_{t+s})]^{1/2}\right\}, \tag{29}$$

where $E_t(.)$ is a short notation for $E(.|\mathcal{F}_{t-1})$. Notice that as in the DCC model, $H_{t+s|t}$ is not equal to $E_t(H_{t+s})$, due to the nonlinearity of the transformation of covariances into correlations and of the square root function.

To obtain $E_t(h_{t+s})$ and $E_t(R_{t+s})$, the conditional expectation expressions of the previous subsection are useful to compute $E_t(\mu_{t+s})$ and $E_t(\Phi_{t+s})$, denoted by $\mu_{t+s|t}$ and $\Phi_{t+s|t}$, respectively.

Starting from (23) leads to $E_t(\mu_{t+s})$. In moving more than one step ahead, $x_{t+s|t}$ is not known and needs to be substituted with its corresponding conditional expectation $\mu_{t+s|t}$. Hence,

$$\begin{aligned} \mu_{t+1|t} &= \omega + Ax_t + B\mu_t, \\ \mu_{t+s|t} &= \omega + (A + B)\mu_{t+s-1|t} \quad \text{for } s \geq 2, \end{aligned} \tag{30}$$

which can be solved recursively, giving the closed-form forecast

$$\mu_{t+s|t} = \tilde{\omega} + C^{s-1}\mu_{t+1|t} \quad \text{for } s \geq 2,$$

where $\tilde{\omega} = (I - C)^{-1}(I - C^{s-1})\omega$ and $C = A + B$.

Proceeding in the same way for $E_t(\Phi_{t+s})$ from (27) gives

$$\begin{aligned} \Phi_{t+1|t} &= \Omega + (\boldsymbol{\alpha} \otimes J_k)Y_t + (\boldsymbol{\beta} \otimes J_k)\Phi_t, \\ \Phi_{t+s|t} &= \Omega + ((\boldsymbol{\alpha} + \boldsymbol{\beta}) \otimes J_k)\,\Phi_{t+s-1|t} \quad \text{for } s \geq 2. \end{aligned} \tag{31}$$

The closed-form forecast is

$$\Phi_{t+s|t} = \tilde{\Omega} + \left(c^{s-1} \otimes J_k\right)\Phi_{t+1|t} \text{ for } s \geq 2,$$

where $\tilde{\Omega} = (I - c)^{-1}(I - c^{s-1})\Omega$ and $c = \boldsymbol{\alpha} + \boldsymbol{\beta}$.

For instance, the DCC-HEAVY model s-step-ahead forecasts $\mu_{t+s|t}$ and $\Phi_{t+s|t}$ are derived from (30) and (31) by setting A, B, $\boldsymbol{\alpha}$, and $\boldsymbol{\beta}$ to the matrices defined after (23) and (27). The s-step-ahead forecast $E_t[h_{t+s}]$ of the conditional variance vector corresponds to the first k elements of $\mu_{t+s|t}$, and the s-step forecast $E_t[R_{t+s}]$ of the conditional correlation corresponds to the $k \times k$ upper block of $\Phi_{t+s|t}$.

3.3. Model extensions

It has long been known that stock markets react differently to positive and negative news. The asymmetric effect is now commonly used to refer to any volatility model, univariate and multivariate alike, in which the (co)variances respond asymmetrically to positive and negative shocks. The DCC-HEAVY model can be extended by incorporating the asymmetric effect into the variance and correlation equations. In the variance equation, the asymmetric effect implies that volatility tends to increase more following negative return shocks than equally sized positive shocks. In the correlation equation, the asymmetric effect implies that the correlation between stock returns tends to increase when the market turns down. The extended model is called the ADCC-HEAVY model.

Define $d_t = (d_{1t}, d_{2t}, \ldots, d_{kt})'$, where $d_{it} = 1$ if $r_{it} < 0$, and $d_{it} = 0$ if $r_{it} \geq 0$, for $i = 1, \ldots, k$, and $D_t = d_t d_t' - \mathrm{diag}(d_t d_t')$ (a matrix with diagonal elements equal to 0). The conditional variance and correlation equations of the ADCC-HEAVY model are then

$$h_t = \omega_h + A_h v_{t-1} + B_h h_{t-1} + \Gamma_h d_{t-1} \odot v_{t-1}, \tag{32}$$

$$R_t = \tilde{R} + \alpha_r RL_{t-1} + \beta_r R_{t-1} + \gamma_r D_{t-1} \odot RL_{t-1}, \tag{33}$$

$$\tilde{R} = (1 - \beta_r)\bar{R} - (\alpha_r \bar{P} + \gamma_r \bar{D} \odot \bar{P}), \tag{34}$$

where Γ_h is a $k \times k$ diagonal matrix, and $\bar{D}$ is the sample mean of D_t. Eq. (32) extends (10); asymmetric effects correspond to positive values on the diagonal of Γ_h. Eqs. (33)–(34) extend (11)–(12). The asymmetric effect corresponds to a positive value of γ_r: the impact of the lagged realized correlation between assets i and j on their current conditional correlation is equal to $\alpha_r + \gamma_r$ only if both $r_{i,t-1}$ and $r_{j,t-1}$ are negative. Otherwise, the impact is reduced to α_r if γ_r is positive. Notice that the diagonal elements of $D_{t-1} \odot RL_{t-1}$ and $\bar{D} \odot \bar{P}$ are equal to one, so that the same holds for R_t. Like in the simpler model where $\gamma_r = 0$, in estimation, $(\alpha_r, \beta_r, \gamma_r)$ are constrained to values such that R_t is PD for all t. In the empirical applications, this did not create any difficulty.

The same asymmetric effects can be included in the realized variance and correlation equations. Furthermore, the heterogeneous autoregressive (HAR) model of Corsi (2009) has emerged as a simple and powerful way to include the long-memory feature of realized volatilities. The model was extended to the multivariate setting by Chiriac and Voev (2011) and Oh and Patton (2016). Adding HAR terms to the realized variance and correlation Eqs. (17) and (18), respectively, results in richer dynamic equations:

$$m_t = \omega_m + A_m v_{t-1} + B_m m_{t-1} + \Gamma_m d_{t-1} \odot v_{t-1} + A_m^w v_{t-1}^w + A_m^m v_{t-1}^m, \tag{35}$$

$$P_t = \tilde{P} + \alpha_p RL_t + \beta_p P_{t-1} + \gamma_p D_{t-1} \odot RL_{t-1} + \alpha_p^w RL_{t-1}^w + \alpha_p^m RL_{t-1}^m, \tag{36}$$

where Γ_m, A_m^w, and A_m^m are $k \times k$ diagonal matrices, $v_{t-1}^w = \frac{1}{5}\sum_{j=1}^{5} v_{t-1-j}$, $v_{t-1}^m = \frac{1}{22}\sum_{j=1}^{22} v_{t-1-j}$, $RL_{t-1}^w = \frac{1}{5}\sum_{j=1}^{5} RL_{t-1-j}$, $RL_{t-1}^m = \frac{1}{22}\sum_{j=1}^{22} RL_{t-1-j}$, and $\tilde{P} = (1 - \alpha_p - \beta_p - \gamma_r \bar{D}^r - \alpha_p^w - \alpha_p^m)\bar{P}$. Although P_t has unit diagonal elements, in estimation, the parameter space must be constrained to ensure that this matrix is PD for all t. HAR terms are not added to the conditional covariance and correlation equations, since these effects are insignificant in the empirical application.

Other ways to define the asymmetric effect have been used. For example, one can define $d_{it} = 1$ when the stock volatility $RC_{ii,t}$ is "high" (above some threshold). In the correlation equation, that implies that the correlation between two stock returns increases more when both stocks are highly volatile than if neither or only one is highly volatile. Another asymmetric effect is to let all correlations increase if the market volatility (for instance measured by VIX) increases; see Bauwens and Otranto (2016).

Section C of the supplemental appendix provides the formulas of multiple-step forecasts of the extended models.

4. Estimation

The DCC- and DECO-HEAVY models are parameterized with a finite-dimensional $p \times 1$ parameter vector $\theta \in \Theta \subset \mathbb{R}^p$. Partitioning θ' into (θ_H', θ_M'), where the $p_H \times 1$ vector θ_H is the parameter vector of the HEAVY model for H_t, and the $p_M \times 1$ vector θ_M is the parameter vector of the HEAVY model for M_t, θ_H and θ_M can be estimated separately, as they are variation-free in the sense of Engle, Hendry, and Richard (1983). Moreover, each of these estimations can be split into two steps, as explained below.

