THE CONTRIBUTION OF REALIZED VARIANCE–COVARIANCE MODELS TO THE ECONOMIC VALUE OF VOLATILITY TIMING

Luc Bauwens, Yongdeng Xu

REPRINT | 3348

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

The contribution of realized variance-covariance models to the economic value of volatility timing

Luc Bauwens * Yongdeng Xu†

November 12, 2024

Abstract

Realized variance-covariance models define the conditional expectation of a realized variance-covariance matrix as a function of past matrices, using a GARCH-type structure. We evaluate the forecasting performance of various such models in terms of economic value, measured through economic loss functions, across two datasets. Our empirical findings reveal that models that incorporate realized volatilities, offer significant economic value and outperform GARCH models that rely solely on daily returns, for both daily and weekly horizons. Among two realized variance-covariance measures, using a directly rescaled intraday measure for full-day estimation provides more economic value than employing overnight returns, which tends to produce noisier estimators of overnight covariance, thereby diminishing their predictive effectiveness. The HEAVY-H model for the variance-covariance matrix of the daily return demonstrates superior or comparable performance to the best-performing realized variance-covariance models, making it a preferred choice for empirical analysis.

Keywords: volatility timing, realized volatility, high-frequency data, HEAVY model, forecasting.

JEL Classification: G11, G17, C32, C58.

*LIDAM/CORE, UCLouvain,Voie du Roman Pays, 34 B-1348 Louvain-La-Neuve, Belgium. Email: luc.bauwens@uclouvain.be

†Correspondence Author: Cardiff Business School, Cardiff University, CF10 3EU, UK. Email: xuy16@cf.ac.uk.

1 Introduction

Forecasting the multivariate volatility (in the sense of the variance-covariance matrix) of daily returns is important for risk management. The current consensus, based on a large variety of empirical studies, is that, to this end, the use of a realized measure of the daily volatility, based on intra-daily returns, is more beneficial than the exclusive use of daily data. The intuitive explanation of this difference is that, under suitable but realistic conditions, a realized variance-covariance matrix is a more precise measure of the daily volatility than a measure constructed using only daily returns.

The benefits gained thanks to the use of a realized variance-covariance matrix are evaluated through measures of forecasting performance, which can be of statistical or economic nature. In an influential paper, Fleming et al. (2003) provide a method to measure "the economic value of volatility timing using realized volatility" and show, empirically, that the economic value of switching from daily to intra-daily returns can be substantial. A large number of authors have adopted the method of Fleming et al. (2003), using different models and data.[1]

Whether the basic data is daily or intra-daily, volatility forecasts are based on a time series econometric model. Such a model specifies the conditional variance-covariance (i.e., the conditional expectation of the multivariate volatility) matrix of the daily return vector as a function of past data. When the data is daily, the conditional variance-covariance matrix of day t is typically specified as a moving average of the outer-products of the past returns; this can be achieved through a multivariate GARCH model. When the data is intra-daily, the conditional variance-covariance matrix of day t can be specified as a moving average of the past realized variances and covariances. This is what Fleming et. al. (2003) did explicitly, with a parsimonious way to define and estimate the coefficients of the moving average terms as functions of a scalar parameter.

A critical aspect of realized variance-covariance matrix modeling is its typical reliance on "intraday" returns, i.e. returns from the trading period (typically, market open to close times), to compute realized variance-covariance matrices. By omitting possible return variations during the non-trading (say, overnight) period, one may bias the estimation of the variances and

[1] A Google search, conducted on June 25, 2023, using the title of the paper of Fleming et al. (2003), identified 778 citations;, among which e.g., Marquering and Verbeek (2004) and Bollerslev et al. (2018a). Fleming et al. (2003) extends Fleming et al. (2001) where the authors show the economic value of switching from unconditionally efficient static portfolios to portfolios based upon volatility models based on daily returns.

covariances of the entire day returns. Two methods exist to correct this potential bias. In the first method, one adds the squared overnight returns to the trading period realized variances, and the cross products of overnight returns to the trading period realized covariances; in a univariate context, Hansen and Lunde (2005) find that adding overnight squared returns is effective. The second method involves rescaling the intraday realized variance-covariance matrix to a daily scale while ensuring the variance-covariance matrix positivity. Notably, the rescaling technique by Fleming et al. (2003) does not meet this criterion, but an alternative proposed by Sheppard and Xu (2019) ensures the positivity of the variance-covariance matrix.

Recent developments have seen sophisticated extensions to the multivariate GARCH model for forecasting daily variance-covariance matrices of returns. A notable advancement involves integrating a realized measure of volatility in place of the outer product of past return vectors in the conditional variance equation of such models, prominently featured in the high-frequency-based volatility (HEAVY) class of models.[2] This progress includes the HEAVY-BEKK model by Noureldin et al. (2012), the HEAVY β-Factor-GARCH model by Sheppard and Xu (2019), and the HEAVY-DCC (dynamic conditional correlation) model by Bauwens and Xu (2023).[3] In these frameworks, the conditional variance-covariance matrix of daily returns is defined in the HEAVY model as a function of lagged realized covariances, in contrast to the lagged outer products of daily returns used in MGARCH models. A key advantage of the HEAVY model is that it directly models the variance-covariance matrix of daily returns, circumventing the need to adjust the intraday realized variance-covariance matrix to a daily scale.

Empirical findings consistently show that HEAVY models outperform traditional multivariate GARCH models in forecasting the conditional variance-covariance matrix of daily returns. Statistical performance measures used in these evaluations vary, including loss functions and statistical tests. In Noureldin et al. (2012) and Sheppard and Xu (2019), the forecasts of different models are compared by equal predictive accuracy (EPA) Diebold-Mariano tests, enabling only pairwise comparisons, while Bauwens and Xu (2023) use the model confidence set (MCS) approach of Hansen et al. (2011) which enables to compare jointly a set of models. Forecast performance comparisons can also be based on economic loss functions. Sheppard and Xu (2019)

[2]This includes the independently developed univariate HEAVY model by Shephard and Sheppard (2010), and the realized GARCH model by Hansen et al. (2012).

[3]While the multivariate realized β-GARCH model by Hansen et al. (2014) exists, there are no realized-BEKK and realized-DCC versions that extend the univariate realized GARCH model by Hansen et al. (2012).

compare the relative performance of pairs of models (by EPA tests) in terms of marginal expected shortfall, which is the expected loss of an asset conditional on a factor (such as a market index) being in a state of stress, see Acharya et al. (2017). Bauwens and Xu (2023) use the MCS testing procedure to rank jointly a set of models in terms of optimal (global minimum variance, and minimum variance) portfolio performances. Notably, none of these studies uses the criterion of the economic value of volatility timing of Fleming et al. (2003). Thus, there is plenty of scope for assessing the economic value of HEAVY models. Furthermore, most of these model comparisons involve HEAVY and GARCH models, with direct comparisons between HEAVY models and realized variance-covariance models being surprisingly scarce.

Therefore, our empirical investigation focuses on two primary objectives: 1) To compare the forecasting performance of realized variance-covariance models employing two methods to correct the possible bias of the estimate of the entire day variance-covariance matrix based on the intraday returns; and 2) To compare the forecasting performance of GARCH, HEAVY, and realized variance-covariance models, with a particular emphasis on their economic value in volatility timing.

We implement the comparisons using two datasets. The first one is relative to futures contracts for S&P 500, Treasury bonds and gold, for the period 2003/07/01-2022/08/05; this dataset is an update of the data used by Fleming et al. (2003). The trading of these contracts occurs almost without interruption during each trading calendar day, so that it is possible to measure the daily volatility of a complete day through a realized variance-covariance measure (of dimension 3x3) computed using the intra-daily observed returns.

The second dataset is relative to twenty-nine stocks belonging to the DJIA (Dow Jones Industrial Average) index, for the period 2001/01/03-2018/04/16. On this type of market, the trading period corresponds to a few hours of a day. Since a realized variance-covariance matrix uses the intra-daily returns available during the trading period, it underestimates the complete day variance-covariance matrix. A bias correction method is required. We apply two correction methods: the first one adds the overnight variation (i.e., the outer product of the overnight return vector) to the trading period realized matrix, and then transforms the sum to match the sample average of the outer products of the daily (close-to-close) returns; the second one transforms directly the trading period realized variance-covariance matrix so that it has the same sample average as the average of the outer products of the daily (close-to-close) returns;

Our empirical results primarily confirm that HEAVY-type models, which utilize realized volatilities, provide economic value and significantly outperform GARCH models that rely solely on daily returns. This holds true particularly for daily and weekly forecasts. When considering two realized variance-covariance models using the DJ29 dataset, we observe that the models implementing the bias correction approach incorporating overnight returns to construct a full-day variance-covariance matrix are less effective than those employing the alternative approach where the intraday return variance-covariance matrix is directly rescaled for the full day. The main reason for this disparity lies in the noisy estimates of the overnight variances and covariances provided by overnight returns. In terms of economic value, the HEAVY-H model and the realized variance-covariance model that uses the second bias correction method rank as top performers in their categories, with almost identical results, neither consistently outperforming the other. This trend holds true for both three-asset and twenty-nine stock datasets. Within HEAVY-type models, those designed to distinguish individual conditional variance processes from correlation processes, like DCC-type models, are more effective than BEKK-type models, which apply the same dynamics to variances and covariances. This advantage is more pronounced in the twenty-nine stock dataset, whereas performance is similar across models in the three-asset dataset. Overall, the HEAVY-H model, which effectively captures daily return covariances and relies on realized variances and covariances, performs strongly. It is not affected by intraday bias correction issues, making it a good choice for analysis. The DCC-HEAVY model, developed by Bauwens and Xu (2023), is especially effective for large asset portfolios.

The remainder of the paper is divided in five sections. Section 2 presents the data, the definition of the realized variance-covariance measure and the bias correction methods used for the second dataset. Section 3 defines the HEAVY and realized variance-covariance models used in the empirical applications and discusses the estimation results. Section 4 expounds the economic criteria used to compare the models through out-of-sample forecasts and the related statistical methods. Section 5 contains the empirical results. Section 6 concludes. Additional technical and empirical details are provided in a set of appendices.

2 Realized covariance: definitions, bias correction, and data

2.1 Definitions and bias correction

Let $r_t = (r_{1t}, r_{2t} \dots r_{kt})'$ denote the $k \times 1$ (close-to-close) 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 j-th intra-daily return vector at time j on day t, where $j = 1, 2, ..., N$. The simplest realized variance-covariance measure for the k assets on day t is the $k \times k$ matrix defined as

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

Other types of realized variance-covariance matrix estimators which are somewhat robust to noise have been considered, for example, the realised kernel of Barndorff-Nielsen et al. (2011).

Assuming that $N > k$, RC_t is positive definite (PD). 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.

The realized variance-covariance measure RC_t defined above uses the intra-daily returns available during the trading period, i.e. between the opening and closing times. The overnight variation of returns is not taken into account. Hence RC_t underestimates the whole day covariance. A bias correction method is required if the non-trading period corresponds to a large part of the day. If that is the case, we adopt two correction methods to transform RC_t into a daily covariance. Both correction methods are designed to satisfy the condition that the average value (over the sample period of T observations) of the realized variance-covariance matrices is the same as the average value of the outer products of the daily returns.

