JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. (2011)
Published online in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/jae.1260

MULTIVARIATE HIGH-FREQUENCY-BASED VOLATILITY (HEAVY) MODELS

DIAA NOURELDIN,[a] NEIL SHEPHARD[a]* AND KEVIN SHEPPARD[b]

[a] *Department of Economics, University of Oxford, Oxford, UK*
[b] *Oxford-Man Institute, University of Oxford, UK*

SUMMARY

This paper introduces a new class of multivariate volatility models that utilizes high-frequency data. We discuss the models' dynamics and highlight their differences from multivariate generalized autoregressive conditional heteroskedasticity (GARCH) models. We also discuss their covariance targeting specification and provide closed-form formulas for multi-step forecasts. Estimation and inference strategies are outlined. Empirical results suggest that the HEAVY model outperforms the multivariate GARCH model out-of-sample, with the gains being particularly significant at short forecast horizons. Forecast gains are obtained for both forecast variances and correlations. Copyright © 2011 John Wiley & Sons, Ltd.

Received 9 March 2011; Revised 22 April 2011

1. INTRODUCTION

This paper introduces a new class of multivariate volatility models capable of producing precise multi-step forecasts of the conditional covariance matrix of daily returns. Multivariate volatility models have been the focus of a voluminous literature summarized recently by Bauwens *et al.* (2006) and Asai *et al.* (2006), where the focus in the latter is on multivariate stochastic volatility.

The covariance matrix of daily asset returns is a key input in portfolio allocation, option pricing and financial risk management. An interesting question is whether the increasing availability of high-frequency financial data enables the development of more accurate forecasting models for the conditional covariance of daily returns. We address this question by studying a new class of models which utilize high-frequency data for the objective of multi-step volatility forecasting. We call this class multivariate High-frEquency-bAsed VolatilitY (HEAVY) models.

Volatility forecasts from HEAVY models have some properties that distinguish them from those of multivariate generalized autoregressive conditional heteroskedasticity (GARCH) models. HEAVY models have a relatively short response time, which means they are likely to perform well in periods where the level of volatility or correlation is subject to abrupt changes. HEAVY models also have short-run momentum effects, so that volatility forecasts may exhibit a continuation of upward (or downward) trends before mean reverting. The latter distinction pertains to comparing the HEAVY model to a baseline specification such as the GARCH(1,1) model. More richly parameterized GARCH models could, of course, also exhibit momentum effects.

The univariate HEAVY model was introduced in Shephard and Sheppard (2010), where it is shown—for a wide spectrum of asset classes—that the HEAVY model outperforms the GARCH model in- and out-of-sample. The forecast gains tend to be more pronounced at short forecast horizons, typically the first few days. In the empirical section of this paper, we show similar results in a multivariate setting. The multivariate analysis poses additional interesting questions such as

* Correspondence to: Neil Shephard, Department of Economics, Oxford-Man Institute, Eagle House, Walton Well Road, Oxford OX2 6ED, UK. E-mail: neil.shephard@economics.ox.ac.uk

whether the forecast gains are due to the variance forecasts of individual assets, their correlations or a combination of both. We develop a novel out-of-sample model evaluation strategy to address this question.

To highlight the distinction between HEAVY and GARCH models, and how HEAVY models differ from recently proposed models which also utilize high-frequency data, we start with a brief overview of the univariate HEAVY model of Shephard and Sheppard (2010). Let $\mathcal{F}_t^{\mathrm{LF}}$ and $\mathcal{F}_t^{\mathrm{HF}}$ respectively denote the information set generated by low-frequency (i.e. daily) and high-frequency (i.e. intra-daily) data up to time t, where $t = 1, 2, \ldots,$ indexes days. Also let r_t denote the (demeaned) daily return and v_t denote the realized measure (e.g. realized variance) at time t. The univariate HEAVY model in its linear specification is the two-equation system

$$\mathrm{E}[r_t^2|\mathcal{F}_{t-1}^{\mathrm{HF}}] := h_t = c_h + b_h h_{t-1} + a_h v_{t-1},$$
$$\mathrm{E}[v_t|\mathcal{F}_{t-1}^{\mathrm{HF}}] := m_t = c_m + b_m m_{t-1} + a_m v_{t-1}$$

while the GARCH model is

$$\mathrm{E}[r_t^2|\mathcal{F}_{t-1}^{\mathrm{LF}}] := h_t^* = c_g + b_g h_{t-1}^* + a_g r_{t-1}^2$$

The primary distinction between HEAVY and GARCH models is the conditioning information set used in modelling the conditional variance of daily returns. The first equation of the HEAVY model uses the lagged realized measure, v_{t-1}, to drive to dynamics of h_t, whereas the GARCH model uses the lagged squared return. The second equation of the HEAVY model is needed for multi-step forecasts of h_t.

The HEAVY model utilizes recently developed estimators of *ex post* volatility of daily returns that have proven to be more precise compared to squared returns. Realized variance is the first realized measure to be systematically studied and used in modelling and forecasting the volatility of daily returns. Andersen and Bollerslev (1998) show that the realized variance has a much lower noise-to-signal ratio than the daily squared return when used as proxy for the unobserved variance, while Barndorff-Nielsen and Shephard (2002) formalize the econometrics of the realized variance. In the context of multi-step forecasting, Shephard and Sheppard (2010) show that the use of the realized kernel of Barndorff-Nielsen *et al*. (2008) leads to notable in- and out-of-sample improvements in predicting h_t, especially at short forecast horizons.

Univariate HEAVY models are related to recently proposed models by Engle (2002), Engle and Gallo (2006), Cipollini *et al*. (2007), Brownlees and Gallo (2010) and Hansen *et al*. (2011). Engle (2002) models volatility using a multiplicative error model (MEM).[1] He applies this model to squared returns and realized volatility as separate models, but they were not considered as a system for multi-step forecasting of the conditional variance of daily returns. These models are usually referred to as GARCH-X models when both v_{t-1} and r_{t-1}^2 appear in the h_t equation. Engle and Gallo (2006) model a three-variable system comprising the squared return, the high-minus-low price range and the realized variance in an MEM setup. Cipollini *et al*. (2007) allow for contemporaneous correlations in a four-variable vector MEM including the absolute daily return and three realized measures, and tackle the problem of a suitable multivariate density choice using copulas.

The papers by Brownlees and Gallo (2010) and Hansen *et al*. (2011) are the closest in structure to the univariate HEAVY model. The model in Brownlees and Gallo (2010) has a HEAVY-like structure, with the difference being that it uses a smoothed version of the realized measure to drive

[1] An MEM can be used for any non-negative valued process which can be modelled as i.i.d. innovations from a density with non-negative support scaled by a conditionally deterministic factor.

DOI: 10.1002/jae

h_t by specifying the latter as an affine function of m_t. Hansen *et al*. (2011) treat the dynamics of the realized measure differently. While the HEAVY model postulates GARCH-type dynamics for the realized measure by modelling its conditional expectation, Hansen *et al*. (2011) relate the realized measure itself to h_t and a term that captures leverage effects.

Multivariate volatility models are becoming increasingly important not only because of their direct application in portfolio allocation and asset pricing, but also due to the insights they provide into risk management practices. Using low-frequency data, Brownlees and Engle (2010) portray the importance of modelling conditional correlations for systemic risk management, where they show that a rise in a firm's stock volatility and correlation with the market magnifies its contribution to their proposed measure of systemic risk. Highly leveraged financial companies in the recent financial crisis are a case in point. The work of Hansen *et al*. (2010), which is independent and concurrent, utilizes realized measures in modelling a stock's conditional beta in a GARCH-like framework. Our primary empirical example focuses on the returns of Bank of America and the S&P 500 exchange traded fund (ETF) during the recent financial crisis, which relates to the applications in these papers.

There is some recent research that focuses only on modelling and forecasting the realized covariance matrix; see, for example, Voev (2008), Chiriac and Voev (2011) and Bauer and Vorkink (2011). The focus in these studies is on developing parsimonious models to forecast the realized covariance matrix. In contrast, this paper develops a framework for forecasting the covariance of daily returns which also requires forecasts of the realized measure. We find the realized measure to be a more precise factor to drive the volatility dynamics for daily returns compared to the outer product of daily returns which is used in GARCH models.

Jin and Maheu (2010) pursue an objective similar to ours by utilizing realized measures to improve the density forecasts of multivariate daily returns; however, their model is different from ours as it is cast in the multivariate stochastic volatility framework. In addition, they propose a different nexus between the dynamics of daily returns and the realized measure. The implication of this is that our model is much easier to estimate and allows for straightforward out-of-sample model evaluation since we provide closed-form forecasting formulas.

The structure of the paper is as follows. Section 2 introduces multivariate HEAVY models with some detailed analysis of their properties using a linear specification. Section 3 discusses estimation and inference. In Section 4, we present the out-of-sample model evaluation framework. Section 5 contains the results of our empirical analysis, while Section 6 concludes the paper. Appendix A derives the second moments' structure implied by the model. All proofs are collected in Appendix B. The Web Appendix to this paper includes relevant results from matrix algebra and calculus, an overview of the Wishart distribution related to the discussion in Section 3, as well as additional empirical results.

2. MULTIVARIATE HEAVY MODELS

2.1. Definitions and Notation

Let the multivariate log-price process be given by the $(k \times 1)$ vector Y^*_τ, where $\tau \in \mathbb{R}_+$ represents continuous time. Suppose we observe $m + 1$ intra-daily prices, assumed to be uniformly spaced, so that the jth intra-daily vector of returns on day t is given by

$$R_{j,t} = Y^*_{(t-1)+\frac{j}{m}} - Y^*_{(t-1)+\frac{(j-1)}{m}}, \quad j = 1, \ldots, m, \quad t = 1, 2, \ldots$$

Assuming, for instance, 24-hour trading means $m = 1440$ for 1-minute returns, and $R_{j,t}$ is the vector of returns for the jth minute on day t. The vector of daily returns is $R_t = \sum_{j=1}^{m} R_{j,t}$. The

DOI: 10.1002/jae

outer product of daily returns is the $(k \times k)$ matrix denoted by $P_t = R_t R_t'$. The realized measure on day t is a $(k \times k)$ matrix denoted by V_t. One example of V_t which we use in this paper is the realized covariance (RC_t) matrix defined as

$$\mathrm{RC}_t = \sum_{j=1}^{m} R_{j,t} R_{j,t}'$$

Barndorff-Nielsen and Shephard (2004) show that, in the absence of market microstructure noise, RC_t is a mixed normal consistent estimator of the quadratic covariation of Y_τ^* as $m \to \infty$. In the presence of market microstructure noise, RC_t is a biased estimator. Therefore, in practice one needs to sample sparsely and use subsampling. An alternative is to use a noise-robust estimator such as the realized kernel of Barndorff-Nielsen *et al*. (2008, 2011).

Letting $\mathcal{F}_t^{\mathrm{LF}}$ and $\mathcal{F}_t^{\mathrm{HF}}$ be as defined previously, the HEAVY model is the two-equation system

$$\mathrm{E}[P_t | \mathcal{F}_{t-1}^{\mathrm{HF}}] = \mathrm{E}[R_t R_t' | \mathcal{F}_{t-1}^{\mathrm{HF}}] := H_t \tag{1}$$

$$\mathrm{E}[V_t | \mathcal{F}_{t-1}^{\mathrm{HF}}] := M_t \tag{2}$$

where, for simplicity, we assume $\mathrm{E}[R_t | \mathcal{F}_{t-1}^{\mathrm{HF}}] = 0$ so that H_t is the conditional covariance matrix of daily returns, or alternatively, the conditional expectation of the outer product of daily returns. We will occasionally use $\mathrm{E}_t[\cdot] := \mathrm{E}[\cdot | \mathcal{F}_t^{\mathrm{HF}}]$ to denote the expectation conditional on $\mathcal{F}_t^{\mathrm{HF}}$. Thus the conditional first moments (H_t, M_t) are assumed $\mathcal{F}_{t-1}^{\mathrm{HF}}$-measurable.

We shall call (1)–(2) the HEAVY-P and HEAVY-V equations, respectively. HEAVY models can be equivalently represented as

$$P_t = H_t^{\frac{1}{2}} \varepsilon_t H_t^{\frac{1}{2}} \tag{3}$$

$$V_t = M_t^{\frac{1}{2}} \eta_t M_t^{\frac{1}{2}} \tag{4}$$

where ε_t and η_t are $(k \times k)$ symmetric innovation matrices satisfying $\mathrm{E}_{t-1}[\varepsilon_t] = \mathrm{E}_{t-1}[\eta_t] = I_k$, where I_k is an identity matrix. We have defined the symmetric square root of a generic positive semidefinite matrix A, denoted by $A^{\frac{1}{2}}$, using the spectral decomposition such that $A^{\frac{1}{2}} = U \Lambda^{\frac{1}{2}} U'$, where U is a matrix containing the eigenvectors of A, and $\Lambda^{\frac{1}{2}}$ is a diagonal matrix containing the square root of the eigenvalues of A. The representation (3)–(4) is a matrix-variate generalization of the univariate MEM introduced in Engle (2002) and the vector MEM presented in Cipollini *et al*. (2007).