4.1. Estimation of θ_H

To get a quasi-likelihood function, we add to assumption (8)–(9) the hypothesis that the distribution of the innovation of the return vector is multivariate Gaussian:

$$u_t | \mathcal{F}_{t-1} \sim N(0, R_t), \tag{37}$$

implying that $r_t | \mathcal{F}_{t-1} \sim N(0, H_t)$, with H_t defined in (7).

Neglecting irrelevant constants, the quasi-log-likelihood function for T observations, given initial values, is

$$QL_H(\theta_H) = -\frac{1}{2}\sum_{t=1}^{T}\left(\log|H_t| + r_t' H_t^{-1} r_t\right) = -\frac{1}{2}\sum_{t=1}^{T}\left(2\log|\mathrm{Diag}(h_t^{1/2})| + \log|R_t| + u_t' R_t^{-1} u_t\right). \tag{38}$$

This function can be maximized numerically in a single step, but for large k, the large dimension of the parameter space makes this difficult. A two-step estimation can be based on a partition of θ_H into θ_{H1}, the parameters of the variance Eq. (10), and θ_{H2}, the parameters of the correlation Eq. (11). The two-step procedure was proposed by Engle (2002) for the DCC-GARCH model.

The first step consists in estimating θ_{H1} by maximizing the quasi-log-likelihood obtained by replacing R_t in the second line of (38) by the identity matrix, so that the objective function does not depend on θ_{H2}:

$$QL_{H1}(\theta_{H1}) = -\frac{1}{2}\sum_{t=1}^{T}\left(2\log|\mathrm{Diag}(h_t^{1/2})| + u_t'u_t\right). \tag{39}$$

The matrix A_h of (10) can be non-diagonal to allow spillover effects, but the matrix B_h must be diagonal to allow a separate estimation of the k equations. In the case of GARCH models with the same structure, Francq and Zakoian (2016) provide the asymptotic properties of the separate (equation-by-equation) estimation of the variance equations. In our empirical analysis, both A_h and B_h are restricted to be diagonal.

The second step maximizes (38) with respect to θ_{H2}, fixing θ_{H1} to the value $\hat{\theta}_{H1}$ obtained at the first step. The second step objective function can be written as

$$QL_{H2}(\theta_{H2};\hat{\theta}_{H1}) = -\frac{1}{2}\sum_{t=1}^{T}\left(\log|R_t| + \hat{u}_t'R_t^{-1}\hat{u}_t\right), \tag{40}$$

where $\hat{u}_t = r_t \odot \hat{h}_t^{-1/2}$, and $\hat{h}_t$ means that h_t defined in (10) is evaluated at $\theta_{H1} = \hat{\theta}_{H1}$.

The two-step estimator is presumably a consistent but inefficient estimator of the parameter θ_H under conditions (such as asymptotic identification, strict stationarity, and ergodicity of the time-series processes) similar to those stated in Engle and Sheppard (2001) for the DCC-GARCH model. The two-step estimation method is numerically tractable for large k, if $\bar{R}$ and $\bar{P}$ are targeted, i.e., estimated by their empirical counterparts, as proposed below (12). The impact of the targeting is a loss of efficiency. Noureldin et al. (2012) discuss this issue for the BEKK-HEAVY model, which is estimated in a single step. Further work to derive rigorously the asymptotic distribution of the single- and two-step estimators of the DCC-HEAVY, with and without targeting, is needed and beyond the scope of this paper. In Section 4.3, we provide the results of a simulation study of the properties of the two-step estimator.

4.2. Estimation of θ_M

Since a realized covariance matrix is symmetric positive definite, a natural choice of distribution to form a likelihood function for the realized covariance process is the Wishart distribution. This assumption has been used in several papers, e.g., Bauwens, Braione, and Storti (2016), Bauwens et al. (2012), Chiriac and Voev (2011), Golosnoy, Gribisch, and Liesenfeld (2012), Gouriéroux, Jasiak, and Sufana (2009), and Noureldin et al. (2012).

We assume that conditionally on the past information set, RC_t follows a central Wishart distribution of dimension k, and we denote this assumption by

$$RC_t|\mathcal{F}_{t-1} \sim W_k(\nu, M_t/\nu), \tag{41}$$

where ν is the degrees-of-freedom parameter restricted by $\nu > k-1$. The chosen parameterization implies that

$$E(RC_t|\mathcal{F}_{t-1}) := M_t.$$

Using the expression of a Wishart density function, and of M_t in (16), the quasi-log-likelihood function for a sample of T observations, given initial conditions, is

$$\begin{aligned} QL_M(\theta_M) &= -\frac{\upsilon}{2}\sum_{t=1}^{T}\left[\log|M_t| + \mathrm{tr}(M_t^{-1}RC_t)\right] \\ &= -\frac{\upsilon}{2}\sum_{t=1}^{T}\left\{2\log|D_{m,t}| + \log|P_t| + \mathrm{tr}[(D_{m,t}P_tD_{m,t})^{-1}RC_t]\right\}, \end{aligned} \tag{42}$$

where θ_M is the vector of the parameters that appear in (17) and (18), and $D_{m,t}$ stands for $\mathrm{Diag}(m_t^{1/2})$. In the above expression, terms that depend on ν but do not depend on θ_M are not included. The parameter ν is considered a nuisance parameter that can be neglected to estimate θ_M, and practically it can be set to unity without loss of information, since the score for θ_M is proportional to the value of this parameter.

As shown by Bauwens et al. (2012), the Wishart assumption provides a quasi-likelihood function, which can serve as an objective function to get a single-step estimator. They also show that the DCC-HEAVY model part for the realized covariance matrix can be estimated in two steps. The parameter space θ_M is split into θ_{M1} for the parameters in the realized volatility model and θ_{M2} for the parameters in the realized correlation model. We denote by QL_{M1} the quasi-log-likelihood where P_t in (42) is replaced by the identity matrix and ν is set to unity:

$$QL_{M1}(\theta_{M1}) = -\frac{1}{2}\sum_{t=1}^{T}\left[2\log|D_{m,t}| + \mathrm{tr}(D_{m,t}^{-1}RC_tD_{m,t}^{-1})\right]. \tag{43}$$

This estimation is split into k separate estimations when the matrices A_m and B_m of (17) are restricted to be diagonal.

We denote by QL_{M2} the quasi-log-likelihood where θ_{M1} is fixed at the value $\hat{\theta}_{M1}$ obtained at the first step:

$$\begin{aligned} &QL_{M2}(\theta_{M2};\hat{\theta}_{M1}) \\ &\quad = -\frac{1}{2}\sum_{t=1}^{T}\left\{\log|P_t| + \mathrm{tr}[(P_t^{-1} - I_k)\hat{D}_{m,t}^{-1}RC_t\hat{D}_{m,t}^{-1}]\right\}, \end{aligned} \tag{44}$$

where $\hat{D}_{m,t}$ means that $\hat{D}_{m,t}$ is evaluated at $\theta_{M1} = \hat{\theta}_{M1}$. The parameter vector θ_{M2} includes α_p, β_p, and the elements of $\bar{P}$. The latter can be targeted by the unconditional mean of the realized correlations, as discussed after (18), in which case the second-step maximization is done with respect to two parameters and is therefore feasible for large k.

4.3. Properties of the QML estimator by simulation

Shephard and Sheppard (2010) provide the formula of the asymptotic Gaussian distribution and its covariance matrix for the QML estimator of the univariate HEAVY model, with and without targeting. Noureldin et al. (2012) provide the same type of result for the BEKK-HEAVY model. The asymptotic theory of Shephard and Sheppard (2010) is relevant for the estimators of the parameters of the variance equations of the DCC-HEAVY model, since these equations are univariate models. A complete

Table 1
Simulation results for the two-step QMLE of the DCC-HEAVY parameters.