- <u>Correction 1:</u> Adding to RC_t the outer product $onr_t\, onr'_t$ of the overnight return onr_t (the close-to-open return preceding the trading period) to estimate the overnight variance-covariance matrix provides a possible estimate of the whole day matrix:

$$RCon_t = RC_t + onr_t\, onr_t'. \tag{3}$$

If the addition of the $onr_t\, onr_t'$ to RC_t is not sufficient to ensure that the average of the $RCon_t$ matrices is sufficiently close to the average of the outer products of the daily returns, a rescaling correction can be used to ensure that the condition that the average value of the rescaled variance-covariance matrices is exactly the same as the average value of the outer products of the daily returns. The rescaled realized variance-covariance matrix is defined by

$$RC1_t = \Lambda_1 RCon_t \Lambda_1', \tag{4}$$

where $\Lambda_1 = \bar{\Sigma}^{1/2}\bar{M_1}^{-1/2}$, with $\bar{\Sigma} = T^{-1}\sum_{t=1}^{T} r_t r_t'$ and $\bar{M}_1 = T^{-1}\sum_{t=1}^{T} RCon_t$.

- Correction 2: As an alternative to the previous correction that uses the overnight returns, RC_t can be directly rescaled to

$$RC2_t = \Lambda RC_t \Lambda', \tag{5}$$

where $\Lambda = \bar{\Sigma}^{1/2}\bar{M}^{-1/2}$, $\bar{\Sigma}$ defined as above and $\bar{M} = T^{-1}\sum_{t=1}^{T} RC_t$.

It is straightforward to check that both $RC1_t$ and $RC2_t$ satisfy the condition that their average value (over the sample period of T observations) is exactly equal to the average value of the outer products of the daily returns. This type of rescaling formula is used by Sheppard and Xu (2019) and it guarantees a PD variance-covariance matrix.

2.2 Data

We use two datasets for the empirical analyses. The first is a dataset about the same futures contracts as studied by Fleming et al. (2001): S&P 500 futures (STOCK, hereafter), Treasury bond futures (BOND), and Gold futures (GOLD); it is named SBGF (for Stock-Bond-Gold Futures) hereafter. The sample period is July 1, 2003 to August 5, 2022, with a total of 4864 trading days. We choose July 1, 2003 as starting date, as trading occurs both in the daytime and in the evening (e.g., from 7:20 to 16:00 and from 17:00 to 7:20 for BOND) from that day.

This was not the case in the dataset of Fleming et al. (2001), since the trading occurred only during the daytime. Appendix A provides the exact trading times of the three futures contracts and explains the computation of the realized covariances from the high-frequency data. Since the realized variance-covariance matrices of this dataset are computed using five-minute spaced intra-daily returns based on (approximately) 23 hours per day, they should be close to the whole day covariances. In Table 9 (in Appendix A), it can be checked that the average realized variances are close to the average daily squared returns, so that no bias correction method is applied.

The second dataset consists of realized variance-covariance matrices for twenty-nine stocks belonging to the Dow Jones Industrial Average (DJIA) index; it is named DJ29 hereafter. The sample period is January 3, 2001 to April 16, 2018, with a total of 4092 trading days. The same dataset is used by Bauwens and Xu (2023) and Bauwens and Otranto (2022), where more information is provided about the data source and construction from high-frequency data. The realized variance-covariance matrices are computed using synchronized five-minute returns in the trading session that extends from 9:30 Eastern Standard Time (EST) to 16:00 EST. Overnight returns are computed accordingly as the differences between the opening trade log-price at 9:30 EST and the previous close log-price at 16:00 EST.

Descriptive statistics for the original and bias-corrected variance-covariance matrices are reported in Appendix B. The realized variances (v_t, the diagonal of RC_t) do not account for the overnight variation, hence their time series means, reported in column 3 of Table 10 under the header v), are at about 50 to 60 percent of the corresponding average squared returns (column 2). The time series means (von, in column 4 of the same table) of the realized variances augmented by the squared overnight returns, as defined by (3) but not rescaled by Λ_1 as in (4), are much closer to the average squared returns (r^2). The corrected realized variances ($v1_t$ and $v2_t$, diagonal of $RC1_t$ and $RC2_t$) are on average equal to the average squared returns by construction (column 2 of the table).

The time series standard deviations of the squared returns (r^2), realized variances (v), and bias-corrected realized variances ($v1$ and $v2$) satisfy almost always the following inequalities: $r^2 > v1 > v2 > v$ – only the inequality $v1 > v2$ is (marginally) reversed in six cases. The time series of the daily squared returns are generally featuring more extreme peaks than the realized variances, explaining the larger time series means and standard deviations of the squared returns. The smaller time series means and standard deviations of the trading period are obviously due to the fact that the trading period is a fraction of the day, but the addition of the overnight return to the realized variance of that period and the rescaling by Λ_1 inflates the standard deviations more than directly rescaling by Λ_2, because the overnight variations has also some specific extremes.

Concerning the realized covariances and correlations, the bias corrections increase generally the means and standard deviations (see Tables 11 and 12) with respect to the uncorrected realized measures, but the differences between the average realized correlations ($Rcor$), rescaled overnight return augmented realized correlations ($RCor1$) and directly scaled realized correlations ($RCor2$) are much smaller than for the corresponding average realized variances and covariances.

3 Realized variance-covariance models and estimation results

3.1 Definitions of models

Table 1 – BEKK, FKO and DCC model equations

BEKK-	Equation	Restrictions
GARCH	$G_t = (1-\alpha_G-\beta_G)\bar{H} + \alpha_G r_{t-1}r'_{t-1} + \beta_G G_{t-1}$	$\alpha_G, \beta_G \geq 0, \alpha_G+\beta_G < 1$
HEAVY-H	$H_t = (1-\beta_H)\bar{H} - \alpha_H\bar{M} + \alpha_H RC_{t-1} + \beta_H H_{t-1}$	$\alpha_H, \beta_H \geq 0, \beta_H < 1$
HEAVY-M	$M_t = (1-\alpha_M-\beta_M)\bar{M} + \alpha_M RC_{t-1} + \beta_M M_{t-1}$	$\alpha_M, \beta_M \geq 0, \alpha_M+\beta_M < 1$
FKO-	Equation	Restrictions
GARCH	$G_t = \exp(-\alpha_G)G_{t-1} + \alpha_G \exp(-\alpha_G) r_{t-1}r'_{t-1}$	$\alpha_G \geq 0$
HEAVY-H	$H_t = \exp(-\alpha_H)H_{t-1} + \alpha_H \exp(-\alpha_H) RC_{t-1}$	$\alpha_H \geq 0$
HEAVY-M	$M_t = \exp(-\alpha_M)M_{t-1} + \exp(-\alpha_M) RC_{t-1}$	$\alpha_M \geq 0$
DCC-	Variance equation of asset i	Restrictions
GARCH	$g_{t,i} = \omega_{G,i} + A_{G,i} r^2_{t-1,i} + B_{G,i} g_{t-1,i}$	$A_{G,i}, B_{G,i} \geq 0, A_{G,i}+B_{G,i} < 1$
HEAVY-H	$h_{t,i} = \omega_{H,i} + A_{H,i} v_{t-1,i} + B_{H,i} h_{t-1,i}$	$A_{H,i}, B_{H,i} \geq 0, A_{H,i}+B_{H,i} < 1$
HEAVY-M	$m_{t,i} = \omega_{M,i} + A_{M,i} v_{t-1,i} + B_{M,i} m_{t-1,i}$	$A_{M,i}, B_{M,i} \geq 0, A_{M,i}+B_{M,i} < 1$
	Correlation equation	
GARCH	$Q_t = (1-\alpha_Q-\beta_Q)\bar{R} + \alpha_Q u_{t-1}u'_{t-1} + \beta_Q Q_{t-1}$	$\alpha_Q, \beta_Q \geq 0, \beta_Q < 1$
HEAVY-H	$R_t = (1-\beta_R)\bar{R} - \alpha_R\bar{P} + \alpha_R RL_{t-1} + \beta_R R_{t-1}$	$\alpha_R, \beta_R \geq 0, \beta_R < 1$
HEAVY-M	$P_t = (1-\alpha_P-\beta_P)\bar{P} + \alpha_P RL_{t-1} + \beta_P P_{t-1}$	$\alpha_P, \beta_P \geq 0, \alpha_P+\beta_P < 1$

r_t: daily return vector; RC_t: realized variance-covariance matrix of day t, see (1); $\mathcal{F}_t^{LF} = \{r_s \text{ for } s < t\}$ (past daily returns); $\mathcal{F}_t^{HF} = \{RC_s, r_s \text{ for } s < t\}$ (past realized variance-covariance matrices and daily returns).
-For BEKK and FKO equations: $G_t = E(r_t r'_t|\mathcal{F}_{t-1}^{LF})$, $H_t = E(r_t r'_t|\mathcal{F}_{t-1}^{HF})$, $M_t = E(RC_t|\mathcal{F}_{t-1}^{HF})$, $\bar{H} = E(r_t r'_t) \approx \sum_{t=1}^T r_t r'_t/T$, $\bar{M} = E(RC_t) \approx \sum_{t=1}^T RC_t/T$.
-For DCC-variance equations: $r_{t,i}$: daily return of asset i (i-th element of r_t); $v_{t,i}$: realized variance of asset i (i-th diagonal element of RC_t); $g_{t,i} = E(r^2_{t,i}|\mathcal{F}_{t-1}^{LF})$; $h_{t,i} = E(r^2_{t,i}|\mathcal{F}_{t-1}^{HF})$; $m_{t,i} = E(v_{t,i}|\mathcal{F}_{t-1}^{HF})$.
-For DCC-correlation equations: u_t: vector of degarched returns, with i-th element $u_{t,i=}r_{t,i}/g_{t,i}$; RL_t: realized correlation matrix, see (2); $Q_t = E(u_t u'_t|\mathcal{F}_{t-1}^{LF})$, transformed into a correlation matrix by transforming each element into a correlation coefficient; $R_t = E(RL_t|\mathcal{F}_{t-1}^{HF})$; $P_t = E(M_t|\mathcal{F}_{t-1}^{HF})$; $\bar{R} = E(u_t u'_t) \approx \sum_{t=1}^T u_t u'_t/T$; $\bar{P} = E(RL_t) \approx \sum_{t=1}^T RL_t/T$.

We use the term 'realized variance-covariance model' in the broad sense of a model that specifies the conditional expectation of a variance-covariance matrix (conditional covariance) as a function of past realized variance-covariance matrices. The conditional expectation in question can be that of the variance-covariance matrix of a (daily or other) return vector or of a realized variance-covariance matrix. For the former case, we consider the BEKK-HEAVY-H model of Noureldin et al. (2012), a model of Fleming et al. (2003) – see their equation (15)– which we call FKO-HEAVY-H, and the DCC-HEAVY-H model of Bauwens and Xu (2023); the suffix -H indicates that these models concern the conditional variance-covariance matrix of daily returns, usually named H_t. For the realized variance-covariance case, we consider the BEKK-HEAVY-M of Noureldin et al. (2012), the DCC-HEAVY-M of Bauwens and Xu (2023), to which we add a FKO-HEAVY-M model. In Table 1, we provide the equations defining these models and the restrictions imposed on their parameters to ensure the positivity of the conditional variance-covariance matrices. The table also defines the usual BEKK-GARCH and DCC-GARCH models, to which we add the FKO-GARCH model, which is similar to the equation (14) in Fleming et al. (2003). All models are in scalar version, i.e. each equation of a variance-covariance or correlation matrix depends on scalar parameters.

Regarding the information used in each model group, the GARCH model uses only low-frequency daily data (the daily returns r_t); the HEAVY-M model employs high-frequency intraday data (the realized variance-covariance matrix RC_t); while the HEAVY-H model utilizes both daily and intraday data (incorporating both r_t and RC_t).