Since our focus is on multivariate volatility models, we use the terms HEAVY and GARCH to refer to their multivariate formulation unless otherwise stated. The difference between the HEAVY-P equation and the GARCH model is the conditioning information set. GARCH models condition on $\mathcal{F}_{t-1}^{\mathrm{LF}}$ and thus H_t is influenced by the squares and cross-products of past daily returns (i.e. lags of P_t). In the HEAVY-P equation, we condition on $\mathcal{F}_{t-1}^{\mathrm{HF}}$, which enables us to use lags of V_t to project the path of H_t.

Equations (1)–(2), or equivalently (3)–(4), define a class of models which links the dynamics of H_t to the realized measure. This becomes clear once we specify the dynamic equations for H_t and M_t. Choosing a specification for the dynamics of H_t and M_t yields a particular model within the HEAVY class. For ease of presentation, we will focus in the rest of this paper on one particular specification within the HEAVY class which is akin to a multivariate GARCH(1,1) model, and we shall refer to it simply as the HEAVY model.

DOI: 10.1002/jae

2.2. Model Parameterization

A primary challenge in multivariate volatility modelling is to ensure that the conditional covariance matrix is positive semidefinite. In the GARCH literature, one of the ways this has been approached is the BEKK parameterization introduced by Engle and Kroner (1995). We can adopt that approach to our model, which we call BEKK-type parameterization, although the models are distinct. The BEKK-type parameterization is

$$H_t = \overline{C}_H \overline{C}_H' + \overline{B}_H H_{t-1} \overline{B}_H' + \overline{A}_H V_{t-1} \overline{A}_H' \tag{5}$$

$$M_t = \overline{C}_M \overline{C}_M' + \overline{B}_M M_{t-1} \overline{B}_M' + \overline{A}_M V_{t-1} \overline{A}_M' \tag{6}$$

The $(k \times k)$ matrices $\overline{A}_H$, $\overline{B}_H$, $\overline{A}_M$ and $\overline{B}_M$ each have k^2 free parameters, while $\overline{C}_H$ and $\overline{C}_M$ are $(k \times k)$ lower triangular matrices each with $k^* = k(k+1)/2$ free parameters. The parameterization in (5)–(6) guarantees that H_t and M_t are positive semidefinite for all t assuming H_0 and M_0 are positive semidefinite. If, in addition, $\overline{C}_H$ and $\overline{C}_M$ are full rank matrices, then H_t and M_t are positive definite for all t. We refer to $\overline{A}_H$, $\overline{B}_H$, $\overline{A}_M$ and $\overline{B}_M$ as the dynamic parameters, which are of main interest to us. Sometimes we consider $\overline{C}_H$ and $\overline{C}_M$ to be 'nuisance parameters'.

Although our interest is to obtain multi-step forecasts of H_t, forecasts from (6) are needed due to the presence of V_{t-1} in (5). Forecasting the realized measure itself has been the focus of a number of recent studies, e.g. Andersen *et al.* (2003, 2007, 2011). We note that postulating GARCH-type dynamics for the realized measure is consistent with its empirical properties such as time-varying volatility of realized volatility and evidence of excess kurtosis; see Corsi *et al.* (2008). Therefore, (6) may produce accurate forecasts of M_t.

Of course, other parameterizations for (5)–(6) could be adopted. For instance, a higher-order lag structure akin to GARCH(p,q) processes, or a component model which decomposes the conditional covariance matrix into long-run (permanent) and short-run (transitory) components as in Engle and Lee (1999). Also, a long-memory model could be specified for (6) as proposed in Chiriac and Voev (2011).

The unrestricted BEKK-type parameterization in (5)–(6) has $O(k^2)$ parameters. To avoid the curse of dimensionality one could impose that $\overline{A}_H$, $\overline{B}_H$, $\overline{A}_M$ and $\overline{B}_M$ are scalars or diagonal matrices, which yields the scalar or diagonal HEAVY model, respectively. In either case, the resulting equations for the diagonal elements of H_t and M_t would constitute univariate HEAVY models. The equations for the off-diagonal elements would also have a HEAVY structure in which the conditional covariances are driven by their own lags and the corresponding realized covariances. If the elements of $\overline{A}_H$, $\overline{B}_H$, $\overline{A}_M$ and $\overline{B}_M$ are unrestricted (i.e. a full HEAVY parameterization), the multivariate HEAVY model no longer comprises univariate HEAVY models, since in this case the evolution of every element in H_t and M_t will be influenced by own as well as cross-asset effects.

Example 1. For the H_t equation in the scalar HEAVY model, $\overline{A}_H = \overline{a}_H I_k$ and $\overline{B}_H = \overline{b}_H I_k$ where $\overline{a}_H$ and $\overline{b}_H$ are scalars, which gives the following parameterization:

$$H_t = \overline{C}_H \overline{C}_H' + \overline{b}_H^2 H_{t-1} + \overline{a}_H^2 V_{t-1}$$

In the case of the bivariate diagonal HEAVY model, the H_t equation is given by

$$\begin{pmatrix} h_{11,t} & h_{12,t} \\ h_{21,t} & h_{22,t} \end{pmatrix} = \begin{pmatrix} \overline{c}_{11,H} & 0 \\ \overline{c}_{21,H} & \overline{c}_{22,H} \end{pmatrix} \begin{pmatrix} \overline{c}_{11,H} & 0 \\ \overline{c}_{21,H} & \overline{c}_{22,H} \end{pmatrix}'$$

DOI: 10.1002/jae

$$+\begin{pmatrix} \overline{b}_{11,H} & 0 \\ 0 & \overline{b}_{22,H} \end{pmatrix} \begin{pmatrix} h_{11,t-1} & h_{12,t-1} \\ h_{21,t-1} & h_{22,t-1} \end{pmatrix} \begin{pmatrix} \overline{b}_{11,H} & 0 \\ 0 & \overline{b}_{22,H} \end{pmatrix}$$
$$+\begin{pmatrix} \overline{a}_{11,H} & 0 \\ 0 & \overline{a}_{22,H} \end{pmatrix} \begin{pmatrix} v_{11,t-1} & v_{12,t-1} \\ v_{21,t-1} & v_{22,t-1} \end{pmatrix} \begin{pmatrix} \overline{a}_{11,H} & 0 \\ 0 & \overline{a}_{22,H} \end{pmatrix}$$

where for any matrix A, a_{ij} denotes its $(i,\ j)$th element.

To better understand the dynamics, we express (5)–(6) in vector form. Define $p_t := \text{vech}(P_t)$, $v_t := \text{vech}(V_t)$, $h_t := \text{vech}(H_t)$ and $m_t := \text{vech}(M_t)$, where the vech operator stacks the lower triangular part including the main diagonal of a $(k \times k)$ symmetric matrix into a $(k^* \times 1)$ vector, $k^* = k(k+1)/2$. These $(k^* \times 1)$ vectors retain the unique elements of the matrices of interest to us. An equivalent representation of (3)–(4) is

$$P_t = H_t + H_t^{\frac{1}{2}}(\varepsilon_t - I_k)H_t^{\frac{1}{2}}, \quad V_t = M_t + M_t^{\frac{1}{2}}(\eta_t - I_k)M_t^{\frac{1}{2}}$$

which, using the vech notation, can be expressed as

$$p_t = h_t + \widetilde{\varepsilon}_t, \quad v_t = m_t + \widetilde{\eta}_t$$

where $\widetilde{\varepsilon}_t = \text{vech}(H_t^{\frac{1}{2}}(\varepsilon_t - I_k)H_t^{\frac{1}{2}}) = L_k(H_t^{\frac{1}{2}} \otimes H_t^{\frac{1}{2}})D_k\text{vech}(\varepsilon_t - I_k)$ and $\widetilde{\eta}_t = \text{vech}(M_t^{\frac{1}{2}}(\eta_t - I_k) M_t^{\frac{1}{2}}) = L_k(M_t^{\frac{1}{2}} \otimes M_t^{\frac{1}{2}})D_k\text{vech}(\eta_t - I_k)$.[2] The matrices L_k and D_k are, respectively, the elimination and duplication matrices defined in Web Appendix A. This representation is particularly convenient since $\widetilde{\varepsilon}_t$ and $\widetilde{\eta}_t$ are a vector martingale difference sequence with respect to $\mathcal{F}_{t-1}^{\text{HF}}$.

Similarly, (5)–(6) can be written as

$$h_t = C_H + B_H h_{t-1} + A_H v_{t-1} \tag{7}$$
$$m_t = C_M + B_M m_{t-1} + A_M v_{t-1} \tag{8}$$

where $C_H = L_k(\overline{C}_H \otimes \overline{C}_H)D_k\text{vech}(I_k)$, $B_H = L_k(\overline{B}_H \otimes \overline{B}_H)D_k$ and $A_H = L_k(\overline{A}_H \otimes \overline{A}_H)D_k$. C_M, B_M, and A_M are defined similarly using the parameters of (6). C_H and C_M are $(k^* \times 1)$ vectors, while A_H, B_H, A_M and B_M are $(k^* \times k^*)$ matrices. The elimination and duplication matrices, L_k and D_k, are non-stochastic matrices of zeros and ones, so the parameters in (7)–(8) are uniquely identified from (5)–(6) and vice versa.

By substituting $h_t = p_t - \widetilde{\varepsilon}_t$ and $m_t = v_t - \widetilde{\eta}_t$ into (7)–(8), it is straightforward to show that the HEAVY model has the following VARMA(1,1) representation:

$$\begin{pmatrix} p_t \\ v_t \end{pmatrix} = \begin{pmatrix} C_H \\ C_M \end{pmatrix} + \begin{pmatrix} B_H & A_H \\ 0 & B_M + A_M \end{pmatrix} \begin{pmatrix} p_{t-1} \\ v_{t-1} \end{pmatrix} + \begin{pmatrix} \widetilde{\varepsilon}_t \\ \widetilde{\eta}_t \end{pmatrix} - \begin{pmatrix} B_H & 0 \\ 0 & B_M \end{pmatrix} \begin{pmatrix} \widetilde{\varepsilon}_{t-1} \\ \widetilde{\eta}_{t-1} \end{pmatrix}$$

since $(\widetilde{\varepsilon}_t', \widetilde{\eta}_t')'$ is a vector martingale difference sequence with respect to $\mathcal{F}_{t-1}^{\text{HF}}$, assuming $\text{var}[(\widetilde{\varepsilon}_t', \widetilde{\eta}_t')']$ exists. The coefficient matrix attached to $(p_{t-1}', v_{t-1}')'$ determines the persistence of the HEAVY system. For covariance stationarity, the eigenvalues of this matrix must be less than one in modulus. Since it is block triangular, its eigenvalues are members of the multiset of the eigenvalues of B_H and $(B_M + A_M)$.[3] In the following assumption we explicitly state this covariance stationarity condition,

[2] The second equality in each expression follows from the property that for any $(k \times k)$ matrices A and B, with B being symmetric, $\text{vech}(ABA') = L_k(A \otimes A)D_k\text{vech}(B)$; see Web Appendix A.

[3] A multiset is a set that allows for some or all of its elements to be repeated. This general definition is needed to allow for the case when B_H and $(B_M + A_M)$ have some common eigenvalues.

DOI: 10.1002/jae

where for any $(k \times k)$ matrix A with eigenvalues $\lambda_1, \ldots, \lambda_k$, $\rho(A) := \max_i |\lambda_i|$ denotes the spectral radius of A.

Assumption 1. In the HEAVY model given by (7)-(8), $\rho(B_H) < 1$ and $\rho(B_M + A_M) < 1$.

The covariance stationarity condition in Assumption 1 is analogous to the one given in Engle and Kroner (1995). This can be seen by noting that for any square matrix A, $D_k^+(A \otimes A)D_k$ and $(A \otimes A)$ have the same eigenvalues, where $D_k^+ = (D_k' D_k)^{-1} D_k'$ is the Moore–Penrose inverse of D_k; see Magnus (1988, Theorem 4.10). Also, it holds that for any square matrix A, $D_k^+(A \otimes A)D_k = L_k(A \otimes A)D_k$; see Lutkepohl (1996, Section 9.5.5). Thus $B_H = L_k(\overline{B}_H \otimes \overline{B}_H)D_k$ and $(\overline{B}_H \otimes \overline{B}_H)$ have the same eigenvalues. A similar argument applies to $(B_M + A_M)$.

We can express the unconditional first moments of p_t and v_t in terms of the model parameters. By taking unconditional expectation of (7)–(8), it is straightforward to show that

$$\omega_H := \mathrm{E}[p_t] = (I_{k^*} - B_H)^{-1}[C_H + A_H(I_{k^*} - B_M - A_M)^{-1}C_M] \tag{9}$$

$$\omega_M := \mathrm{E}[v_t] = (I_{k^*} - B_M - A_M)^{-1}C_M \tag{10}$$

In Appendix A, we derive the unconditional second moments of p_t and v_t, which correspond to the fourth moments of the returns (i.e. kurtosis) and second moments of the realized measure (i.e. volatility of volatility).

2.3. Covariance Targeting

The covariance targeting parameterization was introduced by Engle and Mezrich (1996) for the univariate GARCH model. This allows the unconditional moments of the model to be estimated by the empirical moments, and the dynamic parameters would then be estimated using a quasi-likelihood. The HEAVY model differs from ARCH-type models by using a shock other than the outer-product of returns to model the conditional covariance. This has an implication for the covariance targeting specification when the dynamics of the model are restricted from the full specification in (5), as is the case when $\overline{A}_H$ is assumed to be diagonal or scalar. We elaborate on this point after the following proposition, which gives two covariance targeting parameterizations of the HEAVY model.