	$T = 2000$				$T = 4000$			
	A_h	B_h	A_m	B_m	A_h	B_h	A_m	B_m
Min RB	−2.83	−4.89	−0.18	−2.11	−2.02	−2.31	−0.09	−1.13
Med RB	2.15	−3.47	0.23	−0.53	1.28	−1.73	0.12	−0.25
Max RB	13.44	8.04	3.02	−0.27	6.32	6.10	1.61	−0.12
Min RMSE x100	19.28	12.91	1.73	1.84	12.73	8.13	1.22	1.27
Med RMSE x100	25.04	16.41	1.87	2.01	17.69	11.54	1.31	1.43
Max RMSE x100	33.43	19.69	2.68	2.70	23.94	14.51	1.62	1.63
	α_r	β_r	α_m	β_m	α_r	β_r	α_m	β_m
True	0.069	0.866	0.042	0.946	0.069	0.866	0.042	0.946
RB	−3.229	−0.031	−0.139	−0.199	−1.534	−0.037	−0.416	−0.095
RMSE x100	0.463	0.882	0.041	0.196	0.290	0.593	0.033	0.099
N. test	0.001	0.01	0.62	0.82	0.32	0.50	0.91	0.57

Min: minimum; Max; maximum; Med: median; RB: relative bias (in percentages); RMSE: root mean squared error; see the definitions in (45) For the variance parameter vectors (A_h, B_h, A_m, B_m): the values are the minimum, median, and maximum of the RB and RMSE across the 29 parameters in each vector. N. test: p-value of the Jarque–Bera normality test of the sampling distribution. The reported values are obtained from 1000 simulated samples of the data generating process defined in Section 4.3.

asymptotic distribution theory for the single-step and two-step QML estimators of θ_H and θ_M of the DCC-HEAVY model, with or without targeting, is beyond the scope of this paper. Some results of a simulation study are reported below, focusing on the bias, the root mean squared error, and the normality of the sampling distribution of the two-step estimator in the case of targeting, as defined in the previous subsections and used in Section 5.

The parameters of the process from which the data are generated are taken from the empirical results reported in the next section. The data-generating process is defined by (16)–(18) for the M_t part; once m_t is computed as (17) and P_t as (18), M_t is computed as defined by (16). Next, a realized covariance matrix RC_t is simulated according to (41), where ν is set to 50. The corresponding diagonal v_t and correlation matrix RL_t are the "observed" (simulated) data for date t and are also used as inputs to compute m_{t+1}, P_{t+1}, h_{t+1}, and R_{t+1}. For the H_t part, H_t is simulated as $N_k(0, H_t)$, where H_t is computed as defined by (7); for this, h_t is computed as in (8) and R_t as in (11)–(12). The dimension k is equal to 29, like in the empirical application of Section 5. A sample of T observations is generated and used for estimation; we set $T = 2000$ and 4000 to illustrate the impact of increasing the sample size. The largest sample size of 4000 is chosen close to the sample size of the application in Section 5, and the smallest is close to the sample size of the application reported in Section F of the supplemental appendix. For each T, the data simulation and the parameter estimation is repeated 1000 times, delivering 1000 estimates of each parameter.

Table 1 provides a synthetic view of the simulation results. The relative bias (RB) and root mean squared error (RMSE) are defined as

$$RB(\hat{\phi}) = 100 \times \frac{1}{S}\sum_{s=1}^{S}\frac{(\hat{\phi}_s - \phi_0)}{\phi_0},$$
$$RMSE(\hat{\phi}) = \sqrt{\frac{1}{S}\sum_{s=1}^{S}(\hat{\phi}_s - \phi_0)^2}, \tag{45}$$

for the estimator $\hat{\phi}$ of the parameter ϕ (an element of θ) having the true value ϕ_0 and estimated by $\hat{\phi}_s$ for the sth simulated dataset. The RMSE values indicated in the table must be divided by 100 to obtain the actual values. Some comments about the results follow.

The RMSE in absolute values and the relative biases decrease (with two minor exceptions for the RMSE)) as T increases, which indicates the likely consistency of the estimator. Each RMSE for $T = 4000$ is approximately equal to (actually, a bit smaller than) the corresponding one for $T = 2000$ divided by the square root of 4000/2000, which indicates that convergence occurs at the rate $\sqrt{T}$. The Jarque–Bera normality tests of the sampling distributions of the correlation parameter estimators indicate that normality is not rejected (except for α_r and β_r at $T = 2000$). In brief, the results indicate that the consistency and the asymptotic normality are likely properties of the QML estimator; notice that the QML estimator is actually an ML estimator, since the estimated model is correctly specified in the simulation study.

For the estimators of the parameters of the variance equations, the relative median biases and RMSE are generally small in absolute values; for $T = 4000$, the largest median bias is −1.73%. The results for the estimator of each equation are in Tables D1 and D2 of Section D of the supplemental appendix. For the correlation parameter estimators, the largest bias is −1.53% at $T = 4000$, and the other biases are smaller than 1%. The RMSEs are small and of the same order of magnitude as the standard errors of the estimates of Section 5: for example, for $\hat{\alpha}_h$, the RMSE is equal to 0.0029 in Table 1 for $T = 4000$, and the standard error is equal to 0.0065 (based on Table 3) for $T = 4318$.

5. Empirical application

5.1. Data description

High-frequency data for 29 stocks belonging to the Dow Jones Industrial Average (DJIA) index are used; the 30th stock was dropped, since it was not permanently in the index during the sample period. The sample period

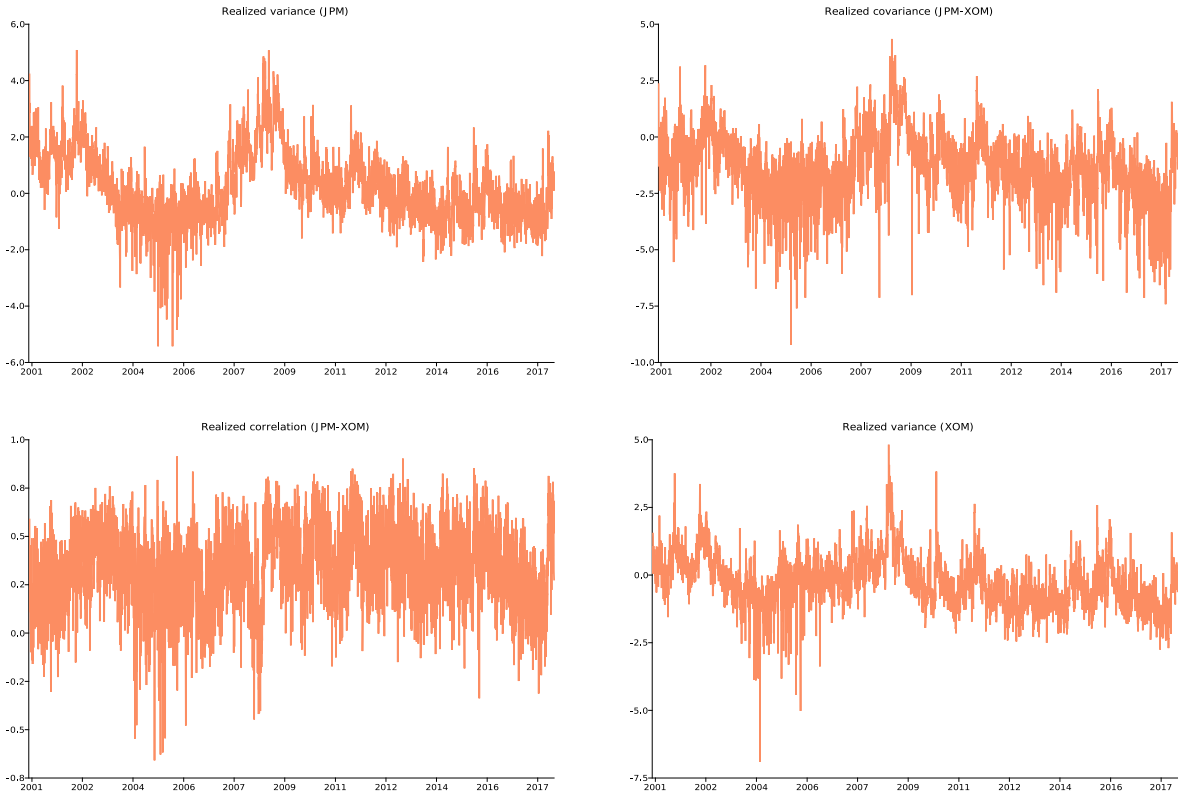

Fig. 1. The figure shows the log realized volatilities of JP Morgan (JPM) and Exxon Mobil (XOM), the log realized covariance, and the realized correlation between these stocks, over the sample period 03/01/2001–16/04/2018.

is 3 January 2001 to 16 April 2018 with a total of 4318 trading days, and the data source is the TAQ database.

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), and Exxon Mobil Corporation (XOM).

The daily realized covariance matrices are computed as explained at the beginning of Section 2, using five-minute returns. The synchronization of intraday prices of the 29 stocks was done using five-minute intervals; the price closest (from the left) to the respective sampling point was taken; and the first and last 15 min of the day (9:30–16:00) were excluded, so that $m = 72$ in (1).