A BEKK model has one parameter more than the corresponding FKO model. Each of these models implies that the conditional variance-covariance matrix V_t (where V_t is one of the matrices G_t, H_t, M_t) is a moving average of the past observed matrices O_t ($r_t r_t'$ for GARCH, RC_t for HEAVY-H and -M), i.e. $V_t = \bar{V} + \sum_{j=1}^{\infty} w_j O_{t-j}$, where $\bar{V}$ is a constant matrix and w_j are scalar declining weights. For FKO, $\bar{V} = 0$ and $w_j = \alpha \exp(-j\alpha)$ (where α is α_G, α_H or α_M). For BEKK, $\bar{V} \neq 0$ and $w_j = \alpha\beta^{j-1}$. For FKO, α determines the first weight, $\alpha \exp(-\alpha)$, and the rate of decline of the weights, $\exp(-\alpha)$; for BEKK, α is the first weight and the rate of decline is β. A DCC model has more parameters, allowing each conditional variance to have its own weights, different from the identical weights of the realized correlations; the weights of each conditional covariance depend on the parameters of the corresponding variances and of the correlation process.

3.2 Full sample estimation results

To distinguish the two HEAVY-M models with the DJ29 dataset, we use the suffixes M1 (instead of M) when $RC1_t$ is used, and M2 when $RC2_t$ is used. For the SBGF dataset, only RC_t is used, so we use HEAVY-M only.

The lagged RC_t matrix appears in the HEAVY-H equations of the BEKK and FKO models, and the lagged realized variances and correlation matrix RL_t in the equations of the DCC models. For the SBGF dataset, RC_t is used in the HEAVY-H model. In the DJ29 dataset, using $RC2_t$ instead of RC_t does not change the estimated models since they differ only by a scaling matrix that is constant. It is notable that Noureldin et al. (2012) and Bauwens and Xu (2023) use RC_t in the HEAVY-H model, while Sheppard and Xu (2019) use $RC2_t$. Effectively, using RC_t or $RC2_t$ does not alter the dynamics of the HEAVY model or its out-of-sample forecasting; the only change is in the targeted intercept of the model. For convenience, our HEAVY-H model in the DJ29 dataset employs $RC2_t$.

The models are estimated by the quasi-maximum likelihood (QML) method explained in Appendix C. The full sample parameter estimates are reported in the tables of Appendix D.

For the SBGF data, the estimates of the parameters of the conditional variance equations of BOND and GOLD of the DCC models are close within each type of model. For STOCK, the $A_\bullet$ coefficient (with $\bullet$ = G, H or M) is much higher than for the other two assets, and the $B_\bullet$ one is smaller; this corresponds to a stronger sensitivity of the STOCK volatility to its lagged realized variance and a less smooth volatility. The HEAVY-H correlation process is more sensitive to the lagged realized correlation than the HEAVY-M, but the reverse holds for the HEAVY-H process of the BEKK model (though not of FKO). The HEAVY-H first weight (w_1) of the moving average form is equal to 0.25 for FKO, 0.27 for BEKK, and the respective decline ratios (ρ) are 0.78 and 0.68. For HEAVY-M, these values are $w_1 = 0.19$, $\rho = 0.83$ for FKO, $w_1 = 0.37$, $\rho = 0.61$ for BEKK. Another noticeable feature is the quasi-identical estimates of $A_\bullet$ and $B_\bullet$ of HEAVY-H and HEAVY-M for STOCK, which is not the case for BOND and GOLD.

For the DJ29 data, in each model class (DCC, BEKK, KKO), the HEAVY-M2 and HEAVY-H models have rather close estimates, but the estimates are different from those of the corresponding HEAVY-M1 model.

In the HEAVY-M2 and HEAVY-H models, the α (β) estimates are larger (smaller) than in the corresponding M1 models, indicating more sensitivity of the conditional moment (variance or correlation) to the lagged realized value and a less smooth conditional process. These differences indicate that the bias correction method used to transform the trading period realized matrix to the whole day matrix matters. Adding the overnight return and rescaling (method 1), instead of directly rescaling (method 2), yields a less reactive and smoother estimated conditional process.

Comparing a specific model across the three classes, the HEAVY-H first weight and decline ratio of the moving average form are equal to 0.038 and 0.96 for FKO, 0.13 and 0.86 for BEKK, 0.047 and 0.95 for the DCC correlation process, but (using the median estimates) 0.39 and 0.60 for the DCC variance process. So, the FKO process is the least sensitive and smoothest process, followed by the DCC correlation process, the BEKK process, and the DCC variance processes. This ranking is the same for the other four types of models.

Further insight can be gained from the partial log-likelihood (PLL) value and Bayesian information criterion (BIC). To obtain comparable measures between the GARCH and HEAVY models, we define, similarly to Hansen et al. (2012), a partial log-likelihood (PLL) function for the time series of daily returns, which is based on the assumption that r_t is Gaussian, with zero mean and variance-covariance matrix V_t, where V_t is G_t for a GARCH model, H_t for a HEAVY-H, and M_t for a HEAVY-M. The PLL function formula is thus

$$PLL_V = -\frac{1}{2}\sum_{t=1}^{T}\left(\log|V_t| + r_t' V_t^{-1} r_t\right) \text{ for } V = G, H, M. \tag{6}$$

From the results reported in Tables 2 and 3 we observe that i) in the FKO and DCC classes, HEAVY-H has the lowest BIC (i.e., best fit) in both datasets; ii) in the BEKK class, HEAVY-H has the lowest BIC for DJ29 data, but HEAVY-M has the lowest BIC for SBGF data; iii) in each model category (column of the table), the DCC model has the lowest BIC, though having a larger number of parameters, than the BEKK and FKO models, and the BEKK is better than FKO (with one minor exception).

4 Model evaluation criteria

For comparing the economic value of the models defined in Section 3, we use three economic loss functions (defined in section 4.1) and compute them for a set of out-of-sample forecasts. Because the loss values may differ between models due to the inevitable estimation imprecision inherent to the use of a finite sample of data, we employ the model confidence set (MCS) statistical method (briefly exposed in section 4.2) to assess the significance of the loss differences between the models. In addition, we use the measure of economic value proposed by Fleming et al. (2001, 2003) for comparing two models, i.e., the return an investor (with a given risk aversion) would be willing to sacrifice to switch from one model to another. Like Bollerslev et al. (2018b), we assess the statistical significance of the sacrificed return via the reality check of White (2000).

4.1 Economic loss functions

We adopt the widely used economic loss functions of global minimum variance (GMV) portfolio; see e.g. Engle and Colacito (2006), Engle and Kelly (2012) and Bollerslev et al. (2018b). We denote by V_t the variance-covariance matrix forecast of a model for day t of the forecast period, and by T_s the number of days of the forecast period. For each day, we compute the optimal GMV portfolio weight vector w_t that is the solution of the minimization of $\hat{w}_t' V_t \hat{w}_t$ under the constraint that the weights add up to 1, allowing short sales, and we compute the implied optimal GMV portfolio return $w_t' r_t$ (where r_t is the observed return). Next we compute the forecast variance of the T_s returns, which is the GMV loss function for the considered model. A superior model according to this loss function produces an optimal portfolio with lower variance.

For each model and portfolio allocation criterion, we compute three characteristics of the optimal weight vectors $w_t = (w_{t,1} w_{t,2} \dots w_{t,k})'$ of the forecast period. These characteristics, introduced by Bollerslev et al. (2018b), are the portfolio concentration CO_t, the portfolio total short positions SP_t, and the portfolio turnover TO_t from day t to day $t+1$, defined as

$$CO_t = \left(\sum_{i=1}^{k} w_{t,i}^2 \right)^{1/2}, \tag{7}$$

$$SP_t = \sum_{i=1}^{k} w_{t,i} 1_{\{w_{t,i}<0\}}, \tag{8}$$

$$TO_t = \sum_{i=1}^{k} \left| w_{t+1,i} - w_{t,i} \frac{1 + r_{t,i}}{1 + w_t' r_t} \right|. \tag{9}$$

Portfolios with less concentration and short positions may be of interest for practical implementation and reducing transaction costs. The turnover measure is relevant if transaction costs are proportional to it, in which case the portfolio return is reduced and becomes

$$r_{wt}(c) = w_t' r_t - c\, TO_t. \tag{10}$$

We compare the models by the MCS procedure using the forecast sample average of each characteristic as "loss" function.

To compare the economic significance of the different forecasting models, we also consider the utility-based framework of Fleming et al. (2001, 2003), also used in Bollerslev et al. (2018b) and many other papers. Assuming that an investor has a quadratic utility function with absolute risk aversion parameter γ, the realized daily utility generated by the optimal portfolio based on the variance-covariance matrix forecasts from model a, is defined as

$$U\left(r_{wt}^{(a)}(c),\gamma\right)=\left(1+r_{wt}^{(a)}(c)\right)-\frac{\gamma}{2(1+\gamma)}\left(1+r_{wt}^{(a)}(c)\right)^{2}, \tag{11}$$

where $r_{wt}^{(a)}(c)$ is the optimal (GMV or MV) portfolio return of day t obtained with model a for transaction cost c. We use as additional loss function the forecast sample average of (11) multiplied by -1.

Following Fleming et al. (2003), the economic value of using model b instead of model a can be measured by solving for Δ_γ the equation

$$\sum_{t=1}^{T_s} U\left(r_{wt}^{(a)}(c),\gamma\right)=\sum_{t=1}^{T_s} U\left(r_{wt}^{(b)}(c)-\Delta_\gamma(c),\gamma\right), \tag{12}$$

where $\Delta_\gamma(c)$ can be interpreted as the return the investor with absolute risk aversion γ would be willing to sacrifice to switch from using model a to using model b. The above equation can be solved analytically as explained in Appendix E.

For robustness, we have incorporated three additional loss functions in the supplementary appendix. Among these, two are statistical loss functions: the first is based on the negative of the Wishart log-density function, and the second employs the Frobenius norm to measure the difference between the forecast and benchmark matrices. The third is an economic loss function grounded in the minimum variance (MV) portfolio strategy. Across these varied loss functions, our main results demonstrate consistent patterns.

4.2 The model confidence set statistical procedure

The MCS procedure allows researchers to compare a set of models through a loss function based on the forecasts of the models of the considered set – in our setup these forecasts are the variance-covariance matrix forecasts. The procedure compares the models jointly, without the need to choose arbitrarily a reference model. It has been developed by Hansen et al. (2003) and further elaborated by Hansen et al. (2011); it has become a widespread tool for model comparison.[4]

For a chosen loss function, and a initial set of models $\mathcal{M}_0$, the loss difference between each

[4] A Google search on "the model confidence set" found 1993 citations on June 27, 2023.

pair of models in the set is computed (at every time point $t = 1, \ldots, T_s$ of the forecast period), so that for models a and b, we get $D_{t,ab} = L_{t,a} - L_{t,b}$, where $L_{t,\cdot}$ is the chosen loss function (e.g., the MV loss). At each step of the procedure, the null hypothesis of equal predictive accuracy, $E[D_{t,ab}] = 0$, is tested for $\forall a > b \in \mathcal{M}$, a subset of models $\mathcal{M} \subset \mathcal{M}_0$, with $\mathcal{M} = \mathcal{M}_0$ at the initial step. If H_0 is rejected at a chosen significance level α (e.g., 0.05 or 5 percent), the worst performing model is removed from $\mathcal{M}$; hence, a model is removed from $\mathcal{M}$ only if it is significantly inferior to the other models. This procedure is continued until a set of models includes no model that can be rejected at the level α, the resulting set being the model confidence set at the $1 - \alpha$ confidence level. The procedure does not necessarily select a single best model, as it may end with a set models of equal forecasting ability, which may be the initial set $\mathcal{M}_0$.

The test statistic for the null hypothesis $E[D_{t,ab}] = 0$ is the range statistic of Hansen et al. (2011):

$$\max_{a,b \in \mathcal{M}} \frac{\mid \overline{D}_{ab} \mid}{\sqrt{\widehat{\mathrm{Var}}(\overline{D}_{ab})}}, \tag{13}$$

where $\overline{D}_{ab} = \frac{1}{T_s} \sum_{t=1}^{T_s} D_{t,ab}$, and the estimated variance $\widehat{\mathrm{Var}}(\overline{D}_{ab})$ is obtained by a bootstrap approach (Hansen et al. (2003)), with 5,000 replications and block length of 22.

