KU LEUVEN

DEPARTMENT OF ECONOMICS

DISCUSSION PAPER SERIES
DPS16.12

JUNE 2016

Mixed-frequency multivariate GARCH

Geert DHAENE and Jianbin WU

Econometrics

Faculty of Economics
and Business

MIXED-FREQUENCY MULTIVARIATE GARCH

Geert Dhaene

KU Leuven, Department of Economics, Naamsestraat 69, B-3000 Leuven, Belgium
geert.dhaene@kuleuven.be

Jianbin Wu

KU Leuven, Department of Economics, Naamsestraat 69, B-3000 Leuven, Belgium
jianbin.wu@kuleuven.be

June 15, 2016

We introduce and evaluate mixed-frequency multivariate GARCH models for forecasting low-frequency (weekly or monthly) multivariate volatility based on high-frequency intra-day returns (at five-minute intervals) and on the overnight returns. The low-frequency conditional volatility matrix is modelled as a weighted sum of an intra-day and an overnight component, driven by the intra-day and the overnight returns, respectively. The components are specified as multivariate GARCH(1,1) models of the BEKK type, adapted to the mixed-frequency data setting. For the intra-day component, the squared high-frequency returns enter the GARCH model through a parametrically specified mixed-data sampling (MIDAS) weight function or through the sum of the intra-day realized volatilities. For the overnight component, the squared overnight returns enter the model with equal weights. Alternatively, the low-frequency conditional volatility matrix may be modelled as a single-component BEKK-GARCH model where the overnight returns and the high-frequency returns enter through the weekly realized volatility (defined as the unweighted sum of squares of overnight and high-frequency returns), or where the overnight returns are simply ignored. All model variants may further be extended by allowing for a non-parametrically estimated slowly-varying long-run volatility matrix. The proposed models are evaluated using five-minute and overnight return data on four DJIA stocks (AXP, GE, HD, and IBM) from January 1988 to November 2014. The focus is on forecasting weekly volatilities (defined as the low frequency). The mixed-frequency GARCH models are found to systematically dominate the low-frequency GARCH model in terms of in-sample fit and out-of-sample forecasting accuracy. They also exhibit much lower low-frequency volatility persistence than the low-frequency GARCH model. Among the mixed-frequency models, the low-frequency persistence estimates decrease as the data frequency increases from daily to five-minute frequency, and as overnight returns are included. That is, ignoring the available high-frequency information leads to spuriously high volatility persistence. Among the other findings are that the single-component model variants perform worse than the two-component variants; that the overnight volatility component exhibits more persistence than the intra-day component; and that MIDAS weighting performs better than not weighting at all (i.e., than realized volatility).

Keywords: *multivariate GARCH, mixed-frequency sampling, overnight returns.*

1. INTRODUCTION

GARCH models are widely used to model conditional variances and covariances of asset returns. The availability of high-frequency financial data in recent decades opened up possibilities for more accurate estimation of return volatilities. At the same time, new challenges for GARCH modeling arose since the variance forecast horizons of interest are often at a lower frequency, e.g., daily, weekly, or monthly, than the frequency at which financial returns are observed. Standard GARCH models are single-frequency models. How to efficiently use high-frequency data to forecast lower-frequency covariance matrices in GARCH-type models is of great interest.

A useful tool for this type of mixed-frequency problem is the mixed-data sampling (MIDAS) approach introduced by Ghysels, Santa-Clara, and Valkanov (2005, 2006). In the univariate case, Chen, Ghysels, and Wang (2015) extended the regression-based method of Ghysels, Santa-Clara, and Valkanov (2006) to GARCH-type models, proposing a HYBRID GARCH process that allows to forecast volatilities at different frequencies than the frequency of the information set. In the multivariate GARCH setting, the literature about mixed sampling frequencies is relatively scarce. As far as we know, the only existing work is Colacito, Engle, and Ghysels (2011) and Golosnoy, Gribisch, and Liesenfeld (2012), where the MIDAS filter has been used to extract long-run variances.

In this paper, we propose multivariate versions of mixed-frequency GARCH models. In particular, we focus on weekly conditional variance forecasts using 5-minute returns. One inevitable issue with this type of weekly variance forecast is how to use the overnight information effectively. Simply treating the overnight return as one extra 5-minute return might not work well, since asst prices are likely to evolve in different ways during trading and non-trading hours; see, e.g., Blanc, Chicheportiche, and Bouchaud (2014) and Ahoniemi and Lanne (2013). To the best of our knowledge, the overnight returns have not been considered in previous work on MIDAS GARCH models, and also has it only rarely been considered in the GARCH literature in general.

In the literature on realized volatility, there are several related approaches of incorporating the overnight returns. The simplest one is to add the squared overnight return to the intra-day volatility as if the overnight return were an extra 5-minute return; see, e.g., Ahoniemi and Lanne (2013) and the references therein. Another approach, called the scaling estimator, is to scale the intra-day realized volatility estimator to a measure of volatility for the whole day, as discussed in Martens (2002) and Hansen and Lunde (2005). Still another approach is to combine the overnight and intra-day volatilities with optimally chosen weight parameters, as proposed by Hansen and Lunde (2005). In this paper, we explore these various approaches in a mixed-frequency multivariate GARCH framework, and compare them empirically.

Section 2 proposes the mixed-frequency GARCH models: one-component, two-component, and local stationary two-component models. Section 3 evaluates the models in and out of sample using return data from 1998 to 2014 on four DJIA stocks: AXP, GE, HD, and IBM. Section 4 concludes.

2. MIXED-FREQUENCY MULTIVARIATE GARCH MODELS

Let r_t be the vector of log returns on n assets between Friday's close of week $t-1$ and that of week t. Also, let $\mathcal{I}_{t-1}$ be the information set at Friday's close of week $t-1$ and define

$$H_t = \mathrm{Var}(r_t|\mathcal{I}_{t-1}). \tag{2.1}$$

Our purpose is to predict H_t, the weekly conditional variance matrix of r_t, using a multivariate GARCH framework that incorporates the 5-minute intra-day returns and the overnight returns prior to week t. Since we work with returns in log form, the weekly returns are additively composed as

$$r_t = \sum_{d=1}^{5} r_{td}, \qquad r_{td} = r_{td}^{\mathrm{co}} + r_{td}^{\mathrm{oc}}, \qquad r_{td}^{\mathrm{oc}} = \sum_{i=1}^{78} r_{tdi}^{\mathrm{5m}}, \tag{2.2}$$

where r_{td}, r_{td}^{co}, and r_{td}^{oc} are the close-to-close, close-to-open (overnight), and open-to-close returns on day d of week t and $r_{tdi}^{5\text{m}}$ is the ith 5-minute return on day d of week t. We shall use the set of variables

$$r_{\tau d}^{\text{co}}, r_{\tau di}^{5\text{m}}, \qquad \tau \leq t-1; d = 1, \ldots, 5; i = 1, \ldots, 78; \tag{2.3}$$

all belonging to $\mathcal{I}_{t-1}$, as predictors of H_t. For simplicity, and given that the weekly conditional mean returns are negligible compared to their volatilities, we assume that $E(r_t|\mathcal{I}_{t-1}) = 0$. Alternatively, we can allow for nonzero but constant conditional mean returns by demeaning the returns at each frequency prior to the analysis. *Doing so gave nearly identical empirical results, so we omit these.*

The models developed below are all models of H_t. They vary in the level of detail in which the intra-day or intra-week returns are used and weighted, the way in which the overnight returns enter the model (as a separate component or not), and the specification of the unconditional variance (constant or slowly varying over time). We group the models under four different headings: two-component models, one-component models, locally stationary models, and garch models.