Proposition 1. Let $\Omega_H := \mathrm{E}[P_t] = \mathrm{E}[H_t]$ and $\Omega_M := \mathrm{E}[V_t] = \mathrm{E}[M_t]$. The covariance targeting parameterization of the HEAVY model in (7)-(8) is

$$h_t = (I_{k^*} - B_H - A_H\kappa)\omega_H + B_H h_{t-1} + A_H v_{t-1} \tag{11}$$

$$m_t = (I_{k^*} - B_M - A_M)\omega_M + B_M m_{t-1} + A_M v_{t-1} \tag{12}$$

where $\kappa = L_k(\overline{\kappa} \otimes \overline{\kappa})D_k$, $\overline{\kappa} = \Omega_M^{\frac{1}{2}} \Omega_H^{-\frac{1}{2}}$, $\omega_H := \mathrm{vech}(\Omega_H)$, $\omega_M := \mathrm{vech}(\Omega_M)$, and L_k and D_k denote respectively the elimination and duplication matrices of order k. An alternative covariance targeting parameterization for (7) is

$$h_t = (I_{k^*} - B_H - A_H^*)\omega_H + B_H h_{t-1} + A_H^* \widetilde{v}_{t-1} \tag{13}$$

where $\widetilde{v}_t = \kappa^{-1} v_t$ is a rotated realized measure such that $\mathrm{E}[\widetilde{v}_t] = \omega_H$.

DOI: 10.1002/jae

While the covariance targeting specification in (11)–(12) is a reparameterization of the original model in (7)–(8), the specification (13)–(12) corresponds to a different model which uses a rotated rather than the original realized measure. This is why the coefficient matrix on $\widetilde{v}_{t-1}$ is now denoted by A_H^*. The two models are equivalent, implying $A_H^* = A_H \kappa$ holds, if and only if both A_H^* and A_H are fully parameterized matrices. When A_H and A_H^* are restricted to be scalar (diagonal), this equivalence does not hold unless $\kappa \propto I_k$ (κ is diagonal).

Using (13)–(12) has the advantage that it is easier to impose the condition $\rho(B_H + A_H^*) < 1$ during estimation; see Assumption 2 below. Imposing the condition $\rho(B_H + A_H \kappa) < 1$ is more involved, particularly in the diagonal and full HEAVY models since κ is a $(k^* \times k^*)$ matrix with non-zero elements. For covariance stationarity in (11)–(12), or alternatively (13)–(12), we replace Assumption 1 with the following assumption.

Assumption 2. In the covariance targeting parameterization of the HEAVY model given by (11)–(12), $\rho(B_H + A_H \kappa) < 1$ and $\rho(B_M + A_M) < 1$. In the covariance targeting parameterization of the HEAVY model given by (13)-(12), $\rho(B_H + A_H^*) < 1$ and $\rho(B_M + A_M) < 1$.

Estimating the model in its covariance targeting specification can be carried out in two steps, and we discuss the appropriate inference method in this case in Section 3.3.

2.4. Multi-step Forecasting

We are primarily interested in forecasting the conditional covariance of daily returns, H_t. One-step forecasts are directly computable using (7), which expresses H_t in its vech form. To compute s-step forecasts for $s = 2, 3, \ldots$, we need the forecasts from (8) as well to compute the s-step conditional expectation of the realized measure appearing in the right-hand side of (7). The s-step forecast of h_t is given in the following proposition.

Proposition 2. Let the model be given by (7)–(8), then the s-step forecast of h_t is

$$\begin{aligned} \mathrm{E}_t[h_{t+s}] &= \sum_{i=1}^{s-1} B_H^{i-1} C_H + B_H^{s-1} h_{t+1} \\ &\quad + \sum_{i=1}^{s-1} B_H^{i-1} A_H \left\{ \sum_{j=1}^{s-i-1} (B_M + A_M)^{j-1} C_M + (B_M + A_M)^{s-i-1} m_{t+1} \right\} \end{aligned} \tag{14}$$

where h_{t+1} and m_{t+1} are $\mathcal{F}_t^{\mathrm{HF}}$-measurable. Alternatively, let the model be given by (11)–(12), then the s-step forecast of h_t is

$$\mathrm{E}_t[h_{t+s}] = \omega_H + B_H^{s-1}(h_{t+1} - \omega_H) + \sum_{i=1}^{s-1} B_H^{i-1} A_H (B_M + A_M)^{s-i-1}(m_{t+1} - \omega_M) \tag{15}$$

The difference between (14) and (15) is that the latter is obtained under a covariance targeting specification in which the constant terms C_H and C_M are replaced with expressions involving ω_H and ω_M; see Section 2.2. In (14), Assumption 1 implies $\mathrm{E}_t[h_{t+s}] \rightarrow \omega_H$ as $s \rightarrow \infty$ since the coefficients on h_{t+1} and m_{t+1} will tend to zero, while the limit of the constant terms including C_H and C_M will be the right-hand side of (9). In (15), we also have that $\mathrm{E}_t[h_{t+s}] \rightarrow \omega_H$ as

DOI: 10.1002/jae

$s \to \infty$; however, in this case Assumption 2 is the operative assumption since the derivation of this equation is based on the covariance targeting specification.

In deriving (15), we focused on the covariance targeting specification given by (11)–(12) since it is more constructive to study the properties of the HEAVY model forecasts. For example, (15) can be used to compute the HEAVY model's half-life (of a deviation of the one-step forecast of h_t from ω_H) and compare it to that of the GARCH model. The presence of the term $(m_{t+1} - \omega_M)$ also indicates that mean reversion of the forecast matrix is not necessarily monotonic. To forecast using the covariance targeting specification in (13)–(12), A_H^* will appear in (15) instead of A_H. Thus the term $(m_{t+1} - \omega_M)$ must be pre-multiplied by κ^{-1} to ensure positive definiteness of $\mathrm{E}_t[H_{t+s}]$.

3. ESTIMATION AND INFERENCE

3.1. The Distribution of ε_t and η_t

For the HEAVY model in (3)–(4)

$$P_t = H_t^{\frac{1}{2}} \varepsilon_t H_t^{\frac{1}{2}}, \quad V_t = M_t^{\frac{1}{2}} \eta_t M_t^{\frac{1}{2}}$$

the natural choice for the density of the innovation matrices, ε_t and η_t, is the Wishart distribution. It is an appropriate choice in models where the support of the random variable of interest is restricted to the space of positive semidefinite matrices.[4] Web Appendix B provides an overview of the Wishart distribution including the definitions and notation used in this section.

In GARCH models, the vector of daily returns is usually modelled as $R_t = H_t^{\frac{1}{2}} \xi_t$ with $\xi_t \overset{\text{i.i.d.}}{\sim} N(0, I_k)$, which motivates quasi-maximum likelihood estimation (QMLE). For the HEAVY-P equation, we have $P_t = R_t R_t' = H_t^{\frac{1}{2}} \varepsilon_t H_t^{\frac{1}{2}}$, where $\varepsilon_t = \xi_t \xi_t'$. The assumption that $\xi_t \overset{\text{i.i.d.}}{\sim} N(0, I_k)$ implies that ε_t follows a Wishart distribution.

One of the key results on the Wishart distribution is that if any matrix $S \sim W_k(n, \Sigma)$, then $ASA' \sim W_k(n, A\Sigma A')$ for any $(k \times k)$ nonsingular matrix A. Assuming a Wishart density for ε_t and η_t implies that P_t and V_t are assumed to be conditionally Wishart distributed. However, one distinction between the densities of ε_t and η_t relates to the differences in the ranks of P_t and V_t. The matrix $P_t = R_t R_t'$ has rank 1 by construction if there is at least one non-zero return in R_t. Whether using the realized covariance estimator or the realized kernel of Barndorff-Nielsen *et al.* (2011), the matrix V_t is guaranteed to be full rank under standard regularity conditions, provided that $k < m$, where m is the number of intra-daily returns. This difference in rank entails that ε_t should have a singular Wishart density and η_t a standardized Wishart density. The discussion in Web Appendix B makes it clear that this distinction is necessary for the two conditional moment assumptions, $\mathrm{E}_{t-1}[\varepsilon_t] = I_k$ and $\mathrm{E}_{t-1}[\eta_t] = I_k$, to be satisfied.

Therefore, we assume $\varepsilon_t \overset{\text{i.i.d.}}{\sim} \mathrm{SINGW}_k(1, I_k)$ and $\eta_t \overset{\text{i.i.d.}}{\sim} \mathrm{SW}_k(n, I_k)$. The densities of ε_t and η_t are given by, respectively, (B.2) and (B.1) in Web Appendix B. Thus $P_t|\mathcal{F}_{t-1}^{\mathrm{HF}} \sim \mathrm{SINGW}_k(1, H_t)$ and $V_t|\mathcal{F}_{t-1}^{\mathrm{HF}} \sim \mathrm{SW}_k(n, M_t)$. The distinction between the densities of ε_t and η_t is of no consequence

[4] Some recent multivariate stochastic volatility models also employ the Wishart distribution to model time-varying correlations; see Chib *et al.* (2009) and the references cited therein, and also Jin and Maheu (2010).

DOI: 10.1002/jae

to QMLE as we show in a moment. However, it is needed to have a correctly specified model satisfying $\mathrm{E}_{t-1}[\varepsilon_t] = \mathrm{E}_{t-1}[\eta_t] = I_k$.[5]

3.2. Quasi-Maximum Likelihood Estimation

The HEAVY model is parameterized with a finite-dimensional $(\delta \times 1)$ parameter vector $\theta \in \Theta \subset \mathbb{R}^{\delta}$. Decompose $\theta = (\theta_H', \theta_M')'$, where the $(\delta_H \times 1)$ vector θ_H and $(\delta_M \times 1)$ vector θ_M denote the parameter vectors of the HEAVY-P and HEAVY-V equations, respectively. Let $\theta_0 = (\theta_{H,0}', \theta_{M,0}')'$ denote the true parameter vector. The log-likelihood for the tth observation will be denoted by $l_{H,t}(\theta_H)$ and $l_{M,t}(\theta_M)$. Inference for the HEAVY model will be based on QMLE of the following two log-likelihood functions:

$$l_{H,t}(\theta_H) = c_H - \frac{1}{2}(\log|H_t| + tr(H_t^{-1}P_t)), \quad l_{M,t}(\theta_M) = c_M - \frac{n}{2}(\log|M_t| + tr(M_t^{-1}V_t))$$

where c_H and c_M are constants with respect to θ_H and θ_M; see, respectively, (B.2) and (B.1) in Web Appendix B. Thus the distinction between the densities of ε_t and η_t is of no consequence for QMLE of the model parameters. Engle and Gallo (2006) argue similarly for the Gamma density where the shape parameter is of no consequence when estimating the scale parameter by QMLE.

We assume the initial values, H_0 and M_0, are known and are positive semidefinite. We also assume that θ_H and θ_M are variation free in the sense of Engle *et al.* (1983), which allows for equation-by-equation estimation. This assumption is not essential and is only used to simplify estimation and inference. The QML estimator is $\widehat{\theta} = (\widehat{\theta}_H', \widehat{\theta}_M')'$, where

$$\widehat{\theta}_H = \underset{\theta_H \in \Theta}{\arg\max} L_H(\theta_H), \quad \widehat{\theta}_M = \underset{\theta_M \in \Theta}{\arg\max} L_M(\theta_M)$$

and $L_H(\theta_H) = \sum_{t=1}^{T} l_{H,t}(\theta_H)$, $L_M(\theta_M) = \sum_{t=1}^{T} l_{M,t}(\theta_M)$.

For the BEKK model, Comte and Lieberman (2003) show strong consistency of QMLE by verifying the conditions given in Jeantheau (1998). Hafner and Preminger (2009) show similar results for the VEC model which nests the BEKK model, and their results also apply to integrated processes. An important condition to establish strong consistency is for the model to admit a strictly stationary and ergodic solution, which we assume for the HEAVY model.

Before discussing the asymptotic distribution of $\widehat{\theta}$, we first give results on the score vector in the following proposition. It will be convenient to consider the score for each equation separately.

Proposition 3. (i) The score vectors, $S_{H,t}(\theta_H) = \dfrac{\partial l_{H,t}(\theta_H)}{\partial \theta_H'}$ and $S_{M,t}(\theta_M) = \dfrac{\partial l_{M,t}(\theta_M)}{\partial \theta_M'}$ of dimensions $(1 \times \delta_H)$ and $(1 \times \delta_M)$, respectively, are given by

$$S_{H,t}(\theta_H) = \frac{\partial l_{H,t}(\theta_H)}{\partial \theta_H'} = \frac{1}{2}[(\mathrm{vec}(P_t))' - (\mathrm{vec}(H_t))'](H_t^{-1} \otimes H_t^{-1})\frac{\partial \mathrm{vec}(H_t)}{\partial \theta_H'} \tag{16}$$

$$S_{M,t}(\theta_M) = \frac{\partial l_{M,t}(\theta_M)}{\partial \theta_M'} = \frac{1}{2}[(\mathrm{vec}(V_t))' - (\mathrm{vec}(M_t))'](M_t^{-1} \otimes M_t^{-1})\frac{\partial \mathrm{vec}(M_t)}{\partial \theta_M'} \tag{17}$$

[5] One can test for the Wishart distribution assumption by making use of the property that if $S \sim W_k(n, \Sigma)$, then $\frac{a'Sa}{a'\Sigma a} \sim \chi^2_{(n)}$ for any $(k \times 1)$ vector $a \neq 0$; see Gupta and Nagar (2000). Also, conditional moment tests can be used to detect misspecification.

