Incorporating overnight and intraday returns into multivariate GARCH volatility models

Geert Dhaene* Jianbin Wu†

January 17, 2019

Abstract

We propose and evaluate mixed-frequency multivariate GARCH models for forecasting low-frequency (weekly) volatility based on high-frequency intraday returns (at 5-minute intervals) and on the overnight returns. The low-frequency conditional volatility matrix is modelled as a weighted sum of an intraday and an overnight component. The components are specified as multivariate GARCH processes of the BEKK type, adapted to the mixed-frequency data setting, and may enter the model as two separate components or as a single one. The models may further be extended by a nonparametrically estimated slowly-varying long-run volatility matrix. We evaluate the models in and out of sample using the 5-minute and overnight returns on four DJIA stocks (AXP, GE, HD, and IBM) from January 1988 to November 2014 and find that they systematically dominate a variety of models that only use lower-frequency data (weekly, daily, or close-to-open and open-to-close returns).

Keywords: Mixed-frequency sampling, overnight returns, intraday returns, multivariate GARCH.

JEL classification C22; C53.

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 has opened up possibilities for more accurate estimation of return volatilities. At the same time, new challenges for GARCH modeling have arisen since the variance forecast horizons of interest are often of a lower frequency (e.g., daily, weekly, or monthly) than the frequency at which financial returns are observed. Standard GARCH models, being single-frequency models, are not tailored to exploit high-frequency data for forecasting lower-frequency covariance matrices. A useful tool for this type of mixed-frequency problem is the mixed-data sampling (MIDAS) regression approach of Ghysels, Santa-Clara, and Valkanov (2005, 2006) and its multivariate extension (Ghysels, Sinko, and Valkanov 2007). Since the introduction of MIDAS regression, mixed-frequency versions have been developed for a wide range of models, including GARCH (Chen, Ghysels, and Wang 2011, 2015; Colacito, Engle, and Ghysels 2011; Ghysels and Valkanov 2012), stochastic volatility (Pettenuzzo, Timmermann, and Valkanov 2016; Gagliardini, Ghysels, and Rubin 2017), copulas (De Lira Salvatierra and Patton 2015; Oh and Patton 2016), and realized volatility (Forsberg

*KU Leuven, Faculty of Economics and Business. Email: geert.dhaene@kuleuven.be.
†Corresponding author. Nanjing University, School of Economics. Email: wujianbin@nju.edu.cn.

and Ghysels 2007; Ghysels and Sohn 2009; Ghysels and Sinko 2011; Golosnoy, Gribisch, and Liesenfeld 2012; Andreou 2016).[1]

We propose and evaluate multivariate versions of mixed-frequency GARCH models—or HYBRID GARCH models, in the terminology of Chen, Ghysels, and Wang (2011, 2015)—for weekly conditional variance forecasts based on 5-minute returns and on overnight returns. The main appeal of the mixed-frequency approach is that the model underlying the forecasts is directly focusing on volatility at the targeted frequency (e.g., weekly), while at the same time using higher-frequency information (intraday and overnight). Directly focusing on the targeted frequency has the advantage that the model parameters are calibrated so as to maximize the predictive power at that frequency, just like a standard GARCH model fitted to return data at the targeted frequency (e.g., weekly). On the other hand, the use of more refined, higher-frequency data makes the forecasts in principle better informed, just like multi-period (i.e., iterated) forecasts derived from a daily GARCH model or from realized volatility. The mixed-frequency approach combines the attractiveness of direct forecasting with the use of high-fequency data; see Chen, Ghysels, and Wang (2015) for further discussion.

For weekly (or longer) forecasting horizons, one question is how to use the overnight information effectively. Simply treating the overnight return as one extra 5-minute return might not work well, since asset prices are likely to evolve in different ways during trading and non-trading hours; see, e.g., Andersen, Bollerslev, and Huang (2011); Ahoniemi and Lanne (2013); and Blanc, Chicheportiche, and Bouchaud (2014). To the best of our knowledge, the overnight returns have not been considered in previous work on MIDAS GARCH models, and also have they rarely been considered in the GARCH literature in general. In the realized volatility literature, several proposals have been made to incorporate overnight returns. The simplest is to add the outer product of the overnight return to the intraday realized variance 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 up the intraday realized variance 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 intraday volatilities with optimally chosen weight parameters, as proposed by Hansen and Lunde (2005). We explore these various approaches in a mixed-frequency multivariate GARCH framework.

This paper is also related to recent work incorporating realized volatility into the forecasting of conditional (co)variances. The multiplicative error model (Engle 2002b; Engle and Gallo 2006; Caporin, Rossi, and Santucci de Magistris 2017) and the high-frequency-based volatility model (Shephard and Sheppard 2010; Noureldin, Shephard, and Sheppard 2012) specify multiple latent volatility processes involving realized volatility. The realized GARCH model (Hansen, Huang, and Shek 2012; Hansen, Lunde, and Voev 2014) specifies a measurement equation linking realized volatility to conditional (co)variances and derives multi-period volatility forecasts from the joint model of realized volatility and returns. In these models, high-frequency intraday information

[1] There are also a growing number of applications in macro, e.g., Clements and Galvão (2008, 2009); Andreou, Ghysels, and Kourtellos (2010, 2013); Marcellino and Schumacher (2010); Kuzin, Marcellino, and Schumacher (2011); Guérin and Marcellino (2013); Bai, Ghysels, and Wright (2013); Engle, Ghysels, and Sohn (2013); Schorfheide and Song (2015); Foroni, Marcellino, and Schumacher (2015); Ghysels (2016); Marcellino and Sivec (2016); Marcellino, Porqueddu, and Venditti (2016).

enters the volatility forecast through the realized volatility. Differently, we utilize both high-frequency intraday information and overnight information, we are not relying on latent volatility processes, and we forecast volatility directly over longer intervals (e.g., weekly) instead of in a multi-period fashion.

We model the low-frequency conditional volatility matrix as a weighted sum of an intraday and an overnight component, driven by the intraday 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 intraday component, the outer product of the high-frequency (5-minute) returns enter the model through a parametrically specified mixed-data sampling (MIDAS) weight function or through the realized volatility. For the overnight component, the outer product of the overnight returns enter the model similarly. Alternatively, the low-frequency conditional volatility matrix may be modelled as a single-component BEKK-GARCH model by imposing a common dynamic structure on the intraday and overnight components. All model variants may further be extended by a nonparametrically estimated slowly-varying long-run volatility matrix.

We evaluate the models using 5-minute and overnight return data on four DJIA stocks (AXP, GE, HD, and IBM) from January 1988 to November 2014. Our focus is on forecasting weekly volatilities (defined as the low frequency). The mixed-frequency GARCH models are found to systematically dominate the low-frequency (i.e., weekly) 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 5-minute frequency. Thus, ignoring the available high-frequency information leads to spuriously high volatility persistence. Other findings are that the two-component model variants perform better than the single-component variants; that the overnight volatility component exhibits more persistence than the daytime component; and that MIDAS weights do not have much effect on the model performance.

Section 2 proposes the mixed-frequency GARCH models: one-component, two-component, and locally stationary two-component models. Section 3 discusses estimation. Section 4 evaluates the models in and out of sample. Section 5 concludes. Technical derivations are given in an appendix. A supplementary appendix, available at https://www.researchgate.net/profile/Jianbin_Wu6, contains additional tables and figures.

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. Let $\mathcal{I}_{t-1}$ be the information set at Friday's close of week $t-1$ and define

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

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 intraday returns and the overnight returns prior to week t. Since we work with returns in log form, the weekly returns are additively

composed as

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

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}^{5m} is the ith 5-minute return on day d of week t. We shall use the set of variables

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

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.

The models developed below are models of H_t. They vary in the level of detail with which the intraday 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). In all models, weekend returns (from Friday's close to Monday's open) are treated as ordinary overnight returns. We group the models under four different headings: two-component models, one-component models, locally stationary models, and GARCH models. We end the section with a discussion of temporal aggregation.

2.1 Two-component models

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

$$H_t = \lambda_1 P_t + \lambda_2 Q_t, \tag{2.1}$$

where λ_1 and λ_2 are scalar weight parameters and P_t and Q_t are modeled as separate processes. Following Chen, Ghysels, and Wang (2011, 2015), we model P_t and Q_t as multivariate HYBRID GARCH 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 intraday returns in week $t-1$:

$$P_t = (I_n - aa' - bb') \odot \overline{P} + aa' \odot \sum_{d,i} \omega_{di}^{\mathrm{5m}} r_{t-1,di}^{\mathrm{5m}} r_{t-1,di}^{\mathrm{5m}\prime} + bb' \odot P_{t-1}. \tag{2.2}$$

Here, $\overline{P} = E(P_t) = E(\sum_{d,i} \omega_{di}^{\mathrm{5m}} r_{tdi}^{\mathrm{5m}} r_{tdi}^{\mathrm{5m}\prime})$, a and b are parameter vectors, $\odot$ is the Hadamard product, and $\omega_{di}^{\mathrm{5m}}$ is a weight function with average value, across d and i, equal to 1. We explore a variety of weight functions. The simplest is equal weights, $\omega_{di}^{\mathrm{5m}} = 1$. An alternative is to specify exponential weights,

$$\omega_{di}^{\mathrm{5m}} = \frac{(78(d-1)+i)^{\gamma^{\mathrm{5m(w)}}}}{\sum_{d,i}(78(d-1)+i)^{\gamma^{\mathrm{5m(w)}}}/390} =: [A],$$

as in Engle, Ghysels, and Sohn (2013). The sign of $\gamma^{\mathrm{5m(w)}}$ is left unrestricted, although it is natural to expect $\gamma^{\mathrm{5m(w)}} > 0$, which gives more weight to the more recent 5-minute periods;

$\gamma^{5\mathrm{m(w)}} = 0$ corresponds to equal weights. It might be argued that smoothly changing weights over the whole week are not very natural because they ignore that there are intervening nights. Therefore, we also consider the weight specifications

$$\omega_{di}^{5\mathrm{m}} = \omega_d^{\mathrm{oc}} \omega_i^{5\mathrm{m}}, \qquad \omega_i^{5\mathrm{m}} = \frac{i^{\gamma^{5\mathrm{m(d)}}}}{\sum_i i^{\gamma^{5\mathrm{m}}}/78} =: [C],$$
$$\omega_d^{\mathrm{oc}} = \frac{d^{\gamma^{\mathrm{oc}}}}{\sum_d d^{\gamma^{\mathrm{oc}}}/5} =: [B_1] \qquad \text{or} \qquad \omega_d^{\mathrm{oc}} = \frac{\gamma_d^{\mathrm{oc}}}{\sum_d \gamma_d^{\mathrm{oc}}/5} =: [B_2]$$

(together with a normalization, e.g., $\gamma_5^{\mathrm{oc}} = 1$). This specification imposes the same set of relative weights for the 5-minute periods within a day, but allows the weights to differ across the days, either through an exponential specification or via day-of-the-week parameters γ_d^{oc}.

By analogy to P_t, we specify the overnight component as

$$Q_t = (I_n - \alpha\alpha' - \beta\beta') \odot \overline{Q} + \alpha\alpha' \odot \sum_d \omega_d^{\mathrm{co}} r_{t-1,d}^{\mathrm{co}} r_{t-1,d}^{\mathrm{co}\prime} + \beta\beta' \odot Q_{t-1}, \tag{2.3}$$

where $\overline{Q} = E(Q_t) = E\left(\sum_d \omega_d^{\mathrm{co}} r_{td}^{\mathrm{co}} r_{td}^{\mathrm{co}\prime}\right)$, α and β are parameter vectors, and the weights are specified as $\omega_d^{\mathrm{co}} = 1$,

$$\omega_d^{\mathrm{co}} = \frac{d^{\gamma^{\mathrm{co}}}}{\sum_d d^{\gamma^{\mathrm{co}}}/5} =: [D_1], \qquad \text{or} \qquad \omega_d^{\mathrm{co}} = \frac{\gamma_d^{\mathrm{co}}}{\sum_d \gamma_d^{\mathrm{co}}/5} =: [D_2],$$

by analogy to ω_d^{oc}. Note that scales of P_t and Q_t are identified because $E(P_t)$ and $E(Q_t)$ are directly linked to the data (given $\omega_{di}^{5\mathrm{m}}$ and ω_d^{co}). Likewise, the scale of H_t is fixed by $E(H_t) = E(\sum_d r_{td} r_{td}')$, so the scales of λ_1 and λ_2 in (2.1) are identified. We shall refer to the model (2.1)–(2.3) as the 2comp(co,5m) model, as it makes use of the overnight returns (close-to-open) and the 5-minute intraday returns.

If we ignore the information in the 5-minute returns and replace (2.2) by

$$P_t = (I_n - aa' - bb') \odot \overline{P} + aa' \odot \sum_d \omega_d^{\mathrm{oc}} r_{t-1,d}^{\mathrm{oc}} r_{t-1,d}^{\mathrm{oc}\prime} + bb' \odot P_{t-1}$$

(with ω_d^{oc} as above), we obtain the 2comp(co,oc) model, which uses the close-to-open and open-to-close returns only.

The weight specifications for the daytime and nighttime returns may be combined in several ways. The combinations that we estimate are listed in the panels (co,5m) and (co,oc) of Table 1 in Section 2.5. The building blocks are $[A]$ to $[D_2]$ as defined earlier and are repeated at the bottom of the table. The number in the first column of the table is the number of parameters in the combined weights. In the specifications 5 and 4a (with 5 and 4 parameters, resp.), the constraint $\gamma_d^{\mathrm{oc}} = \gamma_d^{\mathrm{co}}$ sets the day-of-the-week parameters for the daytime equal to those for the nighttime.

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

$$H_t = (I_n - aa' - bb') \odot \overline{H} + aa' \odot W_{t-1} + bb' \odot H_{t-1}, \tag{2.4}$$

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

$$W_{t-1} = \lambda_1 \sum_{d,i} \omega_{di}^{5\mathrm{m}} r_{t-1,di}^{5\mathrm{m}} r_{t-1,di}^{5\mathrm{m}\prime} + \lambda_2 \sum_{d} \omega_{d}^{\mathrm{co}} r_{t-1,d}^{\mathrm{co}} r_{t-1,d}^{\mathrm{co}\prime}. \tag{2.5}$$

Note that this model arises from imposing common dynamics on the components P_t and Q_t in the two-component model. That is, imposing the restrictions $\alpha = a$ and $\beta = b$ in (2.1)–(2.3) leads to (2.4)–(2.5). We refer to (2.4)–(2.5) as the 1comp(co,5m) model.

By analogy to the two-component model, replacing (2.5) with

$$W_{t-1} = \lambda_1 \sum_{d} \omega_{d}^{\mathrm{oc}} r_{t-1,d}^{\mathrm{oc}} r_{t-1,d}^{\mathrm{oc}\prime} + \lambda_2 \sum_{d} \omega_{d}^{\mathrm{co}} r_{t-1,d}^{\mathrm{co}} r_{t-1,d}^{\mathrm{co}\prime} \tag{2.6}$$

gives the 1comp(co,oc) model.

We maintain the same weight specifications as in the two-component models, i.e., those in the panels (co,5m) and (co,oc) of Table 1.