2.1. Two-component models

As documented in Blanc, Chicheportiche, and Bouchaud (2014), the intra-day and overnight returns behave very differently. One possibility, therefore, is to specify H_t as a weighted sum of an intra-day component P_t and an overnight component Q_t,

$$H_t = \lambda_1 P_t + \lambda_2 Q_t, \tag{2.4}$$

where λ_1 and λ_2 are scalar weight parameters and P_t and Q_t are modeled as separate processes. We specify P_t as a diagonal BEKK model (Engle and Kroner 1995) extended with the MIDAS approach (Ghysels, Santa-Clara, and Valkanov 2005, 2006) to incorporate the variation in the 390 5-minute intra-day returns in week $t-1$:

$$P_t = (I_n - aa' - bb') \odot \overline{P} + aa' \odot \sum_{d,i} \omega_{di} r_{t-1,di}^{5\text{m}} r_{t-1,di}^{5\text{m}\prime} + bb' \odot P_{t-1}. \tag{2.5}$$

Here, $\overline{P} = E(P_t) = E\left(\sum_d r_{td}^{\text{oc}}\right)$, a and b are parameter vectors, $\odot$ is the Hadamard product, and ω_{di} is a weight function with average value, across d and i, equal to 1. We specify exponential weights,

$$\omega_{di} = \frac{(78(d-1)+i)^{\gamma}}{\sum_{d,i}(78(d-1)+i)^{\gamma}/390}, \tag{2.6}$$

as in Engle, Ghysels, and Sohn (2008). We leave the sign of γ unrestricted, although it is natural to expect $\gamma > 0$, giving more weight to the more recent 5-minute periods; $\gamma = 0$ gives equal weights, $\omega_{di} = 1$. For the overnight component we adopt a relatively parsimonious specification,

$$Q_t = (1 - \alpha^2 - \beta^2)\overline{Q} + \alpha^2 \sum_d r_{t-1,d}^{co} r_{t-1,d}^{co\prime} + \beta^2 Q_{t-1}, \tag{2.7}$$

where $\overline{Q} = E(Q_t) = E\left(\sum_d r_{td}^{\text{co}}\right)$ and α and β are scalar parameters. We shall refer to the model (2.4)–(2.7) as the 2comp(co,5m) model, as it makes use of the overnight returns (close-to-open) and the 5-minute intra-day returns.

If we ignore the information in the 5-minute returns and replace (2.5) by

$$P_t = (I_n - aa' - bb') \odot \overline{P} + aa' \odot \sum_d r^{\text{oc}}_{t-1,d} r^{\text{oc}\prime}_{t-1,d} + bb' \odot P_{t-1} \tag{2.8}$$

(where each weekday is given equal weight), we obtain the 2comp(co,oc) model, which uses the close-to-open and open-to-close returns only.

2.2. One-component models

Alternatively, and somewhat more in line with the GARCH literature, we also consider a one-component model, with basic specification

$$H_t = (I_n - aa' - bb') \odot \overline{H} + aa' \odot W_{t-1} + bb' \odot H_{t-1}, \tag{2.9}$$

where $\overline{H} = E(H_t) = E(W_t)$ and W_{t-1} is a combination of the 5-minute intra-day and the overnight squared returns,

$$W_{t-1} = \lambda_1 \sum_{d,i} \omega_{di} r^{5\text{m}}_{t-1,di} r^{5\text{m}\prime}_{t-1,di} + \lambda_2 \sum_d r^{\text{co}}_{t-1,d} r^{\text{co}\prime}_{t-1,d}. \tag{2.10}$$

We refer to the model (2.9)–(2.10) as the 1comp(co,5m) model.

By analogy with the two-component model, replacing (2.10) with

$$W_{t-1} = \lambda_1 \sum_d r^{\text{oc}}_{t-1,d} r^{\text{oc}\prime}_{t-1,d} + \lambda_2 \sum_d r^{\text{co}}_{t-1,d} r^{\text{co}\prime}_{t-1,d} \tag{2.11}$$

gives the 1comp(co,oc) model.

Note, further, that ignoring the information in the overnight returns amounts to setting $\lambda_2 = 0$, which makes the one- and two-component models coincide. For example, with $\lambda_2 = 0$, the models 2comp(co,5m) and 1comp(co,5m) reduce to the same model, which we label 1comp(5m).

2.3. Locally stationary models

The empirical realism of global stationarity as implied by the standard GARCH approach to volatility modeling has been called into question. Several authors have suggested ways of extending GARCH models with a slowly varying variance component, giving rise to locally stationary GARCH processes. Engle and Rangel (2008) specified the slowly varying component by linking it to macro-economic variables using splines. Engle, Ghysels, and Sohn (2008) specified it using a parametric MIDAS filter. Hafner and Linton (2010) proposed a nonparametric specification.

Here, we follow the approach of Hafner and Linton (2010) to obtain locally stationary versions of the two-component models developed above. Consider the main two-component equation, (2.4), with the intra-day component and the overnight component now specified as

$$P_t = M_t^{1/2} P_t^* M_t^{1/2}, \tag{2.12}$$

$$Q_t = N_t^{1/2} Q_t^* N_t^{1/2}, \tag{2.13}$$

where M_t and N_t are nonparametrically specified unconditional variance matrices that are slowly varying as a function of t and P_t^* and Q_t^* are short-run variance matrices with the property that $E(P_t^*) = E(Q_t^*) = I_n$. As in the stationary two-component model, we specify P_t^*

and Q_t^* as stationary BEKK processes with a MIDAS extension,

$$P_t^* = (I_n - aa' - bb') \odot I_n + aa' \odot M_{t-1}^{-1/2} \sum_{d,i} \omega_{di} r_{t-1,di}^{\text{5m}} r_{t-1,di}^{\text{5m}\prime} M_{t-1}^{-1/2} + bb' \odot P_{t-1}^*, \tag{2.14}$$

$$Q_t^* = (1 - \alpha^2 - \beta^2) I_n + \alpha^2 N_{t-1}^{-1/2} \sum_{d} r_{t-1,d}^{\text{co}} r_{t-1,d}^{\text{co}\prime} N_{t-1}^{-1/2} + \beta^2 Q_{t-1}^*. \tag{2.15}$$

We refer to the model (2.4) and (2.12)–(2.15) as ls2comp(co,5m). If M_t and N_t are constant in t, the model reduces to 2comp(co,5m).