DOI: 10.1002/jae

(ii) Under $\mathrm{E}_{t-1}[\varepsilon_t] = I_k$ and $\mathrm{E}_{t-1}[\eta_t] = I_k$, the score vectors evaluated at the true parameter value are a martingale difference sequence with respect to $\mathcal{F}_{t-1}^{\mathrm{HF}}$.

The scores have a similar structure to those of GARCH models (e.g. Bollerslev and Wooldridge, 1992). In analogy with generalized least squares, the terms in square brackets can be considered 'errors', while $(H_t^{-1} \otimes H_t^{-1})$ and $(M_t^{-1} \otimes M_t^{-1})$ are weights and the derivatives $\dfrac{\partial \mathrm{vec}(H_t)}{\partial \theta_H'}$ and $\dfrac{\partial \mathrm{vec}(M_t)}{\partial \theta_M'}$ are instruments which are orthogonal to the errors at the maximum likelihood estimator, which is a condition for consistency.

To discuss the asymptotic distribution of the QML estimator, $\widehat{\theta}$, we define the $(1 \times \delta)$ combined score vector $S_t(\theta) = (S_{H,t}(\theta_H), S_{M,t}(\theta_M))$. Having established that the scores are a martingale difference sequence with respect to $\mathcal{F}_{t-1}^{\mathrm{HF}}$, it can be shown under certain regularity conditions (e.g. Comte and Lieberman, 2003) that

$$\sqrt{T}(\widehat{\theta} - \theta_0) \xrightarrow{d} N(0, \mathcal{I}^{-1}\mathcal{J}\mathcal{I}^{-1})$$

where

$$\mathcal{J} = \mathrm{E}[S_t(\theta)' S_t(\theta)] = \mathrm{E}\begin{bmatrix} \dfrac{\partial l_{H,t}(\theta_H)}{\partial \theta_H}\dfrac{\partial l_{H,t}(\theta_H)}{\partial \theta_H'} & \dfrac{\partial l_{H,t}(\theta_H)}{\partial \theta_H}\dfrac{\partial l_{M,t}(\theta_M)}{\partial \theta_M'} \\ \dfrac{\partial l_{M,t}(\theta_M)}{\partial \theta_M}\dfrac{\partial l_{H,t}(\theta_H)}{\partial \theta_H'} & \dfrac{\partial l_{M,t}(\theta_M)}{\partial \theta_M}\dfrac{\partial l_{M,t}(\theta_M)}{\partial \theta_M'} \end{bmatrix},$$

$$\mathcal{I} = -\mathrm{E}\left[\frac{\partial S_t(\theta)}{\partial \theta}\right] = -\mathrm{E}\begin{bmatrix} \dfrac{\partial^2 l_{H,t}(\theta_H)}{\partial \theta_H \partial \theta_H'} & 0 \\ 0 & \dfrac{\partial^2 l_{M,t}(\theta_M)}{\partial \theta_M \partial \theta_M'} \end{bmatrix}$$

The block diagonality of the Hessian, $\mathcal{I}$, is due to the assumption that θ_H and θ_M are variation free, which implies that equation-by-equation standard errors are correct for the HEAVY system. With covariance targeting, a two-step estimation procedure is adopted and in this case the score vector will no longer be a martingale difference sequence, but it will have mean zero at the true parameter value. Also, the Hessian will not be block diagonal due to accounting for the accumulation of estimation error from the first step. We formalize inference in the case of covariance targeting in the following subsection.

3.3. Two-Step Estimation under Covariance Targeting

With covariance targeting, the parameter vectors θ_H and θ_M are decomposed into $\theta_H = (\omega_H', \widetilde{\theta}_H')'$ and $\theta_M = (\omega_M', \widetilde{\theta}_M')'$ and are to be estimated in two steps. The unconditional moments, ω_H and ω_M, will be estimated in the first step by a moment estimator

$$\widehat{\omega}_H = T^{-1}\sum_{t=1}^{T} p_t, \quad \widehat{\omega}_M = T^{-1}\sum_{t=1}^{T} v_t$$

and then $\widetilde{\theta}_H$ and $\widetilde{\theta}_M$ will be estimated by QMLE in the second step. The asymptotics of the QML estimator in this case is a direct application of two-step generalized method of moments (GMM) estimation discussed in Newey and McFadden (1994). Define $\widetilde{l}_{H,t}(\omega_H, \omega_M, \widetilde{\theta}_H)$ and $\widetilde{l}_{M,t}(\omega_M, \widetilde{\theta}_M)$ to be the tth observation log-likelihoods for the covariance targeting HEAVY model. Two-step

DOI: 10.1002/jae

estimation gives the following $(1 \times \delta)$ vector of moment conditions:

$$\widetilde{S}_t(\widetilde{\theta}) = \left((p_t - \omega_H)', \frac{\partial \widetilde{l}_{H,t}}{\partial \widetilde{\theta}'_H}, (v_t - \omega_M)', \frac{\partial \widetilde{l}_{M,t}}{\partial \widetilde{\theta}'_M} \right), \quad \widetilde{\theta} = (\omega'_H, \widetilde{\theta}'_H, \omega'_M, \widetilde{\theta}'_M)'$$

which is no longer martingale difference sequence with respect to $\mathcal{F}^{\mathrm{HF}}_{t-1}$. In this case

$$\sqrt{T}(\widehat{\theta} - \theta_0) \xrightarrow{d} N(0, \mathcal{I}^{-1}\mathcal{J}(\mathcal{I}^{-1})')$$

where

$$\mathcal{J} = \mathrm{var}\left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \widetilde{S}_t(\widetilde{\theta}) \right],$$

$$\mathcal{I} = -\mathrm{E}\left[\frac{\partial \widetilde{S}_t(\widetilde{\theta})}{\partial \widetilde{\theta}} \right] = -\mathrm{E}\begin{bmatrix} -I_{k^*} & \frac{\partial^2 \widetilde{l}_{H,t}}{\partial \omega_H \partial \widetilde{\theta}'_H} & 0 & 0 \\ 0 & \frac{\partial^2 \widetilde{l}_{H,t}}{\partial \widetilde{\theta}_H \partial \widetilde{\theta}'_H} & 0 & 0 \\ 0 & \frac{\partial^2 \widetilde{l}_{H,t}}{\partial \omega_M \partial \widetilde{\theta}'_H} & -I_{k^*} & \frac{\partial^2 \widetilde{l}_{M,t}}{\partial \omega_M \partial \widetilde{\theta}'_M} \\ 0 & 0 & 0 & \frac{\partial^2 \widetilde{l}_{M,t}}{\partial \widetilde{\theta}_M \partial \widetilde{\theta}'_M} \end{bmatrix}$$

In implementation we use a HAC estimator (e.g. Newey and West, 1987) to estimate $\mathcal{J}$. With covariance targeting, variation freeness between the parameters of the HEAVY-P and HEAVY-V equations no longer holds since κ depends on ω_M. Thus the block $\frac{\partial^2 \widetilde{l}_{H,t}}{\partial \omega_M \partial \widetilde{\theta}'_H}$ now appears in the Hessian to account for this dependence in the second step of estimation.

4. MODEL EVALUATION

For out-of-sample model evaluation, we use a quasi-likelihood (QLIK) loss function of the form

$$L_{t,s}(\Sigma_{t+s}, H^a_{t+s|t}) = \log |H^a_{t+s|t}| + tr((H^a_{t+s|t})^{-1}\Sigma_{t+s}) \tag{18}$$

where Σ_{t+s} is the actual (unobserved) covariance matrix and $H^a_{t+s|t}$ denotes its s-step forecast using model a conditional on time t information. Since Σ_{t+s} is unobservable, our analysis will be based on some proxy denoted by $\widehat{\Sigma}_{t+s}$, which we take to be the realized covariance matrix, V_{t+s}. The loss function (18) evaluates the s-step predicted density from model a using the proxy $\widehat{\Sigma}_{t+s}$ as data,[6] and it provides a consistent ranking of volatility models in the sense of Patton (2011) and Patton and Sheppard (2009) as it is robust to noise in the proxy $\widehat{\Sigma}_{t+s}$; see also Laurent *et al.* (2009).

Note that even if—at time t—the true density of R_{t+1} is normal (i.e. the density of P_{t+1} is Wishart), normality will not hold under temporal aggregation unless the conditional covariance

[6] Note that (18) is the negative of the log-likelihood of a multivariate normal density excluding the constant terms. The switched sign is due to defining (18) as a 'loss' function.

matrix is constant. Therefore the s-step density will not be normal, implying that the density used for the QLIK loss function (18) is misspecified. However, the loss difference between two competing models a and b, $L_{t,s}(\Sigma_{t+s}, H^a_{t+s|t}) - L_{t,s}(\Sigma_{t+s}, H^b_{t+s|t})$, can be interpreted as a Kullback–Leibler distance, which yields a valid assessment even if both models are misspecified. Cox (1961) proposes a likelihood ratio test based on this idea, while Vuong (1989) provides the theoretical framework in the case of nested and non-nested models. Similar approaches are proposed for out-of-sample model selection in Amisano and Giacomini (2007) and Diks *et al.* (2008).

We denote the loss difference between the HEAVY and GARCH models by

$$D_{t,s} = L_{t,s}(\Sigma_{t+s}, H^{\text{HEAVY}}_{t+s|t}) - L_{t,s}(\Sigma_{t+s}, H^{\text{GARCH}}_{t+s|t}), \quad t = Q, Q+1, \ldots, T-s$$

where $L_{t,s}(\cdot)$ is given by (18), T is the size of the full sample and Q is the size of the estimation window. We assume Q is fixed so that we use a rolling window of data to estimate the model parameters, which gives $T - Q - s + 1$ data points for out-of-sample model evaluation. The average loss is denoted by

$$\overline{D}_s = \frac{1}{T - Q - s + 1} \sum_{t=Q}^{T-s} D_{t,s}$$

which is used to test $H_0 : E[D_{t,s}] = 0$, for all s, against a two-sided alternative. Let $\overline{D}^*_s$ denote the average loss evaluated at the true parameter value, then we have

$$\sqrt{T}(\overline{D}_s - \overline{D}^*_s) \xrightarrow{d} N(0, \Lambda_s)$$

where Λ_s is the asymptotic variance of $D_{t,s}$ estimated using a HAC estimator. Significantly negative values of the test statistic indicate superior forecast performance of the HEAVY model. This predictive ability test was first introduced by Diebold and Mariano (1995), and later formalized by West (1996) and Giacomini and White (2006).

We extend this strategy in the context of multivariate volatility models by conducting separate tests for forecasts of the individual variances and also for the dependence structure of the group of assets under consideration. Consider the margins–copula decomposition of the log-likelihood of R_t:

$$\log f(X) = \sum_{i=1}^{k} \log f_i(x_i) + \log c(F_1(x_1), F_2(x_2), \ldots, F_k(x_k)) \tag{19}$$

where $f(X)$ is the joint density of the returns of the k assets, $f_i(x_i)$ and $F_i(x_i)$, $i = 1, \ldots, k$, are respectively the density and cumulative distribution function of asset i returns, and $c(\cdot)$ is the copula density.[7] The normality assumption for R_t implies that $f(X)$, $f_i(x_i)$ and $c(\cdot)$ correspond to the multivariate normal density, normal density and normal copula, respectively.

We decompose the QLIK loss in (18) in a similar fashion to (19). Thus computing the loss in (18) based on the whole forecast matrix $(H^a_{t+s|t})$ corresponds to $\log f(X)$, while computing the loss based on a particular diagonal element of $H^a_{t+s|t}$, say $h^a_{ii,t+s|t}$, corresponds to $\log f_i(x_i)$. The latter corresponds to the loss encountered in forecasting the individual variance for asset i, and we compute it for all k assets. We compute the loss attributed to forecasting the dependence structure (summarized by the copula contribution) as the residual, i.e. corresponding

[7] Nelsen (2006) and Patton (2009) provide recent reviews of copula theory and financial applications.

DOI: 10.1002/jae

to $\log f(X) - \sum_{i=1}^{k} \log f_i(x_i)$. Based on this QLIK loss decomposition, we conduct the predictive ability test, outlined above, separately for each margin as well as the copula. Due to the normality assumption, the copula parameter is the conditional correlation matrix of the daily returns; thus we use the terms margins–copula and variances–correlations interchangeably.

5. EMPIRICAL APPLICATION

We use high-frequency data on Spyder (SPY), the S&P 500 ETF, along with some of the most liquid stocks in the Dow Jones Industrial Average (DJIA) index. These are: Alcoa (AA), American Express (AXP), Bank of America (BAC), Coca Cola (KO), Du Pont (DD), General Electric (GE), International Business Machines (IBM), JP Morgan (JPM), Microsoft (MSFT), and Exxon Mobil (XOM). The sample period is 1 February 2001 to 31 December 2009 with a total of 2242 trading days, and the data source is the TAQ database. We choose the starting date for the sample to be after decimal pricing had been fully implemented in the NYSE, which took place on 29 January 2001.