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. Colacito, Engle, and Ghysels (2011) proposed a DCC-MIDAS model with a short-run and a long-run component, building on Engle and Lee (1999). Engle and Rangel (2008) and Engle, Ghysels, and Sohn (2013) specified the slowly varying component by linking it to macro-economic variables using splines or 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, (2.1), with the intraday component and the overnight component now specified as

$$P_t = M_t^{1/2} P_t^* M_t^{1/2}, \qquad Q_t = N_t^{1/2} Q_t^* N_t^{1/2}, \tag{2.7}$$

where M_t and N_t are nonparametrically specified unconditional variance matrices that are slowly varying as functions of t and P_t^* and Q_t^* are short-run variance matrices with the property that $E(P_t^*) = E(Q_t^*) = I_n$. (The "short run" here refers to the weekly frequency and "slowly varying" refers to variation that is negligible from one week to the next.) As in the stationary two-component model, we specify P_t^* and Q_t^* as stationary BEKK processes with a MIDAS extension,

$$P_t^* = (I_n - aa' - bb') \odot I_n + aa' \odot M_{t-1}^{-1/2} \sum_{d,i} \omega_{di}^{5\mathrm{m}} r_{t-1,di}^{5\mathrm{m}} r_{t-1,di}^{5\mathrm{m}\prime} M_{t-1}^{-1/2} + bb' \odot P_{t-1}^*, \tag{2.8}$$

$$Q_t^* = (1 - \alpha\alpha' - \beta\beta') \odot I_n + \alpha\alpha' \odot N_{t-1}^{-1/2} \sum_{d} \omega_{d}^{\mathrm{co}} r_{t-1,d}^{\mathrm{co}} r_{t-1,d}^{\mathrm{co}\prime} N_{t-1}^{-1/2} + \beta\beta' \odot Q_{t-1}^*, \tag{2.9}$$

with the same weights as earlier (Table 1, panel (co,5m)). We refer to the model (2.1) and (2.7)–(2.9) 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

$$H_t = L_t^{1/2} H_t^* L_t^{1/2}, \tag{2.10}$$

$$H_t^* = (I_n - aa' - bb') \odot I_n + aa' \odot L_{t-1}^{-1/2} r_{t-1} r_{t-1}' L_{t-1}^{-1/2} + bb' \odot H_{t-1}^*, \tag{2.11}$$

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

We also consider two locally stationary models that lie between ls2comp(co,5m) and lsgarch(w). One uses the close-to-open and open-to-close returns, is labeled ls2comp(co,oc), and consists of (2.1), (2.7), (2.9), and

$$P_t^* = (I_n - aa' - bb') \odot I_n + aa' \odot M_{t-1}^{-1/2} \textstyle\sum_d \omega_d^{\text{oc}} r_{t-1,d}^{\text{oc}} r_{t-1,d}^{\text{oc}\prime} M_{t-1}^{-1/2} + bb' \odot P_{t-1}^*.$$

The other one uses the daily returns, is labeled lsgarch(d), and consists of (2.10) and

$$\begin{aligned} H_t^* &= (I_n - aa' - bb') \odot I_n + aa' \odot L_{t-1}^{-1/2} W_{t-1} L_{t-1}^{-1/2} + bb' \odot H_{t-1}^*, \\ W_{t-1} &= \lambda \sum_d \omega_d r_{t-1,d} r_{t-1,d}'. \end{aligned}$$

We estimate the ls2comp(co,oc) and lsgarch(d) models only with equal weights, $\omega_d^{\text{oc}} = \omega_d^{\text{co}} = \omega_d = 1$.

2.4 GARCH models

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

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

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

$$\begin{aligned} H_t &= (I_n - aa' - bb') \odot \overline{H} + aa' \odot W_{t-1} + bb' \odot H_{t-1}, \\ W_{t-1} &= \lambda \sum_d \omega_d r_{t-1,d} r_{t-1,d}', \end{aligned}$$

which we label garch(d) and where the weights are specified as $\omega_d = 1$,

$$\omega_d = \frac{d^\gamma}{\sum_d d^\gamma / 5}, \qquad \text{or} \qquad \omega_d = \frac{\gamma_d}{\sum_d \gamma_d / 5},$$

corresponding to panel (d) in Table 1.

We also estimate a standard diagonal BEKK model for the daily returns, labeled d-garch, which will be used to forecast H_t using multi-period forecasting.

2.5 Overview of weight specifications and models

Table 1 gives an overview of the various weight specifications, grouped by the level of detail of the intraday or intra-week returns being used: (co,5m), (co,co), or (d). The column "no." indicates the number of parameters to be estimated in the corresponding weight specification. All weights are constrained to be nonnegative.

Table 1: Weight specifications

no.	daytime weights	nighttime weights	restrictions
		(co,5m)	
0	$\omega_{di}^{5\mathrm{m}} = 1$	$\omega_d^{\mathrm{co}} = 1$	
1	$\omega_{di}^{5\mathrm{m}} = [A]$	$\omega_d^{\mathrm{co}} = 1$	
2	$\omega_{di}^{5\mathrm{m}} = [A]$	$\omega_d^{\mathrm{co}} = [D_1]$	
3	$\omega_{di}^{5\mathrm{m}} = [B_1] \cdot [C]$	$\omega_d^{\mathrm{co}} = [D_1]$	
5	$\omega_{di}^{5\mathrm{m}} = [B_2] \cdot [C]$	$\omega_d^{\mathrm{co}} = [D_2]$	$\gamma_d^{\mathrm{oc}} = \gamma_d^{\mathrm{co}}$
8	$\omega_{di}^{5\mathrm{m}} = [B_2]$	$\omega_d^{\mathrm{co}} = [D_2]$	
9	$\omega_{di}^{5\mathrm{m}} = [B_2] \cdot [C]$	$\omega_d^{\mathrm{co}} = [D_2]$	
		(co,oc)	
0a	$\omega_d^{\mathrm{oc}} = 1$	$\omega_d^{\mathrm{co}} = 1$	
2a	$\omega_d^{\mathrm{oc}} = [B_1]$	$\omega_d^{\mathrm{co}} = [D_1]$	
4a	$\omega_d^{\mathrm{oc}} = [B_2]$	$\omega_d^{\mathrm{co}} = [D_2]$	$\gamma_d^{\mathrm{oc}} = \gamma_d^{\mathrm{co}}$
8a	$\omega_d^{\mathrm{oc}} = [B_2]$	$\omega_d^{\mathrm{co}} = [D_2]$	
		(d)	
0b	$\omega_d = 1$		
1b	$\omega_d = [B_1]$ (without "oc" superscript)		
4b	$\omega_d = [B_2]$ (without "oc" superscript)		

$$[A] = \frac{(78(d-1)+i)^{\gamma^{5\mathrm{m(w)}}}}{\sum_{d,i}(78(d-1)+i)^{\gamma^{5\mathrm{m(w)}}}/390}, \quad [B_1] = \frac{d^{\gamma^{\mathrm{oc}}}}{\sum_d d^{\gamma^{\mathrm{oc}}}/5}, \quad [B_2] = \frac{\gamma_d^{\mathrm{oc}}}{\sum_d \gamma_d^{\mathrm{oc}}/5}$$

$$[C] = \frac{i^{\gamma^{5\mathrm{m(d)}}}}{\sum_i i^{\gamma^{5\mathrm{m(d)}}}/78}, \quad [D_1] = \frac{d^{\gamma^{\mathrm{co}}}}{\sum_d d^{\gamma^{\mathrm{co}}}/5}, \quad [D_2] = \frac{\gamma_d^{\mathrm{co}}}{\sum_d \gamma_d^{\mathrm{co}}/5}$$

Table 2 gives an overview of the models. In the models with daytime and nighttime returns, the daytime (5-minute or open-to-close) returns always enter the models via a diagonal BEKK specification, as presented above. In the one-component models, the nighttime returns automatically share the same specification and BEKK coefficients a and b as the daytime returns; see equations (2.4)–(2.6). On the other hand, in the two-component models (2comp) there are separate BEKK coefficients for the daytime and the nighttime returns, and one may also specify different BEKK models for the nighttime and daytime components. We estimated two BEKK specifications for the nighttime component: a diagonal version (as presented above) and a scalar version (i.e., with α and β proportional to a vector of ones). In the locally stationary two-component models (ls2comp), we estimated only a scalar BEKK model for the nighttime component (for reasons of stability of the estimates). The next 5 models in Table 2 only use daily or weekly returns and were estimated with a diagonal BEKK specification. The last model is the exponentially weighted moving average for the daily conditional variance, labeled d-ewma and specified as

$$H_{td} = 0.06(RV_{t,d-1} + r_{t,d-1}^{\mathrm{co}} r_{t,d-1}^{\mathrm{co}\prime}) + 0.94 H_{t,d-1}$$

where $H_{td} = \mathrm{Var}(r_{td}|\mathcal{I}_{t,d-1})$, $\mathcal{I}_{t,d-1}$ is the information set at the close of day $d-1$ in week t (and $\mathcal{I}_{t,0} = \mathcal{I}_{t-1,5}$), and $RV_{td} = \sum_i r_{tdi}^{5\mathrm{m}} r_{tdi}^{5\mathrm{m}\prime}$ is the realized variance. The d-ewma forecast of H_t given $\mathcal{I}_{t-1}$ is $H_t = 5H_{t-1,5}$. This gives 50 models in total.

Table 2: Models

	daytime BEKK	nighttime BEKK	weight specifications	restrictions
2comp(co,5m)	diag	diag, scalar	0,1,2,3,5,8,9	
2comp(co,oc)	diag	diag, scalar	0a,2a,4a,8a	
1comp(co,5m)	diag	diag	0,1,2,3,5,8,9	
1comp(co,5m)	diag	diag	0	$\lambda_1 = \lambda_2 = 1$
1comp(co,oc)	diag	diag	0a,2a,4a,8a	
ls2comp(co,5m)	diag	scalar	0,1,2,3,5,8,9	
ls2comp(co,oc)	diag	scalar	0a	
	daily or weekly BEKK			
lsgarch(d)		diag	0b	
lsgarch(w)		diag		
garch(d)		diag	0b,1b,4b	
garch(w)		diag		
d-garch*		diag		
d-ewma*				

*applied with multi-period forecasting.

The stationary models with 5-minute returns and weight specification 0 can be linked to some proposals in the literature for incorporating the overnight returns into realized volatility. The realized variance enters these models via $\sum_{d,i} \omega_{di}^{5\mathrm{m}} r_{t-1,di}^{5\mathrm{m}} r_{t-1,di}^{5\mathrm{m}\prime} = \sum_d RV_{t-1,d}$, the aggregate daytime realized variance of the previous week. The aggregate outer products of the overnight returns of the previous week enter the models similarly. In the 1comp(co,5m) model with weight specification 0 and $\lambda_1 = \lambda_2 = 1$, the prediction of H_t is driven by $W_{t-1} = \sum_d (RV_{t-1,d} + r_{t-1,d}^{\mathrm{co}} r_{t-1,d}^{\mathrm{co}\prime})$, which corresponds to augmenting the realized variance with the outer product of the overnight return as if the overnight return were a 5-minute return. Another approach is to scale up the daytime realized variance to a measure of volatility for the whole day. This can be linked to the 1comp(co,5m) model with weight specification 0 and $\lambda_2 = 0$, where $W_{t-1} = \lambda_1 \sum_d RV_{t-1,d}$.[2] A third approach is to optimally combine the intraday and overnight volatilities. The 1comp(co,5m) and 2comp(co,5m) models with weight specification 0 have this flavor: $\sum_d RV_{t-1,d}$ and $\sum_d r_{t-1,d}^{\mathrm{co}} r_{t-1,d}^{\mathrm{co}\prime}$ drive the components P_t and Q_t, which are then combined to predict H_t or, alternatively, $\sum_d RV_{t-1,d}$ and $\sum_d r_{t-1,d}^{\mathrm{co}} r_{t-1,d}^{\mathrm{co}\prime}$ are directly combined into a single component W_{t-1} that drives the prediction of H_t.

2.6 Temporal aggregation

The components P_t and Q_t in (2.2) and (2.3) are specified as multivariate HYBRID GARCH processes. One motivation for such processes is that they arise from temporal aggregation of high-frequency periodic GARCH processes and ad hoc parameter restrictions. Chen, Ghysels,

[2] This model does not appear explicitly in Table 2, but it will turn out that λ_2 is estimated to be very close to zero in the corresponding unconstrained model.

and Wang (2011) show this for univariate processes. Their arguments extend to the multivariate case although, as shown below, only for a restricted class of processes.

Consider P_t (the analysis for Q_t is analogous). Denote the 5-minute intraday returns in week t as $r^{5\mathrm{m}}_{t,j}$, $j = 1, ..., m$, where $m = 390$. Suppose $r^{5\mathrm{m}}_{t,j}$ follows a periodic weak GARCH(1,1) process (Bollerslev and Ghysels 1996) of the BEKK form. That is, for every t and every $j = 1, ..., m$, defining the space of functions $\mathcal{F}_{t,j-1} = \{1, r^{5\mathrm{m}}_{t',j'}, \mathrm{vec}(r^{5\mathrm{m}}_{t',j'} r^{5\mathrm{m}\prime}_{t',j'}); mt' + j' < mt + j\}$, the best linear predictors of $r^{5\mathrm{m}}_{t,j}$ and $r^{5\mathrm{m}}_{t,j} r^{5\mathrm{m}\prime}_{t,j}$ based on $\mathcal{F}_{t,j-1}$ are, respectively, zero and

$$P_{t,j|t,j-1} = c_{j-1} + a_{j-1} r^{5\mathrm{m}}_{t,j-1} r^{5\mathrm{m}\prime}_{t,j-1} a'_{j-1} + b_{j-1} P_{t,j-1|t,j-2} b'_{j-1},$$

where a_j, b_j, and c_j are $n \times n$ matrices. By convention, any 5-minute period (t, j) is the same as $(t-1, j+m)$, so $r^{5\mathrm{m}}_{t,0} = r^{5\mathrm{m}}_{t-1,m}$ and $P_{t,1|t,0} = P_{t,1|t-1,m}$. Also, by convention, $(a_m, b_m, c_m) = (a_0, b_0, c_0)$. We first write $P_{t,j|t,j-1}$ in vector form,

$$\mathrm{vec}P_{t,j|t,j-1} = \mathrm{vec}c_{j-1} + a^{\otimes 2}_{j-1} \mathrm{vec}(r^{5\mathrm{m}}_{t,j-1} r^{5\mathrm{m}\prime}_{t,j-1}) + b^{\otimes 2}_{j-1} \mathrm{vec}P_{t,j-1|t,j-2}, \tag{2.12}$$

where $M^{\otimes 2} = M \otimes M$ and $\otimes$ is the Kronecker product. We assume all eigenvalues of $a^{\otimes 2}_j + b^{\otimes 2}_j$ are less than 1, for all j. For $j = 1, ..., m$, let $P_{t,j|t-1,m}$ be the best linear predictor of $r^{5\mathrm{m}}_{t,j} r^{5\mathrm{m}\prime}_{t,j}$ based on $\mathcal{F}_{t-1,m}$. From (2.12), for $j = 2, ..., m$, we have

$$\mathrm{vec}P_{t,j|t-1,m} = \mathrm{vec}c_{j-1} + (a^{\otimes 2}_{j-1} + b^{\otimes 2}_{j-1}) \mathrm{vec}P_{t,j-1|t-1,m},$$

and, by recursive substitution,

$$\mathrm{vec}P_{t,j|t-1,m} = g_{j-1} + d_{j-1} \mathrm{vec}P_{t,1|t-1,m}$$

where, for $j = 0, ..., m-1$,

$$g_j = \sum_{r=1}^{j} \left(\prod_{s=0}^{j-1-r} (a^{\otimes 2}_{j-s} + b^{\otimes 2}_{j-s}) \right) \mathrm{vec}c_r, \qquad d_j = \prod_{s=0}^{j-1} (a^{\otimes 2}_{j-s} + b^{\otimes 2}_{j-s}),$$

with the convention that an empty sum is zero and an empty product is unity. Hence, the best linear predictor of $r_t r'_t$ based on $\mathcal{F}_{t-1,m}$ is

$$\mathrm{vec}P_{t|t-1} = \sum_{j=1}^{m} \mathrm{vec}P_{t,j|t-1,m} = G + D \mathrm{vec}P_{t,1|t-1,m}, \tag{2.13}$$

where $G = \sum_{j=0}^{m-1} g_j$ and $D = \sum_{j=0}^{m-1} d_j$. Another recursive substitution gives

$$\begin{aligned} \mathrm{vec}P_{t,1|t-1,m} &= \mathrm{vec}c_m + a^{\otimes 2}_m \mathrm{vec}(r^{5\mathrm{m}}_{t-1,j-1} r^{5\mathrm{m}\prime}_{t-1,j-1}) + b^{\otimes 2}_m \mathrm{vec}P_{t-1,m|t-1,m-1} \\ &= K + \sum_{j=0}^{m-1} A_{m-j} \mathrm{vec}(r^{5\mathrm{m}}_{t-1,m-j} r^{5\mathrm{m}\prime}_{t-1,m-j}) + B \mathrm{vec}P_{t-1,1|t-2,m} \end{aligned} \tag{2.14}$$

where, for $j = 0, ..., m-1$,

$$K = \sum_{r=1}^{m} \left(\prod_{s=0}^{m-1-r} b^{\otimes 2}_{m-s} \right) \mathrm{vec}c_r, \quad B = \prod_{s=0}^{m-1} b^{\otimes 2}_{m-s}, \quad A_j = \left(\prod_{s=0}^{m-1-j} b^{\otimes 2}_{m-s} \right) a^{\otimes 2}_j.$$