By comparison, the locally stationary model of Hafner and Linton (2010) to predict H_t uses weekly returns only and is specified as

$$H_t = L_t^{1/2} H_t^* L_t^{1/2}, \tag{2.16}$$

$$H_t^* = (I_n - aa' - bb') \odot I_n + aa' \odot L_{t-1}^{-1/2} r_{t-1} r_{t-1}' L_{t-1}^{-1/2} + bb' \odot H_{t-1}^*, \tag{2.17}$$

where L_t is the slowly varying unconditional variance matrix of r_t. We refer to this model as lsgarch(w).

2.4. GARCH models

As benchmarks against which to compare the models that use intra-day information, we consider the diagonal BEKK model for weekly returns,

$$H_t = (I_n - aa' - bb') \odot \overline{H} + aa' \odot r_{t-1} r_{t-1}' + bb' \odot H_{t-1}, \tag{2.18}$$

which we label garch(w), and its extension that incorporates the equally-weighted daily returns,

$$H_t = (I_n - aa' - bb') \odot \overline{H} + aa' \odot \sum_d r_{t-1,d} r_{t-1,d}' + bb' \odot H_{t-1}, \tag{2.19}$$

which we label garch(d).

3. ESTIMATION

It is already known since the seminal work of Engle (1982) that (G)ARCH processes generate heavier tails than the normal density. To further account for the heavy tails of asset returns, we model the standardized weekly returns, $H_t^{-1/2} r_t$, as independent standard multivariate Student t-distributed variates with $\nu > 2$ degrees of freedom. Then the density of r_t, given H_t, is

$$\frac{\Gamma(\frac{\nu+n}{2})}{\Gamma(\frac{\nu}{2})(\pi(\nu-2))^{n/2} |H_t|^{1/2}} \left(1 + \frac{r_t' H_t^{-1} r_t}{\nu - 2}\right)^{-\frac{\nu+n}{2}}. \tag{3.1}$$

Each model is now defined by the parametric density (3.1) and the specification of H_t. The log-likelihood corresponding to the series of weekly returns $r_1, \ldots, r_T$, sequentially conditioned on $H_1, \ldots, H_T$, is $L = \sum_{t=1}^T l_t$ where

$$l_t = \log \Gamma\left(\frac{\nu+n}{2}\right) - \log \Gamma\left(\frac{\nu}{2}\right) - \frac{n}{2} \log(\pi(\nu-2)) - \frac{1}{2} \log |H_t| - \frac{\nu+n}{2} \log\left(1 + \frac{r_t' H_t^{-1} r_t}{\nu-2}\right).$$

We estimate the models in two steps. In step 1, we profile out the unconditional variance matrices from L. For the stationary models, we use variance targeting, setting

$$\overline{Q} = T^{-1} \sum\nolimits_{t,d} r_{td}^{\text{co}} r_{td}^{\text{co}\prime},$$

$$\overline{P} = \begin{cases} T^{-1} \sum_{t,d,i} \omega_{di} r_{tdi}^{\text{5m}} r_{tdi}^{\text{5m}\prime} & \text{in 2comp(co,5m)}, \\ T^{-1} \sum_{t,d} r_{td}^{\text{oc}} r_{td}^{\text{oc}\prime} & \text{in 2comp(co,oc)}, \end{cases}$$

$$\overline{H} = \begin{cases} T^{-1} \left(\lambda_1 \sum_{t,d,i} \omega_{di} r_{tdi}^{\text{5m}} r_{tdi}^{\text{5m}\prime} + \lambda_2 \sum_{t,d} r_{td}^{\text{oc}} r_{td}^{\text{oc}\prime}\right) & \text{in 1comp(co,5m)}, \\ T^{-1} \left(\lambda_1 \sum_{t,d} r_{td}^{\text{co}} r_{td}^{\text{co}\prime} + \lambda_2 \sum_{t,d} r_{td}^{\text{oc}} r_{td}^{\text{oc}\prime}\right) & \text{in 1comp(co,oc)}, \\ T^{-1} \sum_{t,d} r_{td} r_{td}' & \text{in garch(d)}, \\ T^{-1} \sum_{t} r_{t} r_{t}' & \text{in garch(w)}. \end{cases}$$

For the locally stationary models, we use the nonparametric smoothing estimator of Rodríguez-Poo and Linton (2001), setting

$$M_t = \frac{\sum_{\tau=1}^{T} K_h(t-\tau) \sum_{d,i} \omega_{\tau i} r_{\tau di}^{\text{5m}} r_{\tau di}^{\text{5m}\prime}}{\sum_{\tau=1}^{T} K_h(t-\tau)},$$

$$N_t = \frac{\sum_{\tau=1}^{T} K_h(t-\tau) \sum_{d} r_{\tau d}^{\text{co}} r_{\tau d}^{\text{co}\prime}}{\sum_{\tau=1}^{T} K_h(t-\tau)},$$

$$L_t = \frac{\sum_{\tau=1}^{T} K_h(t-\tau) r_\tau r_\tau'}{\sum_{\tau=1}^{T} K_h(t-\tau)},$$

where $K_h(\cdot) = K(\cdot/h)/h$, $K(\cdot)$ is a kernel function, and h is a bandwidth parameter. We use a quartic kernel (truncated near the boundaries) with $h = 0.15$. In step 2, we estimate the other parameters, $a, b, \alpha, \beta, \lambda_1, \lambda_2$, and γ, by maximizing the profiled log-likelihood obtained from step 1.

4. APPLICATION TO FOUR DJIA STOCKS, 1998-2014

We applied the models presented above using the 5-minute intra-day and overnight returns between January 2, 1998, and November 28, 2014 (Friday's close to Friday's close) to predict the weekly return volatility of four DJIA stocks: American Express (AXP), General Electric (GE), Home Depot (HD), and International Business Machines (IBM). There are 15 model variants in total. Our selection of stocks is the same as in the study of Golosnoy, Gribisch, and Liesenfeld (2012) for the period 2000 to 2008, except that they also included Citigroup. We excluded Citigroup because it was removed from the DJIA in 2009 after a price drop of around 90%.

4.1. In-sample evaluation

Table 1 reports the model estimates using the entire sample. A very clear picture emerges on how the coefficients a_i and b_i evolve as higher-frequency returns are introduced in the models. In the classic GARCH model for weekly returns, garch(w), the estimated GARCH coefficients b_i are around 0.97 to 0.98. These estimates decrease to about 0.95 in the garch(d) model, where daily returns are incorporated as volatility predictors. They further decrease in the models that use 5-minute returns, to around 0.9 in the 1comp models, to 0.5 to 0.8 in the 2comp models, and even to 0.3 to 0.7 in the ls2comp models. The substantial decrease of the b_i estimates, as higher-frequency returns are incorporated, is compensated by a corresponding increase of the ARCH coefficients a_i from around 0.2 in the garch(w) model to 0.3 in the garch(d) model, 0.3 to 0.5 in the 1comp models, 0.3 to 0.8 in the 2comp models, and 0.6 to above 0.9 in the ls2comp models. Hence, while all estimated models exhibit nearly integrated volatility (with $a_i^2 + b_i^2$ being estimated close to unity), much more weight is given to information in the most recent week if that information is more detailed, i.e., when it consist of high-frequency returns. This is most pronounced in the locally stationary models. There is a corresponding clear-cut tendency in the estimated degree-of-freedom parameter, ν, which varies from around 5 in the garch(w) model to around 8 or 9 in the models with 5-minute returns. Hence, when viewed through the sharper lens of the high-frequency returns, the weekly returns appear somewhat less fat-tailed.