We focus on the realized covariance matrix as our choice for V_t. In computing the realized covariance matrix, we use 5-minute returns with subsampling. We exclude the opening and closing 15 minutes of trading to control for overnight effects. For the daily return, we focus on the open-to-close returns, which of course ignore overnight effects, and for consistency with the realized covariance estimator we compute the open-to-close daily returns over the same interval.[8] Our estimation and forecast evaluation computations were repeated using the noise-robust realized kernel of Barndorff-Nielsen *et al.* (2011) with the results being qualitatively similar in general.[9]

The main focus of our empirical application will be on modelling and forecasting the conditional covariance matrix of a stock (BAC) and an index (S&P 500) using the scalar HEAVY model. Most of the model's features can be readily seen in this bivariate model, which is analysed in Section 5.1. In Section 5.2 we report estimates of the scalar HEAVY model for the 10 DJIA stocks using covariance targeting. In Web Appendix C, we report empirical results for the diagonal HEAVY model for SPY-BAC, as well as scalar and diagonal models for other pairs of assets selected from the 10 DJIA stocks.

5.1. Bivariate Scalar HEAVY Model: S&P 500 and Bank of America

Figure 1 contains the annualized realized volatility of SPY and BAC, their realized correlation and realized beta for BAC over the full sample. The sharp increase in volatility in 2008–2009 is associated with the turmoil in financial markets during the recent financial crisis. The increase in BAC volatility is much more pronounced especially after the collapse of Lehman Brothers in mid September 2008. BAC realized correlation with the market seems to have been relatively high during the crisis, and its realized beta increased sharply and was very volatile during this period.

In Table I, we present the HEAVY and GARCH model estimates. We also report estimates for the GARCH-X model, which is similar to (7) with p_{t-1} included on the right-hand side with coefficient D_{GX}. Thus the GARCH-X model nests both the HEAVY-P equation and the GARCH model. For ease of interpretation, we only report the parameter estimates for the models' vech representation excluding the constant terms.

The estimate of B_H implies that the elements of H_t will be smooth, although less smooth than the corresponding estimates from the GARCH model with the estimate of B_G equal to 0.934.

[8] We also estimated some of the models using close-to-close returns. The differences in results are discussed at the end of Section 5.1.
[9] These are not reported in the interest of parsimony, but are available upon request.

DOI: 10.1002/jae

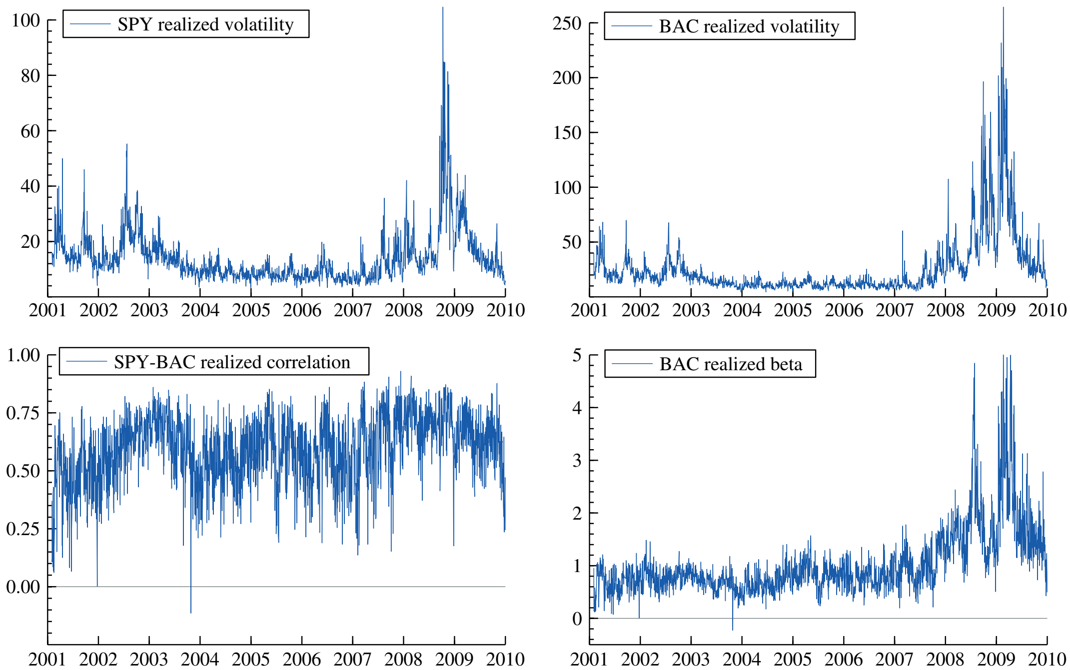

Figure 1. SPY and BAC annualized realized volatility, realized correlation and BAC realized beta. This figure is available in color online at wileyonlinelibrary.com/journal/jae

For the HEAVY-V equation, the B_M coefficient is relatively small, implying that the estimated conditional moments will be somewhat erratic. In terms of magnitude, these estimates are largely in line with those from the univariate HEAVY model in Shephard and Sheppard (2010), and they also suggest a somewhat high level of persistence. Compared to the nesting GARCH-X model, there is no loss of fit when moving to HEAVY-P since the coefficient on p_{t-1} (D_{GX}) is not statistically significant. This is not the case when moving from GARCH-X to GARCH, which suggests that v_{t-1} effectively crowds out p_{t-1}.

The estimates also suggest that the HEAVY model's half-life (of a deviation of the one-step forecast of h_t from its long run) is substantially shorter than that of the GARCH model, suggesting that the former's forecast responds faster to abrupt changes in the level of volatility or correlation.[10]

The log-likelihood and its decomposition into margin and copula likelihoods in the middle panel of Table I indicate an improvement in fit of the HEAVY-P equation compared to the GARCH model. Note that the two models are non-nested, so direct LR tests are not possible; however, we will present below the outcome of the predictive ability tests discussed in Section 4. Although non-nested, the decomposition suggests that the HEAVY-P equation improves on GARCH for both the margins and the copula. The model residuals, $\widehat{\varepsilon}_t$ and $\widehat{\eta}_t$, seem to be centred around the identity matrix, with the exception of two large outliers in $\widehat{\eta}_t$ corresponding to the realized variances of SPY and BAC on 27 February 2007, due to the 9% fall in the Shanghai stock exchange index that day.

An interesting feature from the residual analysis is that it displays evidence of the leverage effect between the returns and the realized measure. This is shown in Figure 2. The upper-left

[10] The half-life can be easily computed from (15) by noting that the two gaps, $(h_{t+1} - \omega_H)$ and $(m_{t+1} - \omega_M)$, tend to have the same sign as our results indicate that the elements of h_t and m_t tend to be very highly correlated. Thus these two gaps can be set, without loss of generality, equal to a $(k^* \times 1)$ vector of ones.

DOI: 10.1002/jae

Table I. Scalar HEAVY estimation and forecast evaluation results for SPY-BAC. Top panel: parameter estimates of HEAVY, GARCH and GARCH-X with standard errors reported in parentheses. Middle panel: decomposition of the log-likelihood (excluding constant terms) at the estimated parameter values. Bottom panel: t-statistics of the predictive ability tests for HEAVY versus GARCH

	HEAVY-P		GARCH		GARCH-X			HEAVY-V	
	A_H	B_H	A_G	B_G	A_{GX}	B_{GX}	D_{GX}	A_M	B_M
SPY-BAC	0.214	0.727	0.062	0.934	0.187	0.741	0.019	0.421	0.574
(SE)	(0.054)	(0.068)	(0.010)	(0.011)	(0.056)	(0.068)	(0.012)	(0.033)	(0.033)

Log-likelihood decomposition (HEAVY-P versus GARCH)

	HEAVY-P	GARCH	HEAVY-P gains
Margin 1 (SPY)	−658	−713	55
Margin 2 (BAC)	−1593	−1648	55
Copula	815	808	7
Joint distribution	−1436	−1553	117

Predictive ability tests at different forecast horizons (days)

	(1)	(2)	(3)	(5)	(10)	(22)
Margin 1 (SPY)	−3.72	−3.03	−2.33	−1.23	0.84	1.87
Margin 2 (BAC)	−3.27	−2.45	−1.70	−0.58	1.06	2.04
Copula	−3.37	−3.22	−3.39	−3.28	−3.26	−3.85
Joint distribution	−4.32	−3.78	−3.23	−2.33	−0.07	1.03

chart shows the scatter-plot of $\widehat{\xi}_{1,t}$ and $\widehat{\eta}_{11,t}$, which are the innovations to the daily return and realized variance of SPY, respectively.[11] The lower-left chart displays the innovations to the daily return and realized variance of BAC. The right panel charts correspond to the same plots but mapped into copula space where the empirical distribution function is used to transform the innovations into probability integral transforms. The leverage effect can be seen in the right panel. For instance, large negative innovations to SPY returns tend to be associated with large positive innovations to its realized variance, indicating higher volatility in response to bad news. The same applies to BAC innovations.

The bottom panel of Table I gives the results of the predictive ability tests. We estimate the model using a rolling window of 1486 observations and then use the parameter estimates to obtain forecasts of H_t at horizons $s = 1, 2, 3, 5, 10, 22$ days using (14). The size of the rolling window is chosen such that our forecasts start at 3 January 2007. The reported figures are t-statistics to test equal predictive ability and significantly negative t-statistics favour the HEAVY model over the GARCH model. The results show that HEAVY outperforms GARCH especially at short forecast horizons. This is true for the whole covariance matrix forecast as well as its decomposition into margins and copula, which provides further insight into the source of forecast gains. The copula gains are maintained at longer forecast horizons, indicating that the realized measure provides valuable information for forecasting the conditional correlation.

As pointed out earlier, the forecast profile of the HEAVY model is distinct from that of the GARCH(1,1) model particularly over short forecast horizons due to momentum effects. This can be seen in Figure 3, which plots the forecasts of the SPY-BAC conditional correlation (implied by the forecasts of H_t) over the period 3 November 2008 to 30 September 2009. This is an

[11] $\widehat{\xi}_{1,t}$ is the first element of the vector $\widehat{\xi}_t = \widehat{H}_t^{-\frac{1}{2}} R_t$, and $\widehat{\eta}_{11,t}$ is the (1, 1) element of the matrix $\widehat{\eta}_t = \widehat{M}_t^{-\frac{1}{2}} V_t \widehat{M}_t^{-\frac{1}{2}}$.

DOI: 10.1002/jae

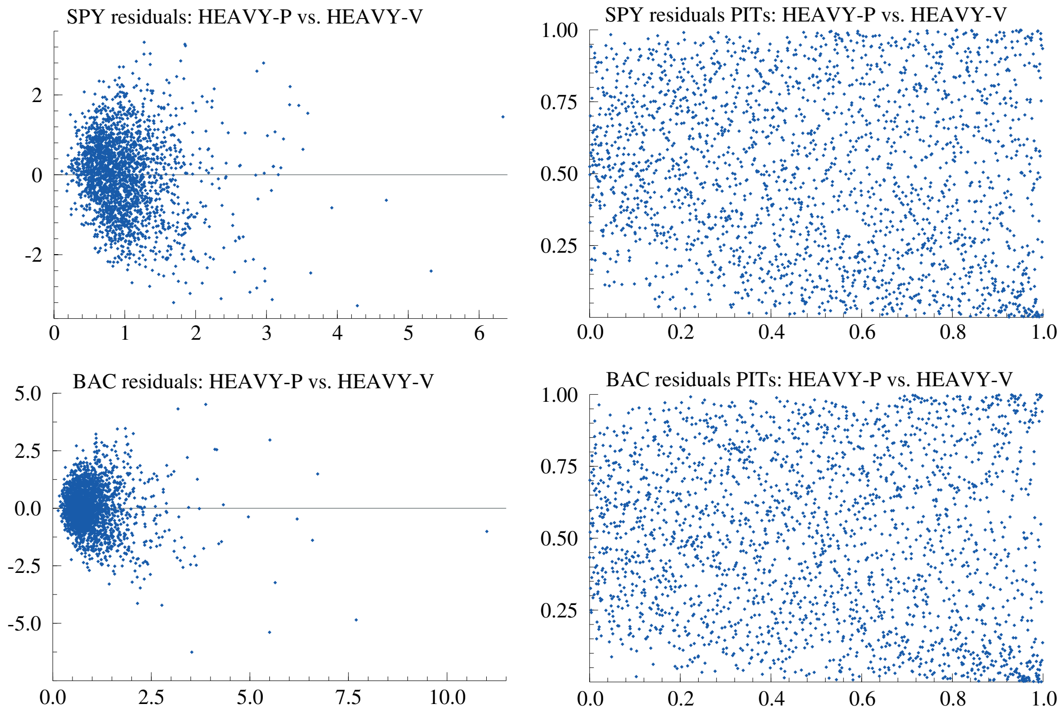

Figure 2. Left: scatter-plots of SPY and BAC residuals in the HEAVY-P and HEAVY-V equations. Right: scatter-plots of the residuals mapped into probability integral transforms (PITs). This figure is available in color online at wileyonlinelibrary.com/journal/jae

interesting period for analysis as it marks a very volatile period during the 2007–2009 financial crisis. The solid lines are the one-step forecasts, and at selected points we plot the forecast profile at this date for 22 days into the future. We do this only for selected peak and trough points for clarity of illustration. The momentum effects in the HEAVY model can be readily seen. Whereas the GARCH correlation forecast monotonically mean reverts, the HEAVY forecast displays some short-run momentum influenced by the deviation of the realized measure from its long run before ultimately mean reverting. Interestingly, the plot also shows how the one-step forecasts from both models diverge in some periods, pointing to important differences in the information content of the realized measure and the outer product of daily returns.