From (2.13), $\mathrm{vec}P_{t-1,1|t-2,m} = D^{-1}(\mathrm{vec}P_{t-1|t-2} - G)$. Hence, from (2.13) and (2.14),

$$\mathrm{vec}P_{t|t-1} = C + D \sum_{j=0}^{m-1} A_{m-j} \mathrm{vec}(r^{5\mathrm{m}}_{t-1,m-j} r^{5\mathrm{m}\prime}_{t-1,m-j}) + DBD^{-1} \mathrm{vec}P_{t-1|t-2}$$

where $C = (I - DBD^{-1})G + DK$. To see the connection with the HYBRID GARCH component P_t specified in (2.2), first write P_t in vector form,

$$\begin{aligned}\text{vec}P_t &= (I_{n^2} - (\text{diag}a)^{\otimes 2} - (\text{diag}b)^{\otimes 2})\text{vec}\overline{P} \\ &\quad + (\text{diag}a)^{\otimes 2} \sum_{j=0}^{m-1} \omega_{m-j}^{5\text{m}} \text{vec}(r_{t-1,m-j}^{5\text{m}} r_{t-1,m-j}^{5\text{m}\prime}) + (\text{diag}b)^{\otimes 2}\text{vec}P_{t-1}, \end{aligned} \tag{2.15}$$

where $\omega_j^{5\text{m}} = j^{\gamma^{5\text{m}}} / \sum_{j=1}^m j^{\gamma^{5\text{m}}}$, diag$a$ and diag$b$ are diagonal matrices with the elements of a and b, respectively, on the diagonal, and note that $P_t = P_{t|t-1}$. Hence, P_t as in (2.2) follows from temporally aggregating the periodic model if certain parameter restrictions on the periodic model yield $DBD^{-1} = (\text{diag}b)^{\otimes 2}$ and $DA_{m-j} = (\text{diag}a)^{\otimes 2}\omega_{m-j}^{5\text{m}}$. (By periodic stationarity, if these two equalities hold, then $C = (I_{n^2} - (\text{diag}a)^{\otimes 2} - (\text{diag}b)^{\otimes 2})\text{vec}\overline{P}$ also holds.) Sufficient restrictions are: (i) a_j and b_j are diagonal; (ii) $a_j = a_0$; (iii) $b_j = \sqrt{\omega_{j-1}^{5\text{m}}/\omega_j^{5\text{m}}} I_n$ for $j = 1, ..., m$; and (iv) a_0 is proportional to I_n. The diagonality restriction implies that D, B, and A_j are diagonal and that $DBD^{-1} = B$, so $DBD^{-1} = (\text{diag}b)^{\otimes 2}$ is satisfied if we set diag$b = \prod_{j=0}^{m-1} b_j$. The additional restriction (ii) implies the recursion $A_{m-j} = A_{m-j+1} b_{m-j+1}^{\otimes 2}$ $(j = 1, ..., m-1)$. By (iii), we can write this as

$$A_{m-j} A_{m-j+1}^{-1} = b_{m-j+1}^{\otimes 2} = \frac{\omega_{m-j}^{5\text{m}}}{\omega_{m-j+1}^{5\text{m}}} I_{n^2}.$$

Hence $DA_{m-j} = (\text{diag}a)^{\otimes 2}\omega_{m-j}^{5\text{m}}$ is satisfied if we can set a such that $Da_0^{\otimes 2} = (\text{diag}a)^{\otimes 2}$. However, this requires D to be of the form $M^{\otimes 2}$, which is the case under the restrictions (i)–(iv) but not under (i)–(iii). So, in the multivariate case the temporal aggregation arguments of Chen, Ghysels, and Wang (2011) hold for the scalar BEKK model (i.e., with a_j and b_j proportional to I_n), but do not seem to extend much further. The restriction (ii) is the same as in the univariate case. The restriction (iii) is to make b_j compatible with the weights $\omega_j^{5\text{m}}$. The choice $\omega_j^{5\text{m}} = j^{\gamma^{5\text{m}}} / \sum_{j=1}^m j^{\gamma^{5\text{m}}}$ gives $b_j = (1 - 1/j)^{\gamma^{5\text{m}}/2} I_n$; a different specification of $\omega_j^{5\text{m}}$ changes b_j accordingly.

One may wonder if the restriction to the scalar BEKK model arises from using scalar weights $\omega_j^{5\text{m}}$. Consider diagonal weights—in line with the diagonal BEKK model—that is, $\omega_{d,i}^{5\text{m}}$ is an n-vector of weights, the summand in (2.2) is $\omega_{d,i}^{5\text{m}}\omega_{d,i}^{5\text{m}\prime} \odot r_{t-1,d,i}^{5\text{m}} r_{t-1,d,i}^{5\text{m}\prime}$, and the scalar $\omega_{m-j}^{5\text{m}}$ in (2.15) is replaced with $(\text{diag}\omega_{m-j}^{5\text{m}})^{\otimes 2}$. We look for parameter restrictions yielding (a) $DBD^{-1} = (\text{diag}b)^{\otimes 2}$ and (b) $DA_{m-j} = (\text{diag}a)^{\otimes 2}(\text{diag}\omega_{m-j}^{5\text{m}})^{\otimes 2}$. As before, (i) suffices to satisfy (a); (ii) can be kept and (iii) becomes $b_j = (\text{diag}\omega_{j-1}^{5\text{m}})(\text{diag}\omega_j^{5\text{m}})^{-1}$; and (i)–(iii) give $A_{m-j}A_{m-j+1}^{-1} = b_{m-j+1}^{\otimes 2} = (\text{diag}\omega_{m-j}^{5\text{m}})(\text{diag}\omega_{m-j+1}^{5\text{m}})^{-1}$. Therefore, (b) can only be satisfied if, again, a can be chosen such that $Da_0^{\otimes 2} = (\text{diag}a)^{\otimes 2}$, requiring $D = M^{\otimes 2}$ for some M. A sufficient condition (and necessary if we exclude the trivial case $a_j = 0$ for all j) is that a_j and b_j are proportional to I_n. Hence, initially nonscalar weights in a diagonal HYBRID BEKK-GARCH and imposing compatibility with a temporally aggregated periodic GARCH lead us back to a scalar HYBRID BEKK-GARCH with scalar weights.

Despite the finding that the restrictions from temporal aggregation lead to a scalar BEKK model, the implications of working with a scalar model are strong and potentially distortive in

terms of forecasting accuracy. Therefore, we will not impose a scalar structure; we shall continue to work with the diagonal model as specified in (2.2).

3 Estimation

We estimate the models by variance targeting or nonparamatric smoothing for the unconditional variance matrices, and by maximizing a Gaussian quasi-likelihood (QML) for the other parameters. Given the series of weekly returns $r_1, \ldots, r_T$, the loglikelihood is $l = \sum_{t=2}^{T} l_t$ where $l_t = -\frac{1}{2}\log\det H_t - \frac{1}{2} r_t' H_t^{-1} r_t - \frac{n}{2}\log(2\pi)$ and H_t is specified by the model (for some initialization H_1). The estimation proceeds in two steps.

In step 1, we profile out the unconditional variance matrices from l. For the stationary models, we use variance targeting, setting

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

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

as functions of parameters whenever these appear on the right-hand side. For the locally stationary models, we use the nonparametric smoothing estimator of Rodríguez-Poo and Linton (2001), setting

$$M_t = \frac{\sum_{\tau=1}^{T} K_h((t-\tau)/T) r_\tau^{\mathrm{oc}} r_\tau^{\mathrm{oc}\prime}}{\sum_{\tau=1}^{T} K_h((t-\tau)/T)}, \qquad N_t = \frac{\sum_{\tau=1}^{T} K_h((t-\tau)/T) r_\tau^{\mathrm{co}} r_\tau^{\mathrm{co}\prime}}{\sum_{\tau=1}^{T} K_h((t-\tau)/T)},$$

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

where $r_t^{\mathrm{oc}} = \sum_d r_{td}^{\mathrm{oc}}$, $r_t^{\mathrm{co}} = \sum_d r_{td}^{\mathrm{co}}$, $K_h(\cdot) = K(\cdot/h)/h$, $K(\cdot)$ is a kernel function, and h is a bandwidth parameter. We use the two-sided quartic kernel, $K(u) = \frac{15}{16}(1-u^2)^2 1_{|u|\leq 1}$, with $h = 0.15T$. To estimate M_t, N_t, L_t we only use the weekly aggregates $r_t^{\mathrm{oc}}, r_t^{\mathrm{co}}, r_t$ because the daily overnight returns, r_{td}^{co}, are often heavy-tailed and may not possess fourth-order moments, as required by the asymptotic theory. Working with weekly aggregates tends to mitigate this problem. An alternative way of dealing with heavy-tailed overnight returns in the estimation of long-run variances is suggested in Linton and Wu (2017), but their models are for single-frequency data. Note that, while the long-run variance estimates at any time $t < T$ also use data from time periods τ later than t, they never use data from outside the estimation window $\{1, \ldots, T\}$.

In step 2, we estimate the other parameters, denoted together as ϕ, by maximizing the profiled loglikelihood obtained in step 1. Because step 1 is not QML, the estimate of ϕ is not a QML estimate either and the usual QML standard errors are incorrect. We calculate standard errors that correct for this, taking the estimation in step 1 into account. In the case of variance targeting, the standard errors follow from standard GMM theory. In the locally stationary

models, they follow along the lines of Hafner and Linton (2010). The derivation is given in the appendix.

Our motivation to use variance targeting for the stationary models is mainly to ease the estimation, in particular to mitigate numerical difficulties in maximizing the quasi-loglikelihood function. As discussed in Francq, Horvath, and Zakoïan (2011), variance targeting in standard GARCH models further guarantees that, even when the model is misspecified, the unconditional variance matrix (the "targeted" matrix) is consistently estimated, an important advantage for long-term (multi-period) forecasting. Since the garch(w) model is a standard GARCH model, this property holds: $T^{-1}\sum_t r_t r_t'$ consistently estimates the targeted unconditional variance matrix, $E(H_t) = E(r_t r_t')$, regardless of the garch(w) specification being correct or not. In the mixed-frequency models, however, this interpretation is generally lost because the unconditional variance matrix of interest, $E(H_t)$, does not coincide with the matrix (or matrices) estimated by variance targeting. This is obvious in the two-component models; it also holds in the one-component models because the variance targeting estimator depends on the specification, so it cannot generally be consistent for $E(H_t)$ without correct specification. Hence, the motivation for variance targeting related to multi-period forecasting does not apply to the mixed-frequency approach. The multi-period forecasting perspective actually reveals a limitation of the mixed-frequency approach: there are no model-implied multi-period forecasts because there is no specification for how the high-frequency and the overnight returns are generated. A mixed-frequency model yields direct forecasts of the conditional variance, H_t, at the chosen (here, weekly) frequency only. Longer-range volatility forecasts, e.g., at monthly horizons, would require estimating a mixed-frequency model with the corresponding low frequency.

4 Application to four DJIA stocks, 1998–2014

We applied the models presented above using the 5-minute intraday 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). 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%.

The average weekly realized correlations and standard deviations (computed from the 5-minute intraday returns) are given in panel (a) of Table 3, and the weekly correlations and standard deviations of the overnight returns are given in panel (b). The intraday part of the returns has larger standard deviations but smaller correlations than the overnight part.

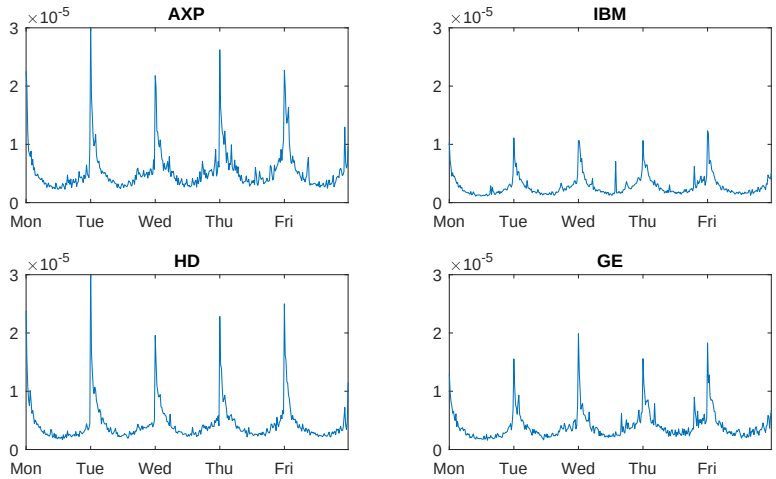
Table 3: Weekly realized correlations and standard deviations

	AXP	IBM	HD	GE	AXP	IBM	HD	GE
	(a) daytime returns				(b) overnight returns			
AXP	1.0000	0.3586	0.3626	0.4300	1.0000	0.4022	0.4713	0.6007
IBM	0.3586	1.0000	0.3476	0.3988	0.4022	1.0000	0.3524	0.4567
HD	0.3626	0.3476	1.0000	0.3786	0.4713	0.3524	1.0000	0.4913
GE	0.4300	0.3988	0.3786	1.0000	0.6007	0.4567	0.4913	1.0000
std. dev.	0.0461	0.0332	0.0416	0.0396	0.0265	0.0238	0.0263	0.0252

Period: January 2, 1998 to November 28, 2014.

All our weight specificatons in Table 1 are "scalar" in that they impose common weights across all n assets. This may be justified, to some extent, by the degree of commonality across assets of the intraday and day-of-the-week seasonal patterns of volatilities and correlations of the four stocks considered. Figures 1 and 2 show, respectively, the intra-week variances and correlations of the returns for each 5-minute period. Clearly, there is a reasonably high degree of commonality in the seasonal patterns across the four stocks. There is much less commonality in the day-of-the-week pattern of the daily variances and correlations of the overnight returns (the figures are provided in the supplementary appendix), but the correlations tend to be relatively constant across the days of the week and there are some outlying overnight returns that influence the results.

Figure 1: Variances of 5-minute intraday returns by time of the week (January 2, 1998 to November 28, 2014)



4.1 In-sample model evaluation and comparison

4.1.1 BEKK-GARCH estimates

Table 4 reports the model estimates using the entire sample. For the BEKK-GARCH parameter vectors (a, b, α, β) the table gives the average coefficient estimates $(\overline{a}, \overline{b}, \overline{\alpha}, \overline{\beta})$ across the four stocks, with the average standard errors in parentheses. (The estimates and standard errors for the indivual coefficients are available in the supplementary appendix.) A very clear picture emerges on how $\overline{a}$ and $\overline{b}$ evolve as higher-frequency returns are introduced in the models. In

Table 4: Estimates of mixed-frequency BEKK-type models of weekly returns for four DJIA stocks

column	1	2	3	4	5	6	7	8	9	10	11	12	13
1/2comp	2	2	2	2	2	2	2	2	2	2	2	2	2
loc. stat.	no	no	no	no	no	no	no	no	no	no	no	no	no
returns	co,5m	co,5m	co,5m	co,5m	co,5m	co,5m	co,5m	co,5m	co,5m	co,5m	co,5m	co,5m	co,5m
BEKK	diag	diag	diag	diag	diag	diag	diag	scalar	scalar	scalar	scalar	scalar	scalar
weight	0	1	2	3	5	8	9	0	1	2	3	5	8
$\overline{a}$	.776	.778	.777	.771	.844	.846	.832	.720	.715	.712	.706	.816	.824
	(.144)	(.114)	(.113)	(.133)	(.103)	(.097)	(.152)	(.237)	(.114)	(.115)	(.114)	(.109)	(.107)
$\overline{b}$	.620	.618	.618	.629	.524	.489	.520	.652	.657	.659	.664	.500	.501
	(.178)	(.145)	(.142)	(.163)	(.175)	(.178)	(.272)	(.263)	(.127)	(.128)	(.128)	(.195)	(.185)
$\overline{\alpha}$	.133	.129	.132	.128	.190	.143	.144	.157	.149	.148	.150	.169	.138
	(.050)	(.045)	(.045)	(.048)	(.064)	(.041)	(.037)	(.034)	(.026)	(.026)	(.026)	(.025	(.035)
$\overline{\beta}$	.987	.987	.987	.987	.964	.987	.986	.985	.986	.986	.986	.983	.988
	(.009)	(.009)	(.009)	(.010)	(.027)	(.007)	(.007)	(.005)	(.004)	(.004)	(.004)	(.004)	(.005)
λ_1 or λ	.867	.860	.855	.959	.976	.894	1.024	.919	.919	.917	1.051	1.066	.934
	(.096)	(.094)	(.090)	(.107)	(.123)	(.104)	(.132)	(.099)	(.088)	(.087)	(.120)	(.160)	(.117)
λ_2	.689	.744	.747	.718	.761	.667	.632	.553	.594	.577	.518	.553	.584
	(.213)	(.190)	(.174)	(.163)	(.165)	(.156)	(.180	(.209)	(.177)	(.173)	(.173)	(.195)	(.193)
logl	7106	7109	7109	7114	7118	7120	7123	7100	7103	7103	7109	7110	7114
BIC	−14090	−14089	−14083	−14085	−14080	−14064	−14063	−14118	−14117	−14112	−14115	−14106	−14092
LL stat	5.05	4.65	4.73	4.76	5.94	5.78	5.96	5.57	5.03	5.08	5.25	5.62	6.03
LL p val.	.89	.91	.91	.91	.82	.83	.82	.85	.89	.89	.87	.85	.81
L_e '98-'01	15.41	15.32	15.32	15.27	15.33	15.31	15.21	15.45	15.36	15.35	15.30	15.33	15.33
'02-'05	2.46	2.51	2.51	2.45	2.36	2.45	2.38	2.50	2.55	2.56	2.48	2.36	2.46
'06-'09	9.96	1.16	1.16	9.92	9.88	1.34	1.27	9.96	1.05	1.04	9.78	9.95	1.34
'10-'14	.93	.95	.95	.97	1.03	.92	.92	.94	.94	.94	.96	1.04	.93
'98-'14	6.84	6.89	6.89	6.81	6.81	6.91	6.85	6.87	6.88	6.88	6.79	6.83	6.91

Continued on the next page.