Another observation, in line with Blanc, Chicheportiche, and Bouchaud (2014), is that the dynamics of the overnight and intra-day components are very different from each other. In the 2comp(co,oc) model, which treats the close-open and open-close returns nearly symmetrically, the estimated intra-day coefficients a_i and b_i are on the order of 0.3 to 0.4 and 0.9 to 0.95, respectively, while the corresponding overnight coefficients α and β are estimated as 0.07 and 0.99, respectively. This suggests that the overnight component decays more slowly than the intra-day component, a property that is also borne out by the other 2comp and ls2comp models, although there is a natural tendency for the intra-day decay rate to go up as more refined intra-day return information is used.

In the 2comp and ls2comp models, the weights given to the intra-day and overnight components, λ_1 and λ_2, are both estimated to be around 1, with those for λ_2 being somewhat bigger than for λ_1, and those for the 2comp models being somewhat bigger than for the ls2comp models. In the 1comp models, the estimates of λ_1 and λ_2 are quite different, with λ_1 systematically exceeding 1 and λ_2 being close to 0. Because of the much lesser weight given to the overnight returns, the estimated 1comp(co,5m) and 1comp(5m) models are almost identical. This is rather unexpected and, in particular, difficult to reconcile with the 2comp model estimates, even though the 1comp and 2comp models are very different in the way the intra-day and overnight returns enter the model.

The estimates of γ range between 1.35 and 1.7, suggesting that the most recent 5-minute returns (near to Friday's close) have considerably more predictive power for next week's volatility than the earlier 5-minute returns. For example, when $\gamma = 1.5$, the aggregate relative weights of Monday to Friday's 5-minute returns are 0.02, 0.08, 0.18, 0.29, and 0.43.

Table 1. Estimates of mixed-frequency BEKK-type models of weekly returns for four DJIA stocks

											locally stationary				
	2comp			1comp							ls2comp		lsgarch	garch	garch
	(co,5m)		(co,oc)	(co,5m)			(5m)		(co,oc)		(co,5m)		(w)	(d)	(w)
ω_{di}	expon	equal	–	expon	equal	equal	expon	equal	–	–	expon	equal	–	–	–
a_1	.657	.671	.365	.443	.437	.383	.445	.441	.335	.322	.815	.893	.196	.304	.207
	(.087)	(.095)	(.032)	(.062)	(.066)	(.039)	(.061)	(.065)	(.031)	(.003)	(.087)	(.089)	(.034)	(.027)	(.002)
a_2	.849	.841	.421	.411	.429	.324	.414	.434	.308	.276	.909	.880	.223	.269	.174
	(.100)	(.089)	(.074)	(.057)	(.068)	(.030)	(.056)	(.065)	(.038)	(.002)	(.101)	(.118)	(.085)	(.028)	(.002)
a_3	.695	.656	.389	.472	.450	.379	.476	.454	.336	.322	.689	.650	.184	.302	.204
	(.085)	(.080)	(.041)	(.065)	(.061)	(.035)	(.064)	(.059)	(.034)	(.003)	(.101)	(.101)	(.046)	(.029)	(.002)
a_4	.536	.558	.305	.370	.370	.307	.372	.373	.291	.295	.919	.944	.191	.283	.209
	(.091)	(.098)	(.028)	(.048)	(.049)	(.029)	(.047)	(.048)	(.023)	(.003)	(.082)	(.087)	(.036)	(.023)	(.003)
b_1	.739	.732	.926	.887	.886	.921	.886	.884	.937	.942	.528	.366	.963	.948	.974
	(.191)	(.195)	(.091)	(.121)	(.119)	(.098)	(.122)	(.121)	(.079)	(.074)	(.207)	(.201)	(.037)	(.067)	(.035)
b_2	.506	.541	.892	.905	.897	.940	.903	.894	.939	.955	.291	.340	.760	.957	.981
	(.169)	(.207)	(.114)	(.109)	(.116)	(.075)	(.110)	(.118)	(.070)	(.058)	(.160)	(.228)	(.195)	(.055)	(.026)
b_3	.719	.755	.908	.880	.882	.922	.878	.880	.931	.938	.659	.720	.945	.943	.972
	(.205)	(.196)	(.101)	(.132)	(.124)	(.096)	(.133)	(.125)	(.080)	(.074)	(.201)	(.190)	(.051)	(.067)	(.034)
b_4	.800	.799	.949	.905	.905	.940	.904	.903	.953	.952	.395	.329	.974	.956	.974
	(.162)	(.167)	(.068)	(.095)	(.095)	(.069)	(.096)	(.096)	(.063)	(.064)	(.174)	(.151)	(.032)	(.060)	(.035)
α	.062	.053	.068								.093	.085			
	(.025)	(.017)	(.040)								(.047)	(.052)			
β	.995	.995	.992								.975	.976			
	(.003)	(.002)	(.006)								(.023)	(.028)			
λ_1	.948	.897	1.053	1.340	1.328	1*	1.348	1.337	1.195	1*	.752	.740			
	(.083)	(.074)	(.090)	(.112)	(.109)		(.103)	(.100)	(.082)		(.074)	(.073)			
λ_2	1.175	1.242	1.274	.019	.021	1*	0*	0*	.269	1*	.987	.949			
	(.269	(.273)	(.403)	(.095)	(.100)				(.134)		(.163)	(.163)			

Continued on the next page.

Table 1. Estimates of mixed-frequency BEKK-type models of weekly returns for four DJIA stocks (cont.)

											locally stationary				
	2comp			1comp							ls2comp		lsgarch	garch	garch
	(co,5m)		(co,oc)	(co,5m)			(5m)		(co,oc)		(co,5m)		(w)	(d)	(w)
ω_{di}	expon	equal	–	expon	equal	equal	expon	equal	–	–	expon	equal	–	–	–
γ	1.473	0*		1.355	0*	0*	1.345	0*			1.712	0*			
	(.722)			(.862)			(.730)				(.891)				
ν	8.302	8.206	7.088	7.770	7.708	6.978	7.780	7.721	7.071	6.100	8.804	8.725	5.315	6.360	5.548
	(.932)	(.916)	(.732)	(.835)	(.830)	(.713)	(.835)	(.830)	(.731)	(.045)	(1.033)	(1.023)	(.404)	(.547)	(.039)
log$L$	7222	7218	7203	7198	7194	7176	7198	7194	7198	7187	7232	7228	7172	7183	7115
BIC	−14348	−14348	−14317	−14316	−14314	−14281	−14322	−14324	−14327	−14313	−14369	−14368	−14282	−14305	−14170
LL stat	5.53	5.91	14.91	6.22	6.60	6.52	6.20	6.60	5.76	6.48	5.75	6.29	44.09	7.60	60.21
LL p-val	.85	.82	.14	.80	.76	.77	.80	.76	.84	.77	.84	.79	.00	.67	.00