It is interesting to track the model's performance in relation to the accuracy of the realized measure. For this purpose, we report in Table II the parameter estimates, log-likelihood gains and out-of-sample performance using various sampling intervals for the realized covariance estimator. The table also includes results when using the realized kernel as the realized measure. In general, the results indicate that when sampling between 5 and 15 minutes the parameter estimates of the HEAVY and GARCH-X models are rather stable, implying similar persistence levels, and indeed the estimates become very close when sampling at 30 minutes. At 1-minute sampling, there is substantial drop in the estimate of B_H and a moderate increase in A_H. Using the realized kernel leads to a noticeable decline in the smoothing parameters in both equations of the HEAVY model as well as the GARCH-X model. In terms of forecasting performance, the results are similar.

To investigate the sensitivity of the results to including overnight effects, we also estimated the scalar HEAVY model using close-to-close returns for SPY-BAC and also for other asset pairs selected from the 10 DJIA stocks and analyzed in Web Appendix C. The primary difference when

DOI: 10.1002/jae

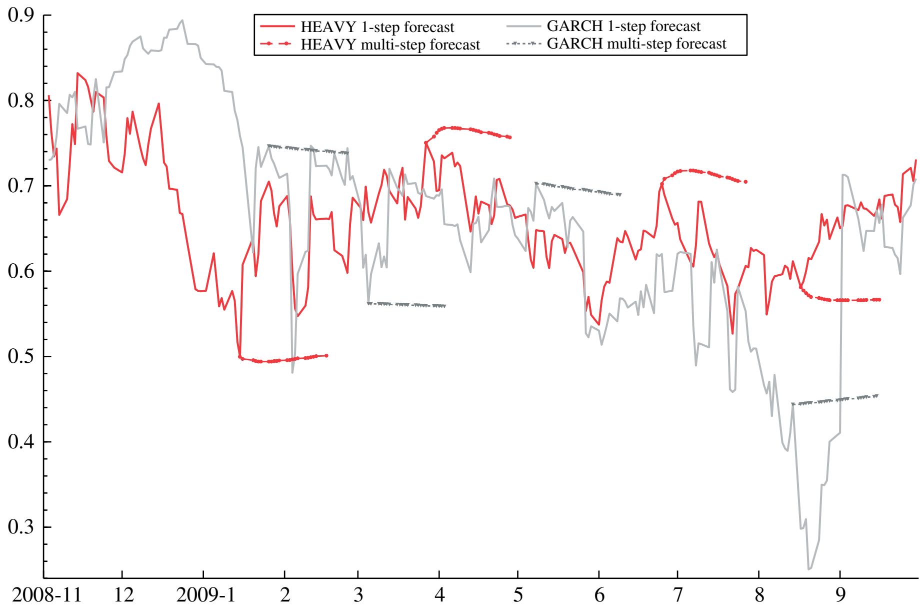

Figure 3. One-step and multi-step forecasts for the SPY-BAC conditional correlation. This figure is available in color online at wileyonlinelibrary.com/journal/jae

Table II. Scalar HEAVY estimation and forecast evaluation results for SPY-BAC using different realized measures. Top panel: scalar HEAVY and GARCH-X parameter estimates using different sampling intervals in computing the realized covariance and also using the realized kernel. Log-likelihood gains from the HEAVY model are reported in the last column. Bottom panel: *t*-statistics of the predictive ability tests for HEAVY versus GARCH

	HEAVY-P		GARCH-X			HEAVY-V		
	A_H	B_H	A_{GX}	B_{GX}	D_{GX}	A_M	B_M	Joint LL gain
RC 1-min	0.256	0.597	0.128	0.741	0.052	0.527	0.471	65
RC 5-min	0.214	0.727	0.187	0.741	0.019	0.421	0.574	117
RC 10-min	0.202	0.760	0.189	0.764	0.011	0.362	0.633	120
RC 15-min	0.185	0.787	0.169	0.793	0.012	0.300	0.696	116
RC 30-min	0.143	0.842	0.134	0.843	0.009	0.236	0.759	107
Realized kernel	0.213	0.677	0.194	0.689	0.015	0.508	0.488	128

Joint distribution predictive ability tests at different forecast horizons (days)

	(1)	(2)	(3)	(5)	(10)	(22)
RC 1-min	−3.59	−3.08	−2.47	−1.82	0.30	1.50
RC 5-min	−4.32	−3.78	−3.23	−2.33	−0.08	1.03
RC 10-min	−4.08	−3.70	−3.21	−2.35	−0.23	0.99
RC 15-min	−4.26	−3.87	−3.34	−2.64	−0.26	1.11
RC 30-min	−3.81	−3.34	−2.75	−2.01	−0.13	1.12
Realized kernel	−4.25	−3.76	−3.23	−2.63	−0.46	1.00

using close-to-close returns is an increase in the loadings on the shock terms in both the HEAVY and GARCH models through A_H, A_M and A_G, and particularly so for the GARCH model. The HEAVY model still provides gains for the joint and marginal log-likelihoods. The copula gains

DOI: 10.1002/jae

are obtained only for the pairs IBM-MSFT, AXP-DD and GE-KO. Interestingly, the predictive ability test results indicate that the HEAVY model gains for the joint log-likelihood are sustained at all horizons in most cases, which is also the case for some of the margins. The copula gains are significant at all horizons for the pairs IBM-MSFT and AXP-DD, only at longer horizons for SPY-BAC and BAC-JPM, and insignificant for XOM-AA and GE-KO.

5.2. Covariance Targeting Scalar HEAVY Model

In this subsection, we estimate the scalar HEAVY model including all 10 DJIA assets. We show the estimation results for both the original HEAVY specification and the covariance targeting model given by (13)–(12). We focus on this covariance targeting specification since it is easier to handle the parameter restrictions required for covariance stationarity and positive definiteness of the target. For the GARCH model, we also estimate its covariance targeting parameterization, which has a similar structure to (12). With covariance targeting, the number of parameters to be estimated through numerical optimization is reduced from 57 to 2 parameters per equation, where the latter are the dynamic parameters of interest.

Table III presents the estimates of the dynamic parameters for the HEAVY and GARCH models. The parameter estimates show some differences compared to the average estimate from bivariate models for the same assets; see Web Appendix C. The estimates of the smoothing parameters (B_H, B_M and B_G) have all increased, especially B_M, while the estimates of A_H, A_M and A_G are now smaller. The log-likelihood decomposition results show uniform gains for the HEAVY model in all margins and the copula. The copula gains seem particularly impressive. In terms of parameter estimates and the log-likelihood decomposition, the covariance targeting model (bottom panel) shows only slight differences compared to the non-targeting specification.

In Figure 4, we present summary results of the predictive ability tests for the covariance targeting scalar HEAVY and GARCH models. The figure shows the t-statistics for tests of the joint distribution and copula, as well as the minimum, maximum and median t-statistics for the 10 margins. In the first 3 days, the HEAVY model gains are confirmed for the joint distribution, all margins and the copula. The gains of the joint distribution are maintained up to 11 days ahead, then it falls into the insignificance region before improving again towards the end of the forecast horizon. For the margins, the median t-statistics show gains up to 7 days ahead. The copula gains are maintained throughout until the end of the forecast horizon, which is consistent with the substantial overall gain in the copula log-likelihood.

6. CONCLUSION

This paper introduces a new class of multivariate volatility models with robust performance in out-of-sample prediction of the covariance matrix for a collection of financial assets. While GARCH models—in their many variations—have proved successful in the past two decades, the increasing availability of high-frequency data provides important additional information. Utilizing this information to forecast the conditional variance of daily asset returns has already borne fruit in the univariate case, as documented by several recent studies.

Our study is one of the first to document this feature in the multivariate case using a relatively large group of assets. We present our results in the framework of the multivariate HEAVY class of models. Using a linear specification, we discuss in some detail the model's dynamic properties, its covariance targeting representation, and provide closed-form forecasting formulas. We show how the profile of forecasts from HEAVY models differs from GARCH models, in particular with

Table III. Scalar HEAVY estimates for 10 DJIA assets. Top panel: parameter estimates and log-likelihood (excluding constant terms) decomposition for scalar HEAVY and GARCH without covariance targeting. Bottom panel: parameter estimates and log-likelihood (excluding constant terms) decomposition for the covariance targeting scalar HEAVY and GARCH

Scalar models

	HEAVY-P		GARCH		HEAVY-V	
	A_H	B_H	A_G	B_G	A_M	B_M
Dynamic parameters	0.141	0.792	0.024	0.973	0.247	0.744
(SE)	(0.021)	(0.037)	(0.002)	(0.001)	(0.011)	(0.010)

Log-likelihood decomposition

	HEAVY-P	GARCH	HEAVY-P gains
Margin 1 (BAC)	−1611	−1696	85
Margin 2 (JPM)	−1999	−2098	99
Margin 3 (IBM)	−1267	−1323	56
Margin 4 (MSFT)	−1471	−1525	54
Margin 5 (XOM)	−1331	−1420	89
Margin 6 (AA)	−2332	−2381	49
Margin 7 (AXP)	−1957	−2034	77
Margin 8 (DD)	−1530	−1595	65
Margin 9 (GE)	−1532	−1590	58
Margin 10 (KO)	−911	−956	45
Copula	4861	4661	200
Joint distribution	−11080	−11958	878

Covariance targeting scalar models

	HEAVY-P		GARCH		HEAVY-V	
	A_H	B_H	A_G	B_G	A_M	B_M
Dynamic parameters	0.177	0.818	0.022	0.977	0.234	0.761
(SE)	(0.022)	(0.023)	(0.001)	(0.001)	(0.009)	(0.010)

Log-likelihood decomposition

	HEAVY-P	GARCH	HEAVY-P gains
Margin 1 (BAC)	−1616	−1753	138
Margin 2 (JPM)	−1985	−2119	133
Margin 3 (IBM)	−1257	−1327	69
Margin 4 (MSFT)	−1464	−1525	61
Margin 5 (XOM)	−1340	−1424	84
Margin 6 (AA)	−2324	−2379	55
Margin 7 (AXP)	−1940	−2046	106
Margin 8 (DD)	−1528	−1592	64
Margin 9 (GE)	−1521	−1595	74
Margin 10 (KO)	−911	−954	43
Copula	4781	4629	151
Joint distribution	−11105	−12084	978

regard to persistence and short-run momentum effects. We also discuss QMLE of HEAVY models under the assumption of a Wishart distribution for the innovation matrices.

In an application to the S&P 500 ETF and 10 stocks from the DJIA index, we compare the HEAVY and GARCH models in the challenging environment of the financial crisis. We show that

DOI: 10.1002/jae

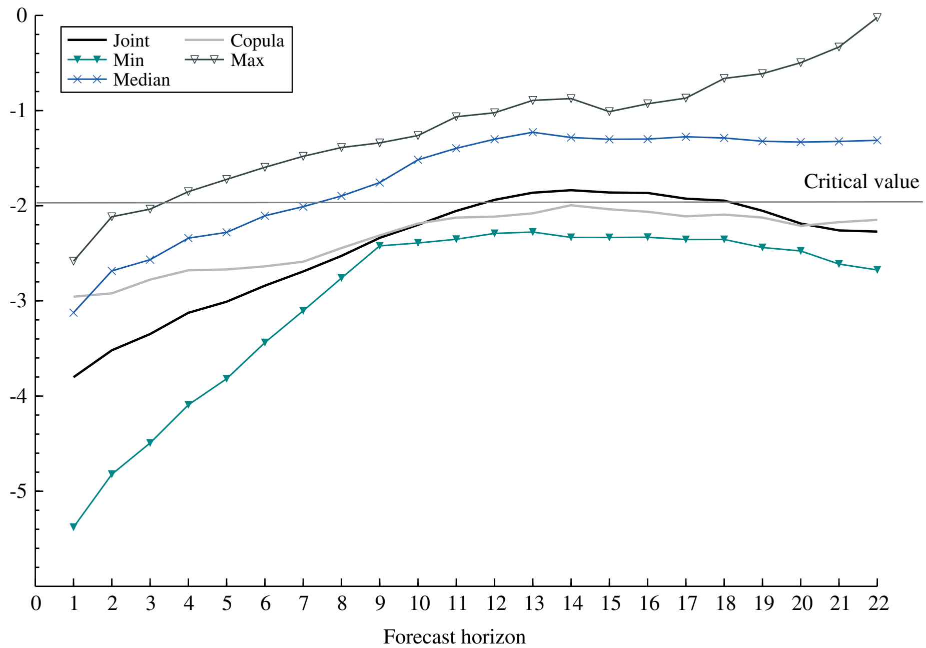

Figure 4. Predictive ability tests' t-statistics for the covariance targeting scalar HEAVY and GARCH models. This figure is available in color online at wileyonlinelibrary.com/journal/jae

forecasts from the HEAVY model dominate GARCH forecasts, with the gains being particularly significant at short forecast horizons. The results seem consistent across different pairs of assets and also when using all 10 DJIA stocks in a covariance targeting model. The HEAVY model's relatively short response time compared to GARCH seems to enable it to efficiently track sudden changes in asset return volatilities and correlations. With regard to the latter, our results for log-likelihood decompositions and predictive ability tests strongly suggest that high-frequency data provide timely and important information for modelling and forecasting conditional correlations.