Table 4: Estimates of mixed-frequency BEKK-type models of weekly returns for four DJIA stocks (cont.)

column	14	15	16	17	18	19	20	21	22	23	24	25	26
1/2comp	2	1	1	1	1	1	1	1	1	2	2	2	2
loc. stat.	no	no	no	no	no	no	no	no	no	no	no	no	no
returns	co,5m	co,5m	co,5m	co,5m	co,5m	co,5m	co,5m	co,5m	co,5m	co,oc	co,oc	co,oc	co,oc
BEKK	scalar									diag	diag	diag	diag
weight	9	0	1	2	3	5	8	9	0	0a	2a	4a	8a
$\overline{a}$	.818	.626	.598	.598	.615	.634	.655	.627	.700	.371	.346	.361	.355
	(.115)	(.107)	(.119)	(.119)	(.124)	(.161)	(.131)	(.159)	(.178)	(.066)	(.048)	(.089)	(.063)
$\overline{b}$	.508	.742	.764	.764	.750	.731	.710	.738	.667	.926	.929	.927	.928
	(.202)	(.102)	(.108)	(.108)	(.119)	(.160)	(.134)	(.154)	(.186)	(.028)	(.020)	(.035)	(.029)
$\overline{\alpha}$	.139									.199	.219	.238	.260
	(.034)									(.079)	(.119)	(.250)	(.139)
$\overline{\beta}$	.988									.954	.871	.897	.934
	(.005)									(.045)	(.164)	(.188)	(.088)
λ_1 or λ	1.056	1.178	1.193	1.194	1.342	1.383	1.196	1.367	1 (fixed)	.975	1.125	1.032	1.088
	(.139)	(.058)	(.058)	(.058)	(.084)	(.113)	(.076)	(.125)		(.101)	(.073)	(.110)	(.075)
λ_2	.559	.001	.001	.001	.001	.001	.065	.023	1 (fixed)	.779	.407	.641	.512
	(.211)	(.002)	(.002)	(.001)	(.001)	(.002)	(.204)	(.073)		(.188)	(.108)	(.183)	(.130)
logl	7117	7089	7090	7090	7097	7099	7094	7099	7046	7080	7085	7085	7096
BIC	−14092	−14109	−14106	−14099	−14106	−14097	−14065	−14070	−14039	−14038	−14035	−14020	−14016
LL stat	5.97	6.42	5.97	5.97	5.95	6.10	6.76	6.31	7.64	7.07	6.50	6.47	6.80
LL p val.	.82	.78	.82	.82	.82	.81	.75	.79	.66	.72	.77	.77	.74
L_e '98-'01	15.24	15.47	15.42	15.42	15.36	15.32	15.52	15.32	17.09	15.77	15.82	15.85	16.34
'02-'05	2.40	2.51	2.56	2.56	2.46	2.43	2.49	2.44	2.53	2.90	2.95	2.98	2.97
'06-'09	1.25	9.80	10.09	1.09	9.87	9.88	9.69	9.82	1.63	11.75	11.85	11.58	11.98
'10-'14	.94	.94	.95	.95	.97	.91	.93	.92	1.02	1.08	1.08	1.08	1.08
'98-'14	6.86	6.84	6.91	6.91	6.82	6.79	6.81	6.78	7.44	7.50	7.55	7.49	7.71

Continued on the next page.

Table 4: Estimates of mixed-frequency BEKK-type models of weekly returns for four DJIA stocks (cont.)

column	27	28	29	30	31	32	33	34	35	36	37	38	39
1/2comp	2	2	2	2	1	1	1	1					
loc. stat.	no	no	no	no	no	no	no	no	no	no	no	no	no
returns	co,oc	co,oc	co,oc	co,oc	co,oc	co,oc	co,oc	co,oc	d	d	d	w	d-garch
BEKK	scalar	scalar	scalar	scalar									
weight	0a	2a	4a	8a	0a	2a	4a	8a	0b	1b	4b		
$\overline{a}$	.369	.367	.369	.374	.329	.329	.332	.331	.302	.302	.301	.218	.155
	(.054)	(.052)	(.053)	(.058)	(.044)	(.043)	(.044)	(.053)	(.035)	(.035)	(.037)	(.022)	(.035)
$\overline{b}$	.922	.923	.922	.921	.934	.933	.932	.933	.944	.944	.945	.970	.987
	(.024)	(.023)	(.024)	(.026)	(.019)	(.019)	(.019)	(.024)	(.013)	(.013)	(.014)	(.007)	(.007)
$\overline{\alpha}$	.143	.142	.143	.122									
	(.035)	(.036)	(.035)	(.048)									
$\overline{\beta}$	.982	.983	.982	.986									
	(.007)	(.007)	(.007)	(.009)									
λ_1 or λ	1.067	1.075	1.077	1.052	1.226	1.230	1.238	1.194	.962	.962	.958		
	(.112)	(.106)	(.107)	(.107)	(.096)	(.089)	(.091)	(.098)	(.042)	(.041)	(.041)		
λ_2	.569	.529	.531	.554	.246	.222	.220	.308					
	(.185)	(.168)	(.176)	(.249)	(.171)	(.153)	(.161)	(.214)					
logl	7072	7075	7076	7082	7069	7072	7074	7079	7052	7052	7052	6945	7040
BIC	−14064	−14055	−14044	−14028	−14071	−14063	−14052	−14035	−14043	−14036	−14016	−13836	−14025
LL stat	7.12	6.79	6.75	7.02	6.72	6.41	6.30	7.12	9.43	9.42	9.57	47.50	6.06
LL p val.	.71	.74	.75	.72	.75	.78	.79	.71	.49	.49	.48	.00	.81
L_e '98-'01	15.80	15.81	15.80	15.74	15.81	15.82	15.81	15.89	16.08	16.08	16.11	16.50	16.08
'02-'05	2.89	2.90	2.95	2.95	2.90	2.92	2.96	2.97	3.10	3.10	3.12	3.35	3.05
'06-'09	11.76	11.77	11.63	11.99	11.76	11.70	11.57	11.78	12.60	12.60	12.62	15.13	11.80
'10-'14	1.08	1.08	1.07	1.08	1.08	1.08	1.08	1.08	1.10	1.10	1.09	1.17	1.08
'98-'14	7.51	7.52	7.49	7.56	7.51	7.50	7.48	7.55	7.83	7.83	7.84	8.61	7.62

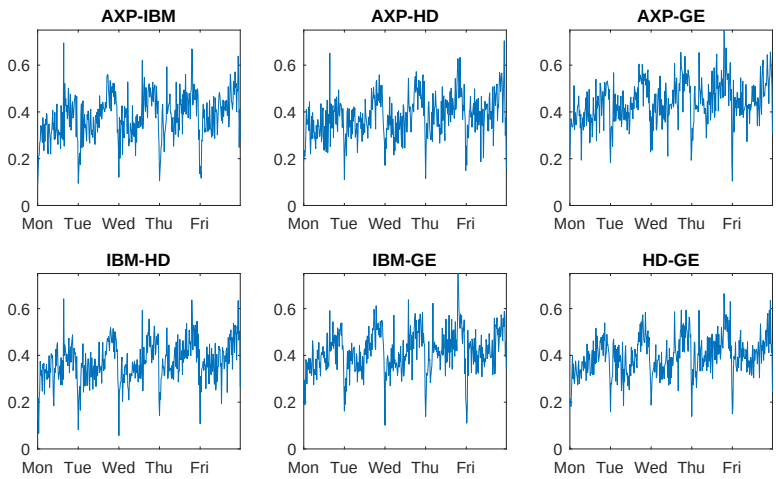
Continued on the next page.

Table 4: Estimates of mixed-frequency BEKK-type models of weekly returns for four DJIA stocks (cont.)

column	40	41	42	43	44	45	46	47	48	49	50
1/2comp	2	2	2	2	2	2	2	2			
loc. stat.	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	
returns	co,5m	co,5m	co,5m	co,5m	co,5m	co,5m	co,5m	co,oc	d	w	d-ewma
BEKK	scalar	scalar	scalar	scalar	scalar	scalar	scalar	scalar			
weight	0	1	2	3	5	8	9	0	0	0	
$\overline{a}$	.669	.662	.662	.679	.691	.747	.741	.392	.247	.173	
	(.274)	(.300)	(.300)	(.305)	(.315)	(.226)	(.152)	(.254)	(.274)	(.188)	
$\overline{b}$	.295	.317	.318	.373	.380	.376	.393	.799	.792	.231	
	(.094)	(.123)	(.125)	(.294)	(.238)	(.302)	(.207)	(.226)	(.290)		
$\overline{\alpha}$	.001	.003	.004	.011	.001	.006	.001	.019			
	(.010)	(.001)	(.003)	(.050)	(.002)	(.006)	(.002)	(.024)			
$\overline{\beta}$	.677	.817	.825	.886	.728	.856	.774	.898			
	(5.603)	(.199)	(.197)	(.375)	(.626)	(.230)	(1.265)	(.271)			
λ_1 or λ	.791	.823	.824	.840	.866	.642	.754	.852	.802		
	(.139)	(.162)	(.162)	(.188)	(.209)	(.267)	(.079)	(.163)	(.277)		
λ_2	.883	.866	.865	.914	.904	.565	.533	.956			
	(.238)	(.235)	(.235)	(.226)	(.237)	(1.244)	(1.821)	(.249)			
logl	7177	7182	7182	7183	7185	7181	7184	7154	7184	7173	6998
BIC	−14272	−14275	−14268	−14263	−14255	−14225	−14226	−14226	−14308	−14291	−13983
LL stat	7.96	7.46	7.46	7.71	7.29	7.67	7.71	1.98	14.19	27.29	7.78
LL p val.	.63	.68	.68	.66	.70	.66	.66	.36	.16	.00	.65
L_e '98-'01	15.22	15.12	15.12	15.04	15.01	15.16	15.05	15.43	15.46	15.62	16.44
'02-'05	2.83	2.86	2.86	2.85	2.79	2.85	2.79	2.98	3.10	3.23	2.70
'06-'09	11.90	11.73	11.72	11.45	11.08	11.40	10.99	12.59	13.30	14.64	1.38
'10-'14	.99	.97	.97	.96	.94	.95	.93	1.09	1.08	1.15	1.03
'98-'14	7.36	7.30	7.30	7.21	7.10	7.22	7.08	7.64	7.84	8.25	7.27

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 to predict the weekly return volatility, H_t. The estimates $\overline{a}, \overline{b}, \overline{\alpha}, \overline{\beta}$ are QML estimates with variance targeting or nonparametric smoothing, averaged across the four stocks; average standard errors in parentheses; 1/2comp: one or two components; BEKK: specification for nighttime component (the daytime component is always diagonal BEKK); weight: see specifications in Table 1; LL: Ling and Li (1997) portmanteau test with 10 lags; L_e: average Euclidean loss, $L_e = T^{-1} \sum_t \text{vech}(R_t - H_t)' \text{vech}(R_t - H_t)$ where $R_t = \sum_d r_{td} r_{td}'$ and H_t is the volatility forecast.

Figure 2: Correlations of 5-minute intraday returns by time of the week (January 2, 1998 to November 28, 2014)



the classic GARCH model for weekly returns, garch(w) (col. 38), the average of the estimated GARCH coefficients b_i is 0.97. This average estimate decreases to about 0.94 in the garch(d) models (cols. 35–37), where daily returns are incorporated as volatility predictors. It further decreases in the models that use 5-minute returns, to around 0.75 in the 1comp models where λ_1 and λ_2 are unconstrained (cols. 15–21), 0.5 to 0.65 in the 2comp models (cols. 1–14), and even 0.3 to 0.4 in the ls2comp models (cols. 40–45). The substantial decrease of the b_i estimates, as higher-frequency returns are incorporated, is mirrored by an increase of the ARCH coefficients a_i from an average of around 0.2 in the garch(w) model (col. 38) to 0.3 in the garch(d) models (cols. 35–37), 0.6 in the 1comp models (cols. 15–21), 0.7 to 0.85 in the 2comp models (cols. 1–14), and 0.65 to 0.75 in the ls2comp models (cols. 40–45). Hence, while all estimated models except the ls2comp 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 high-frequency returns are used. This is most pronounced in the two-component models.[3]

Another observation, in line with Blanc, Chicheportiche, and Bouchaud (2014), is that the dynamics of the overnight and intraday components are very different from each other. In the 2comp(co,oc) models (cols. 23–30), which treat the close-to-open and open-to-close returns symmetrically, the average estimated daytime coefficients $\overline{a}$ and $\overline{b}$ are all on the order of 0.35 and 0.92, respectively, while the corresponding average nighttime coefficient $\overline{\alpha}$ is estimated to be around 0.2 (with some mild variation) and $\overline{\beta}$ typically lies between 0.95 and 0.98 (with a few exceptions that have higher standard errors). This suggests that the overnight component decays more slowly than the intraday component, a property that is also borne out by the other 2comp models (cols. 1–14), with $\overline{b}$ nearly always between 0.5 and 0.65 and $\overline{\beta}$ between 0.98 and

[3]In an earlier version of this paper, we estimated a subset of the models with a multivariate Student t likelihood for the standardized returns, $H_t^{-1/2} r_t$, instead of the Gaussian quasi-likelihood. The results were broadly similar to those presented here. The degree-of-freedom parameter of the t distribution was estimated to be around 5 in the garch(w) model and 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 standardized returns appeared somewhat less fat-tailed.

0.99, although there is a natural tendency for the intraday decay rate, $1 - \overline{b}$, to go up as more refined intraday return information is used. In the ls2comp models (cols. 40–47) the estimated $\overline{b}$ and $\overline{\beta}$ are much smaller, typically between 0.3 and 0.4 for $\overline{b}$ and between 0.7 and 0.9 for $\overline{\beta}$, but the standard errors are often quite large, in particular for $\overline{\beta}$. The large standard errors for $\overline{\beta}$ (actually, for the β_i) in the locally stationary models may be explained by the slow decay rate of the overnight component, as evidenced by the large $\overline{\beta}$ in the 2comp models (cols. 1–14 and 23–30), which makes it to harder to separate identify the weekly dynamics from the slowly-varying unconditional variance.