Estimation period: January 2, 1998, to November 28, 2014 (Friday's close to Friday's close). Stocks: AXP, GE, HD, and IBM (in this order). The models use close-to-open (co), open-to-close (oc), 5-minute (5m), daily (d), or weekly (w) returns, as indicated, to predict weekly return volatility, H_t. All models assume the standardized weekly returns, $H_t^{-1/2} r_t$, to be multivariate t_ν distributed. The model equations are as follows. 2comp(co,5m): (2.4)–(2.7); 2comp(co,oc): (2.4)–(2.7); 1comp(co,5m): (2.9)–(2.10); 1comp(5m): (2.9)–(2.10) with $\lambda_2 = 0$; 1comp(co,oc): (2.9) and (2.11); ls2comp(co,5m): (2.9) and (2.12)–(2.15); lsgarch(w): (2.16)–(2.16); garch(d): (2.19); garch(w): (2.18). The 5-minute return weights, ω_{di}, are specified exponentially (expon, Eq. (2.6)) or equally (equal, $\omega_{di} = 1$). LL: Ling and Li (1997) portmanteau test with 10 lags. * indicates fixed parameter values.

In terms of model fit, the locally stationary two-component models yield the greatest $\log L$ values, but to some extent this is because the long-run variance at any time t is estimated by a two-sided kernel (only truncated near the boundaries of the sample period), thus using information on returns that occur after time t. The next best fit is given by the 2comp(co,5m) model with exponential weights, followed by the 2comp(co,5m) model with equal weights. These models dominate the 1comp, garch(d), garch(w), and lsgarch(w) models by a large margin. Surprisingly, the 1comp(co,5m) model with equal intra-day weights and $\lambda_1 = \lambda_2 = 1$, which corresponds to using realized volatility augmented with the squared overnight return as volatility predictor, fits rather poorly, even more poorly than the garch(d) model and the 1comp(co,oc) model with $\lambda_1 = \lambda_2 = 1$. Apart from this, using high-frequency return information improves the model. Compare, for example, the garch(w), garch(d), 2comp(co,oc), and 2comp(co,5m) sequence of models. Another important conclusion is that incorporating the overnight returns generally improves the model, as a comparison between the garch(d) and 2comp(co,oc) models shows (and, to a lesser extent, between the garch(d) and 1comp(co,oc) models). Because the number of estimated parameters is relatively small and does not vary much across the models, roughly the same picture emerges on comparing the BIC values. The locally stationary two-component models yield the least BIC value, but in addition to the earlier remark it should be mentioned that the reported BIC value does not penalize for the nonparametric long-run variance estimation step. The next least BIC value is reached, again, by the 2comp(co,5m) model with exponential weights, but very closely followed by the 2comp(co,5m) model with equal weights.

As a diagnostic model check, Table 1 reports the portmanteau statistic of Ling and Li (1997), which is based on the sum of squared (standardized) residual autocorrelations. Under the null hypothesis of no conditional heteroskedasticity of the standardized residuals, the statistic is asymptotically distributed as χ^2_ℓ, where ℓ is the lag length chosen to compute the sample residual autocorrelation matrix. We only report the statistics and asymptotic p-values for $\ell = 10$, given that the results for other values of ℓ are very similar. The garch(w) and lsgarch(w) models are the only models that strongly reject the null of no conditional heteroskedasticity in the standardized residuals.

Table 2 reports likelihood ratio tests for several sets of parameter restrictions. The null hypothesis that the overnight returns do not matter in the 2comp(co,5m) model corresponds to $\alpha = \beta = \lambda_2 = 0$ and is strongly rejected by the data. In the 1comp(co,5m) model this hypothesis corresponds to $\lambda_2 = 0$ and is not rejected, but this model fits considerably less well. The hypothesis of equal 5-minute return weights, $\gamma = 0$, is rejected, although only marginally so in the 2comp(co,5m) model. Thus, MIDAS weights help to improve the in-sample model fit, albeit only to a modest degree. Finally, in the 1comp(co,5m) model, the hypothesis of equal intra-day and overnight weights, $\lambda_1 = \lambda_2 = 1$, is very strongly rejected.

Table 2. Likelihood ratio tests

model	null hypothesis	LR statistic	p-value
2comp(co,5m)	$\alpha = \beta = \lambda_2 = 0$	32.09	0.000
2comp(co,5m)	$\gamma = 0$	3.01	0.083
2comp(co,5m)	$\alpha = \beta = \lambda_2 = \gamma = 0$	39.76	0.000
1comp(co,5m)	$\gamma = 0$	8.62	0.003
1comp(co,5m)	$\gamma = 0, \lambda_1 = \lambda_2 = 1$	55.76	0.000
1comp(co,5m)	$\lambda_2 = 0$	0.04	0.842
1comp(co,5m)	$\gamma = \lambda_2 = 0$	8.66	0.034
1comp(co,oc)	$\lambda_1 = \lambda_2 = 1$	20.98	0.000

We present two figures that visualize the return (co)variance forecasts implied by the various models and compare them with the corresponding realized (co)variance, viewed as an approximation to the true unobserved conditional (co)variance. Figure 1 presents the time series of realized variances at weekly frequency for the returns on the AXP stock (in blue; the same in each subplot) along with the one-week-ahead conditional variance forecasts implied by the estimated models (in red; one subplot for each model). The weekly realized variance is defined as the sum of squares of all 5-minute and all overnight returns of the corresponding week. The variance forecasts of the garch(w) and lsgarch(w) models exhibit the least variation, in line with the low estimates of the ARCH coefficients α_i. In periods of high volatility, the realized variance tends to be underpredicted by these models. As we move to models that use daily returns or close-open and open-close returns, the variance forecasts change more rapidly, matching the realized variance series more closely. This pattern is further reinforced in models that use 5-minute return data, and is somewhat more pronounced in the 2comp and ls2comp models than in the 1comp models. Figure 2 presents a similar set of graphs as in Figure 1, but now for the realized covariances between the returns on the AXP and IBM stocks, and the corresponding one-week-ahead conditional covariance forecasts. Roughly the same patterns emerge as in Figure 1, with more rapidly changing forecasts as more detailed return information enters the models.