For future research, a number of extensions could potentially add to our understanding of how best to model and forecast multivariate volatility. It would be interesting to add asymmetric terms to the HEAVY model to explicitly capture the leverage effect and see how this improves its forecast performance. It might also be beneficial to use a long-run/short-run component model in the dynamic equations to separate out transitory movements in volatility.

REFERENCES

Amisano G, Giacomini R. 2007. Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business and Economic Statistics* **25**: 177–190.

Andersen TG, Bollerslev T. 1998. Answering the skeptics: yes, standard volatility models do provide accurate forecasts. *International Economic Review* **39**: 885–905.

Andersen TG, Bollerslev T, Diebold FX, Labys P. 2003. Modeling and forecasting realized volatility. *Econometrica* **71**: 579–625.

Andersen TG, Bollerslev T, Diebold FX. 2007. Roughing it up: including jump components in the measurement, modeling and forecasting of return volatility. *Review of Economics and Statistics* **89**: 701–720.

Andersen TG, Bollerslev T, Meddahi N. 2011. Realized volatility forecasting and market microstructure noise. *Journal of Econometrics* **160**: 220–234.

DOI: 10.1002/jae

Asai M, McAleer M, Yu J. 2006. Multivariate stochastic volatility: a review. *Econometric Reviews* **25**: 145–175.
Barndorff-Nielsen OE, Shephard N. 2002. Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B* **64**: 253–280.
Barndorff-Nielsen OE, Shephard N. 2004. Econometric analysis of realised covariation: high frequency based covariance, regression and correlation in financial economics. *Econometrica* **72**: 885–925.
Barndorff-Nielsen OE, Hansen PR, Lunde A, Shephard N. 2008. Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica* **76**: 1481–1536.
Barndorff-Nielsen OE, Hansen PR, Lunde A, Shephard N. 2011. Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics* **162**: 149–169.
Bauer GH, Vorkink K. 2011. Forecasting multivariate realized stock market volatility. *Journal of Econometrics* **160**: 93–101.
Bauwens L, Laurent S, Rombouts JVK. 2006. Multivariate GARCH models: a survey. *Journal of Applied Econometrics* **21**: 79–109.
Bollerslev T, Wooldridge JM. 1992. Quasi maximum likelihood estimation and inference in dynamic models with time varying covariances. *Econometric Reviews* **11**: 143–172.
Brownlees CT, Engle RF. 2010. *Volatility, correlation and tails for systemic risk measurement*. Working paper, Stern School of Business, New York University.
Brownlees CT, Gallo GM. 2010. Comparison of volatility measures: a risk management perspective. *Journal of Financial Econometrics* **8**: 29–56.
Chib S, Omori Y, Asai M. 2009. Multivariate stochastic volatility. In *Handbook of Financial Time Series*, Andersen TG, Davis RA, Kreiss JP, Mikosch T (eds). Springer: Berlin; 365–400.
Chiriac R, Voev V. 2011. Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics* (forthcoming).
Cipollini F, Engle RF, Gallo GM. 2007. *A model for multivariate non-negative valued processes in financial econometrics*. Working Paper: Stern School of Business, New York University.
Comte F, Lieberman O. 2003. Asymptotic theory for multivariate GARCH processes. *Journal of Multivariate Analysis* **84**: 61–84.
Corsi F, Mittnik S, Pigorsch C, Pigorsch U. 2008. The volatility of realized volatility. *Econometric Reviews* **17**: 46–78.
Cox DR. 1961. Tests of separate families of hypotheses. In *Proceedings of the Berkeley Symposium 4*; 105–123.
Diebold FX, Mariano RS. 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* **13**: 253–263.
Diks C, Panchenko V, van Dijk D. 2008. *Out-of-sample comparison of copula specifications in multivariate density forecasts*. Working paper, Australian School of Business, University of New South Wales.
Engle RF. 2002. New frontiers for ARCH models. *Journal of Applied Econometrics* **17**: 425–446.
Engle RF, Gallo JP. 2006. A multiple indicator model for volatility using intra daily data. *Journal of Econometrics* **131**: 3–27.
Engle RF, Kroner KF. 1995. Multivariate simultaneous generalized ARCH. *Econometric Theory* **11**: 122–150.
Engle RF, Lee GGJ. 1999. A permanent and transitory component model of stock return volatility. In *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W.J. Granger*, Engle RF, White H (eds). Oxford University Press: Oxford; 475–497.
Engle RF, Mezrich J. 1996. GARCH for groups. *Risk* **9**: 36–40.
Engle RF, Hendry DF, Richard JF. 1983. Exogeneity. *Econometrica* **51**: 277–304.
Giacomini R, White H. 2006. Tests of conditional predictive ability. *Econometrica* **74**: 1545–1578.
Gupta AK, Nagar DK. 2000. *Matrix Variate Distributions*. Chapman & Hall: New York.
Hafner CM. 2003. Fourth moment structure of multivariate GARCH models. *Journal of Financial Econometrics* **1**: 26–54.
Hafner CM, Preminger A. 2009. On asymptotic theory for multivariate GARCH models. *Journal of Multivariate Analysis* **100**: 2044–2054.
Hansen PR, Lunde A, Voev V. 2010. *Realized beta GARCH: a multivariate GARCH model with realized measures of volatility and covolatility*. Working paper, CREATES, Aarhus University.
Hansen PR, Huang ZA, Shek HH. 2011. Realized GARCH: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics* (forthcoming).
Jeantheau T. 1998. Strong consistency of estimators for multivariate ARCH models. *Econometric Theory* **14**: 70–86.

Jin X, Maheu JM. 2010. *Modelling realized covariances and returns*. Working paper, Department of Economics, University of Toronto.
Laurent S, Rombouts JVK, Violante F. 2009. *On loss functions and ranking forecasting performances of multivariate volatility models*. Working paper: Centre Interuniversitaire sur le Risque, les Politiques Economiques et l'Emploi (CIRPEE).
Lutkepohl H. 1996. *Handbook of Matrices*. Wiley: Chichester.
Magnus JR. 1988. *Linear Structures*. Griffin: London.
Nelsen RB. 2006. *An Introduction to Copulas* (2 edn). Springer: New York.
Newey WK, McFadden D. 1994. Large sample estimation and hypothesis testing. In *The Handbook of Econometrics*, Vol. 4, Engle RF, McFadden D (eds). North-Holland: Amsterdam; 2111–2245.
Patton AJ. 2009. Copula-based models for financial time series. In *Handbook of Financial Time Series*, Andersen TG, Davis RA, Kreiss JP, Mikosch T (eds). Springer: Berlin; 767–785.
Patton AJ. 2011. Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics* **160**: 246–256.
Patton AJ, Sheppard KK. 2009. Evaluating volatility and correlation forecasts. In *Handbook of Financial Time Series*, Andersen TG, Davis RA, Kreiss JP, Mikosch T (eds). Springer: Berlin; 801–838.
Shephard N, Sheppard KK. 2010. Realising the future: forecasting with high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics* **25**: 197–231.
Voev V. 2008. Dynamic modelling of large-dimensional covariance matrices. In *High Frequency Financial Econometrics*, Bauwens L, Pohlmeier W, Veredas D (eds). Physica-Verlag HD: Berlin; 293–312.
Vuong QH. 1989. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* **57**: 307–333.
West KD. 1996. Asymptotic inference about predictive ability. *Econometrica* **64**: 1067–1084.

DOI: 10.1002/jae

APPENDIX A: SECOND MOMENTS' STRUCTURE

Since the model is expressed for p_t (i.e. for the squares and cross-products of daily returns), we are able to obtain explicit expressions for the fourth moment of returns by deriving $\text{var}[p_t]$. Similarly, by deriving $\text{var}[v_t]$, we are able to analyse the second moment of the realized measure, which gives an expression for the volatility of volatility; see Engle (2002) and Corsi *et al*. (2008) for a discussion of modelling the volatility of volatility using the VIX and realized volatility, respectively.

The following proposition gives the structure of the second moments of p_t and v_t, which is derived under the assumption $\text{E}_{t-1}[\varepsilon_t] = \text{E}_{t-1}[\eta_t] = I_k$. The expressions in (A.1)–(A.2) can be simplified further by assuming a Wishart distribution for the innovations, which gives (A.3)–(A.4).

Proposition 4.

(i) Under the assumption that $\text{E}_{t-1}[\varepsilon_t] = \text{E}_{t-1}[\eta_t] = I_k$, the second moments of p_t and v_t are given by

$$\text{var}[p_t] = \text{E}[Z_{H,t}\text{var}_{t-1}[\text{vech}(\varepsilon_t)]Z'_{H,t}] + \text{var}[Z_{H,t}\text{vech}(I_k)] \tag{A.1}$$

$$\text{var}[v_t] = \text{E}[Z_{M,t}\text{var}_{t-1}[\text{vech}(\eta_t)]Z'_{M,t}] + \text{var}[Z_{M,t}\text{vech}(I_k)] \tag{A.2}$$

where $Z_{H,t} = L_k(H_t^{\frac{1}{2}} \otimes H_t^{\frac{1}{2}})D_k$, $Z_{M,t} = L_k(M_t^{\frac{1}{2}} \otimes M_t^{\frac{1}{2}})D_k$ and $\text{var}_{t-1}[\cdot]$ denotes the variance conditional on $\mathcal{F}_{t-1}^{\text{HF}}$.

(ii) Under the additional assumption that ε_t and η_t are i.i.d. Wishart distributed, the second moments of p_t and v_t are given by

$$\text{E}[p_t p_t'] = 2D_k^{+}\text{E}[(H_t \otimes H_t)]L_k' + \text{E}[h_t h_t'] \tag{A.3}$$

$$\text{E}[v_t v_t'] = 2n^{-1}D_k^{+}\text{E}[(M_t \otimes M_t)]L_k' + \text{E}[m_t m_t'] \tag{A.4}$$

where $D_k^{+} = (D_k' D_k)^{-1}D_k'$ is the Moore–Penrose inverse of D_k.

Dropping the t subscripts to avoid cluttered notation, the second moment structure of p_t given in (A.3) will have the following structure in the two-dimensional case:

$$\text{E}[pp'] = \text{E}\begin{pmatrix} r_1^4 & r_1^3 r_2 & r_1^2 r_2^2 \\ r_1^3 r_2 & r_1^2 r_2^2 & r_1 r_2^3 \\ r_1^2 r_2^2 & r_1 r_2^3 & r_2^4 \end{pmatrix} = \text{E}\begin{pmatrix} 3h_{11}^2 & 3h_{11}h_{12} & 2h_{12}^2 + h_{11}h_{22} \\ 3h_{11}h_{21} & 2h_{21}^2 + h_{11}h_{22} & 3h_{12}h_{22} \\ 2h_{21}^2 + h_{11}h_{22} & 3h_{21}h_{22} & 3h_{22}^2 \end{pmatrix}$$

where r_1 and r_2 denote the daily returns for assets 1 and 2, respectively, and h_{ij}, $i, j = 1, 2$, are the elements of H_t. Applying a vec operator to (A.3) gives a similar result to (10) in Hafner (2003), which discusses the fourth moment structure of GARCH models when H_t follows a GARCH specification and daily returns are assumed to be normally distributed.

The result in (A.4) seems novel in the context of realized measures. In the univariate case, Corsi *et al*. (2008) estimate the volatility of realized volatility by utilizing consistent estimators of the integrated quarticity of returns, such as realized quarticity, realized quad-power quarticity and realized tri-power quarticity. In an application to S&P 500 index futures, they show that the unconditional distributions of these three measures are skewed and leptokurtic even after applying a log transformation. The three measures also exhibit clustering, which prompts the authors to develop a GARCH-type model for realized volatility. Engle (2002) also discusses different models for volatility of volatility using the VIX time series.

DOI: 10.1002/jae

APPENDIX B: TECHNICAL PROOFS

Proof of Proposition 1

By taking unconditional expectation of (7) and (8) we have

$$\omega_H = C_H + B_H\omega_H + A_H\omega_M, \quad \omega_M = C_M + B_M\omega_M + A_M\omega_M \tag{B.1}$$

By definition, $\overline{\kappa} = \Omega_M^{\frac{1}{2}}\Omega_H^{-\frac{1}{2}}$, which implies $\Omega_M^{\frac{1}{2}} = \overline{\kappa}\Omega_H^{\frac{1}{2}}$, and $\Omega_M = \overline{\kappa}\Omega_H\overline{\kappa}'$. Thus $\omega_M :=$ $\mathrm{vech}(\Omega_M) = \mathrm{vech}(\overline{\kappa}\Omega_H\overline{\kappa}') = L_k(\overline{\kappa}\otimes\overline{\kappa})D_k\omega_H$ using (A.1) in Web Appendix A for the last equality.

Let $\kappa = L_k(\overline{\kappa}\otimes\overline{\kappa})D_k$ and substitute the last result for ω_M in the first expression in (B.1). By collecting terms we have that the intercept coefficients are given by $C_H = (I_{k^*} - B_H - A_H\kappa)\omega_H$ and $C_M = (I_{k^*} - B_M - A_M)\omega_M$, which when substituted in (7) and (8) gives the stated result.