4.1.2 Daytime and nighttime weights

In the 2comp models (Table 4, cols. 1–14 and 23–30), the weights given to the intraday and overnight components, λ_1 and λ_2, are estimated to be around 1 (ranging from 0.85 to 1.1) and 0.7 (ranging from 0.5 to 0.8), respectively. The ls2comp (cols. 40–47) models tend to give more equal weights to the daytime and nighttime components, with λ_1 estimated to be between 0.65 and 0.85, and λ_2 between 0.55 and 0.95. In the 1comp models (cols. 15–21 and 31–34), the estimates of λ_1 and λ_2 are quite different, with λ_1 systematically exceeding 1 and λ_2 in many cases (cols. 15–19) being estimated to be at the zero lower bound (in the estimation, we imposed a small positive lower bound); in those cases the estimated 1comp(co,5m) model reduces to the 1comp(5m) model. This is rather unexpected and, in particular, difficult to reconcile with the 2comp model estimates, even though the 1comp and 2comp models differ in the way the intraday and overnight returns enter the model. Apparently, imposing common dynamics on the intraday and overnight components tends to suppress the latter.[4]

4.1.3 Model fit

In terms of model fit, the locally stationary two-component models (Table 4, cols. 40–49) yield the greatest quasi-loglikelihood values, but this is primarily 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) models (cols. 1–14). These models dominate all other models, often by a large margin. Surprisingly, the 1comp(co,5m) model with equal weights (weight specificaton 0) and $\lambda_1 = \lambda_2 = 1$, which corresponds to using realized volatility augmented with the outer product of the overnight return as volatility predictor, fits rather poorly (col. 22), even more poorly than the 1comp(co,oc) model with equal weights (col. 31) and the garch(d) model with equal weights (col. 35). Apart from this, using high-frequency return information substantially improves the model fit. For example, for the equal-weight specification (0, 0a, 0b), the sequence of models garch(w) (col. 38), garch(d) (col. 35), 2comp(co,oc) with scalar BEKK (col. 27), and 2comp(co,5m) with scalar BEKK (col. 8), which progressively use higher-frequency information, has corresponding quasi-loglikelihood values of 6945, 7052, 7072, and 7100. Another important conclusion is that incorporating the overnight returns generally improves the model fit, as

[4] In the models that we estimated with t-distributed standardized returns, we obtained similar estimates of λ_1 and, in the 1comp models, we similarly found λ_2 to be estimated at the zero lower bound. In the 2comp and ls2comp models, however, we found higher estimates of λ_2, exceeding those of λ_1 by 0.2 to 0.3; thus, the QML estimates tend to downweight the (more heavy-tailed) overnight component.

shown by comparing the garch(d) models (cols. 35–37) with the 2comp(co,oc) (cols. 23–30) and 1comp(co,oc) (cols. 31–34) models. Because the number of estimated parameters is not excessively large and does not vary too 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 four least BIC values are attained by the most parsimoniously parameterized 2comp(co,5m) models: those with a scalar BEKK specification for the overnight component and with weight specifications 0 to 3 (cols. 8–11). According to both criteria, quasi-loglikelihood and BIC, the garch(w) model (col. 38) has the least fit, followed by d-ewma (col. 50); also the garch(d) models (cols. 35–37) fit poorly.

As a diagnostic model check, Table 4 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) (col. 38) and ls-garch(w) (col. 49) models are the only models that strongly reject the null of no conditional heteroskedasticity in the standardized residuals.

The bottom rows of Table 4 report, for the whole sample and for four sub-samples, the average Euclidean loss, defined as $L_e = T^{-1} \sum_t \text{vech}(R_t - H_t)'\text{vech}(R_t - H_t)$ where $R_t = \sum_d r_{td} r'_{td}$ and H_t is the volatility forecast. There is considerable variation across the sub-samples. The first sub-sample (1998–2001) exhibits the greatest loss for all models, and the implied model ranking for this sub-sample correlates the least with the implied rankings for the other sub-samples, with Spearman rank correlations between 0.47 and 0.68, while among the other sub-samples the rank correlations are between 0.80 and 0.89. For the first sub-sample, the models with the least Euclidean loss are locally stationary models (cols. 40–46) while for the other sub-samples and for the whole sample, the models with the least loss are always 1comp(co,5m) and 2comp(co,5m) models (in cols. 1–21).

4.1.4 Weight parameters

The parameters in the various weight specifications (Table 1) turned out to be hard to pin down empirically. Table 5, as an example, presents the estimates for the 2comp(co,5m) model with diagonal BEKK specification for the nighttime component (cols. 1–7 in Table 4); the results for the other models are qualitatively similar and are available in the supplementary appendix.

Table 5: Weight parameter estimates, 2comp(co,5m), diagonal BEKK for nighttime

no.	$\gamma^{5m(w)}$	$\gamma^{5m(d)}$	γ^{oc}	γ_1^{oc}	γ_2^{oc}	γ_3^{oc}	γ_4^{oc}	γ^{co}	γ_1^{co}	γ_2^{co}	γ_3^{co}	γ_5^{co}
1	1.36											
	(.76)											
2	1.36							.42				
	(.76)							(.76)				
3		.36	1.17					.47				
		(.14)	(1.09)					(.87)				
5		−1.46		1.72	.01	.81	1.38					
		(.61)		(1.57)	(.03)	(.85)	(.99)					
8				1.35	.00	.38	.61		.00	.31	.54	.00
				(.95)	(.00)	(.38)	(.37)		(.09)	(1.28)	(1.81)	(.00)
9		.26		1.20	.00	.35	.52		.00	.01	.54	.00
		(.16)		(.91)	(.00)	(.46)	(.34)		(.04)	(1.23)	(79.2)	(.05)

First column: weight specification (see Table 1). Normalizations: $\gamma_5^{oc} = 1, \gamma_4^{co} = 1$.

The estimates of $\gamma^{5m(w)}$, $\gamma^{5m(d)}$, γ^{oc}, and γ^{co} are positive (with one exception, $\gamma^{5m(d)}$ in weight specification 5) suggesting that more recent returns (near to Friday's close) have more predictive power for next week's volatility than earlier returns. This is in line with what one would expect. However, the estimates come with standard errors that are too large to draw strong conclusions. The estimates of the day-of-the-week parameters γ_d^{oc} and γ_d^{co} are also imprecise—even more so—and some of them are driven to zero, the lower bound of the parameter space. This occurs for γ_2^{oc} (relative to the normalization $\gamma_5^{oc} = 1$) and for γ_1^{co}, γ_5^{co}, and sometimes γ_2^{co} (relative to $\gamma_4^{co} = 1$; we chose this normalization because the more natural normalization $\gamma_5^{co} = 1$ drives some estimates of γ_d^{oc} to infinity). A possible explanation is the presence of large outliers in the returns (in particular in the overnight returns), which do not fit a Gaussian quasi-likelihood well and may lead to the implicit exclusion of the certain days of the week. To conclude, the more refined but highly parameterized weight specifications do not yield reliably estimated weights, and even the parsimonious weight specifications are, at best, only marginally better than equal weights (specification 0).

4.1.5 Fitted variances and correlations

We present two figures that visualize the return variance and correlation in-sample forecasts implied by (a selection of) the models and compare them with the corresponding weekly realized variances and correlations, viewed as a proxy for the true unobserved conditional variances and correlations. The selected models are those in Table 1 with weight specification 0, 0a, 0b, or no weight specification at all. As weekly realized variance matrix we use the sum of the outer products of the daily returns. In the literature on volatility forecasting, another widely used proxy is the realized variance computed from the intraday returns. See, e.g., Hansen and Lunde (2006) and Laurent, Rombouts, and Violante (2012, 2013). However, these authors focused on intraday (open-to-close) variance forecasting while our paper aims at weekly variance forecasting, with the compounding issue of how to incorporate the overnight returns. The sum of the outer products of the daily returns is less informative than the realized variance from intraday returns but is unbiased and free of assumptions about the overnight returns. Figure 3 presents the

time series of weekly realized variances 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 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-to-open and open-to-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 more pronounced in the 1comp and 2comp models than in the ls2comp model. Figure 4 presents a similar set of graphs as in Figure 3, but now for the realized corelation between the returns on the AXP and IBM stocks (computed from daily returns), and the corresponding one-week-ahead conditional correlation forecasts. Roughly the same patterns emerge as in Figure 3, with more rapidly changing forecasts as more detailed return information enters the models. An important difference is that, while most correlation forecasts tend to follow the overall pattern of evolution of the realized correlation, the forecasts evolve in a much more limited range than the realized correlation. This likely limits the usefulness of the models for portfolio optimization purposes, although we have not studied this.

We now examine in greater detail two periods that are marked by high volatility: the period around 9/11 and the financial crisis of 2008. We take the AXP stock as an example.

Figure 5 is the same as Figure 3, but zooms in on the period from August 10, 2001 to January 4, 2002 (21 weeks). The weekly realized variance peaked in the week of 9/11, but the high volatility was short-lived. Unsurprisingly, all models fail to forecast the volatility jump. In the weeks following 9/11, the predicted variances rise substantially. The variance forecasts of the garch(d) and garch(w) models remain high for a long period. On the other hand, the variance forecasts of the models incorporating intraday and overnight returns are more in line with the

Figure 3: AXP returns: weekly realized variance and in-sample variance forecasts (January 2, 1998 to November 28, 2014)

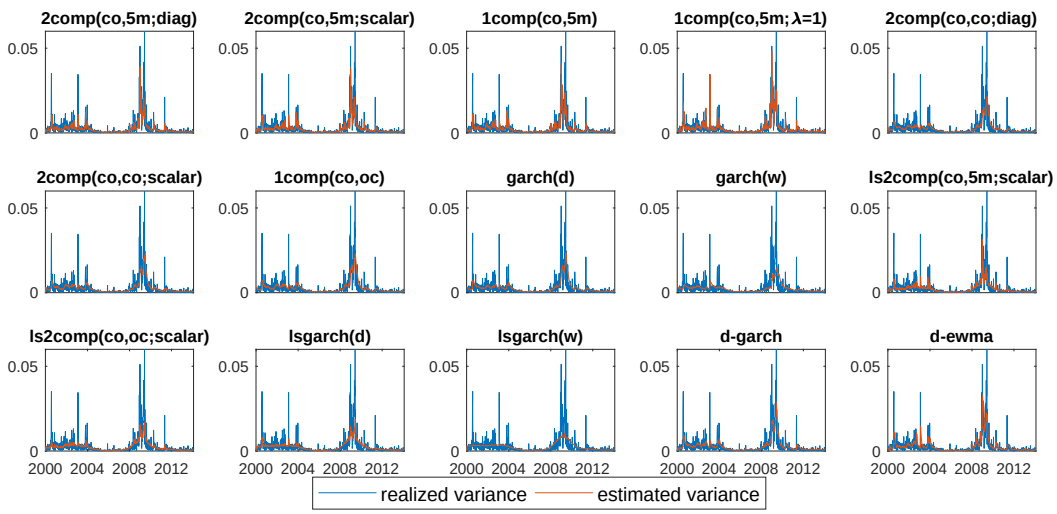


Figure 4: AXP and IBM returns: weekly realized correlation and in-sample correlation forecasts (January 2, 1998 to November 28, 2014)

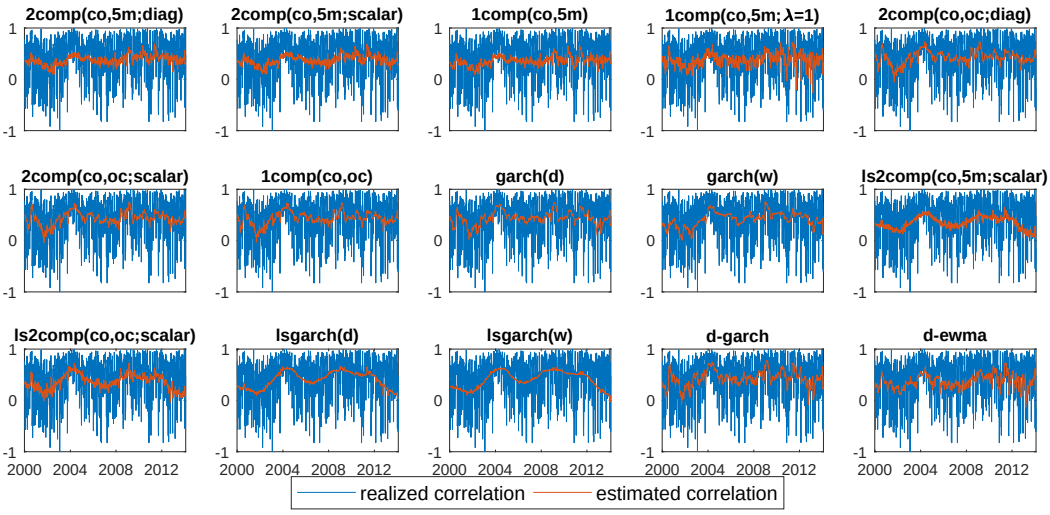
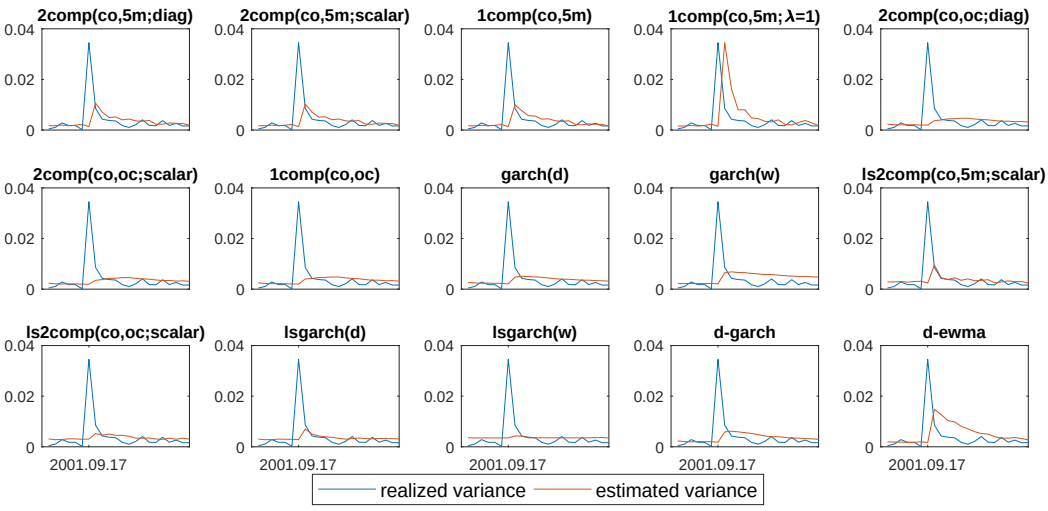


Figure 5: AXP returns: weekly realized variance and in-sample variance forecasts around 9/11 (August 10, 2001 to January 4, 2002)

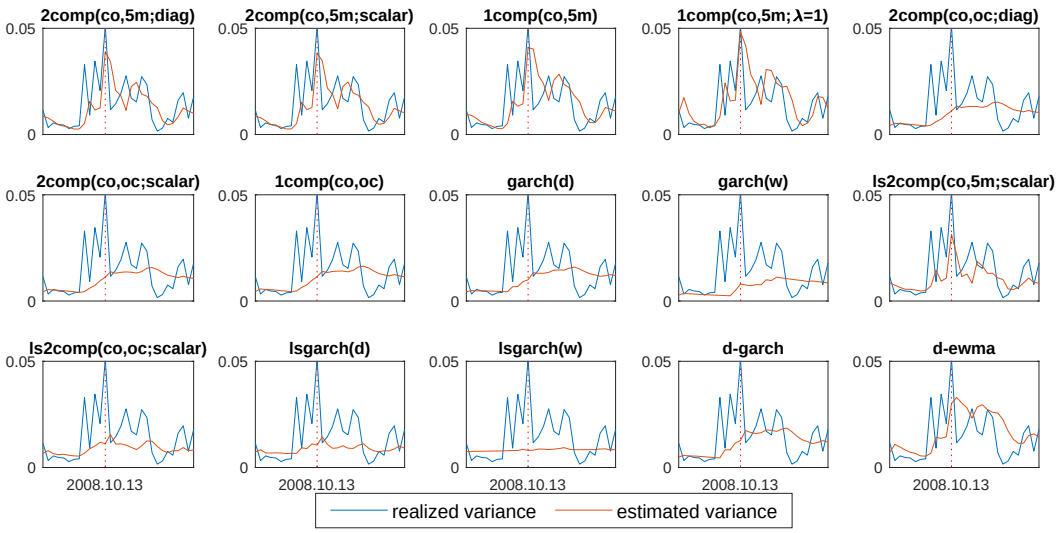


realized variance, falling back more quickly to pre-9/11 levels. This is especially the case for the two-component models.