Descriptive statistics are provided in Table E1 of the supplemental appendix. For each stock, it reports the time-series averages and standard deviations of its squared returns and realized variances, and the means and standard deviations of the time-series averages of its realized covariances and correlations with the other 28 stocks. Each average realized variance does not account for the overnight variation and is therefore a fraction (in most cases 50 to 60 percent) of the corresponding average squared return.

Fig. 1 shows a representative example of time-series plots of the realized variances of two stocks (JPM and XOM) and the corresponding realized covariances and correlations.

The focus of the empirical application is a forecasting comparison of the conditional covariance and correlation matrices, and the conditional variances of the 29 stocks using a set of models. Before reporting the results of the comparisons, estimation results are reported for the DCC and DECO models.

5.2. Estimation results for the full period

Table 2 presents summary statistics (median, minimum, and maximum) of the first-step parameter estimates (except the constant terms) of the 29 variance

Table 2
Summary of parameter estimates of the variance equations for 29 stocks.

	GARCH (23)-(25)		GARCHX (23)-(26)			HEAVY-h (10)		HEAVY-m (17)	
	A_h	B_h	A_h	A_{hm}	B_h	A_h	B_h	A_m	B_m
Min	0.024	0.830	0.002	0.103	0.199	0.353	0.199	0.297	0.514
Med	0.072	0.916	0.028	0.698	0.507	0.781	0.481	0.369	0.606
Max	0.117	0.971	0.068	1.748	0.885	1.748	0.775	0.470	0.695
λ_{max}	0.995		0.996				0.996		
λ_{min}	0.927		0.199				0.199		

Minimum (Min), median (Med), and maximum (Max) are the summary statistics of the estimates; λ_{max} and λ_{min} are, respectively, the largest and smallest eigenvalues of the corresponding $A+B$ matrix; see (23)–(26) for their definition for the GARCH, GARCHX, and HEAVY models. All estimates are provided in Table E2 of the supplementary appendix. For A_h of GARCHX, the statistics are for the 21 non-zero estimates (8 estimates are equal to 0).

equations of the GARCH, GARCHX, HEAVY-h, and HEAVY-m models. The estimates for each stock, and the associated robust t-statistics, are given in Table E2 of the supplemental appendix.

The estimates of the B_h parameters in the HEAVY-h model are smaller than in the GARCH model (median of 0.481 versus 0.916), while the estimates of the A_h parameters are much larger (medians of 0.781 versus 0.072). The influence of these differences is visible in the bottom panel of Fig. 2, which shows (in log) the time series of the average (over the 29 stocks) of the corresponding fitted conditional variances. The GARCH path is smoother than the HEAVY-h path, and the latter fluctuates locally more strongly than the former, responding faster to recent changes of volatility. Similar differences occur for each stock. The GARCH parameter estimates are much more homogeneous across the different stocks than the HEAVY-h estimates. The values of the HEAVY-m parameter estimates are more similar to HEAVY-h than to GARCH, though they are much more homogeneous than for HEAVY-h.

In the nesting GARCHX model, the coefficient (in A_h) of the lagged squared return is set to zero for eight stocks out of 29 because a non-negativity constraint is imposed and is binding; for these stocks, GARCHX estimates are identical to HEAVY-h. For the 21 other stocks, the estimate is positive (between 0.002 and 0.068), with t-statistics below 1.5 in 12 cases (out of 21), and in four cases larger than 2. Even though the t-statistic has a non-standard distribution, since the null hypothesis of zero is on the boundary of the parameter space, these results suggest that for almost all stocks, the estimate is not significant. On the contrary, the estimate of the lagged realized variance coefficient (in A_{mh}) is positive (between 0.103 and 1.748, with 0.698 as the median value) and the associated t-statistics are usually large enough to suggest they are significant (only three are below 2.5). The log-likelihood gain of GARCHX over HEAVY-h is equal to 5 for the 21 additional parameters, and hence it appears to be minor. On the contrary, the gains of GARCHX and HEAVY-h over GARCH are substantial (41 and 36, respectively). Notice that GARCH and HEAVY-h are not nested, but they have the same number of parameters, so choosing between them using their log-likelihood values is equivalent to a choice based on model choice criteria. In brief, these results suggest that the conditional variance dynamics are better captured by the lagged realized variance than by the lagged squared return, confirming the findings of Shephard and Sheppard (2010).

Table 2 also reports the smallest and largest eigenvalues of the matrix $A+B$ (see Section 3.1) corresponding to each model (GARCH, GARCHX, and HEAVY). The largest eigenvalue is smaller than 1, and hence the covariance stationarity condition is satisfied for the variance equations of each model. For the HEAVY and GARCHX models, the largest eigenvalue comes from the M_t part of the model.

Table 3 presents the second-step parameter estimates of the correlation models: DCC-GARCH, DCCX-GARCH, DCC-HEAVY-R (Eqs. (11)–(12)), DCC-HEAVY-P (Eq. (18)), and the corresponding DECO versions. For the DCC models, these estimates are broadly in line with those of the variance equations. The estimate of β_r in the DCC-HEAVY-R model is smaller than that of β_q in the DCC-GARCH model (0.869 versus 0.988), and the estimate of α_r is larger than that of α_q (0.068 versus 0.003), implying less smooth and more reactive fitted correlations. The paths of average fitted correlations and covariances of the three models are shown on Fig. 2. The paths of the DECO-HEAVY-R model are much less smooth and more reactive to recent information than for DCC-HEAVY-R, due to a smaller estimate of β_r (0.552 versus 0.869) and a larger one of α_r (0.447 versus 0.068). The paths for the DCC-GARCH are smoother and less reactive than in the HEAVY models. The described path differences are stronger for each pair of stocks, since averaging reduces the variability.

In the nesting DCCX- and DECOX-GARCH models, the coefficient estimates of the lagged realized correlation (0.068 and 0.447) are of the same magnitude as in the DCC- and DECO-HEAVY-R (0.061 and 0.369), with large t-statistics (9.40 for DCCX and 5.09 for DECOX). The coefficient estimate of the lagged return cross-product is very close to zero (with a t-statistic of 10.59) in DCCX, while it is equal to 0.029 (with a t-statistic of 1.94) in DECOX. The maximized second-step log-likelihood values of DCCX-GARCH (−5115) and DCC-HEAVY (−5119) are slightly different, but they are much larger than the value of DCC-GARCH (−5143). For the DECO models: DECO-HEAVY (−5307) and DECOX-GARCH (−5306) are very close, but DECO-GARCH (−5319) is lower. Thus, lagged realized correlations can be considered as more important