The proof for (13) follows by noting that $\mathrm{E}[\widetilde{v}_t] = \kappa^{-1}\mathrm{E}[v_t] = \kappa^{-1}\omega_M = \omega_H$, where the last equality follows from above by defining $\kappa = L_k(\overline{\kappa}\otimes\overline{\kappa})D_k$. The rest follows by collecting terms and substituting for C_H in the first expression in (B.1).

Proof of Proposition 2

We start with the proof of (14). The one-step forecast of h_t is $\mathrm{E}_t[h_{t+1}] = h_{t+1}$, since h_{t+1} is $\mathcal{F}_t^{\mathrm{HF}}$-measurable. From (7), the two-step forecast is

$$\mathrm{E}_t[h_{t+2}] = \mathrm{E}_t[C_H + B_H h_{t+1} + A_H v_{t+1}] = C_H + B_H h_{t+1} + A_H\mathrm{E}_t[v_{t+1}]$$

The three-step forecast is

$$\begin{aligned}\mathrm{E}_t[h_{t+3}] &= \mathrm{E}_t[C_H + B_H h_{t+2} + A_H v_{t+2}] = C_H + B_H\mathrm{E}_t[h_{t+2}] + A_H\mathrm{E}_t[v_{t+2}]\\ &= (I_{k^*} + B_H)C_H + B_H^2 h_{t+1} + A_H\mathrm{E}_t[v_{t+2}] + B_H A_H\mathrm{E}_t[v_{t+1}]\end{aligned}$$

where the last equality follows by substituting for $\mathrm{E}_t[h_{t+2}]$ from above and collecting terms. By forward iteration, it is straightforward to show that

$$\mathrm{E}_t[h_{t+s}] = \sum_{i=1}^{s-1} B_H^{i-1}C_H + B_H^{s-1}h_{t+1} + \sum_{i=1}^{s-1} B_H^{i-1}A_H\mathrm{E}_t[v_{t+s-i}] \tag{B.2}$$

Now find an expression for $\mathrm{E}_t[v_{t+s-i}]$ in terms of m_{t+1}, which is $\mathcal{F}_t^{\mathrm{HF}}$-measurable. We start with the one-step forecast of v_t, which is $\mathrm{E}_t[v_{t+1}] = m_{t+1}$ by definition. The two-step forecast is

$$\begin{aligned}\mathrm{E}_t[v_{t+2}] &= \mathrm{E}_t[\mathrm{E}_{t+1}[v_{t+2}]] = \mathrm{E}_t[m_{t+2}] = \mathrm{E}_t[C_M + B_M m_{t+1} + A_M v_{t+1}]\\ &= C_M + (B_M + A_M)m_{t+1}\end{aligned}$$

since $\mathrm{E}_t[v_{t+1}] = m_{t+1}$. The three-step forecast is

$$\begin{aligned}\mathrm{E}_t[v_{t+3}] &= \mathrm{E}_t[\mathrm{E}_{t+1}[v_{t+3}]] = \mathrm{E}_t[C_M + (B_M + A_M)m_{t+2}] = C_M + (B_M + A_M)\mathrm{E}_t[m_{t+2}]\\ &= (I_{k^*} + (B_M + A_M))C_M + (B_M + A_M)^2 m_{t+1}\end{aligned}$$

where the second equality follows by substitution from above with a one-period forward iteration, and the last equality follows by substituting for m_{t+2}, applying the conditional expectation operator

DOI: 10.1002/jae

and collecting terms. By forward iteration, we have the following formula for the s-step forecast of v_t:

$$\mathrm{E}_t[v_{t+s}] = \sum_{j=1}^{s-1}(B_M + A_M)^{j-1}C_M + (B_M + A_M)^{s-1}m_{t+1} \tag{B.3}$$

Using (B.3) to substitute for $\mathrm{E}_t[v_{t+s-i}]$ in (B.2), while adapting the summation limit by replacing s in (B.3) with $(s-i)$ gives the stated result.

The proof of (15) follows similar steps. We start by taking unconditional expectations of (7) and (8), which gives

$$\omega_H = C_H + B_H\omega_H + A_H\omega_M, \quad \omega_M = C_M + B_M\omega_M + A_M\omega_M$$

so that the constant terms can be expressed as $C_H = \omega_H - B_H\omega_H - A_H\omega_M$ and $C_M = \omega_M - B_M\omega_M - A_M\omega_M$. Substituting these expressions in (7) and (8) gives

$$\begin{aligned} h_t &= \omega_H - B_H\omega_H - A_H\omega_M + B_H h_{t-1} + A_H v_{t-1} \\ &= \omega_H + B_H(h_{t-1} - \omega_H) + A_H(v_{t-1} - \omega_M), \\ m_t &= \omega_M - B_M\omega_M - A_M\omega_M + B_M m_{t-1} + A_M v_{t-1} \\ &= \omega_M + B_M(m_{t-1} - \omega_M) + A_M(v_{t-1} - \omega_M) \end{aligned}$$

Forward iteration of these equations as illustrated in the proof of (7) yields (8).

Proof of Proposition 3

We derive the score vector and prove that it is a martingale difference sequence only for the HEAVY-P equation. The derivation for the HEAVY-V equation is analogous. We derive the $(1 \times \delta_H)$ score vector $\dfrac{\partial l_{H,t}(\theta_H)}{\partial \theta_H'}$ from the log-likelihood equation which gives

$$\begin{aligned} \frac{\partial l_{H,t}(\theta_H)}{\partial \theta_H'} &= -\frac{1}{2}\frac{\partial \log|H_t|}{\partial \theta_H'} - \frac{1}{2}\frac{\partial tr(H_t^{-1}P_t)}{\partial \theta_H'} \\ &= -\frac{1}{2}\frac{\partial \log|H_t|}{\partial H_t}\frac{\partial H_t}{\partial \theta_H'} - \frac{1}{2}\frac{\partial tr(H_t^{-1}P_t)}{\partial H_t^{-1}}\frac{\partial H_t^{-1}}{\partial H_t}\frac{\partial H_t}{\partial \theta_H'} \\ &= -\frac{1}{2}(\mathrm{vec}(H_t^{-1}))'\frac{\partial \mathrm{vec}(H_t)}{\partial \theta_H'} + \frac{1}{2}(\mathrm{vec}(P_t))'(H_t^{-1} \otimes H_t^{-1})\frac{\partial \mathrm{vec}(H_t)}{\partial \theta_H'} \\ &= \frac{1}{2}[(\mathrm{vec}(P_t))'(H_t^{-1} \otimes H_t^{-1}) - (\mathrm{vec}(H_t^{-1}))']\frac{\partial \mathrm{vec}(H_t)}{\partial \theta_H'} \\ &= \frac{1}{2}[(\mathrm{vec}(P_t))'(H_t^{-1} \otimes H_t^{-1}) - (\mathrm{vec}(H_t^{-1}H_tH_t^{-1}))']\frac{\partial \mathrm{vec}(H_t)}{\partial \theta_H'} \\ &= \frac{1}{2}[(\mathrm{vec}(P_t))'(H_t^{-1} \otimes H_t^{-1}) - ((H_t^{-1} \otimes H_t^{-1})\mathrm{vec}(H_t))']\frac{\partial \mathrm{vec}(H_t)}{\partial \theta_H'} \\ &= \frac{1}{2}[(\mathrm{vec}(P_t))' - (\mathrm{vec}(H_t))'](H_t^{-1} \otimes H_t^{-1})\frac{\partial \mathrm{vec}(H_t)}{\partial \theta_H'} \end{aligned}$$

DOI: 10.1002/jae

where in the second equality we used the chain rule and the matrix derivatives in the third equality are obtained using the rules stated in Web Appendix A.

The score vector is a martingale difference sequence such that $\mathrm{E}_{t-1}\left[\frac{\partial l_{H,t}(\theta_H)}{\partial \theta_H'}\right] = 0$ as the conditional expectation of the term in square brackets is 0 since

$$\mathrm{E}_{t-1}[(\mathrm{vec}(P_t))' - (\mathrm{vec}(H_t))'] = \mathrm{E}_{t-1}[(\mathrm{vec}(P_t))'] - \mathrm{E}_{t-1}[(\mathrm{vec}(H_t))'] = 0$$

where we use $\mathrm{E}_{t-1}[(\mathrm{vec}(P_t))'] = (\mathrm{vec}(H_t))'$, which follows directly from the conditional moment assumption $\mathrm{E}_{t-1}[\varepsilon_t] = I_k$.

Proof of Proposition 4

For the first part of the proposition, we only show the proof for (A.1) as (A.2) follows similar arguments. We start from $p_t := \mathrm{vech}(P_t) = \mathrm{vech}(H_t^{\frac{1}{2}}\varepsilon_t H_t^{\frac{1}{2}}) = L_k(H_t^{\frac{1}{2}} \otimes H_t^{\frac{1}{2}})D_k \mathrm{vech}(\varepsilon_t)$, where the last result follows from (A.1) in Web Appendix A. Let $Z_{H,t} = L_k(H_t^{\frac{1}{2}} \otimes H_t^{\frac{1}{2}})D_k$, which is $\mathcal{F}_{t-1}^{\mathrm{HF}}$-measurable. Also, let $\mathrm{var}_{t-1}[\cdot]$ denote the variance conditional on $\mathcal{F}_{t-1}^{\mathrm{HF}}$. Using the variance decomposition we obtain

$$\begin{aligned}
\mathrm{var}[p_t] &= \mathrm{E}[\mathrm{var}_{t-1}[p_t]] + \mathrm{var}[\mathrm{E}_{t-1}[p_t]] \\
&= \mathrm{E}[\mathrm{var}_{t-1}[Z_{H,t}\mathrm{vech}(\varepsilon_t)]] + \mathrm{var}[\mathrm{E}_{t-1}[Z_{H,t}\mathrm{vech}(\varepsilon_t)]] \\
&= \mathrm{E}[Z_{H,t}\mathrm{var}_{t-1}[\mathrm{vech}(\varepsilon_t)]Z_{H,t}'] + \mathrm{var}[Z_{H,t}\mathrm{E}_{t-1}[\mathrm{vech}(\varepsilon_t)]]
\end{aligned}$$

as $Z_{H,t}$ is $\mathcal{F}_{t-1}^{\mathrm{HF}}$-measurable. As $\mathrm{E}_{t-1}[\varepsilon_t] = I_k$ by assumption, it follows that $\mathrm{E}_{t-1}[\mathrm{vech}(\varepsilon_t)] = \mathrm{vech}(I_k)$, which gives (A.1).

For (A.3), $\varepsilon_t \overset{\text{i.i.d.}}{\sim} SINGW_k(1, I_k)$ implies $P_t|\mathcal{F}_{t-1}^{\mathrm{HF}} \sim SINGW_k(1, H_t)$ and also implies $R_t|\mathcal{F}_{t-1}^{\mathrm{HF}} \sim N(0, H_t)$ since $P_t = R_t R_t'$. Thus $\mathrm{var}_{t-1}[\mathrm{vec}(P_t)] = \mathrm{var}_{t-1}[(R_t \otimes R_t)] = 2D_k D_k^{+}(H_t \otimes H_t)$, where the second equality follows from the conditional normality of R_t by Magnus (1988, Theorem 10.2) noting that conditioning on $\mathcal{F}_{t-1}^{\mathrm{HF}}$ enables us to treat H_t as a nonstochastic matrix. Therefore

$$\begin{aligned}
\mathrm{var}_{t-1}[p_t] &= \mathrm{var}_{t-1}[L_k \mathrm{vec}(P_t)] = L_k \mathrm{var}_{t-1}[\mathrm{vec}(P_t)]L_k' \\
&= 2L_k D_k D_k^{+}(H_t \otimes H_t)L_k' = 2D_k^{+}(H_t \otimes H_t)L_k'
\end{aligned}$$

where the last equality follows since $L_k D_k = I_{k^*}$ by Magnus (1988, Theorem 5.5). We obtain the unconditional second moment of p_t using the variance decomposition

$$\begin{aligned}
\mathrm{var}[p_t] &= \mathrm{E}[\mathrm{var}_{t-1}[p_t]] + \mathrm{var}[\mathrm{E}_{t-1}[p_t]] = \mathrm{E}[2D_k^{+}(H_t \otimes H_t)L_k'] + \mathrm{var}[h_t] \\
&= 2D_k^{+}\mathrm{E}[(H_t \otimes H_t)]L_k' + \mathrm{var}[h_t]
\end{aligned}$$

We can write $\mathrm{var}[p_t] = \mathrm{E}[p_t p_t'] - \mathrm{E}[p_t]\mathrm{E}[p_t]'$ and $\mathrm{Var}[h_t] = \mathrm{E}[h_t h_t'] - \mathrm{E}[h_t]\mathrm{E}[h_t]'$. By noting that $\mathrm{E}[p_t] = \mathrm{E}[\mathrm{E}_{t-1}[p_t]] = \mathrm{E}[h_t]$, the last equation for $\mathrm{var}[p_t]$ can be simplified to give the stated result. The proof of (A.4) is similar except in the intermediate step of deriving $\mathrm{var}_{t-1}[\mathrm{vec}(V_t)]$, where in this case Theorem 10.3 of Magnus (1988) directly applies since η_t has a non-singular Wishart distribution. Thus we have $\mathrm{var}_{t-1}[\mathrm{vec}(V_t)] = 2n^{-1}D_k D_k^{+}(M_t \otimes M_t)$ and the rest of the proof follows as in (A.3).

DOI: 10.1002/jae