5 Empirical results: comparisons of out-of-sample forecasts

Out-of-sample s-step-ahead forecasts of the variance-covariance matrices are computed for $s = 1$ (one day ahead), $s = 5$ (one week ahead), and $s = 22$ (one month ahead). To compute out-of-sample forecasts, each model is first estimated on the sample of 3,000 observations starting at the first date of the available data, and out-of-sample forecasts for the next five observations (3001-3005) are computed from the estimated model. Next, each model is re-estimated on the window of 3,000 observations from observation 6 to 3005, and the five forecasts (3006-3010) are computed. The re-scaling is updated within the rolling window scheme. The procedure is continued until the end of the sample and results in a total of $T_s = 1092$ out-of-sample forecasts for $s = 1$, 1088 for $s = 5$, and 1071 for $s = 22$ for the DJ29 dataset, and $T_s = 1864$ out-of-sample forecasts for $s = 1$, 1860 for $s = 5$, and 1,843 for $s = 22$ for the SBGF dataset.

5.1 SBGF data

5.1.1 Loss functions and portfolio features

For the GMV loss functions (see the StDev columns in Table 4), at horizons 1 and 5, in each case the six realized variance-covariance models form the MCS95 (i.e., the model confidence set, at the 95% level of confidence, obtained by comparing the nine models); thus, the three models that use only daily data (DCC-GARCH, BEKK-GARCH, FKO) are excluded. At horizon 22, the MCS95 consists only of the four BEKK and DCC realized variance-covariance models.

Considering the average turnover (see the TO columns in the tables), the average return values (see the Return columns) and the (opposite of the) utility function [5] (see the Utility columns) as loss criteria, the nine models are included in the MCS95, whatever the horizon. The values do not differ much between the models (except outliers).

The average concentration (column CO in the tables) and the the importance of short positions (SP column, in absolute value) of the HEAVY-M portfolio is lower than or equal to the value for the corresponding HEAVY-H. For both criteria, each of the six MCS95 includes from one to four models (except for SP at horizon 22, where the SP values are all very close to zero so that eight models are in the MCS95).

Considering all these results, we can conclude that i) the realized variance-covariance models are superior to the GARCH models, and ii) the HEAVY-M and HEAVY-H models perform similarly.

5.1.2 Economic value of model switching

The second column of the Table 5 reports the economic gain $\Delta_\gamma(c)$ – see (12) – of using the HEAVY-H model instead of the HEAVY-M and the GARCH model of the same class, when the risk aversion parameter γ is equal to 1 and the transaction cost is set to zero; for example in the DCC class, for horizon 1, using HEAVY-H instead of GARCH, is worth 10.9 basis points, which is a statistically significant value at the 5% level based on the reality check of White (2000), while switching from HEAVY-M to HEAVY-H entails an insignificant loss of 3.1 points. Considering the three model classes and three horizons, the results confirm i) the positive economic value of each realized variance-covariance (HEAVY-H) model with respect to the corresponding model that uses only daily data (with two exceptions, at horizon 22, for

[5] with risk aversion parameter γ equal to 1 and transaction cost set to zero

FKO), this being statistically significant in ten among the eighteen comparisons, and ii) the equivalence of the HEAVY-M and HEAVY-H models, since the gains or losses are small and insignificant.

The last three columns provide a view of the impact of increasing the risk aversion parameter from 1 to 10, and the transaction cost from 0 to 0.01.

For a given value of the transaction cost, the increase of the risk aversion parameter amplifies the economic gain or loss of using the HEAVY-H model instead of the HEAVY-M and the GARCH model of the same class; for example, at horizon 1, for $\gamma = 10$ and $c = 0$, the gain of switching from DCC-GARCH to DCC-HEAVY-H is 90.4 instead of 10.9 for $\gamma = 1$ and $c = 0$. The number of significant gains or losses increases slightly when the risk aversion increases (from 10 to 13 for the changes from GARCH to HEAVY-H, and from 0 to 3 for HEAVY-M to HEAVY-H).

For a given value of the risk aversion parameter, the increase of the transaction cost (from 0 to 0.01) has a small impact on the economic gains or losses: in almost all comparisons, a gain remains a gain, a significant gain remains a significant gain, and likewise for losses. The differences of values of the gains or losses are very small.

In each forecast horizon, the last block (of three rows) shows the results for switches between the HEAVY-H models of the three classes (BEKK, DCC, FKO), since these are the best performing models in their own class. Like for the comparisons commented above, the gains or losses are stable with respect to the transaction cost, and they are boosted by the higher risk aversion. Nevertheless, the differences between the models are mostly small and insignificant for horizons 1 and 5, whereas at horizon 22 they are big and significant, especially for $\gamma = 10$. Except for the monthly horizon where DCC and BEKK models outperform the FKO model, the three HEAVY-H models can be considered as equivalent.

5.2 DJ29 data

5.2.1 Loss functions and portfolio features

Table 6 presents comparisons of the economic loss functions. About the two ways to transform the realized variance-covariance matrix of the trading period into a daily one, no model (HEAVY-M1) that uses the first correction ($RC1_t$) is included in one of the MCS95 for the GMV loss functions (see the StDev columns in the tables). In contrast, the models that use the second correction ($RC2_t$) often belong to the MCS95; in particular the DCC-HEAVY-M2 and -H models are in each of the three MCS95; the FKO-HEAVY-M2 model is in two of them, FKO-HEAVY-H in three, BEKK-HEAVY-M2 in two. In brief, i) the models (-H and -M2) using the $RC2_t$ realized covariances perform better than the models (-M1) using $RC1_t$, and ii) the models that use only the daily returns are rarely included in the StDev MCS95, except the FKO model (included in the three StDev MCS95), which has smaller losses than DCC-GARCH and BEKK-GARCH.

Using the (opposite of the) utility function as loss criterion (see the Utility columns) with risk aversion parameter γ equal to 1 and transaction cost set to zero, the twelve models are considered as equivalent at horizon 22 and at horizon 1, the included models are the HEAVY-M2 and HEAVY-H models, i.e., the same as for the corresponding StDev losses (with one exception: BEKK-HEAVY-H is included for GMV Utility, not for StDev). Considering the average turnover (see the TO columns) and average return values (see the Return columns), the twelve models are in each MCS95. The values do not differ much between the models.

The average concentration (column CO in the tables) and the importance (in absolute value) of short positions (SP column) of the HEAVY-M2 and -H portfolios are lower than or equal to the values for the corresponding GARCH and HEAVY-M1 models. For each measure, the six MCS95 contain the DCC-HEAVY-H model, plus two other models at most.

Considering all these results, we can conclude that: i) the realized variance-covariance models using the second bias correction (HEAVY-M2) outperform both the realized variance-covariance models using the first bias correction (HEAVY-M1) and the daily GARCH models; and ii) the two best-performing models are HEAVY-M2 and HEAVY-H, with similar performances.

5.2.2 Economic value of model switching

The second column of the Table 7 reports the economic gain $\Delta_\gamma(c)$ – see (12) – of using the HEAVY-H model instead of each other model of the same class; for example, in the DCC class, for horizon 1, using HEAVY-H instead of GARCH is worth 98.5 basis points, which is a significant value at the 5% level based on the reality check of White (2000), and switching from HEAVY-M2 to HEAVY-H entails a insignificant gain of 5.9 basis points. The main findings from these economic gain comparisons are summarized in three items:

i) The HEAVY-H model entails a gain with respect to the HEAVY-M1 models of the same class in 8 cases (over 9); only one loss occurs (at horizon 22, for BEKK), but it is not statistically significant (at the 5% level); on the contrary, several gains are statistically significant (3 at horizon 1, 2 at horizon 5, and 1 at horizon 22). This set of results confirms the better performance of the models using $RC2_t$ with respect to the models using $RC1_t$, already mentioned for the GMV loss functions.

ii) Switching from HEAVY-M2 to HEAVY-H provides no significant gain or loss in the three possible comparisons at horizon 1, one significant loss at horizon 5 (for BEKK), and one significant gain at horizon 22 (for DCC). The performances of these two types of models can be considered as broadly equivalent.

iii) Switching from the DCC-GARCH model to the DCC-HEAVY-H corresponding model entails a statistically significant gain in the three possible comparisons, i.e. at each horizon and for each loss function; the gain values (Δ_1) are important, being between 40.5 and 98.5 basis points. Switching from BEKK-GARCH to BEKK-HEAVY-H creates a significantly positive gain in one case (at horizon 22), and switching from FKO to FKO-HEAVY-H entails insignificant gains or losses. Using realized variances and covariances as inputs in the DCC class is more valuable than in the other two classes.

The last three columns of the table provide a view of the impact of increasing the risk aversion parameter from 1 to 10, and the transaction cost from 0 to 0.01.

For a given value of the transaction cost, the increase of the risk aversion parameter amplifies the economic gain or loss of using the HEAVY-H model instead of the other HEAVY models (M2, M1) and the GARCH model of the same class; for example, at horizon 22, for $\gamma = 10$ and $c = 0$, the gain of switching from DCC-GARCH to DCC-HEAVY-H is 265.7 basis points instead of 40.5 for $\gamma = 1$ and $c = 0$. The significant gains or losses for $\gamma = 1$ remain significant for $\gamma = 10$ in most cases.

For a given value of the risk aversion parameter, the increase of the transaction cost (from 0 to 0.01) has in many comparisons a strong effect on the economic gains or losses: there are sign switches, but less so when both values are significant. The differences of values of the gains or losses vary a lot, and several are important.

At each forecast horizon, the last block (of three rows) shows the results for switches between the HEAVY-H models of the three classes (BEKK, DCC, FKO), since these are the best performing models in their own class. Like for the comparisons commented above, the gains or losses are amplified by the higher transaction cost and risk aversion parameter. Switching from BEKK to DCC creates significant gains when they are positive; losses occur only for horizon 5 and transaction cost 0.01, but they are statistically insignificant. Switching from FKO to BEKK results in 11 losses (out of 12 cases), among which six are significant; the single gain is insignificant. Switching from DCC to FKO entails losses in 7 (out of 12) cases but they are statistically insignificant (except at horizon 5, $\gamma = 10$ and $c = 0$). In brief, the BEKK-HEAVY-H is less valuable than the corresponding DCC and FKO models, and DCC seems more valuable than FKO, although there is no statistical support.

Table 8 provides an economic comparison between HEAVY-M2 and the other two models of each class. The results of the HEAVY-M2 comparison are in line with the results of the HEAVY-H comparison. These two models are the best performing models and are not distinguishable from each other using economic loss functions. However, it is important to note that if the HEAVY-M1 model is used in place of HEAVY-M2, the forecasting performance deteriorates significantly.

Our empirical findings shows that simply adding overnight return variations to the realized variance-covariance matrix (HEAVY-M1 models) does not improve forecasting performance. Hansen and Lunde (2005) proposed using different weights for overnight return variations and realized variances in univariate models. For thoroughness, we also tested weighting the overnight returns outer product and the realized variance-covariance matrix in the first bias correction method. However, the optimal weight for the overnight returns turned out to be zero during estimation, effectively making this method similar to the second bias correction approach. As a result, the forecasts using this optimal weight method matched those from the second bias correction method. This confirms that the second bias correction, which involves rescaling the intraday realized variance-covariance matrix to a full-day one, is optimal, at least for the dataset utilized in our empirical analysis.