Figure 1. Weekly realized variance and model-implied variance forecast of AXP returns

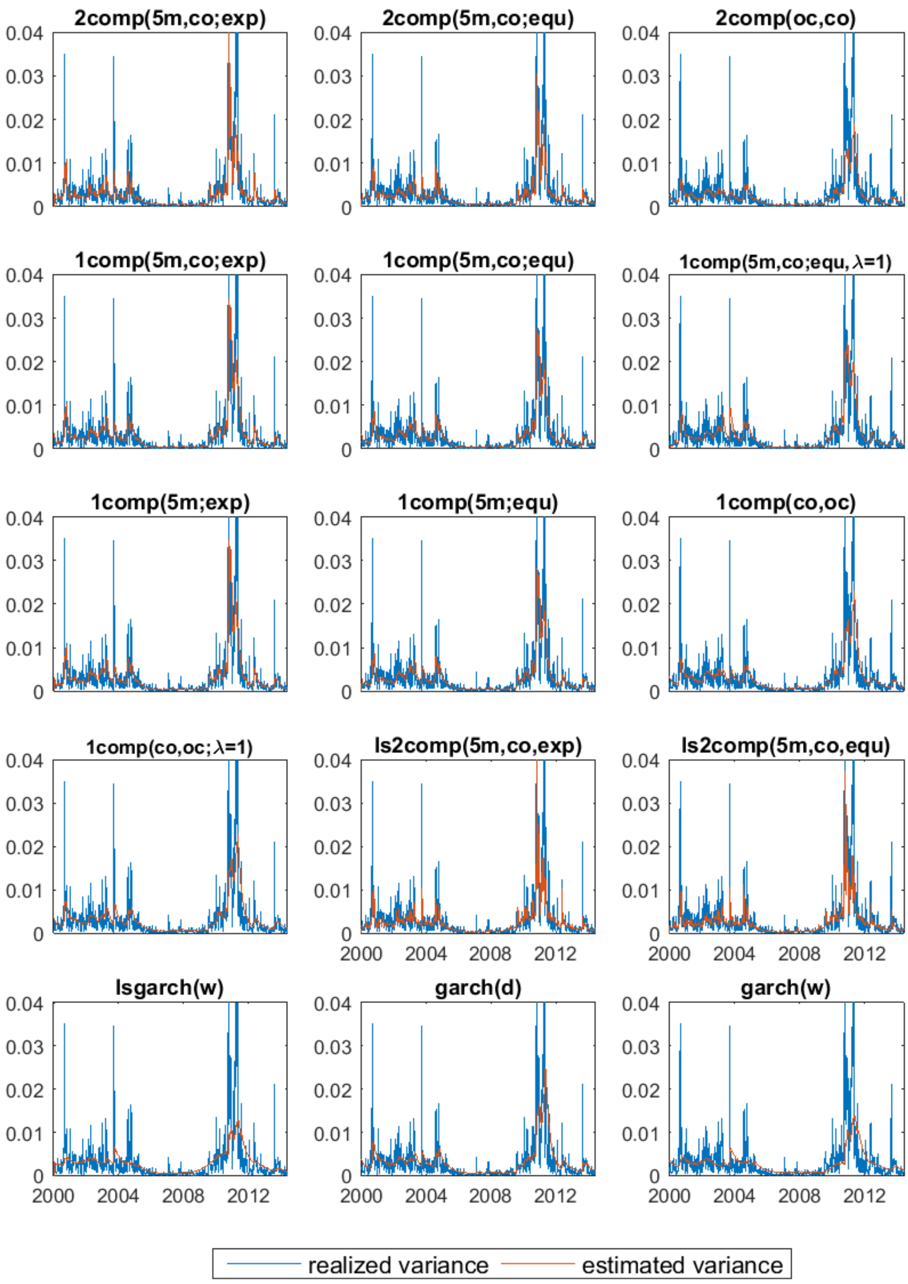
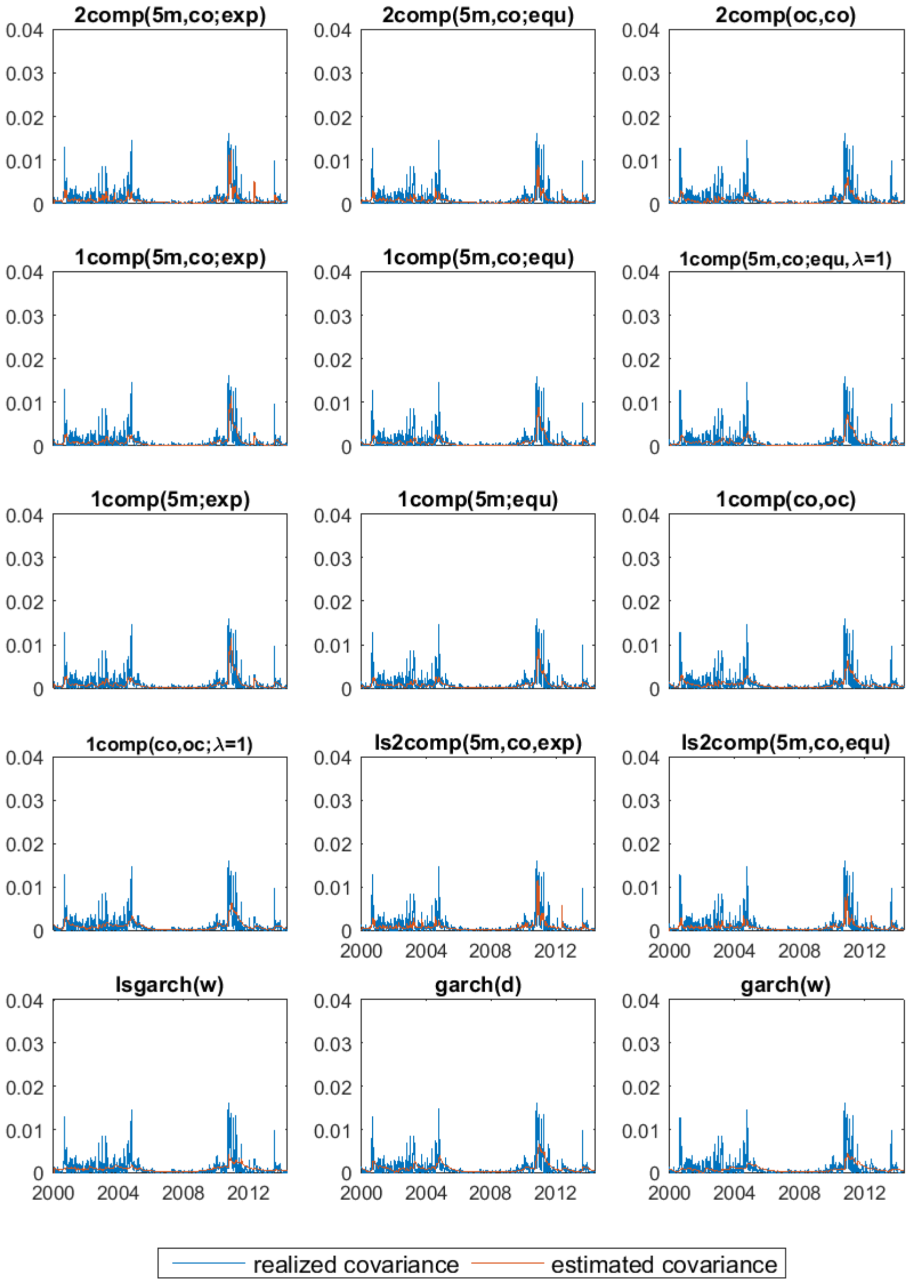


Figure 2. Weekly realized covariance and model-implied covariance forecast of AXP and IBM returns



4.2. Out-of-sample evaluation

We examine and compare the out-of-sample performance of the models using two methods: the Giacomini and White (2006) (GW) test and the model confidence set (MCS) approach of Hansen, Lunde, and Nason (2011). We re-estimated the models using an expanding window covering at least 722 weeks of data, each expansion consisting of 20 weeks of additional data. For each estimation window, the corresponding estimated parameters were held fixed and the next block of 20 weeks was used as the out-of-sample evaluation period, with the forecast horizon being one week throughout. With 8 different estimation windows, the total length of the out-of-sample evaluation period is 160 weeks. As argued in Laurent, Rombouts, and Violante (2012), the combination of an expanding estimation window and a fixed-size rolling out-of-sample evaluation window satisfies the assumptions required by the MCS procedure.