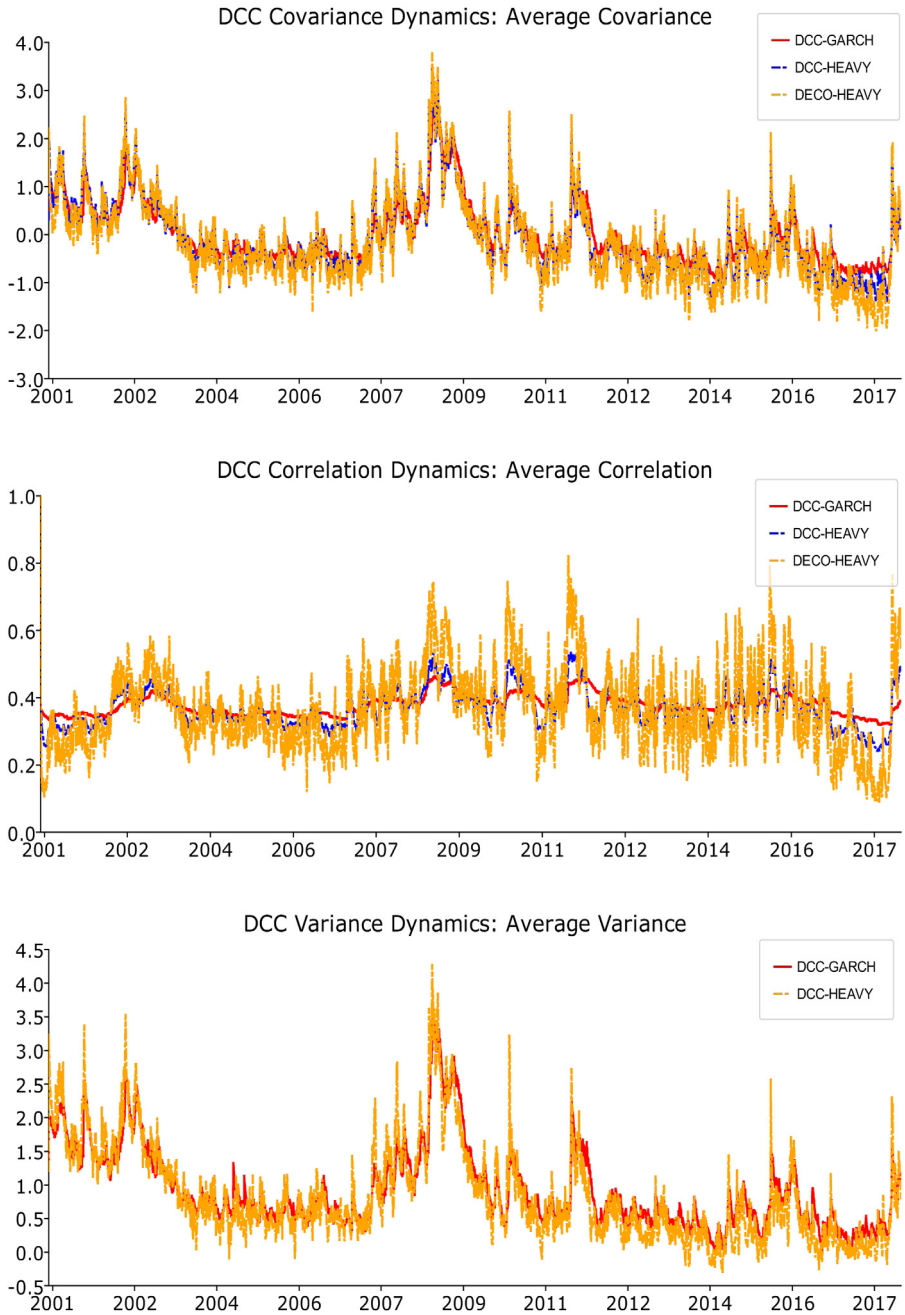

Fig. 2. The figure shows the pairwise average of log-covariances and correlations estimated in the second step with DCC-GARCH and DCC/DECO-HEAVY-H, over the sample period 03/01/2001–16/04/2018. The bottom graph shows the average of the log-variances of the 29 stocks, estimated in the first step of DCC-GARCH and DCC-HEAVY.

drivers of the conditional correlations rather than return cross-products.

The maximized log-likelihood values and their decomposition into the variance and correlation parts are reported in Table 3. The decompositions suggest that the DCC-HEAVY and DECO-HEAVY dominate the DCC-GARCH model in both the variance and correlation parts. However, most of the in-sample gain comes from the variance part. The overall improvement is substantial. DCC-HEAVY improves slightly more than DECO-HEAVY. Notice that these models have the same number of parameters, so comparisons using log-likelihood values are equivalent to comparisons using model choice criteria.

5.3. Forecasting comparisons

A comparison of models can be made by evaluating the in-sample and out-of-sample forecasting performances

Table 3
Parameter estimates (and robust t-statistics) of the correlation equations.

DCC-GARCH (14)		DCCX-GARCH (27)-(28)			DCC-HEAVY-R (11)-(12)		DCC-HEAVY-P (18)	
α_q	β_q	α_q	β_q	α_r	α_r	β_r	α_m	β_m
0.003	0.988	0.006	0.876	0.061	0.068	0.869	0.041	0.948
(8.56)	(457.39)	(10.59)	(56.76)	(9.40)	(10.40)	(54.65)	(30.59)	(506.6)

Log-likelihood decomposition (DCC-HEAVY vs. DCC-GARCH)

	DCC-GARCH	DCCX-GARCHX	DCC-HEAVY	HEAVY Gains
Variance	−7785	−7644	−7649	136
Correlation	−5143	−5115	−5119	24
Total	−12927	−12759	−12768	159

DECO-GARCH		DECOX-GARCH			DECO-HEAVY-R		DECO-HEAVY-P	
α_q	β_q	α_q	β_q	α_r	α_r	β_r	α_m	β_m
0.046	0.937	0.029	0.615	0.369	0.447	0.552	0.347	0.607
(3.25)	(42.66)	(1.94)	(10.02)	(5.09)	(7.97)	(9.86)	(16.33)	(23.80)

Log-likelihood decomposition (DECO-HEAVY vs. DECO-GARCH)

	DECO-GARCH	DECOX-GARCHX	DECO-HEAVY	HEAVY Gains
Variance	−7785	−7644	−7649	136
Correlation	−5319	−5306	−5307	12
Total	−13104	−12950	−12956	148

The HEAVY Gains are the differences of log-likelihood between the DCC/DECO-HEAVY model to the DCC/DECO GARCH model.

of the models using the model confidence set (MCS) of Hansen, Lunde, and Nason (2011). An MCS identifies a set of models having the best forecasting performance at a chosen confidence level, based on a loss function. Six models are compared: DCC-GARCH, DCC-HEAVY, DECO-GARCH, DECO-HEAVY, BEKK-GARCH, and BEKK-HEAVY. Out-of-sample s-step-ahead forecasts of the 29-dimensional covariance and correlation matrices are computed, for $s = 1, 5$, and 22; for horizons 5 and 22, they are iterated forecasts. For DECO models, the correlations are computed from DECO itself, not from the underlying DCC.

5.3.1. Loss functions

Statistical and economic loss functions are adopted along the lines proposed by Becker, Clements, Doolan, and Hurn (2015). For the covariance matrix forecasts, two statistical loss functions are used, which compare the covariance matrix forecasts with respect to the actual (unobserved) covariance matrix Σ_{t+s}. The first one is based on the negative of the Wishart log-density function:

$$QLIK^a_{t,s}(\Sigma_{t+s}, H^a_{t+s|t}) = \mathrm{tr}[(H^a_{t+s|t})^{-1}\Sigma_{t+s}] + \log|H^a_{t+s|t}|, \tag{46}$$

where $H^a_{t+s|t}$ denotes the s-step forecast using model a conditional on time t information. The second loss function is based on the square of the Frobenius norm of the difference between the forecast and benchmark matrices (see, e.g., Golosnoy et al. (2012)), defined by

$$FN^a_{t,s} = \|\Sigma_{t+s} - H^a_{t+s|t}\| = \sum_{i,j}(\sigma_{ij,t+s} - h^a_{ij,t+s})^2 \tag{47}$$

Since Σ_{t+s} is unobservable, the observed realized covariance matrix RC_{t+s} is used as a proxy for it. These statistical loss functions provide a consistent ranking of volatility models in the sense of Patton (2011) and Patton and Sheppard (2009), as they are robust to noise in the proxy; see also Laurent, Rombouts, and Violante (2013).

For the correlation matrix forecasts, the QLIK and FN losses are computed from the same formulas as for covariances, that is (46) and (47). The only difference is that forecasted correlations and realized correlations replace forecasted covariances and realized covariances, respectively.

Once a time series of $T_{h,s}$ (covariance or correlation) forecasts is obtained for a model, the corresponding losses and their time-series average are computed:

$$QLIK^a_s = \frac{1}{T_{h,s}}\sum_{t=1}^{T_{h,s}} QLIK^a_{t,s}, \text{ and } FN^a_s = \frac{1}{T_{h,s}}\sum_{t=1}^{T_{h,s}} FN^a_{i,t,s}. \tag{48}$$

This is performed for each model and each forecast horizon, so that models can be ranked by the MSC procedure and an MCS at a chosen confidence level can be identified.

For the variance forecasts of different models, we use the univariate loss functions

$$UQLIK^a_{i,t,s} = \frac{v_{i,t+s}}{h^a_{i,t+s|t}} - \log\left(\frac{v_{i,t+s}}{h^a_{i,t+s|t}}\right) - 1, \tag{49}$$

and

$$MSE^a_{i,t,s} = (v_{i,t+s} - h^a_{i,t+s|t})^2, \tag{50}$$

where $v_{i,t+s}$ is the observed realized variance of stock i at date $t+s$, and $h^a_{i,t+s|t}$ is the corresponding s-step forecast of model a, based on information available at date t. Once the time-series average of each loss function has been computed for each stock, the mean across stocks is taken

Table 4
MCS for loss functions of out-of-sample covariance forecasting.