5.3 Synthesis of empirical findings

The main findings for both datasets can be summarized in a few items:

1. The models utilizing the high-frequency intraday data, namely, the HEAVY-M and HEAVY-H models for the SBGF dataset and the HEAVY-M2 and HEAVY-H models for the DJ29

dataset, exhibit a significant improvement in forecasting performance for the daily variances and covariances when the forecasting horizon is short (i.e., daily or weekly). However, for forecasting one month ahead, the gains or losses are marginal and statistically insignificant. This observation aligns with the findings in univariate volatility forecasting, as, for example, in Lyocsa et al. (2021).

2. For the DJ29 dataset, the models using the first bias correction method (i.e., HEAVY-M1), which employs the overnight returns to construct the full-day varianc-covariance matrix, are not as effective as the models using the second bias correction method (i.e., HEAVY-M2), where the variance-covariance matrix of the intraday returns is directly rescaled to the full-day one. The primary reason for this is that the overnight returns yield noisy estimators of the overnight variances and covariances.

3. For the portfolios of three assets of the SBGF dataset, the DCC, BEKK, and FKO models exhibit similar performances. Conversely, for the larger portfolio of twenty-nine assets of the DJ29 dataset, the DCC model outperforms the BEKK and FKO models. The reason is that the the DCC model is able to capture the heterogeneity of the individual variance processes, which is more important for twenty-nine assets than for three.

4. From both the SBGF and DJ29 datasets, the two best performing models are HEAVY-M2 (HEAVY-M in the case of SBGF) and HEAVY-H, across nearly all evaluation criteria. These two models perform equally well, and their differences are statistically insignificant.

5. The results from the economic value of volatility timing show that an increase in transaction cost has a very limited impact on the economic losses or gains for the SBGF dataset, while it has a bigger impact for the DJ29 dataset. Transaction costs decrease the portfolio return by $c = 0.01$ times the turnover measure, see (10). In the three asset case (SBGF dataset), the average turnover for the DCC-HEAVY-H model is 0.928 at horizon 1, and the corresponding average return is 2.388 (see Table 4, columns 2 and 5). This values does not change much between different models and horizons. The impact of the transaction cost is very small: 0.01x0.928 = 0.00928; the reduced return is equal to 2.379. In the 29 asset case (DJ29 dataset), the average turnover is much higher than for the other dataset: 3.715 for s=1, with an average return of 1.596 for the DCC-HEAVY-H model

(see Table 6), hence the impact of the transaction cost (0.03715) is more important than in the case of three assets (by a factor of four); the reduced return is equal to 1.559. It is quite normal that the portfolio turnover increases with the number of assets. Having more assets increases the complexity and frequency of rebalancing the portfolio, which can significantly amplify the impact of transaction costs. With more assets, there are more potential trades to execute when adjusting the portfolio in response to volatility changes, leading to higher cumulative transaction costs.

Overall, the HEAVY-H and HEAVY-M (if rescaling is not necessary) or -M2 (in case of rescaling) models emerge as the most promising option for analyzing a wide range of assets. For empirical applications, we suggest using the HEAVY-H model as preferred choice. This model solely relies on intraday returns and directly models the daily variance-covariance matrix, adopting a GARCH model structure while using the lagged realized variance-covariance matrix as its key input variable. However, the HEAVY-M1 (in case of DJ29) model may perform poorly if overnight returns are included in the whole day variance-covariance matrix calculation.

6 Conclusions

It is known, since Fleming et al. (2003), that realized variance-covariance models, i.e., multivariate volatility models that employ high-frequency data to measure multivariate volatility, yield significant additional economic value with respect to GARCH models that employ only daily data.

For the datasets we use, we find that the benefit of the realized variance-covariance models is important for one to five days ahead forecast horizons, not for a monthly horizon. Attention should be paid to the way by which the full day variance-covariance matrix is constructed, when the data do not allow to measure the latter completely from the trading period intraday returns. Adding the outer product of overnight returns to correct the trading period variance-covariance matrix before rescaling the latter does not seem advantageous, with respect to directly rescaling the trading period variance-covariance matrix to the full day one. More empirical research, based on different data, is needed to add further evidence on such issues.

In conclusion, we recommend the HEAVY-H model as the preferred choice in empirical analysis due to several reasons: 1) It directly models the entire day covariance; 2) It utilizes high-frequency intraday data; 3) It consistently performs best, or at least equally well, compared to the top-performing realized variance-covariance models.

References

Acharya, V., Pedersen, L., Philippon, T., Richardson, M., 2017. Measuring systemic risk. Review of Financial Studies 30, 2–47.

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

Bauwens, L., Otranto, E., 2022. Modeling realized covariance matrices: a class of Hadamard exponential models. Journal of Financial Econometrics https://doi.org/10.1093/jjfinec/nbac007.

Bauwens, L., Xu, Y., 2023. Dcc-and deco-heavy: Multivariate garch models based on realized variances and correlations. International Journal of Forecasting 39, 938–955.

Bollerslev, T., Hood, B., Huss, J., Pedersen, L., 2018a. Risk everywhere: modeling and managing volatility. Review of Financial Studies 31, 2728–2773.

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

Engle, R., Colacito, R., 2006. Testing and valuing dynamic correlations for asset allocation. Journal of Business & Economic Statistics 24, 238–253.

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

Fleming, J., Kirby, C., Ostdiek, B., 2001. The economic value of volatility timing. Journal of Finance 56, 329–352.

Fleming, J., Kirby, C., Ostdiek, B., 2003. The economic value of volatility timing using realized volatility. Journal of Financial Economics 67, 473–509.

Hansen, P., 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, P., Lunde, A., 2005. A realized variance for the whole day based on intermittent high-frequency data. Journal of Financial Econometrics 3, 525–554.

Hansen, P., Lunde, A., Nason, J., 2003. Choosing the best volatility models: the model confidence set approach. Oxford Bulletin of Economic and Statistics 65, 839–861.

Hansen, P.R., Lunde, A., Nason, J.M., 2011. The model confidence set. Econometrica 79, 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.

Lyocsa, S., Molnar, P., Vyrost, T., 2021. Stock market volatility forecasting: Do we need high-frequency data? International Journal of Forecasting 37, 1092–1110.

Marquering, W., Verbeek, M., 2004. The economic value of predicting stock index returns and volatility. Journal of Financial and Quantitative Analysis 39, 407–429.

Noureldin, D., Shephard, N., Sheppard, K., 2012. Multivariate high-frequency-based volatility (HEAVY) models. Journal of Applied Econometrics 27, 907–933.

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

Sheppard, K., Xu, W., 2019. Factor high-frequency-based volatility (HEAVY) models. Journal of Financial Econometrics 17, 33–65.

Taylor, N., Xu, Y., 2017. The logarithmic vector multiplicative error model: an application to high frequency nyse stock data. Quantitative Finance 17, 1021–1035.

White, H., 2000. A reality check for data snooping. Econometrica 68, 1097–1126.

Table 2 – PLL and BIC on full sample of SBGF data

	p	GARCH	HEAVY-M	HEAVY-H
		Partial log-likelihood (PLL)		
DCC-	11	-17138	-16837	-16712
BEKK-	2	-17266	-16845	-16915
FKO-	1	-17497	-16920	-16900
		Bayesian information criterion (BIC)		
DCC-		<u>7.066</u>	<u>6.942</u>	<u>**6.891**</u>
BEKK-		7.103	**6.930**	6.959
FKO-		7.196	6.959	**6.951**

p: number of estimated parameters; PLL: see eq. (6). Bold values: minimum BIC in row. Underlined values: minimum BIC in column.

Table 3 – PLL and BIC on full sample of DJ29 data

	p	GARCH	HEAVY-M1	HEAVY-M2	HEAVY-H
		Partial log-likelihood (PLL)			
DCC-	89	-187724	-188125	-186205	-184187
BEKK-	2	-191475	-188526	-186711	-185452
FKO-	1	-193771	-190610	-189158	-188699
		Bayesian information criterion (BIC)			
DCC-		<u>91.93</u>	<u>92.13</u>	<u>91.01</u>	<u>**90.03**</u>
BEKK-		93.59	92.15	91.26	**90.65**
FKO-		94.71	93.16	92.45	**92.23**

p: number of estimated parameters; PLL: see eq. (6). Bold values: minimum BIC in row. Underlined values: minimum BIC in column.

Table 4 – Economic Loss function comparisons for SBGF data

	TO	CO	SP	Return	StDev	Utility
s=1						
DCC-GARCH	**0.724**	0.682	-0.056	**2.018**	6.919	**0.707**
DCC-HEAVY-M	**0.753**	0.665	**-0.009**	**2.375**	**6.692**	**0.710**
DCC-HEAVY-H	**0.928**	0.692	-0.083	**2.388**	**6.738**	**0.710**
BEKK-GARCH	**0.682**	0.685	-0.069	**1.742**	6.960	**0.705**
BEKK-HEAVY-M	**0.805**	**0.653**	-0.026	**2.268**	**6.743**	**0.709**
BEKK-HEAVY-H	**0.790**	0.687	-0.079	**1.828**	**6.787**	**0.708**
FKO	**0.941**	0.682	-0.073	**1.617**	6.976	**0.705**
FKO-HEAVY-M	**0.656**	**0.653**	**-0.012**	**2.077**	**6.742**	**0.709**
FKO-HEAVY-H	**0.832**	0.654	-0.018	**2.149**	**6.739**	**0.709**
s=5						
DCC-GARCH	**0.885**	0.682	-0.057	**1.921**	6.933	**0.706**
DCC-HEAVY-M	**0.773**	**0.653**	**-0.009**	**2.520**	**6.741**	**0.710**
DCC-HEAVY-H	**1.456**	0.689	-0.067	**2.228**	**6.768**	**0.709**
BEKK-GARCH	**0.728**	0.685	-0.061	**1.876**	6.869	**0.707**
BEKK-HEAVY-M	**0.853**	0.655	-0.025	**2.358**	**6.744**	**0.710**
BEKK-HEAVY-H	**0.989**	0.688	-0.065	**2.082**	**6.765**	**0.709**
FKO	**0.979**	0.685	-0.065	**1.690**	6.916	**0.706**
FKO-HEAVY-M	**0.827**	**0.653**	-0.014	**2.489**	**6.828**	**0.709**
FKO-HEAVY-H	**0.982**	**0.653**	-0.018	**2.516**	**6.825**	**0.709**
s=22						
DCC-GARCH	**1.288**	0.688	**-0.001**	**1.723**	7.160	**0.703**
DCC-HEAVY-M	**0.972**	0.673	**0.000**	**2.107**	**7.037**	**0.705**
DCC-HEAVY-H	**0.641**	0.710	**0.000**	**1.875**	**7.013**	**0.705**
BEKK-GARCH	**1.610**	0.687	**-0.001**	**1.706**	7.314	**0.700**
BEKK-HEAVY-M	**0.822**	0.665	**0.000**	**2.710**	**7.068**	**0.706**
BEKK-HEAVY-H	**1.765**	0.692	**0.000**	**2.587**	**7.044**	**0.706**
FKO	**1.519**	0.686	-0.004	**1.995**	7.351	**0.700**
FKO-HEAVY-M	**1.705**	**0.653**	**0.000**	**2.965**	7.574	**0.699**
FKO-HEAVY-H	**1.709**	**0.653**	**0.000**	**2.965**	7.574	**0.699**

TO: average (over forecast sample) turnover;
CO: average portfolio concentration;
SP: average of short positions;
Return: average of optimal portfolio annualized returns for transaction cost = 0;
StDev: corresponding standard deviation;
Utility: mean of utilities defined by (11), with $\gamma = 1$, $c = 0$ and (a) the model indicated in the first column.
Bold values in columns 2-7 identify the models included in the 95% MCS computed for the nine models of each portfolio type.