The GW test allows for a unified treatment of nested and non-nested models. Since the true conditional variance matrix is unobservable, we use the outer product of the weekly returns, $r_t r_t'$, as a proxy for the conditional variance. In the literature on volatility forecast comparisons, another widely used proxy is the realized variance matrix. See, e.g., Hansen and Lunde (2006) and Laurent, Rombouts, and Violante (2012, 2013). However, these authors mainly focused on intra-day (open-to-close) variance forecasting while our paper aims at weekly variance forecasting, with the compounding issue of how to incorporate the overnight variance. Therefore, we use the outer product of the weekly returns as a proxy, which is less informative than realized variance but unbiased (and robust). As loss function in the GW test we use the negative out-of-sample log-likelihood value. Table 3 reports the p-values of the GW tests of the hypothesis of equal unconditional predictive accuracy of any pair of models. The bottom row of the table gives the average (per observation) loss corresponding to each model. The 2comp(co,5m) models with exponential or equal 5-minute return weights significantly dominate all other models except the locally stationary versions of these models. The garch(w) and garch(d) models are outperformed by the models that use 5-minute returns and the dominations are statistically significant for the two-component models. In line with the in-sample results, the 1comp(co,5m) model with equal 5-minute weights and equal weights on intra-day and overnight returns (i.e., with $\gamma = 0$ and $\lambda_1 = \lambda_2 = 1$) exhibits very poor out-of-sample performance, even worse than the garch(d) model.

Table 3. Out-of-sample Giacomini-White tests

	2comp			1comp							locally stationary				
											ls2comp		lsgarch	garch	garch
	(co,5m)		(co,oc)	(co,5m)			(5m)		(co,oc)		(co,5m)		(w)	(d)	(w)
ω_{di}	expon	equal	–	expon	equal	equal	expon	equal	–	–	expon	equal	–	–	–
2co(5m,co;exp)	0.000	0.400	0.072	0.041	0.042	0.001	0.040	0.038	0.080	0.007	0.157	0.201	0.022	0.005	0.018
2co(5m,co;equ)	0.400	0.000	0.077	0.084	0.057	0.001	0.091	0.055	0.087	0.007	0.245	0.270	0.031	0.005	0.021
2co(oc,co)	0.072	0.077	0.000	0.413	0.387	0.361	0.451	0.423	0.703	0.228	0.425	0.435	0.832	0.131	0.171
1co(5m,co;exp)	0.041	0.084	0.413	0.000	0.966	0.004	0.820	0.866	0.444	0.118	0.988	0.993	0.292	0.081	0.129
1co(5m,co;equ)	0.042	0.057	0.387	0.966	0.000	0.001	0.885	0.845	0.417	0.089	0.980	0.983	0.251	0.061	0.108
1co(5m,co;equ,$\lambda = 1$)	0.001	0.001	0.361	0.004	0.001	0.000	0.012	0.006	0.282	0.670	0.040	0.022	0.342	0.846	0.868
1co(5m;exp)	0.040	0.091	0.451	0.820	0.885	0.012	0.000	0.990	0.488	0.150	0.928	0.930	0.305	0.110	0.151
1co(5m;equ)	0.038	0.055	0.423	0.866	0.845	0.006	0.990	0.000	0.460	0.121	0.932	0.932	0.271	0.087	0.130
1co(co,oc)	0.080	0.087	0.703	0.444	0.417	0.282	0.488	0.460	0.000	0.061	0.475	0.482	0.702	0.039	0.084
1co(co,oc;$\lambda = 1$)	0.007	0.007	0.228	0.118	0.089	0.670	0.150	0.121	0.061	0.000	0.131	0.134	0.344	0.521	0.339
ls2co(5m,co,exp)	0.157	0.245	0.425	0.988	0.980	0.040	0.928	0.932	0.475	0.131	0.000	0.986	0.203	0.110	0.129
ls2co(5m,co,equ)	0.201	0.270	0.435	0.993	0.983	0.022	0.930	0.932	0.482	0.134	0.986	0.000	0.216	0.116	0.130
lsgarch(w)	0.022	0.031	0.832	0.292	0.251	0.342	0.305	0.271	0.702	0.344	0.203	0.216	0.000	0.254	0.127
garch(d)	0.005	0.005	0.131	0.081	0.061	0.846	0.110	0.087	0.039	0.521	0.110	0.116	0.254	0.000	0.526
garch(w)	0.018	0.021	0.171	0.129	0.108	0.868	0.151	0.130	0.084	0.339	0.129	0.130	0.127	0.526	0.000
average loss	−9.578	−9.571	−9.514	−9.543	−9.543	−9.478	−9.542	−9.542	−9.517	−9.491	−9.544	−9.544	−9.508	−9.485	−9.471

Entries: p-values of the (unconditional) Giacomini-White test of the null that the corresponding row and column models have equal expected loss, with the negative out-of-sample log-likelihood as the loss function.

The MCS approach produces a set of models that contains the best model, for a given loss function and a given confidence level α. It is based on a sequence of equivalence tests. Let $\mathcal{M}^0$ be the initial set of models, indexed by $i = 1, \ldots, m^0$, and let $L_{i,t}$ the forecast loss of model i at time t. The null hypothesis is that all models in $\mathcal{M}^0$ have equal expected forecast loss. Since the asymptotic distribution of the test statistic is non-standard, a block bootstrap scheme is used to obtain the distribution under the null. If the null of equal expected forecast loss is rejected, an elimination rule is used to remove the model with the greatest loss. The procedure is repeated until the null is not rejected anymore and the MCS is the set of models that have not been removed. We implemented the MCS procedure using the Matlab MFE toolbox. Following Laurent et al. (2012), we used several types of loss functions: (i) the Euclidean loss function $L_{e,t} = \text{vech}(r_t r_t' - H_t)'\text{vech}(r_t r_t' - H_t)$, where the outer product of the weekly returns, $r_t r_t'$, is used as a proxy for the conditional variance matrix; (ii) the negative log-likelihood value, $L_{l,t} = -l_t$, with l_t the multivariate Student t log-likelihood given above; and (iii) the negative Gaussian quasi log-likelihood value, $L_{q,t} = (\log(2\pi) + \log|H_t| + r_t' H_t^{-1} r_t)/2 = -l_t|_{\nu=\infty}$. Further, we set the confidence level equal to $\alpha = 0.25$ and the average block length and the number of bootstrap replications to 10 and 1000, respectively. Table 4 reports the average out-of-sample loss values and the corresponding ranks of the models. Entries in boldface identify the models contained in the MCS. For each of the three loss functions, the garch(w), garch(d), lsgarch(w), 2comp(co,oc), 1comp(co,oc), and 1comp(co,oc) model with $\gamma = 0$ and $\lambda_1 = \lambda_2 = 1$ are excluded from the model confidence set. Among the included models, the 2comp(co,5m) models with exponential and equal 5-minute weights rank first and second. Their locally stationary versions rank third and fourth. Again, the one-component models perform relatively poorly compared to the two-component models, even when 5-minute returns are incorporated into the model.