	FN	MCS Rank	QLIK	MCS Rank	GMV	MCS Rank	MV	MCS Rank
$s = 1$								
DCC-GARCH	16.792	5	12.007	6	0.226	7	0.309	7
DCC-HEAVY	**13.292**	1	**8.821**	1	**0.193**	1	**0.272**	1
DECO-GARCH	16.513	6	14.182	9	0.237	10	0.330	10
DECO-HEAVY	**13.667**	4	11.059	4	0.214	3	0.308	4
BEKK-GARCH	23.140	9	15.016	8	0.213	4	0.302	3
BEKK-HEAVY	28.312	10	13.559	7	0.227	8	0.329	8
ADCC-GARCH	16.768	8	12.077	5	0.228	6	0.311	5
ADCC-HEAVY	13.304	2	8.842	2	0.196	2	0.276	2
ADECO-GARCH	16.780	7	14.125	10	0.239	9	0.336	9
ADECO-HEAVY	13.697	3	11.084	3	0.215	5	0.312	6
$s = 5$								
DCC-GARCH	18.008	5	13.268	6	0.228	7	0.314	7
DCC-HEAVY	**15.907**	4	**11.703**	2	**0.203**	2	**0.286**	2
DECO-GARCH	17.824	6	15.242	9	0.237	10	0.332	9
DECO-HEAVY	**16.135**	3	**11.850**	3	**0.206**	3	0.292	3
BEKK-GARCH	23.637	9	15.298	7	0.216	5	0.307	5
BEKK-HEAVY	29.791	10	14.711	8	0.234	9	0.337	10
ADCC-GARCH	18.126	8	13.395	4	0.229	6	0.316	6
ADCC-HEAVY	15.299	2	11.384	1	0.203	1	0.286	1
ADECO-GARCH	18.209	7	15.255	10	0.239	8	0.337	8
ADECO-HEAVY	15.062	1	12.937	5	0.211	4	0.302	4
$s = 22$								
DCC-GARCH	**20.720**	5	**15.798**	4	0.230	6	0.319	5
DCC-HEAVY	21.863	8	17.310	7	**0.220**	3	**0.313**	3
DECO-GARCH	**20.426**	4	17.327	9	0.235	9	0.329	8
DECO-HEAVY	22.029	6	17.217	8	**0.221**	4	**0.315**	4
BEKK-GARCH	25.201	9	**16.109**	3	0.224	5	0.320	6
BEKK-HEAVY	32.171	10	16.463	5	0.243	10	0.351	10
ADCC-GARCH	21.204	7	16.327	6	0.233	7	0.328	7
ADCC-HEAVY	18.487	2	14.890	1	0.215	1	0.305	1
ADECO-GARCH	20.879	3	18.081	10	0.236	8	0.337	9
ADECO-HEAVY	18.418	1	15.251	2	0.219	2	0.313	2

Values of loss functions in bold identify models in the 90% level MCS when the comparison is limited to the first six models. Underlined values identify models in the 90% level MCS when the comparison is done for all models. A value in bold but not underlined is thus in the MCS of the first six models, but is excluded when considering all models. The MCS rankings are for the global comparison.

and used in the MCS procedure. That is,

$$UQLIK_s^a = \frac{1}{T_{h,s}k}\sum_{t=1}^{T_{h,s}}\sum_{i=1}^{k} QLIK_{i,t,s}^a, \text{ and}$$
$$MSE_s^a = \frac{1}{T_{h,s}k}\sum_{t=1}^{T_{h,s}}\sum_{i=1}^{k} MSE_{i,t,s}^a. \tag{51}$$

The economic loss functions are relevant for the covariance matrix forecasts. They are based on forecasted portfolio performances. The same economic loss functions as Engle and Kelly (2012) are used: global minimum variance portfolio (GMV), and minimum variance portfolio (MV); see also Engle and Colacito (2006). These loss functions are based on the variances of the forecasts of portfolio returns. A superior model produces optimal portfolios with lower forecast variance. Given a covariance matrix forecast $H_{t+s|t}^a$, the GMV portfolio weight vector $\hat{w}_{t+s}^a$ is computed as the minimizer of the portfolio variance $(w_{t+s}^a)' H_{t+s|t}^a w_{t+s}^a$ subject to the constraint that the weights add to unity. Once this is done for each forecast date, the GMV loss function is the average of the portfolio variances over the forecast period:

$$GMV_s^a = \frac{1}{T_{h,s}}\sum_{t=1}^{T_{h,s}} (\hat{w}_{t+s}^a)' H_{t+s|t}^a \hat{w}_{t+s}^a. \tag{52}$$

The MV portfolio is obtained by minimizing the portfolio variance subject to the additional constraint that the expected portfolio return be larger than a chosen value. Following (Engle & Kelly, 2012), this value is fixed at $q = 10\%$ and the expected portfolio return (μ) at the mean of the data. The MV loss is defined like (52) but with the optimal weight vectors corresponding to the MV minimizations. The optimal GMV and MV weights are analytically known functions of $H_{t+s|t}^a$, and of μ and q for MV (see, e.g., Engle & Kelly, 2012).

5.3.2. Results

To compute out-of-sample forecasts, each model is re-estimated every fifth observation based on rolling sample windows of 3000 observations, resulting in a total of $T_h = 1318$ out-of-sample forecasts for $s = 1$, 1314 for $s = 5$, and 1297 for $s = 22$.

Table 4 reports the out-of-sample forecast losses, defined by (48), and the economic loss functions for the

Table 5
MCS for loss functions of out-of-sample correlation and variance forecasting.

	Correlation				Variance			
	FN	MCS Rank	QLIK	MCS Rank	MSE	MCS Rank	UQLIK	MCS Rank
$s = 1$								
DCC-GARCH	6.119	8	20.833	3	2.544	3	0.509	4
DCC-HEAVY	**5.505**	4	**20.450**	2	**2.210**	2	**0.386**	2
DECO-GARCH	5.787	6	21.283	8				
DECO-HEAVY	**5.364**	1	21.081	5				
ADCC-GARCH	6.037	7	20.817	4	2.752	4	0.504	3
ADCC-HEAVY	5.478	3	20.417	1	2.205	1	0.385	1
ADECO-GARCH	5.719	5	21.266	7				
ADECO-HEAVY	5.379	2	21.121	6				
$s = 5$								
DCC-GARCH	6.161	8	20.852	4	2.838	3	0.561	3
DCC-HEAVY	**5.741**	4	**20.527**	3	**2.659**	2	**0.501**	2
DECO-GARCH	5.959	7	21.023	7				
DECO-HEAVY	**5.574**	1	21.713	8				
ADCC-GARCH	6.100	6	20.844	5	3.097	4	0.562	4
ADCC-HEAVY	5.728	3	20.480	2	2.347	1	0.488	1
ADECO-GARCH	5.915	5	20.982	6				
ADECO-HEAVY	5.609	2	20.383	1				
$s = 22$								
DCC-GARCH	6.258	8	20.851	7	**3.359**	2	**0.661**	2
DCC-HEAVY	**6.061**	2	**20.623**	5	3.916	4	0.710	4
DECO-GARCH	6.237	7	**20.536**	1				
DECO-HEAVY	**6.112**	4	21.111	8				
ADCC-GARCH	6.224	6	20.830	6	3.706	3	0.676	3
ADCC-HEAVY	6.025	1	20.560	2	2.865	1	0.620	1
ADECO-GARCH	6.112	3	20.564	4				
ADECO-HEAVY	6.154	5	20.564	3				

Values of loss functions in bold identify models in the 90% level MCS when the comparison is limited to the first six models. Underlined values identify models in the 90% level MCS when the comparison is done for all models. A value in bold but not underlined is thus in the MCS of the first six models, but is excluded when considering all models. The MCS rankings are for the global comparison.

different models. The boldface values identify the models that belong to the 90% model confidence set (MCS90, hereafter) for each loss function when the comparison is limited to the six symmetric models mentioned in the first paragraph of this subsection. The comparisons, including the four asymmetric models included in the table, are presented in Section 5.4, and the MCS ranks in the table pertain to the comparisons of all models.

At forecast horizons 1 and 5, DCC-HEAVY belongs to the MCS90 for all loss functions, DECO-HEAVY belongs to it for a subset of loss functions, and all the other models are out of each MCS90. At horizon 22, no HEAVY model is in the MCS90 of FN and QLIK; MCS90 includes DCC-GARCH and DECO-GARCH for FN, and DCC-GARCH and BEKK-GARCH for QLIK For the GMV and MV loss functions, MCS90 consists of DCC-HEAVY and DECO-HEAVY.