Figure 6 zooms in on the period from July 25, 2008 to February 13, 2009 (29 weeks), covering a long period of high volatility due to the 2008 financial crisis. In general, the models that incorporate intraday and overnight returns predict the very large realized variance in the week of October 13, 2008 quite well. Actually, the intraday volatility was already very high one week earlier, in the week of October 6. But the daily and weekly returns did not well reflect the intraday information, leading almost all models that do not use intraday returns, such as the garch(w) and garch(d) models, to severely underpredict the variance.

The forecasts obtained from the two-component models are, most of the time, close to those

Figure 6: AXP returns: weekly realized variance and in-sample variance forecasts during the financial crisis (July 25, 2008 to February 13, 2009)



obtained from the one-component models. Recall that the overnight component is almost suppressed in the one-component models where λ_2 is estimated to be near zero. When $\lambda_1 = \lambda_2 = 1$ is imposed in the one-component model, and thus the overnight returns effectively enter the forecasts, the model tends to overshoot the variance for a longer time after big shocks, for example, in the weeks following 9/11 (see the 1comp model with $\lambda = 1$ in Figure 5). In the one-component model, the overnight returns and the intraday returns are combined linearly, assuming the same decay rate for overnight and intraday returns after a shock occurs. Apparently this leads to estimated decay rates in the one-component model (with $\lambda_1 = \lambda_2 = 1$) that are too small. This appears to be closely related to the problem documented by Harvey (2013) that GARCH models tend to overshoot the variance for a long time after big shocks. He suggests to use the Beta-t-EGARCH model to shrink the big shocks. In our two-component models, we allow for different decay rates for intraday and overnight shocks, which seems to mitigate the problem.

4.2 Out-of-sample model evaluation and comparison

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. Note also that, while the locally stationary models are estimated with a two-sided kernel to estimate the long-run variances,

the kernel estimates never use data from outside the estimation window, so the corresponding forecasts are genuine out-of-sample forecasts.

The GW test allows for a unified treatment of nested and non-nested models. As loss function in the GW test we use the negative out-of-sample Gaussian quasi-loglikelihood value. The results are presented in Tables 6–7. Table 6 reports the average (per observation) loss for each model and the implied ranking. Of the ten highest ranking models, the first nine are 2comp(co,5m) models, with scalar BEKK specifications for the overnight component occupying the highest ranks. The first-ranked model (col. 14) is of this type, and has the highly parametrized weight specification 9. The first non-2comp(co,5m) model is the 1comp(co,5m) model with weight specification 0 (col. 15), at the tenth place. The other 1comp(co,5m) models are also ranked in the best half. The 1comp(co,5m) model with weight specification 0 and $\lambda_1 = \lambda_2 = 1$ (col. 22) is ranked 25th, at the median. The ls2comp(co,5m) models rank below the median. The models that do not use 5-minute returns also rank below the median, except in one case (col. 26, at rank 19). The garch(d), garch(w), d-garch, and d-ewma models (in cols. 35–39 and 50) are all among the lowest-ranked models.

For any pair of models, the GW test comes with a statistic and a p-value of the hypothesis of equal unconditional predictive accuracy (for the chosen loss function). The 50×50 matrices of pairwise GW test statistics and p-values are available in the supplementary appendix. Here we condense the information in the p-value matrix, P, to get a sense of the statistical significance of the model ranking that was obtained. We order the rows and columns of P according to the model ranking, then partition P into blocks of 10×10, and compute the average of each block (with the diagonal elements of P excluded). This gives the matrix in Table 7. We see that, within each block, the GW test typically cannot reject the null of equal predictive accuracy. This is hardly surprising since there are many closely related models. Looking across blocks, the statistical significance naturally increases. The statistical significance of the ranking of the models ranked 1–10 vs. those ranked 21–30, 31–40, and 41–51 is, on average, 11%, 6%, and 2%. Thus, at conventional significance levels, roughly one-third of the models is statistically dominated by the best one-fifth.

Table 7: Out-of-sample Giacomini-White tests, p-values averaged by block

rank	1–10	11–20	21–30	31–40	41–50
1–10	.50				
11–20	.34	.71			
21–30	.11	.36	.59		
31–40	.06	.16	.60	.66	
41–50	.02	.04	.22	.30	.40

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 $m = 1, \ldots, m^0$, and let L_{mt} be the forecast loss of model m 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

Table 6: Out-of-sample Giacomini-White tests, model ranking

column	1	2	3	4	5	6	7	8	9	10	11	12	13
1/2comp	2	2	2	2	2	2	2	2	2	2	2	2	2
loc. stat.	no	no	no	no	no	no	no	no	no	no	no	no	no
returns	co,5m	co,5m	co,5m	co,5m	co,5m	co,5m	co,5m	co,5m	co,5m	co,5m	co,5m	co,5m	co,5m
BEKK	diag	diag	diag	diag	diag	diag	diag	scalar	scalar	scalar	scalar	scalar	scalar
weight	0	1	2	3	5	8	9	0	1	2	3	5	8
average loss	−9.32	−9.30	−9.29	−9.31	−9.35	−9.37	−9.36	−9.35	−9.33	−9.33	−9.34	−9.37	−9.37
rank	12	20	22	15	6	3	5	7	9	11	8	2	4
column	14	15	16	17	18	19	20	21	22	23	24	25	26
1/2comp	2	1	1	1	1	1	1	1	1	2	2	2	2
loc. stat.	no	no	no	no	no	no	no	no	no	no	no	no	no
returns	co,5m	co,5m	co,5m	co,5m	co,5m	co,5m	co,5m	co,5m	co,5m	co,oc	co,oc	co,oc	co,oc
BEKK	scalar									diag	diag	diag	diag
weight	9	0	1	2	3	5	8	9	0	0a	2a	4a	8a
average loss	−9.38	−9.33	−9.32	−9.32	−9.31	−9.31	−9.31	−9.30	−9.26	−9.22	−9.21	−9.24	−9.30
rank	1	10	14	13	18	16	17	21	25	38	41	28	19
column	27	28	29	30	31	32	33	34	35	36	37	38	39
1/2comp	2	2	2	2	1	1	1	1					
loc. stat.	no	no	no	no	no	no	no	no	no	no	no	no	no
returns	co,oc	co,oc	co,oc	co,oc	co,oc	co,oc	co,oc	co,oc	d	d	d	w	d-garch
BEKK	scalar	scalar	scalar	scalar									
weight	0a	2a	4a	8a	0a	2a	4a	8a	0b	1b	4b		
average loss	−9.24	−9.25	−9.23	−9.24	−9.24	−9.24	−9.23	−9.22	−9.20	−9.20	−9.19	−9.18	−9.14
rank	34	27	35	32	33	29	37	40	42	43	44	45	46
column	40	41	42	43	44	45	46	47	48	49	50		
1/2comp	2	2	2	2	2	2	2	2					
loc. stat.	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes			
returns	co,5m	co,5m	co,5m	co,5m	co,5m	co,5m	co,5m	co,oc	d	w	d-ewma		
BEKK	scalar	scalar	scalar	scalar	scalar	scalar	scalar	scalar					
weight	0	1	2	3	5	8	9	0	0	0			
average loss	−9.26	−9.27	−9.27	−9.24	−9.23	−9.24	−9.22	−9.13	−9.10	−9.12	−9.14		
rank	26	24	23	30	36	31	39	48	50	49	47		

Table 8: Out-of-sample model confidence sets

col.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
L_e	**9**	**11**	**13**	**4**	**1**	20	**2**	**16**	**15**	**17**	**5**	**6**	23	**14**	**12**	**7**	**8**	**10**	**3**	19	18	36	42	49	45
L_e^{hf}	**2**	**3**	**5**	**13**	**1**	**14**	**4**	**6**	**10**	**9**	**8**	**7**	19	**17**	**16**	**11**	**12**	**18**	**15**	20	22	48	45	43	46
L_f	**10**	**13**	**14**	**4**	**1**	21	**2**	**15**	**16**	**17**	**5**	**7**	23	**11**	**9**	**6**	**8**	**12**	**3**	19	18	36	49	46	48
L_f^{hf}	**1**	**6**	**5**	**14**	**2**	**18**	**3**	**4**	**11**	**10**	**12**	**9**	21	**17**	**13**	**7**	**8**	**16**	**15**	19	22	47	48	44	46
L_b	**9**	**14**	**15**	**5**	**2**	25	**13**	**12**	**16**	**18**	**6**	**11**	23	**19**	**3**	**7**	**8**	**4**	**1**	**17**	**10**	46	47	43	45
L_b^{hf}	**1**	**11**	**9**	**7**	**2**	18	**5**	**10**	**16**	**15**	**13**	**14**	25	19	**6**	**3**	**4**	**12**	**8**	17	20	47	45	40	43
L_s^{hf}	**14**	23	25	36	20	**8**	27	**2**	**5**	**4**	**7**	**3**	**1**	**6**	**9**	**13**	**12**	37	40	**16**	43	39	24	22	31
L_q	**10**	**14**	20	**16**	**6**	**2**	**5**	**7**	**12**	17	**8**	**4**	**3**	**1**	**9**	19	18	**15**	**11**	21	22	26	35	38	25

col.	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
L_e	50	47	40	41	43	46	44	38	39	32	34	35	48	37	27	22	24	21	25	29	26	33	28	30	31
L_e^{hf}	49	41	34	37	39	42	36	35	40	33	38	47	50	44	27	21	23	24	25	28	26	32	29	31	30
L_f	50	47	42	40	43	44	41	37	39	33	34	35	45	38	27	24	25	20	22	28	26	32	29	30	31
L_f^{hf}	50	43	35	37	40	41	36	34	39	33	38	45	49	42	27	23	24	20	25	28	26	31	29	32	30
L_b	50	42	39	36	41	40	35	34	38	31	33	44	49	37	27	**22**	24	**20**	**21**	28	26	32	30	48	29
L_b^{hf}	49	37	34	35	39	38	33	32	41	36	42	46	50	30	28	22	23	21	24	29	26	31	44	48	27
L_s^{hf}	33	18	**11**	28	30	**15**	**10**	26	42	35	38	44	45	47	21	19	17	29	32	34	41	50	48	46	49
L_q	**13**	32	24	33	37	29	23	31	41	42	44	45	43	46	30	28	27	34	36	39	40	47	49	50	48

Entries: model ranking according to the standardized relative loss, for a given loss function and volatility proxy. Loss functions: Euclidean loss (L_e), Frobenius (L_f), b-loss (L_b), Stein loss (L_s), negative loglikelood loss (L_q). Volatility proxys based on daily returns or on 5-minute and overnight returns (superscript hf). The models are ordered in the same way as in the other tables. The models with entries in boldface are included in the MCS for the corresponding loss function and confidence level $\alpha = 0.25$. The MCS is computed with 1000 block bootstrap replications and an average block length of 10.

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 standardized relative 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. In addition to the negative Gaussian quasi-loglikelihood loss, we also consider four other types of loss functions (Euclidean loss, Frobenius loss, Stein loss, and B-loss), as specified in Laurent, Rombouts, and Violante (2012). We implemented the loss functions with two volatility proxies: one based on the daily returns and the other, labeled hf-loss, based on the 5-minute returns and the overnight returns. Letting

$$R_t = \sum_d r_{td} r_{td}', \qquad RV_t = \sum_{d,i} r_{tdi}^{5\mathrm{m}} r_{tdi}^{5\mathrm{m}\prime} + \sum_d r_{td}^{\mathrm{co}} r_{td}^{\mathrm{co}},$$

the loss specifications are as follows:

(i) Euclidean loss: $L_{e,t} = \mathrm{vech}(R_t - H_t)'\mathrm{vech}(R_t - H_t)$.

(ii) Euclidean hf-loss: $L_{e,t}^{\mathrm{hf}} = \mathrm{vech}(RV_t - H_t)'\mathrm{vech}(RV_t - H_t)$.

(iii) Frobenius loss: $L_{f,t} = \mathrm{tr}[(R_t - H_t)'(R_t - H_t)]$.

(iv) Frobenius hf-loss: $L_{f,t}^{\mathrm{hf}} = \mathrm{tr}[(RV_t - H_t)'(RV_t - H_t)]$.

(v) b loss: $L_{b,t} = \frac{1}{b(b-1)}\mathrm{tr}(R_t^b - H_t^b) - \frac{1}{b-1}\mathrm{tr}[H_t^{b-1}(R_t - H_t)]$, where we set $b = 3$.

(vi) b hf-loss: $L_{b,t}^{\mathrm{hf}} = \frac{1}{b(b-1)}\mathrm{tr}(RV_t^b - H_t^b) - \frac{1}{b-1}\mathrm{tr}[H_t^{b-1}(RV_t - H_t)]$, where we set $b = 3$.

(vii) Stein hf-loss: $L_{s,t}^{\mathrm{hf}} = \mathrm{tr}(H_t^{-1}RV_t) - \log\det H_t^{-1}RV_t - n$.

(viii) Negative loglikelihood loss: $L_{q,t} = (\log\det H_t + r_t' H_t^{-1} r_t + n\log(2\pi))/2$.

For the Stein loss function, we only use RV_t as volatility proxy because $\log\det H_t^{-1}R_t$ is not defined in weeks with two or more non-trading days.

We ran the MCS procedure with confidence level $\alpha = 0.25$ and the average block length and the number of bootstrap replications set to 10 and 1000, respectively. Table 8 reports the model ranking and MCS (indicated as entries in boldface) for each loss function. To keep the table concise we omit the loss function values (they are available in the supplementary appendix) and model descriptors. The models are identified by the same column header as in the other tables.

A glance at Table 8 shows that the model ranking varies considerably across the various loss functions. Nevertheless, for all loss functions, the models in the MCS are nearly all in the cols. 1–21, which correspond to 2comp(co,5m) and 1comp(co,5m) models. By contrast, the 2comp(co,co) and 1comp(co,co) models (cols. 23–34), which use the open-to-close returns instead of the 5-minute returns, are almost never in the MCS. The ls2comp(co,5m) models (cols. 40–46) rank somewhat in the middle but are not in the MCS, apart from a few exceptions. Finally, the 1comp(co,5m) model with $\lambda_1 = \lambda_2 = 1$ (col. 22) is never in the MCS, just as the garch(d), garch(w), d-garch, and d-ewma models (cols. 35–39 and 50).

Several of the loss functions assume the existence of higher moments of the returns, and are thus more sensitive to outliers. For Stein's loss and the negative loglikelihood loss, which are relatively more robust to outliers, the 2comp(co,5m) models with scalar BEKK specification for the overnight component (cols. 8–14) tend to perform best. Among these models, there is no clearcut ranking with regard to the weight specification. The loss functions that are most sensitive to outliers (in particular L_b and L_b^{hf}, which assume finite sixth-order moments) tend

to favor the 1comp(co,5m) models (cols. 15–21) relatively more. This may be explained by the presence of outliers in the overnight retuns and the fact that the contribution of the overnight returns is largely suppressed in the estimated 1comp(co,5m) models.

5 Concluding remarks

We explored several ways to incorporate mixed-frequency returns and overnight returns into multivariate HYBRID BEKK-GARCH models for forecasting the conditional variance matrix of weekly asset returns. The intraday and overnight returns may enter the model either as two separate components or, perhaps closer in spirit to the GARCH approach, as a single component; and they may be weighted using a variety of MIDAS-style weights, or remain unweighted.

We chose the BEKK model for its simplicity. In principle, mixed-frequency and overnight returns might be incorporated in similar ways into other multivariate GARCH models (see Bauwens, Laurent, and Rombouts 2006 for a survey). A natural candidate would be the DCC model proposed by Engle (2002a). This model is somewhat more flexible than the BEKK model, but there is an inconsistency problem with the correlation targeting approach (see Caporin and McAleer 2012 and Aielli 2013) that would have to be overcome.