Table 5 – Economic values of switching between models for two risk aversion parameters γ and two transaction costs c for SBGF data

	γ=1		γ=10	
	c=0	c=0.01	c=0	c=0.01
s=1				
from DCC-GARCH to DCC-HEAVY-H	10.9	11.9	90.4	87.8
from DCC-HEAVY-M to DCC-HEAVY-H	-3.1	-3.4	-37.1	-39.7
from BEKK-GARCH to BEKK-HEAVY-H	8.7	7.5	86.0	84.9
from BEKK-HEAVY-M to BEKK-HEAVY-H	-7.4	-7.5	-39.5	-39.2
from FKO to FKO-HEAVY-H	13.6	14.1	101.2	101.9
from FKO-HEAVY-M to FKO-HEAVY-H	1.1	0.6	3.0	0.8
from BEKK-HEAVY-H to DCC-HEAVY-H	9.1	8.7	39.2	37.2
from FKO-HEAVY-H to BEKK-HEAVY-H	-6.6	-6.5	-42.2	-41.2
from DCC-HEAVY-H to FKO-HEAVY-H	-2.5	-2.4	-3.2	-1.7
s=5				
from DCC-GARCH to DCC-HEAVY-H	9.6	5.3	64.8	58.4
from DCC-HEAVY-M to DCC-HEAVY-H	-5.0	-6.3	-24.2	-34.6
from BEKK-GARCH to BEKK-HEAVY-H	4.1	3.7	41.9	39.2
from BEKK-HEAVY-M to BEKK-HEAVY-H	-4.2	-4.7	-19.3	-21.1
from FKO to FKO-HEAVY-H	6.2	6.1	31.8	13.5
from FKO-HEAVY-M to FKO-HEAVY-H	0.5	0.2	2.5	0.2
from BEKK-HEAVY-H to DCC-HEAVY-H	1.3	0.4	-0.4	-7.1
from FKO-HEAVY-H to BEKK-HEAVY-H	-0.5	-0.4	37.0	37.0
from DCC-HEAVY-H to FKO-HEAVY-H	-1.2	-0.2	-44.6	-36.7
s=22				
from DCC-GARCH to DCC-HEAVY-H	12.1	14.3	15.0	17.2
from DCC-HEAVY-M to DCC-HEAVY-H	-0.8	0.2	15.5	19.3
from BEKK-GARCH to BEKK-HEAVY-H	28.3	30.4	14.6	28.1
from BEKK-HEAVY-M to BEKK-HEAVY-H	0.5	-2.0	16.9	2.0
from FKO to FKO-HEAVY-H	-7.2	-6.0	-132.9	-123.3
from FKO-HEAVY-M to FKO-HEAVY-H	-0.2	-0.2	-0.3	-0.3
from BEKK-HEAVY-H to DCC-HEAVY-H	-4.9	-2.3	16.5	31.4
from FKO-HEAVY-H to BEKK-HEAVY-H	35.3	34.9	268.2	266.3
from DCC-HEAVY-H to FKO-HEAVY-H	-30.3	-32.7	-272.3	-278.7

The numerical values are the $\Delta_\gamma(c)$ as defined by (12) for γ and c indicated in the headers, i.e. the economic gain (if positive) or loss (if negative) of switching from one model to another.
Underlined values indicate statistical significance at the 5% level.

Table 6 – Economic value comparisons for DJ29 data

	TO	CO	SP	Return	StDev	Utility
s=1						
DCC-GARCH	**4.134**	0.499	-0.423	**1.106**	11.640	0.618
DCC-HEAVY-M1	**4.316**	0.502	-0.473	**-1.084**	11.451	0.618
DCC-HEAVY-M2	**3.410**	**0.451**	-0.366	**1.333**	**10.822**	**0.637**
DCC-HEAVY-H	**3.715**	**0.441**	**-0.348**	**1.596**	**10.793**	**0.638**
BEKK-GARCH	**3.883**	0.482	-0.411	**-3.086**	11.111	0.621
BEKK-HEAVY-M1	**4.502**	0.511	-0.496	**-2.584**	11.350	0.617
BEKK-HEAVY-M2	**3.242**	0.470	-0.407	**-0.535**	**10.765**	**0.634**
BEKK-HEAVY-H	**4.221**	0.510	-0.480	**-2.757**	10.990	**0.625**
FKO	**7.240**	0.452	-0.428	**0.296**	**10.848**	**0.634**
FKO-HEAVY-M1	**9.461**	0.512	-0.527	**-2.127**	11.453	0.616
FKO-HEAVY-M2	**2.832**	0.463	-0.415	**0.348**	**10.833**	**0.634**
FKO-HEAVY-H	**2.857**	**0.452**	-0.392	**1.037**	**10.875**	**0.635**
s=5						
DCC-GARCH	**2.997**	0.485	-0.409	**-0.042**	11.466	**0.620**
DCC-HEAVY-M1	**5.143**	0.504	-0.463	**-1.101**	11.430	0.618
DCC-HEAVY-M2	**3.087**	0.447	-0.329	**0.630**	**11.004**	**0.631**
DCC-HEAVY-H	**3.594**	**0.430**	**-0.312**	**1.468**	**10.948**	**0.634**
BEKK-GARCH	**3.145**	0.482	-0.409	**-3.643**	11.174	**0.619**
BEKK-HEAVY-M1	**4.811**	0.511	-0.492	**-1.713**	11.450	0.617
BEKK-HEAVY-M2	**2.648**	0.471	-0.397	**0.077**	**11.074**	**0.629**
BEKK-HEAVY-H	**2.881**	0.505	-0.459	**-1.864**	11.231	**0.621**
FKO	**3.243**	0.451	-0.428	**-0.452**	**10.986**	**0.629**
FKO-HEAVY-M1	**4.118**	0.512	-0.527	**-1.467**	11.550	0.615
FKO-HEAVY-M2	**3.734**	0.462	-0.413	**1.330**	**11.164**	**0.629**
FKO-HEAVY-H	**3.813**	0.452	-0.393	**1.775**	**11.081**	**0.632**
s=22						
DCC-GARCH	**2.848**	0.462	-0.357	**-0.855**	11.482	**0.618**
DCC-HEAVY-M1	**2.971**	0.507	-0.424	**-1.339**	11.416	**0.618**
DCC-HEAVY-M2	**3.837**	0.458	-0.331	**-2.094**	**11.198**	**0.622**
DCC-HEAVY-H	**2.463**	**0.445**	**-0.289**	**-0.924**	**11.114**	**0.626**
BEKK-GARCH	**4.929**	0.483	-0.402	**-4.913**	11.603	**0.607**
BEKK-HEAVY-M1	**4.764**	0.511	-0.479	**-1.114**	11.574	**0.615**
BEKK-HEAVY-M2	**3.185**	0.478	-0.383	**-0.047**	11.484	**0.619**
BEKK-HEAVY-H	**3.629**	0.507	-0.439	**-1.667**	11.562	**0.614**
FKO	**4.140**	0.451	-0.428	**-0.290**	**11.260**	**0.624**
FKO-HEAVY-M1	**5.061**	0.512	-0.526	**-0.735**	11.701	**0.613**
FKO-HEAVY-M2	**5.244**	0.460	-0.408	**2.284**	11.537	**0.623**
FKO-HEAVY-H	**3.234**	**0.453**	**-0.393**	**2.807**	**11.379**	**0.627**

For explanations, see note below Table 4.

Table 7 – Economic values of switching between HEAVY-H and other models for two risk aversion parameters γ and two transaction costs c for DJ29 data

	γ=1		γ=10	
	c=0	c=0.01	c=0	c=0.01
s=1				
from DCC-GARCH to DCC-HEAVY-H	98.5	164.1	496.2	671.7
from DCC-HEAVY-M1 to DCC-HEAVY-H	77.4	152.3	152.9	532.0
from DCC-HEAVY-M2 to DCC-HEAVY-H	5.9	11.5	33.7	93.8
from BEKK-GARCH to BEKK-HEAVY-H	16.7	-366.0	104.9	-509.7
from BEKK-HEAVY-M1 to BEKK-HEAVY-H	37.6	-128.8	228.6	-166.4
from BEKK-HEAVY-M2 to BEKK-HEAVY-H	-26.7	-323.1	-279.7	-365.8
from FKO to FKO-HEAVY-H	4.5	513.6	-23.9	713.7
from FKO-HEAVY-M1 to FKO-HEAVY-H	91.4	114.7	379.8	417.0
from FKO-HEAVY-M2 to FKO-HEAVY-H	2.4	472.2	-45.1	298.7
from BEKK-HEAVY-H to DCC-HEAVY-H	64.6	424.9	195.2	539.8
from FKO-HEAVY-H to BEKK-HEAVY-H	-50.6	-336.0	-179.6	-365.5
from DCC-HEAVY-H to FKO-HEAVY-H	-14.6	0.4	-136.8	-19.8
s=5				
from DCC-GARCH to DCC-HEAVY-H	72.5	-199.8	359.3	-324.2
from DCC-HEAVY-M1 to DCC-HEAVY-H	42.8	169.5	238.6	475.0
from DCC-HEAVY-M2 to DCC-HEAVY-H	14.6	-179.3	65.3	-324.3
from BEKK-GARCH to BEKK-HEAVY-H	11.6	24.3	-44.6	48.1
from BEKK-HEAVY-M1 to BEKK-HEAVY-H	21.1	41.1	153.9	218.5
from BEKK-HEAVY-M2 to BEKK-HEAVY-H	-36.9	118.6	-270.7	240.3
from FKO to FKO-HEAVY-H	11.8	25.6	-119.9	-138.3
from FKO-HEAVY-M1 to FKO-HEAVY-H	81.3	86.6	188.7	309.1
from FKO-HEAVY-M2 to FKO-HEAVY-H	13.8	66.5	87.7	307.1
from BEKK-HEAVY-H to DCC-HEAVY-H	64.2	-121.8	198.1	-224.4
from FKO-HEAVY-H to BEKK-HEAVY-H	-44.8	-11.3	-44.8	93.3
from DCC-HEAVY-H to FKO-HEAVY-H	-11.7	179.2	-234.4	149.9
s=22				
from DCC-GARCH to DCC-HEAVY-H	40.5	16.9	265.7	152.4
from DCC-HEAVY-M1 to DCC-HEAVY-H	19.6	-3.8	179.2	42.2
from DCC-HEAVY-M2 to DCC-HEAVY-H	21.1	27.4	88.0	94.6
from BEKK-GARCH to BEKK-HEAVY-H	37.2	136.3	70.1	472.9
from BEKK-HEAVY-M1 to BEKK-HEAVY-H	-4.7	54.6	-9.4	267.3
from BEKK-HEAVY-M2 to BEKK-HEAVY-H	-25.1	-36.2	-131.0	-251.8
from FKO to FKO-HEAVY-H	17.5	105.5	-224.0	277.5
from FKO-HEAVY-M1 to FKO-HEAVY-H	70.4	136.1	269.4	455.0
from FKO-HEAVY-M2 to FKO-HEAVY-H	23.4	91.8	155.0	392.9
from BEKK-HEAVY-H to DCC-HEAVY-H	57.4	47.4	311.4	238.6
from FKO-HEAVY-H to BEKK-HEAVY-H	-66.0	-77.1	-268.7	-305.8
from DCC-HEAVY-H to FKO-HEAVY-H	7.4	29.0	-224.0	-65.6

For explanations, see note below Table 5.