Table 5 reports the MCS90 sets for correlation and variance forecasts separately, excluding the BEKK models for these comparisons, and using only the statistical loss functions, since the economic loss functions use the covariance matrix.

For correlations, DCC-HEAVY is in the MCS90 of both loss functions at the three horizons. DECO-HEAVY is also in the MCS90 of FN at the three horizons, and DECO-GARCH is in the QLIK MCS90 set at horizon 22.

For variances, the loss values defined by (49) are reported in the "Variance" part of Table 5. Notice that DECO models are irrelevant (being the same as DCC in the first step of estimation). The results reveal that DCC-HEAVY is alone in MCS90 for both MSE and QLIK at horizons 1 and 5. At horizon 22, only DCC-GARCH is in both MCS90 sets. Shephard and Sheppard (2010) report that the performance of HEAVY with respect to GARCH deteriorates as the forecast horizon increases.

To trace where the forecast gains occur, Table 6 reports the ratios between the losses of the DCC-HEAVY and DCC-GARCH models, and likewise for the DECO models. One can see that the DCC/DECO-HEAVY models outperform the DCC/DECO-GARCH models both in covariance, correlation and variance forecast losses at the forecast horizons 1 and 5 (with a single exception for DECO). The improvements are larger for horizon 1 than 5, and smaller for DECO than for DCC (except in one case). They can be important: e.g., DCC-HEAVY reduces the covariance and variance QLIK losses by 10% at least and up to 25%. At horizon 22, the GARCH models have the smallest losses, except for correlations where the differences are less than 2.5%. The loss improvements of GARCH with respect to HEAVY are between 7% and 17% when they occur.

In brief, the forecast comparisons clearly favor the DCC-HEAVY model at the short forecast horizons for all loss functions, and to a lesser extent the DECO-HEAVY model, at the expense of the BEKK-HEAVY and the three GARCH models. At the longest forecast horizon, the results

Table 6
Loss ratios between DCC-HEAVY and DCC-GARCH: Out-of-sample forecasting.

		DCC-HEAVY vs. DCC-GARCH		DECO-HEAVY vs. DECO-GARCH	
		FN	QLIK	FN	QLIK
Covariance	$s = 1$	0.792	0.735	0.828	0.780
	$s = 5$	0.883	0.882	0.905	0.777
	$s = 22$	1.055	1.096	1.078	0.994
Correlation	$s = 1$	0.900	0.982	0.927	0.991
	$s = 5$	0.932	0.984	0.935	1.033
	$s = 22$	0.968	0.989	0.980	1.028
		MSE	UQLIK		
Variance	$s = 1$	0.869	0.759		
	$s = 5$	0.937	0.892		
	$s = 22$	1.166	1.074		

For the Variance panel, the results of DCC and DECO are identical.

depend on the loss function; moreover, there is a clear worse performance of DCC-HEAVY relative to DCC-GARCH for variance and covariance forecasts at this horizon.

The in-sample forecasting results are reported in the supplemental appendix (Tables E3–E4). They do not differ much from the out-of-sample results.

5.4. Asymmetric and HAR terms in DCC/DECO-HEAVY

The out-of-sample forecast performance of the DCC-HEAVY model deteriorates with respect to DCC-GARCH when the forecast horizon increases, switching from better to worse at some horizon between 5 and 22. DCC-HEAVY makes use of forecasts of realized variances and correlations. If the realized variance or correlation equations are incorrectly specified, the forecast error is brought to the conditional covariance and correlation forecasts. These forecast errors get larger when the forecast horizon is longer. To improve the forecast performance of DCC-HEAVY in multiple-step-ahead forecasts, the ADCC-HEAVY model defined by (32)–(34)–(35)–(36) is worth trying. The HAR terms are useful to capture the long-memory feature of realized variances and correlations.

5.4.1. Estimation results

The parameter estimates of the asymmetric variance equations are reported in the supplemental appendix (Tables E5 and E6). For the conditional variances, the coefficient estimates of the asymmetric term (the diagonal elements of Γ_h in (32)) are positive for 22 stocks, with t-statistics above 2 for eight of them. For the realized variances, the coefficient estimates of the asymmetric term (the diagonal elements of Γ_m in (35)) are all positive, with t-statistics larger than 2.5 for 28 stocks. The coefficient estimates of the weekly HAR term are positive (with a single, insignificant exception) but only 14 of them have t-statistics above 2, indicating a moderate impact of this term. On the contrary, the monthly HAR term is positive for all stocks and appears to be strongly significant (with one exception).

The estimates of the asymmetric correlation equations are reported in Table 7. For the conditional correlations, the estimate of γ_r is positive in the ADCC (0.025) and very significant (t-statistic, 6.51). This means that the impact of each lagged realized correlation on the next conditional correlation is stronger (being equal to $0.062 + 0.025$) when both lagged returns of the corresponding assets are negative than otherwise (being then equal to 0.062). Nevertheless, the additional statistically significant impact of 0.025 is not very important on the next-period correlation, since it increases it by 2.5% of the previous period correlation (if both lagged returns are negative): that means an increase of 0.0125 if the previous correlation is 0.5. For the ADECO model, the impact is of 2% and statistically insignificant.

For the realized correlations, the estimate of γ_p is positive and significant in both models (HAR-ADCC and HAR-ADECO), which means that the impact of each lagged realized correlation on the next conditional mean of the realized correlation is stronger (being equal to $0.063 + 0.010$ in the HAR-ADCC) when both lagged returns of the corresponding assets are negative than otherwise (being then equal to 0.063). The same remark applies as above, about the limited value of the expected correlation change this implies between two consecutive days.

Concerning the HAR terms, the weekly term impact is negative (-0.033) and significant in HAR-ADCC, and positive (0.075) and insignificant in HAR-ADECO. The monthly term impact is negative in HAR-ADCC but positive in HAR-ADECO, being significant in both models. Anyway, the effective daily changes of expected correlations these terms imply are small in HAR-ADCC and moderate in HAR-ADECO.

5.4.2. Forecast comparisons

For covariance forecasts, Table 4 reports the loss function values for the four asymmetric models (ADCC-HEAVY, ADECO-HEAVY, ADCC-GARCH, and ADECO-GARCH) in addition to the six symmetric models. The MCS90 for each loss function and horizon is identified by the underlined values; a bold underlined value is thus in the MCS90 for the 10 models, and in that of the first six models, while a bold value, not underlined, is in the MCS90 of the first six models but is removed from the MCS90 when the 10 models are compared.

The changes of the MCS90 due to the inclusion of the four asymmetric models in the comparisons are in favor of HEAVY models: the asymmetric HEAVY models are added

Table 7
Parameter estimates (and robust t-statistics) of the asymmetric correlation equations.

ADCC-HEAVY-R (33)-(34)			HAR-ADCC-HEAVY-M (36)				
α_r	β_r	γ_r	α_p	β_p	γ_p	α_p^w	α_p^m
0.062	0.857	0.025	0.063	0.964	0.010	−0.033	−0.005
(11.05)	(52.60)	(6.51)	(40.27)	(353.68)	(12.05)	(−18.36)	(−3.06)
ADECO-HEAVY-R			HAR-ADECO-HEAVY-P				
α_r	β_r	γ_r	α_p	β_p	γ_p	α_p^w	α_p^m
0.438	0.573	0.020	0.195	0.511	0.173	0.075	0.100
(6.78)	(9.70)	(0.54)	(9.40)	(8.53)	(13.53)	(1.55)	(5.32)

Table 8
Loss ratios between ADCC-HEAVY and DCC-GARCH: Out-of-sample forecasting.

		ADCC-HEAVY vs. DCC-GARCH		ADECO-HEAVY vs. DECO-GARCH	
		FN	QLIK	FN	QLIK
Covariance	$s = 1$	0.792	0.736	0.830	0.782
	$s = 5$	0.850	0.858	0.845	0.849
	$s = 22$	0.892	0.943	0.902	0.880
Correlation	$s = 1$	0.895	0.980	0.929	0.992
	$s = 5$	0.930	0.982	0.941	0.970
	$s = 22$	0.963	0.986	0.987	1.001
		MSE	UQLIK		
Variance	$s = 1$	0.867	0.758		
	$s = 5$	0.827	0.869		
	$s = 22$	0.853	0.937		

For the Variance panel, the results of DCC and DECO are identical.

Table 9
Loss ratios between ADCC-HEAVY and ADCC-GARCH: Out-of-sample forecasting.