We did not investigate how to choose the intraday return frequency optimally. Our choice of 5-minute returns is in line with much of the realized volatility literature; see, e.g., Liu, Patton, and Sheppard (2015). To guard against contamination by market microstrucure noise, it is important not to choose a frequency that is too high. For large, frequently traded assets, 5-minute returns appear to be on the safe side and potentially this frequency can be increased. In addition, the optimal sampling frequency may depend the forecasting horizon. It is well possible, for example, that the optimal intraday frequency for forecasting monthly volatility is somewhat lower than for forecasting weekly volatility.

The models that we explored are symmetric in that the volatility responses are independent of the signs of the (intraday and overnight) returns. In the univariate HYBRID GARCH framework, Chen, Ghysels, and Wang (2015) found that symmetric models are always dominated by asymmetric counterparts. An important extension, therefore, would be to allow for asymmetric responses in multivariate mixed-frequency volatility models, possibly using semivariances and semicovariances as recently explored by Bollerslev, Patton, and Quaedvlieg (2018) in the context of multivariate realized GARCH.

In our application to forecast the weekly return variance matrix of four DJIA stocks in the period 1998–2014, we evaluated 50 different models in and out of sample. We found that (i) the two-component versions of the model tend to perform better than the one-component versions according to most criteria; (ii) models that exploit 5-minute return information significantly improve on models that use only weekly, daily, or open-to-close and close-to-open returns, and also on the daily BEKK model applied with multi-period forecasting; (iii) incorporating overnight returns as a separate component improves the model performance; (iv) overnight returns exhibit more persistence, as measured by the GARCH coefficients, than the intraday return component (either as 5-minute returns or as open-to-close returns); (v) MIDAS-style weights tend to have little impact on the model performance. 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 is in the spirit of some earlier suggestions on how to treat overnight returns, tends to perform poorly. The estimated weight given to the 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 particularly poorly (worst of all one- and two-component models that use 5-minute returns).

Our tentative conclusion is that models with two components, one for high-frequency intraday returns and one for the overnight returns, appear a sensible way to make progress on how to forecast low-frequency volatility matrices using the available mixed-frequency data. Stated otherwise, the intraday and overnight returns appear 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.

Appendix

Here we derive the asymptotic variance of $\arg\max_\phi l$ when the unconditional variances are profiled out from l by variance targeting (in the stationary models) or nonparametric smoothing (in the locally stationary models).

A.1 Stationary models

The large-sample properties of variance targeting estimators have been studied, among others, by Francq, Horvath, and Zakoïan (2011) in the univariate case and by Pedersen and Rahbek (2014) in the multivariate case. Here, we view the two-step estimator, $\widehat{\phi}$, as a GMM estimator, so its limit distribution follows from standard GMM theory; see Newey and McFadden (1994). Write the loglikelihood as $l(\psi,\phi) = \sum_t l_t(\psi,\phi)$, where ψ is the vectorized lower triangular part of the unconditional variance matrix (or matrices, i.e., either $\overline{P}$ and $\overline{Q}$, or $\overline{H}$, depending on the model). Let ψ_0 and ϕ_0 be the true values of ψ and ϕ. The estimators $\widehat{\psi}$ and $\widehat{\phi}$ solve the exactly identified moment conditions

$$g(\psi,\phi) = \begin{pmatrix} g_1(\psi,\phi) \\ g_2(\psi,\phi) \end{pmatrix} = 0$$

where $g_2(\psi,\phi) = \partial_\phi l(\psi,\phi)$ (with ∂_ϕ denoting the partial derivative relative to ϕ), $g_1(\psi,\phi) = \sum_t g_{1t}(\psi,\phi)$, and $g_{1t}(\psi,\phi)$ is given below. Then, under regularity conditions,

$$\sqrt{T}\begin{pmatrix} \widehat{\psi}-\psi_0 \\ \widehat{\phi}-\phi_0 \end{pmatrix} \to N(0, F^{-1}VF'^{-1})$$

where $V = \lim_{T\to\infty} \mathrm{Var}(g(\psi_0,\phi_0)/\sqrt{T})$ and

$$F = \operatorname*{plim}_{T\to\infty} \frac{1}{T}\begin{pmatrix} \partial_{\psi'} g_1(\psi_0,\phi_0) & \partial_{\phi'} g_1(\psi_0,\phi_0) \\ \partial_{\psi'} g_2(\psi_0,\phi_0) & \partial_{\phi'} g_2(\psi_0,\phi_0) \end{pmatrix}.$$

It is of interest to note that the use of mixed-frequency data does not complicate the asymptotics. Let $\overline{A}$ by any of the first-step estimates as defined in Section 3. Every such $\overline{A}$ is of the form $\overline{A} =$

$\overline{A}(\phi) = T^{-1}\sum_t W_t(\phi)$ where $W_t(\phi)$ depends on data from week t and possibly on parameters ϕ. Since $\overline{A}$ depends on the intra-week and intraday data only through $W_t(\phi)$, for the theory of inference we are essentially back in the equal-frequency case. Note, further, that $\overline{A}(\phi) = \arg\max_A \sum_t q_t(A,\phi)$ where $q_t(A,\phi) = -\frac{1}{2}\log\det A - \frac{1}{2}\mathrm{tr}(A^{-1}W_t(\phi))$, so $\overline{A}(\phi)$ solves

$$0 = \sum\nolimits_t \partial_A q_t(A,\phi) = \sum_t \left(-\frac{1}{2}A^{-1} + \frac{1}{2}A^{-1}W_t(\phi)A^{-1}\right).$$

For the 1comp models, $\overline{A}$ corresponds to $\overline{H}$ and $g_{1t}(\psi,\phi)$ to $\mathrm{vech}\partial_A q_t(A,\phi)$. For the 2comp models, $\overline{A}$ corresponds to $\overline{P}$ and $\overline{Q}$, and $g_{1t}(\psi,\phi)$, correspondingly, has two subvectors of the form $\mathrm{vech}\partial_A q_t(A,\phi)$.

A.2 Locally stationary models

Consider the locally stationary two-component model based on the 5-minute and the overnight returns (the analysis for the other locally stationary models is similar). In this model,

$$r_{tdi}^{5\mathrm{m}} = M_t^{1/2}u_{tdi}^{5\mathrm{m}}, \qquad r_t^{\mathrm{oc}} = \sum_{d,i} r_{tdi}^{5\mathrm{m}} = M_t^{1/2}\sum_{d,i} u_{tdi}^{5\mathrm{m}} = M_t^{1/2}u_t^{\mathrm{oc}},$$
$$r_{td}^{\mathrm{co}} = N_t^{1/2}u_{td}^{\mathrm{co}}, \qquad r_t^{\mathrm{co}} = \sum_{d} r_{td}^{\mathrm{co}} = N_t^{1/2}\sum_{d} u_{td}^{\mathrm{co}} = N_t^{1/2}u_t^{\mathrm{co}},$$

where $M_t = M(t/T)$ and $N_t = N(t/T)$ for some deterministic functions $M(u)$ and $N(u)$, $u \in [0,1]$, and $u_{tdi}^{5\mathrm{m}}$ and u_{td}^{co} are serially uncorrelated random vectors. For identification we impose

$$E\left(\sum_{d,i} u_{tdi}^{5\mathrm{m}}u_{tdi}^{5\mathrm{m}\prime}\right) = E\left(\sum_{d} u_{td}^{\mathrm{co}}u_{td}^{\mathrm{co}\prime}\right) = I_n.$$

The specification of H_t is

$$H_t = \lambda_1 M_t^{1/2}P_t^* M_t^{1/2} + \lambda_2 N_t^{1/2}Q_t^* N_t^{1/2}$$

where

$$P_t^* = I - A_1A_1' - B_1B_1' + A_1M_{t-1}^{-1/2}RV_{t-1}^{\mathrm{oc}}M_{t-1}^{-1/2}A_1' + B_1P_{t-1}^*B_1',$$
$$Q_t^* = I - A_2A_2' - B_2B_2' + A_2N_{t-1}^{-1/2}RV_{t-1}^{\mathrm{co}}N_{t-1}^{-1/2}A_2' + B_2Q_{t-1}^*B_2',$$

(we allow for a non-diagonal BEKK structure with parameter matrices A_1, B_1, A_2, B_2) and RV_t^{oc} and RV_t^{co} are weight-dependent realized variance measures,

$$RV_t^{\mathrm{oc}} = \sum_{d,i}\omega_{di}^{5\mathrm{m}}r_{tdi}^{5\mathrm{m}}r_{tdi}^{5\mathrm{m}\prime}, \qquad RV_t^{\mathrm{co}} = \sum_{d}\omega_d^{\mathrm{co}}r_{td}^{\mathrm{co}}r_{td}^{\mathrm{co}\prime}.$$

We assume that the weight functions satisfy

$$E\left(\sum_{d,i}\omega_{di}^{5\mathrm{m}}u_{tdi}^{5\mathrm{m}}u_{tdi}^{5\mathrm{m}\prime}\right) = E\left(\sum_{d}\omega_d^{\mathrm{co}}u_{td}^{\mathrm{co}}u_{td}^{\mathrm{co}\prime}\right) = I_n,$$

so that $E(M_t^{-1/2}RV_t^{\mathrm{oc}}M_t^{-1/2}) = E(N_t^{-1/2}RV_t^{\mathrm{co}}N_t^{-1/2}) = I_n$. The long-run variance matrices are estimated as $\widetilde{M}_t = \widetilde{M}(t/T)$ and $\widetilde{N}_t = \widetilde{N}(t/T)$, where

$$\widetilde{M}(u) = \frac{\sum_{t=1}^T K_h(u-t/T)r_t^{\mathrm{oc}}r_t^{\mathrm{oc}\prime}}{\sum_{t=1}^T K_h(u-t/T)}, \qquad \widetilde{N}(u) = \frac{\sum_{t=1}^T K_h(u-t/T)r_t^{\mathrm{co}}r_t^{\mathrm{co}\prime}}{\sum_{t=1}^T K_h(u-t/T)},$$

for $u \in [0,1]$. Given $\widetilde{M}_t$ and $\widetilde{N}_t$, let

$$\begin{aligned}
\widetilde{P}_t^* &= I - A_1A_1' - B_1B_1' + A_1\widetilde{M}_{t-1}^{-1/2}RV_{t-1}^{\text{oc}}\widetilde{M}_{t-1}^{-1/2}A_1' + B_1\widetilde{P}_{t-1}^*B_1', \\
\widetilde{Q}_t^* &= I - A_2A_2' - B_2B_2' + A_2\widetilde{N}_{t-1}^{-1/2}RV_{t-1}^{\text{co}}\widetilde{N}_{t-1}^{-1/2}A_2' + B_2\widetilde{Q}_{t-1}^*B_2', \\
\widetilde{H}_t &= \lambda_1\widetilde{M}_t^{1/2}\widetilde{P}_t^*\widetilde{M}_t^{1/2} + \lambda_2\widetilde{N}_t^{1/2}\widetilde{Q}_t^*\widetilde{N}_t^{1/2}.
\end{aligned}$$

The parameters appearing in the short-run variances, ϕ, are now estimated in the same way as in the models with constant long-run variances. The estimator is $\widetilde{\phi} = \arg\max_\phi \widetilde{l}(\phi)$ where $\widetilde{l}(\phi) = \sum_t \widetilde{l}_t(\phi)$ and $\widetilde{l}_t(\phi)$ is $l_t(\phi)$ but with $\widetilde{H}_t$ instead of H_t (and I_n instead of $\overline{Q}, \overline{P}, \overline{H}$).

We make the same assumptions as in Hafner and Linton (2010), adapted to two locally stationary processes. Define $\varepsilon_t^{\text{oc}} = P_t^{*-1/2}u_t^{\text{oc}}$ and $\varepsilon_t^{\text{co}} = Q_t^{*-1/2}u_t^{\text{co}}$.

(A1) $M(u)$ and $N(u)$ are uniformly positive definite and twice continuously differentiable on $[0,1]$.

(A2) $\varepsilon_t^{\text{oc}}$ and $\varepsilon_t^{\text{co}}$ have positive lower semi-continuous densities in a neighborhood of zero. The initialization of $\{r_t^{\text{oc}}, r_t^{\text{co}}, P_t, Q_t\}$ is independent of $\{\varepsilon_t^{\text{oc}}\}$ and $\{\varepsilon_t^{\text{co}}\}$.

(A3) $\det A_j \neq 0$ and $\rho(B_j) < 1$ for $j = 1, 2$, where $\rho(\cdot)$ is the spectral radius.

(A4) The parameter space Φ is compact.

(A5) $\{u_t^{\text{oc}}\}$ and $\{u_t^{\text{co}}\}$ are strictly stationary and ergodic, and $E||u_t^{\text{oc}}||^6 < \infty$ and $E||u_t^{\text{co}}||^6 < \infty$.

(A5) $E||\varepsilon_t^{\text{oc}}||^4 < \infty$, $E||\varepsilon_t^{\text{co}}||^4 < \infty$, and $\text{var}(\varepsilon_t^{\text{oc}}) = \text{var}(\varepsilon_t^{\text{co}}) = I_n$.

(A7) The model is identifiable: for any $\phi \in \Phi$, if $H_t(\phi) = H_t(\phi_0)$ a.s., then $\phi = \phi_0$.

(A8) ϕ_0 is an interior point of Φ.

(A9) $K(\cdot)$ is symmetric about zero and has compact support.

(A10) $h \to 0$ as $T \to \infty$ such that $Th^2 \to \infty$ and $Th^4 \to 0$.

Under these assumptions, we show that $\widetilde{\phi}$ is consistent and asymptotically normal. The derivation is similar to the proof of Theorem 1 in Hafner and Linton (2010).

Since

$$\begin{aligned}
\sup_{\phi\in\Phi}\left|\frac{1}{T}\widetilde{l}(\phi) - \frac{1}{T}El(\phi)\right| &\leq \sup_{\phi\in\Phi}\left|\frac{1}{T}\widetilde{l}(\phi) - \frac{1}{T}l(\phi)\right| + \sup_{\phi\in\Phi}\left|\frac{1}{T}l(\phi) - \frac{1}{T}El(\phi)\right| \\
&= \sup_{\phi\in\Phi}\left|\frac{1}{T}\widetilde{l}(\phi) - \frac{1}{T}l(\phi)\right| + o_p(1),
\end{aligned}$$

the consistency of $\widetilde{\phi}$ follows if it holds that $\sup_{\phi\in\Phi}\left|\frac{1}{T}\widetilde{l}(\phi) - \frac{1}{T}l(\phi)\right| = o_p(1)$. We have

$$\begin{aligned}
\sup_{\phi\in\Phi}\left|\frac{1}{T}\widetilde{l}(\phi) - \frac{1}{T}l(\phi)\right| &\leq \sup_{\phi\in\Phi}\left|\frac{1}{2T}\sum_{t=1}^T \log\det(\widetilde{H}_tH_t^{-1})\right| + \sup_{\phi\in\Phi}\left|\frac{1}{2T}\sum_{t=1}^T r_t'(\widetilde{H}_t^{-1} - H_t^{-1})r_t\right| \\
&= \sup_{\phi\in\Phi}\left|\frac{1}{2T}\sum_{t=1}^T \text{tr}[(\widetilde{H}_t - H_t)H_t^{-1}]\right| \\
&\quad + \sup_{\phi\in\Phi}\left|\frac{1}{2T}\sum_{t=1}^T r_t'H_t^{-1}(\widetilde{H}_t - H_t)H_t^{-1}r_t\right| + o_p(1).
\end{aligned}$$

Further,

$$\max_{1\leq t\leq T}\sup_{\phi\in\Phi}\left\|\widetilde{H}_t - H_t\right\| \leq \max_{1\leq t\leq T}\sup_{\phi\in\Phi}\left\|\lambda_1(\widetilde{M}_t^{1/2}\widetilde{P}_t^*\widetilde{M}_t^{1/2} - M_t^{1/2}P_t^*M_t^{1/2})\right\|$$

$$+ \max_{1\le t\le T} \sup_{\phi\in\Phi} \left\| \lambda_2(\widetilde{N}_t^{1/2}\widetilde{Q}_t^*\widetilde{N}_t^{1/2} - N_t^{1/2}Q_t^*N_t^{1/2}) \right\|,$$

where each term on the right-hand side is $o_p(1)$, and $\min_{1\le t\le T}\inf_{\phi\in\Phi}\mathrm{eig}_{\min}(H_t) > 0$, where $\mathrm{eig}_{\min}(\cdot)$ denotes the smallest eigenvalue. Therefore, $\sup_{\phi\in\Phi}\left|\frac{1}{T}\widetilde{l}(\phi) - \frac{1}{T}l(\phi)\right| = o_p(1)$ and $\widetilde{\phi}$ is consistent.