Table 8 – Economic values of switching between HEAVY-M2 and other models for two risk aversion parameters γ and two transaction costs c for DJ29 data

	γ=1		γ=10	
	c=0	c=0.01	c=0	c=0.01
s=1				
from DCC-GARCH to DCC-HEAVY-M2	96.1	269.6	146.9	340.5
from DCC-HEAVY-M1 to DCC-HEAVY-M2	59.3	79.3	115.2	268.1
from BEKK-GARCH to BEKK-HEAVY-M2	51.4	157.8	75.6	199.7
from BEKK-HEAVY-M1 to BEKK-HEAVY-M2	72.9	206.6	130.8	294.8
from FKO to DCC-HEAVY-M2	2.1	16.8	476.6	743.1
from FKO-HEAVY-M1 to BEKK C-HEAVY-M2	79.5	222.6	481.6	176.6
from DCC-HEAVY-M2 to BEKK-HEAVY-M2	44.2	139.0	140.0	546.5
from BEKK-HEAVY-M2 to FKO-HEAVY-M2	-42.9	-149.3	-381.2	-440.5
from FKO-HEAVY-H2 to DCC-HEAVY-M2	-15.8	-25.2	-48.2	-36.6
s=5				
from DCC-GARCH to DCC-HEAVY-M2	67.6	224.9	123.1	327.6
from DCC-HEAVY-M1 to DCC-HEAVY-M2	51.7	157.1	353.8	455.7
from BEKK-GARCH to BEKK-HEAVY-M2	64.6	186.1	122.1	282.5
from BEKK-HEAVY-M1 to BEKK-HEAVY-M2	68.8	200.5	301.8	534.5
from FKO to DCC-HEAVY-M2	22.1	26.7	176.6	143.1
from FKO-HEAVY-M1 to BEKK C-HEAVY-M2	79.6	222.9	181.8	276.7
from DCC-HEAVY-M2 to BEKK-HEAVY-M2	66.4	157.8	-160.1	-59.4
from BEKK-HEAVY-M2 to FKO-HEAVY-M2	-30.3	31.4	-168.9	-226.4
from FKO-HEAVY-H2 to DCC-HEAVY-M2	-16.6	-143.5	281.6	176.6
s=22				
from DCC-GARCH to DCC-HEAVY-M2	61.5	116.1	188.4	257.9
from DCC-HEAVY-M1 to DCC-HEAVY-M2	15.4	159.9	17.8	150.0
from BEKK-GARCH to BEKK-HEAVY-M2	90.3	129.6	199.0	502.8
from BEKK-HEAVY-M1 to BEKK-HEAVY-M2	15.2	37.8	28.9	66.5
from FKO to DCC-HEAVY-M2	14.8	37.3	105.9	165.1
from FKO-HEAVY-M1 to BEKK C-HEAVY-M2	78.4	149.8	107.1	170.3
from DCC-HEAVY-M2 to BEKK-HEAVY-M2	128.8	403.6	73.6	138.9
from BEKK-HEAVY-M2 to FKO-HEAVY-M2	-46.6	-114.1	-82.3	-139.5
from FKO-HEAVY-H2 to DCC-HEAVY-M2	3.3	14.8	-107.0	-17.0

For explanations, see note below Table 5.

Appendices

A SGBF data description

The data are obtained from TickData, Inc. for the futures contracts S&P 500 futures (ES: CME GROUP), Treasury bond futures (US:CCBOT/CME GROUP), and gold futures (GC: COMEX/CME GROUP). The sample period is July 1, 2003 to August 5, 2022.

The trading close times (US central standard times) of the three futures contracts are as follows: 1) Gold futures contract closes at 13:30 between 01/07/2003 and 03/12/2006, and closes at 17:00 between 04/12/2006 and 05/08/2022; 2) Bond futures contract closes at 16:00 between 01/07/2003 and 05/08/2022; 3) Stock futuresA contract closes at 15:15 between 01/07/2003 and 17/11/2012, and closes at 16:00 between 18/11/2012 and 05/08/2022. Therefore, we assume that portfolios are rebalanced at 13:30 each day between 01/07/2003 and 03/12/2006, at 15:15 each day between 04/12/2006 and 17/11/2012, and at 16:00 each day between between 18/11/2012 and 05/08/2022. The last transaction prices before the chosen close time as the close price. This procedure is the same as in Fleming et al. (2003).

Table 9 – Descriptive statistics - SBGF data

	r^2			v			Mean Cov. & *Mean Cor.*		
	Mean	StDev	Max	Mean	StDev	Max	STOCK	BOND	GOLD
STOCK	3.544	14.931	453	3.571	9.240	205	*1*	*-0.271*	*0.022*
BOND	1.059	2.286	69.6	1.163	1.313	28.6	-0.599	*1*	*0.118*
GOLD	3.081	8.742	294	3.347	4.299	62.2	0.171	0.191	*1*

Column 1: asset type; Columns 2 and 5: time-series (ts) means; Columns 3 and 6: ts standard deviations; Columns 4 and 7: ts maximum; r^2: squared returns; v: realized variances; Columns 8-10: ts means of realized covariances and (in *italics*) realized correlations. The means, standard deviations and maximum are annualized values, in percentage, i.e., multiplied by 252 and by 100.

To compute the realized measures, we follow the method of Fleming et al. (2003). We use the five-minute returns from close time on day t-1 to close time on day t: We construct the realized variances using all of these returns. For the realized covariances, however, we can use only the returns that are contemporaneous across markets. For example, the returns after 1:30 pm in the stock and bond markets cannot be used to compute the realized covariances with gold because the gold market is closed. Since this reduces the number of available observations, we construct the realized covariances using the the two-step procedure of Fleming et al. (2003). First, we use the contemporaneous returns to compute a preliminary set of realized variances and covariances, from which we compute the realized correlations. Second, we convert these correlations back into covariances using the realized variances based on the entire set of five-minute returns.

Table 9 provides summary descriptive statistics of the data. The realized variances are computed based on 23 hours (approximately, depending on the years) of intra-daily returns, hence they should be close to the daily squared return, which can be observed from Table 9. However, one can notice that that average values of the realized variances (v) are slightly higher than the average of the daily squared returns. On the contrary, their standard deviations are much smaller, the reason is that large squared returns are very often more extreme than large realized variances. This can be seen on Figure 1, which shows the squared returns and realized variances of each asset during two periods: 2008, a year of low volatility until September and extreme volatility afterwards (the 'subprime mortgage' crisis), and 2021/08/06-2022/08/05, a period of low and intermediate volatility level.

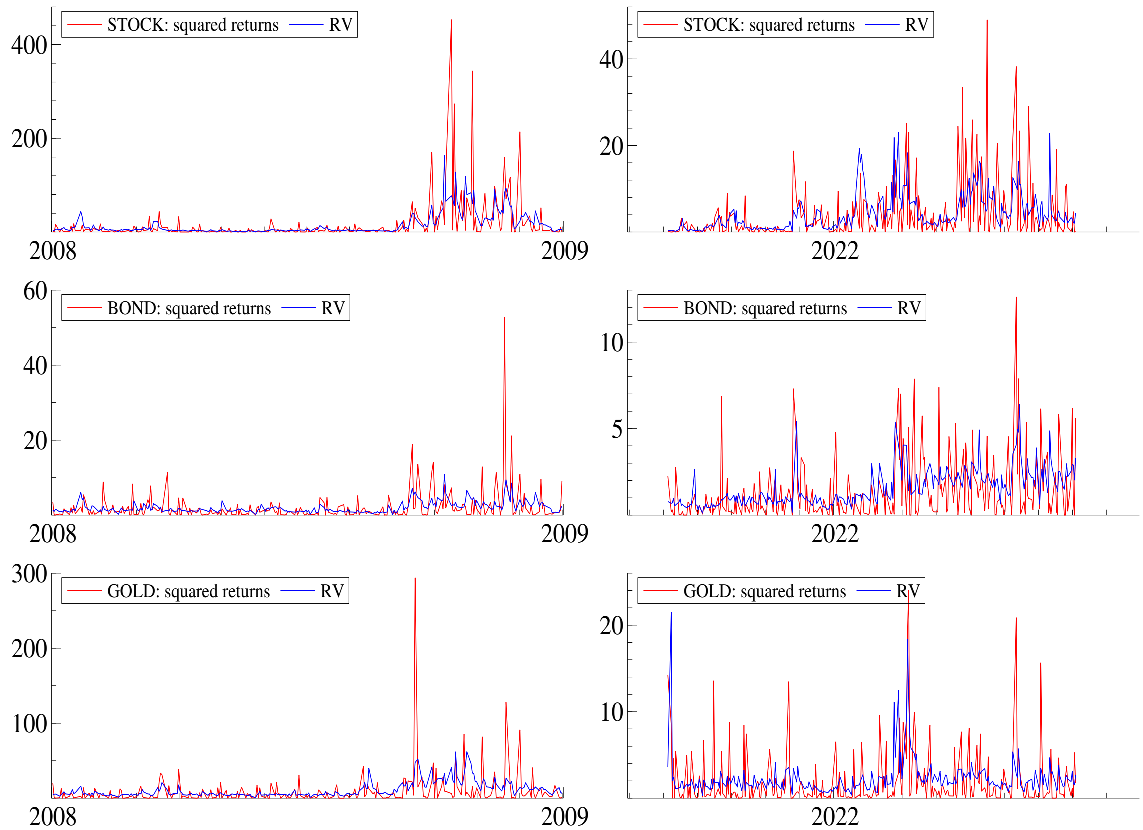

Figure 1 – Annualized realized variances (RV) and squared returns

B DJ29 data description

Table 10 – Descriptive statistics of realized variances - DJ29 data

	Means			Standard deviations			
Stock	$r^2, v1, v2$	v	von	r^2	v	$v1$	$v2$
AAPL	14.234	8.245	13.282	38.982	13.685	30.774	23.518
AXP	13.076	6.954	10.287	47.122	16.866	33.420	31.377
BA	8.840	4.778	7.224	25.816	7.509	25.357	14.339
CAT	10.185	5.533	8.659	26.481	9.073	19.289	17.456
CSCO	14.258	6.757	12.111	49.812	11.748	35.429	24.869
CVX	6.626	3.821	5.406	24.317	7.842	12.517	14.494
DIS	8.780	4.982	7.794	29.172	9.612	27.002	16.823
DWDP	12.212	4.472	8.055	39.113	7.083	29.715	19.246
GE	9.325	5.443	8.657	31.200	13.038	23.675	22.089
GS	13.384	7.237	11.184	53.465	24.863	49.495	42.384
HD	8.693	5.126	7.617	25.575	8.983	18.212	15.473
IBM	5.962	3.294	5.428	17.718	6.665	14.236	11.880
INTC	12.664	6.639	11.069	39.711	10.013	32.934	19.395
JNJ	3.312	2.200	3.311	15.043	4.162	13.106	6.472
JPM	16.069	8.037	12.266	65.176	20.037	38.165	39.965
KO	3.670	2.386	3.421	12.768	4.270	6.619	6.682
MCD	4.934	3.325	4.949	14.781	6.226	9.708	9.314
MMM	5.010	3.082	4.412	13.578	6.715	9.726	10.740
MRK	7.651	4.036	6.697	44.891	11.360	41.629	21.284
MSFT	8.476	4.386	7.077	26.000	6.698	18.103	13.108
NKE	8.030	4.239	6.666	29.404	6.486	23.422	12.436
PFE	5.966	3.700	6.003	19.551	5.469	16.303	9.377
PG	3.303	2.166	3.140	9.719	3.919	6.032	6.195
TRV	8.294	4.426	6.892	37.886	11.525	25.514	21.335
UNH	9.918	5.063	7.641	49.433	8.985	22.767	18.015
UTX	6.889	3.589	5.333	40.373	6.405	19.416	12.782
VZ	5.795	3.841	5.374	16.791	7.173	10.420	10.989
WMT	4.651	3.066	4.408	13.598	5.453	8.455	8.223
XOM	5.906	3.594	5.084	21.656	8.155	11.951	13.775
Min	3.303	2.166	3.140	9.719	3.919	6.032	6.195
Max	16.069	8.245	13.282	65.176	24.863	49.495	42.384
Med	8.294	4.386	6.892	26.481	7.842	19.416	14.494

Column 1: stock tickers. Columns 2-4: time-series averages; Columns 5-9: time-series standard deviations; r^2: squared returns; $v1$: rescaled realized variances augmented by overnight squared returns, see (4); $v2$: rescaled realized variances, see (5) ; v: realized variances, see (1); von: realized variances augmented by overnight squared returns, not rescaled, see (3). Rows "Min", "Max", and "Med" report the minimum, maximum, and median across the stocks. All values are annualized, in percentage.