		ADCC-HEAVY vs. ADCC-GARCH		ADECO-HEAVY vs. ADECO-GARCH	
		FN	QLIK	FN	QLIK
Covariance	$s = 1$	0.793	0.732	0.816	0.785
	$s = 5$	0.844	0.850	0.827	0.848
	$s = 22$	0.872	0.912	0.882	0.843
Correlation	$s = 1$	0.907	0.981	0.940	0.993
	$s = 5$	0.939	0.983	0.948	0.971
	$s = 22$	0.968	0.987	1.007	1.000
		MSE	UQLIK		
Variance	$s = 1$	0.801	0.765		
	$s = 5$	0.758	0.868		
	$s = 22$	0.773	0.916		

For the Variance panel, the results of DCC and DECO are identical.

to the MCS90 of some loss functions and horizons, but no asymmetric GARCH model is added for any loss function and horizon. For example, at horizons 5 and 22, ADCC-HEAVY is in the MCS90 sets for all loss functions; only the ADECO-HEAVY model also belongs to these sets in the case of the FN loss. The asymmetric HEAVY models attenuate or even reverse the worse performance of their symmetric counterparts relative to DCC at the three horizons, as can be seen by comparing the covariance loss ratios reported in Tables 8 and 9 and the corresponding values in Table 6.

For correlation forecasting, Table 5 also includes the asymmetric models in the comparisons. The changes of the MCS90 due to this are mixed: at horizon 1, only the asymmetric versions of the HEAVY models that were in the initial MCS90 are added; at horizon 5, ADECO-HEAVY is added for both losses and ADCC-HEAVY for QLIK loss); at horizon 22, ADCC-HEAVY is added for FN loss, being alone in MCS90; for QLIK loss, ADCC-HEAVY, ADECO-GARCH, and ADECO-HEAVY are added, DCC-HEAVY being removed and DECO-GARCH kept. The correlation loss ratios reported in Tables 8 and 9 differ slightly from those of Table 6, indicating that the asymmetric versions of the models do not much improve the symmetric versions.

For variance forecasting (Table 5), the main change in the MCS90 due to the inclusion of asymmetric models is

that ADCC-HEAVY is added for both loss functions and all horizons, even being the only model in the sets at horizons 5 and 22. This model ranked first in all comparisons. Clearly, ADCC-HEAVY improves variance forecasting, especially at horizons larger than 1. Loss ratios show that this improvement is important at horizon 22; for example, for MSE loss, it goes from −16.6% in Table 6 to +14.7% in Table 8 and +22.7% in Table 9.

6. Conclusions

Multivariate volatility models that specify the dynamics of the daily conditional covariance matrix as a function of realized covariances have emerged in the literature since 2012. They are a valuable alternative to multivariate GARCH models wherein the dynamics depend on lagged squared returns and their cross-products, because realized variances and covariances are more precise measures of daily volatility.

Perhaps surprisingly, with the partial exception of Braione (2016), no dynamic conditional correlation formulation of a HEAVY model has been proposed in the literature, where BEKK-type formulations are used in the papers of Noureldin et al. (2012) and Opschoor et al. (2018). Our contribution fills this gap by developing DCC-type HEAVY models. Such models have the advantage, with respect to BEKK models, of separating the specification of the conditional variances from the specification of the conditional correlations. The same advantage occurs in the specifications of the expected realized variances and correlations. As for GARCH DCC models, this results in more flexible models, in the sense that the dynamics of variances differ between assets and from the dynamics of correlations. Sticking to scalar models for correlations, the models remain parsimonious in parameters.

An illustrative empirical application for 29 assets illustrated the value of the flexibility of DCC and DECO versions of HEAVY models. These models, including extensions to include asymmetric effects, have superior forecasting performance with respect to BEKK formulations. As always in this type of empirical exercise, this finding is contingent on the dataset used and cannot be claimed to be valid in general. A robust conclusion is that HEAVY models dominate GARCH versions in terms of forecasting performance. An additional application to the data of Noureldin et al. (2012) is reported in Section F of the supplemental appendix and broadly confirms the empirical conclusions.

This research did not develop the asymptotic theory of the QML estimator of the parameters of the DCC- and DECO-HEAVY models. A small-scale simulation study illustrated that the usual properties of consistency and asymptotic normality are likely to hold.

Declaration of competing interest

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

Acknowledgments

We thank the referees for their comments. Thanks to Professors Lyudmila Grigoryeva (University of Konstanz) and Juan-Pablo Ortega (University of St. Gallen) for providing the datasets for the DJIA companies, and to Oleksandra Kukharenko (University of Konstanz) for handling the data and computing the realized covariance matrices from the TAQ data.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ijforecast.2022.03.005.

References

Aielli, G. P. (2013). Dynamic conditional correlation: On properties and estimation. *Journal of Business & Economic Statistics*, *31*, 282–299.

Bauwens, L., Braione, M., & Storti, G. (2016). Forecasting comparison of long term component dynamic models for realized covariance matrices. *Annals of Economics and Statistics*, *123/124*, 103–134.

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

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

Bauwens, L., Storti, G., & Violante, F. (2012). *Dynamic conditional correlation models for realized covariance matrices*: *CORE discussion paper 2012–60*, Louvain La Neuve: Université catholique de Louvain.

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

Braione, Manuela (2016). A time-varying long run HEAVY model. *Statistics & Probability Letters*, *119*, 36–44.

Chiriac, R., & Voev, V. (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*(2), 174–196.

Engle, R. F. (2002). Dynamic conditional correlation - a simple class of multivariate GARCH models. *Journal of Business & Economic Statistics*, *20*, 339–350.

Engle, Robert, & Colacito, Riccardo (2006). Testing and valuing dynamic correlations for asset allocation. *Journal of Business & Economic Statistics*, *24*(2), 238–253.

Engle, Robert F., & Gallo, Giampiero M. (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics*, *131*, 3–27.

Engle, Robert F., Hendry, D. F., & Richard, J.-F. (1983). Exogeneity. *Econometrica*, *21*, 277–304.

Engle, R., & Kelly, B. (2012). Dynamic equicorrelation. *Journal of Business & Economic Statistics*, *30*, 212–228.

Engle, R., & Kroner, F. K. (1995). Multivariate simultaneous generalized ARCH. *Economic Theory*, *11*, 122–150.

Engle, R. F., & Sheppard, K. (2001). Theorical and empirical properties of dynamic conditional correlation multivariate GARCH. Mimeo, UCSD.

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

Gorgi, P., Hansen, P. R., Janus, P., & Koopman, S. J. (2019). Realized Wishart-GARCH: A score-driven multi-asset volatility model. *Journal of Financial Econometrics*, *17*(1), 1–32.

Gouriéroux, C., Jasiak, J., & Sufana, R. (2009). The Wishart Autoregressive process of multivariate stochastic volatility. *Journal of Econometrics*, *150*, 167–181.

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

Hansen, Peter R, Lunde, Asger, & Nason, James M (2011). The model confidence set. *Econometrica*, *79*(2), 453–497.

Hansen, P. R., Lunde, A., & Voev, V. (2014). Realized beta GARCH: A multivariate GARCH model with realized measures of volatility. *Journal of Business & Economic Statistics*, *29*, 774–799.

Jin, Xin, & Maheu, John M. (2013). Modeling realized covariances and returns. *Journal of Financial Econometrics*, *11*(2), 335–369.

Laurent, Sébastien, Rombouts, Jeroen V. K., & Violante, Francesco (2013). On loss functions and ranking forecasting performances of multivariate volatility models. *Journal of Econometrics*, *173*(1), 1–10.

Noureldin, Diaa, Shephard, Neil, & Sheppard, Kevin (2012). Multivariate high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics*, *27*(6), 907–933.

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

Opschoor, Anne, Janus, Pawel, Lucas, André, & Van Dijk, Dick (2018). New HEAVY models for fat-tailed realized covariances and returns. *Journal of Business & Economic Statistics*, *36*(4), 643–657.

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

Patton, Andrew, & Sheppard, Kevin (2009). Evaluating volatility and correlation forecasts. In T. G. Andersen, R. A. Davis, J.-P. Kreiss, & T. Mikosch (Eds.), *Handbook of financial time series*. Springer Verlag.

Shephard, Neil, & Sheppard, Kevin K. (2010). Realising the future: forecasting with high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics*, *25*, 197–231.

Sheppard, Kevin, & Xu, Wen (2019). Factor high-frequency-based volatility (HEAVY) models. *Journal of Financial Econometrics*, *17*, 33–65.