Table 4. Out-of-sample model confidence sets

	2comp			1comp							locally stationary				
											ls2comp		lsgarch	garch	garch
	(co,5m)		(co,oc)	(co,5m)			(5m)		(co,oc)		(co,5m)		(w)	(d)	(w)
ω_{di}	expon	equal	–	expon	equal	equal	expon	equal	–	–	expon	equal	–	–	–
$L_{e,t}$	**0.3116**	**0.3137**	0.3340	0.3217	0.3223	0.3449	**0.3191**	**0.3183**	0.3440	0.3798	**0.3175**	**0.3164**	0.3807	0.4056	0.4810
rank	**1**	**2**	10	7	8	9	**6**	**5**	13	14	**3**	**4**	11	12	15
$L_{l,t}$	**-9.5776**	**-9.5712**	-9.5136	-9.5433	-9.5430	-9.4782	-9.5416	-9.5417	-9.5169	-9.4910	**-9.5437**	**-9.5435**	-9.5080	-9.4846	-9.4710
rank	**1**	**2**	10	5	6	15	7	8	9	13	**4**	**3**	11	14	12
$L_{q,t}$	**-9.3588**	**-9.3607**	-9.0269	**-9.2426**	**-9.2720**	-9.1389	**-9.2718**	**-9.2835**	**-9.2405**	-9.2107	**-9.3528**	**-9.3561**	-9.2467	-9.1890	-9.1710
rank	**2**	**1**	13	**7**	**5**	15	**8**	**6**	**9**	12	**4**	**3**	10	14	11

Entries: average Euclidean ($L_{e,t}$), negative log-likelihood ($L_{l,t}$), and negative quasi log-likelihood ($L_{q,t}$) loss, and corresponding ranks of the models. The models with entries in boldface are included in the model confidence set for the corresponding loss function and confidence level $\alpha = 0.25$. The model confidence sets were computed with 1000 block bootstrap replications and an average block length of 10.

5. CONCLUDING REMARKS

In this paper, we explored several ways to incorporate mixed-frequency returns and overnight returns into multivariate BEKK(1,1) GARCH models. The intra-day and overnight returns may enter the model either as two components or, perhaps closer in spirit to the GARCH approach, as a single component. Several proposals in the literature of how to deal with overnight returns are special cases of the models studied here.

In our application to forecast the weekly return variance matrix of four DJIA stocks in the period 1998–2014, we estimated 15 different models and evaluated them in and out of sample. We found that (i) the two-component versions of the model performed systematically better than the one-component versions according to a range of criteria; (ii) models that exploit 5-minute return information significantly improve on GARCH models that use weekly or daily returns only; (iii) incorporating overnight returns as a separate component considerably improves the model performance; (iv) overnight returns exhibit more persistence, as measured by the GARCH coefficients, than the 5-minute intra-day return component; (v) MIDAS weights improve the model but not dramatically so. Our findings hold in and out of sample.

One surprising finding is that the one-component model, even though its specification appears natural and it corresponds to some earlier suggestions of how to treat overnight returns, performs badly. The estimated weight given to overnight returns is close to zero. One special case of the one-component model, which corresponds to treating the overnight return as if it were an extra 5-minute return in a realized variance computation, performs extremely badly, even worse than a GARCH model that uses daily returns only.

Our tentative conclusion is that models with two components, one for high-frequency intra-day returns and one for the overnight returns, appear a sensible way to make progress on how to forecast low-frequency return variance matrices using the available mixed-frequency data. Stated otherwise, the intra-day and overnight returns seem to behave quite differently, and cramming the two into a single component does not seem to work well. Finally, it should be noted that there are potentially many other ways to model the two components (P_t and Q_t in our notation) than those considered here, and that they can also be specified in a dynamically interdependent way.

REFERENCES

Ahoniemi, K. and M. Lanne (2013). Overnight stock returns and realized volatility. *International Journal of Forecasting 29*, 592–604.

Blanc, P., R. Chicheportiche, and J.-P. Bouchaud (2014). The fine structure of volatility feedback II: Overnight and intra-day effects. *Physica A: Statistical Mechanics and its Applications 402*, 58–75.

Chen, X., E. Ghysels, and F. Wang (2015). HYBRID-GARCH: a generic class of models for volatility predictions using mixed frequency data. *Statistica Sinica 25*, 759–786.

Colacito, R., R. F. Engle, and E. Ghysels (2011). A component model for dynamic correlations. *Journal of Econometrics 164*, 45–59.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica 50*, 987–1007.

Engle, R. F., E. Ghysels, and B. Sohn (2008). On the economic sources of stock market volatility. Technical report, AFA 2008 New Orleans Meetings Paper. Available at SSRN: http://dx.doi.org/10.2139/ssrn.971310.

Engle, R. F. and K. F. Kroner (1995). Multivariate simultaneous generalized ARCH. *Econometric theory 11*, 122–150.

Engle, R. F. and J. G. Rangel (2008). The spline-GARCH model for low-frequency volatility and its global macroeconomic causes. *Review of Financial Studies 21*, 1187–1222.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2005). There is a risk-return trade-off after all. *Journal of Financial Economics 76*, 509–548.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2006). Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics 131*, 59–95.

Giacomini, R. and H. White (2006). Tests of conditional predictive ability. *Econometrica 74*, 1545–1578.

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

Hafner, C. M. and O. Linton (2010). Efficient estimation of a multivariate multiplicative volatility model. *Journal of Econometrics 159*, 55–73.

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

Hansen, P. R. and A. Lunde (2006). Consistent ranking of volatility models. *Journal of Econometrics 131*, 97–121.

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

Laurent, S., J. V. Rombouts, and F. Violante (2012). On the forecasting accuracy of multivariate GARCH models. *Journal of Applied Econometrics 27*, 934–955.

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

Ling, S. and W.-K. Li (1997). Diagnostic checking of nonlinear multivariate time series with multivariate ARCH errors. *Journal of Time Series Analysis 18*, 447–464.

Martens, M. (2002). Measuring and forecasting S&P 500 index-futures volatility using high-frequency data. *Journal of Futures Markets 22*, 497–518.

Rodríguez-Poo, J. M. and O. Linton (2001). Nonparametric factor analysis of residual time series. *Test 10*, 161–182.