Table 11 – Descriptive statistics of realized covariances - DJ29 data

	Means		Standard deviations		
Stock	RC	$RC1$ & $RC2$	RC	$RC1$	$RC2$
AAPL	2.135	3.775	4.589	9.113	8.202
AXP	2.239	5.139	5.573	13.015	12.448
BA	1.754	3.691	4.096	10.709	8.598
CAT	2.040	4.267	4.806	9.586	9.966
CSCO	2.156	4.454	4.535	9.753	9.283
CVX	1.625	3.314	4.638	8.566	9.357
DIS	1.806	4.095	4.241	10.782	9.402
DWDP	1.867	4.468	4.319	10.330	10.191
GE	2.061	4.272	5.096	10.816	10.374
GS	2.210	4.935	6.009	14.036	12.724
HD	1.934	3.759	4.649	9.486	8.966
IBM	1.631	3.167	3.963	7.066	7.545
INTC	2.200	4.450	4.321	9.261	8.733
JNJ	1.114	2.049	2.798	5.497	5.301
JPM	2.398	5.466	5.811	13.017	13.202
KO	1.184	2.073	3.020	5.175	5.405
MCD	1.295	2.204	3.354	5.800	5.846
MMM	1.592	3.162	3.755	6.958	7.459
MRK	1.433	2.726	3.618	7.020	7.106
MSFT	1.847	3.765	3.847	7.907	7.858
NKE	1.520	3.201	3.595	7.932	7.443
PFE	1.453	2.840	3.396	7.117	6.823
PG	1.130	1.988	2.909	5.177	5.294
TRV	1.544	3.670	4.191	11.009	9.499
UNH	1.422	2.949	4.020	8.846	8.422
UTX	1.658	3.738	4.091	9.888	8.876
VZ	1.515	2.775	3.975	6.823	7.222
WMT	1.379	2.359	3.205	5.676	5.544
XOM	1.636	3.231	4.635	8.274	9.007
Min	1.114	1.988	2.798	5.175	5.294
Max	2.398	5.466	6.009	14.036	13.202
Med	1.636	3.670	4.096	8.846	8.598

Column 1: stock tickers. Columns 2-4: means of the time series averages of the realized covariances of each stock with the other 28 stocks; Columns 5-7: standard deviations of the time series averages of the realized covariances of each stock with the other 28 stocks; RC: realized covariances, see (1); $RC1$: rescaled realized covariances augmented by overnight squared returns, see (4); $RC2$: rescaled realized covariances, see (5). Rows "Min", "Max", and "Med" report the minimum, maximum, and median across the stocks. All values are annualized, in percentage.

Table 12 – Descriptive statistics of realized correlations - DJ29 data

	Means			Standard deviations		
Stock	$RCor$	$RCor1$	$RCor2$	$RCor$	$RCor1$	$RCor2$
AAPL	0.326	0.349	0.326	0.152	0.213	0.152
AXP	0.447	0.462	0.447	0.145	0.183	0.145
BA	0.378	0.396	0.378	0.159	0.203	0.159
CAT	0.398	0.422	0.398	0.156	0.198	0.156
CSCO	0.367	0.405	0.367	0.148	0.192	0.148
CVX	0.377	0.393	0.377	0.174	0.217	0.174
DIS	0.422	0.447	0.422	0.160	0.196	0.160
DWDP	0.385	0.416	0.385	0.157	0.204	0.157
GE	0.426	0.441	0.426	0.150	0.189	0.150
GS	0.418	0.425	0.418	0.143	0.184	0.143
HD	0.375	0.399	0.375	0.156	0.198	0.156
IBM	0.382	0.418	0.382	0.154	0.193	0.154
INTC	0.382	0.413	0.382	0.146	0.188	0.146
JNJ	0.349	0.380	0.349	0.162	0.206	0.162
JPM	0.429	0.438	0.429	0.143	0.182	0.143
KO	0.318	0.350	0.318	0.160	0.203	0.160
MCD	0.303	0.330	0.303	0.163	0.210	0.163
MMM	0.435	0.460	0.435	0.160	0.194	0.160
MRK	0.304	0.347	0.304	0.162	0.210	0.162
MSFT	0.382	0.419	0.382	0.149	0.188	0.149
NKE	0.336	0.370	0.336	0.162	0.204	0.162
PFE	0.347	0.386	0.347	0.160	0.204	0.160
PG	0.324	0.354	0.324	0.160	0.203	0.160
TRV	0.397	0.413	0.397	0.159	0.200	0.159
UNH	0.284	0.313	0.284	0.162	0.212	0.162
UTX	0.435	0.453	0.435	0.155	0.191	0.155
VZ	0.326	0.351	0.326	0.164	0.205	0.164
WMT	0.317	0.352	0.317	0.159	0.196	0.159
XOM	0.395	0.413	0.395	0.170	0.210	0.170
Min	0.284	0.313	0.284	0.143	0.182	0.143
Max	0.447	0.462	0.447	0.174	0.217	0.174
Med	0.378	0.405	0.378	0.159	0.200	0.159

Column 1: stock tickers. Columns 2-4: means of the time series averages of the realized correlations of each stock with the other 28 stocks;
Columns 5-7: standard deviations of the time series averages of the realized correlations of each stock with the other 28 stocks; $RCor$: realized correlations (2) from realized covariances (1); $RCor1$: correlations from rescaled realized covariances (4); $RCor2$: correlations from rescaled realized covariances (5). Rows "Min", "Max", and "Mean" report the minimum, maximum, and mean across the stocks. .

C Quasi-maximum likelihood estimation

To define a quasi-likelihood function for the GARCH and HEAVY-H models, we add the assumption that the conditional distribution of the return vector is multivariate Gaussian, $N_k(0, V_t)$, where V_t is G_t for a GARCH equation and H_t for a HEAVY-H equation. The quasi-log-likelihood function for T observations, given initial values, is

$$QLN_V(\theta_V) = -\frac{1}{2}\sum_{t=1}^{T}\left(\log|V_t| + r_t' V_t^{-1} r_t\right), \tag{C1}$$

where θ_V is the parameter vector of the corresponding model. This function can be maximized numerically with respect to θ_V for the BEKK and FKO models. For the DCC models, a two-step estimation procedure is used, whereby the parameters of the variance equations are estimated in the first step, and the parameters of the correlation equation are estimated in the second step, see Engle (2002).

For the HEAVY-M equations, we assume that the conditional distribution of RC_t is a central Wishart distribution of dimension k , $W_k(\nu, M_t/\nu)$, where ν is the degrees of freedom parameter (restricted by $\nu > k-1$) and M_t is the conditional mean of RC_t. The quasi-log-likelihood function for a sample of T observations, given initial conditions, is

$$QLW_M(\theta_M) = -\frac{\nu}{2}\sum_{t=1}^{T}\left[\log|M_t| + \mathrm{tr}(M_t^{-1} RC_t)\right] \tag{C2}$$

This function can be maximized numerically with respect to θ_M (independently of the value of ν, which can be set equal to one) in a single step for the BEKK and FKO models. For the DCC models, a two-step estimation procedure is used, similarly to the two-step procedure for the Gaussian – see Bauwens and Xu (2023). For the asymptotic properties of quasi-likelihood estimators, see Taylor and Xu (2017).

D Estimation results on full samples

Table 13 – Parameter estimates (and robust t-statistics) for SBGF data

	GARCH		HEAVY-H		HEAVY-M	
DCC	Conditional Variance					
	A_G	B_G	A_H	B_H	A_M	B_M
STOCK	0.140	0.833	0.606	0.379	0.601	0.384
BOND	0.044	0.946	0.110	0.876	0.235	0.728
GOLD	0.043	0.949	0.154	0.813	0.304	0.673
	Conditional Correlation					
	α_Q	β_Q	α_R	β_R	α_P	β_P
	0.033	0.949	0.435	0.564	0.222	0.758
	(6.32)	(99.14)	(6.78)	(8,80)	(24.53)	(73.00)
BEKK	α_G	β_G	α_H	β_H	α_M	β_M
	0.057	0.934	0.272	0.684	0.372	0.606
	(11.90)	(130.49)	(8.38)	(17.42)	(36.46)	(13.26)
FKO	α_G		α_H		α_M	
	0.034		0.253		0.186	
	(10.52)		(13.57)		(26.22)	

Table 14 – Parameter estimates and robust t-statistics) for DJ29 data

	GARCH		HEAVY-M1		HEAVY-M2		HEAVY-H	
DCC	Conditional Variance							
	A_G	B_G	A_M	B_M	A_M	B_M	A_H	B_H
Min	0.025	0.826	0.056	0.381	0.296	0.508	0.174	0.139
Med	0.068	0.918	0.233	0.738	0.387	0.597	0.401	0.568
Max	0.123	0.970	0.538	0.939	0.479	0.683	0.794	0.788
	Conditional Correlation							
	α_Q	β_Q	α_P	β_P	α_P	β_P	α_R	β_R
	0.003	0.988	0.019	0.976	0.047	0.942	0.089	0.866
	(9.30)	(475.08)	(25.68)	(881.2)	(29.79)	(439.64)	(10.74)	(53.78)
BEKK	α_G	β_G	α_M	β_M	α_M	β_M	α_H	β_H
	0.009	0.988	0.033	0.961	0.134	0.856	0.191	0.755
	(25.96)	(1893.80)	(6.83)	(154.05)	(137.42)	(30.21)	(14.02)	(37.81)
FKO	α_G		α_M		α_M		α_H	
	0.0055		0.018		0.086		0.038	
	(17.05)		(14.71)		(23.81)		(6.71)	

For DCC Conditional Variance: Min is the minimum of the 29 estimates, Med the median, and Max the maximum.

E Solving equation (11)

We solve (11) after dividing both sides by T_s. For lighter notations, we use the symbol Y for Δ_γ , we set $a = 1$, $b = 2$, we denote by U_1 the utility function divided by T_s on the left side of (11), and by $U_2(Y)$ the function on the right side divided by T_s. So, we must find Y such that $U_1 = U_2(Y)$. By direct developments, we get $U_1 = 1 - A + (1 - 2A)S_1 - AV_1$, and

$U_2 = 1 - A + (1-2A)S_2 - AV_2 + (2A-1+2AS_2)Y - AY^2$, where $A = 0.5\gamma/(1+\gamma)$, $S_i = \sum_{t=1}^{T_s} r_{it}$, and $V_i = \sum_{t=1}^{T_s} r_{it}^2$, for $i = 1, 2$, r_{it} being the optimal portfolio return of period t for model i, i.e., r_{it} stands for $r_{wt}^{(i)}$ of (11).

By subtracting $U_2(Y)$ from U_1 using the expressions above, we get the quadratic equation $AY^2 + BY + C = 0$, where A is defined above, $B = 1 - 2A(1 + S_2)$, and $C = (1 - 2A)(S_1 - S_2) + A(V_2 - V_1)$. This equation has two real roots if $B^2 - 4AC \geq 0$. The solution of interest as measure of economic gain (the maximum fee an investor would sacrifice to switch from model 1 to model 2) is the smallest root, if it is positive. If the smallest root is negative, its opposite is the gain of switching from model 2 to 1.