We now show asymptotic normality. A standard expansion gives

$$\sqrt{T}(\widetilde{\phi} - \phi_0) = -\left(\frac{1}{T}\frac{\partial^2\widetilde{l}(\phi_0)}{\partial\phi\partial\phi'}\right)^{-1}\frac{1}{\sqrt{T}}\frac{\partial\widetilde{l}(\phi_0)}{\partial\phi} + o_p(1),$$

where $\frac{1}{T}\frac{\partial^2\widetilde{l}(\phi_0)}{\partial\phi\partial\phi'} \xrightarrow{p} J = \frac{1}{T}\frac{\partial^2 El(\phi_0)}{\partial\phi\partial\phi'}$ and $\frac{1}{\sqrt{T}}\frac{\partial\widetilde{l}(\phi_0)}{\partial\phi}$ will be shown to be asymptotically normal. We have

$$\begin{aligned}
\frac{\partial l_t}{\partial\phi_i} &= -\frac{1}{2}\mathrm{tr}\left[\left(I_n - r_t r_t' H_t^{-1}\right)\frac{\partial H_t}{\partial\phi_i}H_t^{-1}\right],\\
\frac{\partial\widetilde{l}_t}{\partial\phi_i} &= -\frac{1}{2}\mathrm{tr}\left[(I_n - r_t r_t'\widetilde{H}_t^{-1})\frac{\partial\widetilde{H}_t}{\partial\phi_i}\widetilde{H}_t^{-1}\right],
\end{aligned}$$

and so

$$\begin{aligned}
\frac{1}{\sqrt{T}}\left(\frac{\partial\widetilde{l}}{\partial\phi_i} - \frac{\partial l}{\partial\phi_i}\right) &= -\frac{1}{2\sqrt{T}}\sum_{t=1}^{T}\mathrm{tr}\left[\left(I_n - r_t r_t' H_t^{-1}\right)\left(\frac{\partial\widetilde{H}_t}{\partial\phi_i} - \frac{\partial H_t}{\partial\phi_i}\right)H_t^{-1}\right]\\
&\quad + \frac{1}{2\sqrt{T}}\sum_{t=1}^{T}\mathrm{tr}\left[\left(I_n - r_t r_t' H_t^{-1}\right)\frac{\partial H_t}{\partial\phi_i}H_t^{-1}(\widetilde{H}_t - H_t)H_t^{-1}\right]\\
&\quad - \frac{1}{2\sqrt{T}}\sum_{t=1}^{T}\mathrm{tr}\left[r_t r_t' H_t^{-1}(\widetilde{H}_t - H_t)H_t^{-1}\frac{\partial H_t}{\partial\phi_i}H_t^{-1}\right] + o_p(1).
\end{aligned}$$

The first two terms on the right-hand side are $o_p(1)$ because $I_N - r_t r_t' H_t^{-1}$ is a martingale difference sequence and $\frac{\partial\widetilde{H}_t}{\partial\phi_i} - \frac{\partial H_t}{\partial\phi_i}$ and $\widetilde{H}_t - H_t$ are $o_p(1)$. Hence, letting $\Gamma_t = H_t^{-1}\frac{\partial H_t}{\partial\phi_i}H_t^{-1}r_t r_t' H_t^{-1}$ and using

$$\begin{aligned}
\widetilde{H} - H_t &= \lambda_1(\widetilde{M}_t^{1/2} - M_t^{1/2})P_t^* M_t^{1/2} + \lambda_1 M_t^{1/2}(\widetilde{P}_t^* - P_t^*)M_t^{1/2}\\
&\quad + \lambda_1 M_t^{1/2}P_t^*(\widetilde{M}_t^{1/2} - M_t^{1/2}) + \lambda_2(\widetilde{N}_t^{1/2} - N_t^{1/2})Q_t^* N_t^{1/2}\\
&\quad + \lambda_2 N_t^{1/2}(\widetilde{Q}_t^* - Q_t^*)N_t^{1/2} + \lambda_2 N_t^{1/2}Q_t^*(\widetilde{N}_t^{1/2} - N_t^{1/2}) + o_p(1),
\end{aligned}$$

it follows that

$$\begin{aligned}
\frac{1}{\sqrt{T}}\left(\frac{\partial\widetilde{l}}{\partial\phi_i} - \frac{\partial l}{\partial\phi_i}\right) &= -\frac{1}{2\sqrt{T}}\sum_{t=1}^{T}\mathrm{tr}\left[\Gamma_t(\widetilde{H}_t - H_t)\right] + o_p(1)\\
&= -\frac{\lambda_1}{2\sqrt{T}}\sum_{t=1}^{T}\mathrm{tr}\left[(P_t^* M_t^{1/2}\Gamma_t + \Gamma_t M_t^{1/2}P_t^*)(\widetilde{M}_t^{1/2} - M_t^{1/2})\right]\\
&\quad - \frac{\lambda_1}{2\sqrt{T}}\sum_{t=1}^{T}\mathrm{tr}\left[M_t^{1/2}\Gamma_t M_t^{1/2}(\widetilde{P}_t^* - P_t^*)\right]\\
&\quad - \frac{\lambda_2}{2\sqrt{T}}\sum_{t=1}^{T}\mathrm{tr}\left[(Q_t^* N_t^{1/2}\Gamma_t + \Gamma_t N_t^{1/2}Q_t^*)(\widetilde{N}_t^{1/2} - N_t^{1/2})\right]
\end{aligned}$$

$$
\begin{aligned}
&- \frac{\lambda_2}{2\sqrt{T}} \sum_{t=1}^{T} \operatorname{tr}\left[N_t^{1/2}\Gamma_t N_t^{1/2}(\widetilde{Q}_t^* - Q_t^*)\right] + o_p(1) \\
&= D_1 + D_2 + D_3 + D_4 + o_p(1) \qquad \text{(say)}.
\end{aligned}
$$

We now approximate D_1 and D_2. Let $V_t^{\mathrm{oc}} = r_t^{\mathrm{oc}} r_t^{\mathrm{oc}\prime} - M_t$ and $V_t^{\mathrm{co}} = r_t^{\mathrm{co}} r_t^{\mathrm{co}\prime} - N_t$. Write $\widetilde{M}_t - M_t = Z_1 + Z_2$ where

$$
Z_1 = \frac{\sum_{s=1}^{T} K_h((t-s)/T)\left(r_s^{\mathrm{oc}} r_s^{\mathrm{oc}\prime} - M(s/T)\right)}{\sum_{s=1}^{T} K_h((t-s)/T)} = \frac{1}{T}\sum_{s=1}^{T} K_h((t-s)/T) V_s^{\mathrm{oc}} + o_p(T^{-1/2}h^{-1/2}),
$$

$$
Z_2 = \frac{\sum_{s=1}^{T} K_h((t-s)/T)\left(M(s/T) - M(t/T)\right)}{\sum_{s=1}^{T} K_h((t-s)/T)} = \frac{h^2}{2} M''(t/T) \int u^2 K(u) du + o(h^2),
$$

using $T^{-1}\sum_{s=1}^{T} K_h((t-s)/T) = 1 + o(1)$, $E(Z_1) = 0$, and $Z_1 = O_p(T^{-1/2}h^{-1/2})$; Z_2 is the kernel estimation bias. It follows that

$$
\widetilde{M}_t - M_t = \frac{1}{T}\sum_{s=1}^{T} K_h((t-s)/T) V_s^{\mathrm{oc}} + o_p(T^{-1/2}h^{-1/2}),
$$

since $Th^4 \to 0$. Hence, using

$$
\widetilde{M}_t^{1/2} - M_t^{1/2} = \frac{1}{2} M_t^{-1/4}(\widetilde{M}_t - M_t) M_t^{-1/4} + o_p(T^{-1/2}),
$$

we obtain

$$
\begin{aligned}
D_1 &= -\frac{\lambda_1}{4\sqrt{T}} \sum_{t=1}^{T} \operatorname{tr}\left[(P_t^* M_t^{1/2}\Gamma_t + \Gamma_t M_t^{1/2} P_t^*) M_t^{-1/4}\left(\frac{1}{T}\sum_{s=1}^{T} K_h((t-s)/T) V_s^{\mathrm{oc}}\right) M_t^{-1/4}\right] + o_p(1) \\
&= -\frac{\lambda_1}{4T^{3/2}} \sum_{s=1}^{T}\sum_{t=1}^{T} \operatorname{tr}\left[M_t^{-1/4}(P_t^* M_t^{1/2}\Gamma_t + \Gamma_t M_t^{1/2} P_t^*) M_t^{-1/4} K_h((t-s)/T) V_s^{\mathrm{oc}}\right] + o_p(1) \\
&= -\frac{\lambda_1}{4\sqrt{T}} \sum_{s=1}^{T} \Psi_1^{\mathrm{oc}}(s)' \mathrm{vec} V_s^{\mathrm{oc}} + o_p(1)
\end{aligned}
$$

where

$$
\Psi_1^{\mathrm{oc}}(s) = \frac{1}{T}\sum_{t=1}^{T} K_h((t-s)/T) \mathrm{vec} M_t^{-1/4}\left(P_t^* M_t^{1/2}\Gamma_t + \Gamma_t M_t^{1/2} P_t^*\right) M_t^{-1/4}.
$$

From Hafner and Linton (2010, Eq. (36)),

$$
\begin{aligned}
\widetilde{P}_t^* - P_t^* = &-\frac{1}{2}\sum_{j=1}^{t-1} B_1^{j-1} A_1 M_{t-j}^{-3/4}(\widetilde{M}_{t-j} - M_{t-j}) M_{t-j}^{-3/4} R V_{t-j}^{\mathrm{oc}} M_{t-j}^{-1/2} A_1' B_1^{j-1\prime} \\
&- \frac{1}{2}\sum_{j=1}^{t-1} B_1^{j-1} A_1 M_{t-j}^{-1/2} R V_{t-j}^{\mathrm{oc}} M_{t-j}^{-3/4}(\widetilde{M}_{t-j} - M_{t-j}) M_{t-j}^{-3/4} A_1' B_1^{j-1\prime} + o_p(T^{-1/2}),
\end{aligned}
$$

so

$$
D_2 = -\frac{\lambda_1}{2\sqrt{T}} \sum_{t=1}^{T} \operatorname{tr} M_t^{1/2}\Gamma_t M_t^{1/2}(\widetilde{P}_t^* - P_t^*)
$$

$$
\begin{aligned}
&= \frac{\lambda_1}{4\sqrt{T}} \sum_{t=1}^{T} \sum_{j=1}^{t-1} \Psi_{2,t,j}^{\mathrm{oc}\prime} \mathrm{vec}(\widetilde{M}_{t-j} - M_{t-j}) + o_p(1) \\
&= \frac{\lambda_1}{4\sqrt{T}} \sum_{t=1}^{T} \sum_{j=1}^{t-1} \Psi_{2,t,j}^{\mathrm{oc}\prime} \left(\frac{1}{T} \sum_{s=1}^{T} K_h \left(\frac{t-j-s}{T} \right) \mathrm{vec} V_s^{\mathrm{oc}} \right) + o_p(1) \\
&= \frac{\lambda_1}{4T^{3/2}} \sum_{s=1}^{T} \sum_{t=1}^{T} \sum_{j=1}^{t-1} K_h \left(\frac{t-j-s}{T} \right) \Psi_{2,t,j}^{\mathrm{oc}\prime} \mathrm{vec} V_s^{\mathrm{oc}} + o_p(1) \\
&= \frac{\lambda_1}{4\sqrt{T}} \sum_{s=1}^{T} \Psi_2^{\mathrm{oc}}(s)' \mathrm{vec} V_s^{\mathrm{oc}} + o_p(1)
\end{aligned}
$$

where

$$
\begin{aligned}
\Psi_{2,t,j}^{\mathrm{oc}} &= \mathrm{vec} M_{t-j}^{-3/4} R V_{t-j}^{\mathrm{oc}} M_{t-j}^{-1/2} A_1' B_1^{j-1\prime} M_t^{1/2} \Gamma_t' M_t^{1/2} B_1^{j-1} A_1 M_{t-j}^{-3/4} \\
&\quad + \mathrm{vec} M_{t-j}^{-3/4} A_1' B_1^{j-1\prime} M_t^{1/2} \Gamma_t' M_t^{1/2} B_1^{j-1} A_1 M_{t-j}^{-1/2} R V_{t-j}^{\mathrm{oc}} M_{t-j}^{-3/4}, \\
\Psi_2^{\mathrm{oc}}(s) &= \frac{1}{T} \sum_{t=1}^{T} \sum_{j=1}^{t-1} K_h \left(\frac{t-j-s}{T} \right) \Psi_{2,t,j}^{\mathrm{oc}}.
\end{aligned}
$$

The terms D_3 and D_4 are analogous to D_1 and D_2 and can be approximated similarly. Hence the score can be approximated as

$$
\frac{1}{\sqrt{T}} \frac{\partial \widetilde{l}}{\partial \phi} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} g_t + o_p(1)
$$

where g_t is a vector with ith element

$$
g_{it} = \frac{\partial l_t}{\partial \phi_i} - \frac{\lambda_1}{4} (\Psi_{1,i}^{\mathrm{oc}}(t) - \Psi_{2,i}^{\mathrm{oc}}(t))' \mathrm{vec} V_t^{\mathrm{oc}} - \frac{\lambda_2}{4} (\Psi_{1,i}^{\mathrm{co}}(t) - \Psi_{2,i}^{\mathrm{co}}(t))' \mathrm{vec} V_t^{\mathrm{co}},
$$

and $\Psi_{1,i}^{\mathrm{oc}}$ is Ψ_1^{oc} corresponding to ϕ_i etc. It follows that

$$
\begin{aligned}
\frac{1}{\sqrt{T}} \frac{\partial \widetilde{l}}{\partial \phi} &\xrightarrow{d} N(0, \Omega), \qquad \Omega = \sum_{j=-\infty}^{\infty} E(g_t g_{t-j}), \\
\sqrt{T}(\widetilde{\phi} - \phi) &\xrightarrow{d} N(0, J^{-1} \Omega J^{-1}),
\end{aligned}
$$

where J can be estimated by $\frac{1}{T} \frac{\partial^2 \widetilde{l}(\widetilde{\phi})}{\partial \phi \partial \phi'}$ and Ω by a Newey-West estimator.

Acknowledgements We are grateful to two referees and the editors for helpful comments. Wu acknowledges the financial support from the National Natural Science Foundation of China (project code 71803080).

References

Ahoniemi, K. and M. Lanne (2013). Overnight stock returns and realized volatility. *International Journal of Forecasting 29*(4), 592–604.

Aielli, G. P. (2013). Dynamic conditional correlation: On properties and estimation. *Journal of Business & Economic Statistics 31*(3), 282–299.

Andersen, T. G., T. Bollerslev, and X. Huang (2011). A reduced form framework for modeling volatility of speculative prices based on realized variation measures. *Journal of Econometrics 160*(1), 176–189.

Andreou, E. (2016). On the use of high frequency measures of volatility in MIDAS regressions. *Journal of Econometrics 193*(2), 367–389.

Andreou, E., E. Ghysels, and A. Kourtellos (2010). Regression models with mixed sampling frequencies. *Journal of Econometrics 158*(2), 246–261.

Andreou, E., E. Ghysels, and A. Kourtellos (2013). Should macroeconomic forecasters use daily financial data and how? *Journal of Business & Economic Statistics 31*(2), 240–251.

Bai, J., E. Ghysels, and J. H. Wright (2013). State space models and MIDAS regressions. *Econometric Reviews 32*(7), 779–813.

Bauwens, L., S. Laurent, and J. V. Rombouts (2006). Multivariate GARCH models: a survey. *Journal of applied econometrics 21*, 79–109.

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.

Bollerslev, T. and E. Ghysels (1996). Periodic autoregressive conditional heteroscedasticity. *Journal of Business & Economic Statistics 14*(2), 139–151.

Bollerslev, T., A. J. Patton, and R. Quaedvlieg (2018). Multivariate leverage effects and realized semicovariance GARCH models. Available at SSRN: https://ssrn.com/abstract=3164361.

Caporin, M. and M. McAleer (2012). Do we really need both BEKK and DCC? A tale of two multivariate GARCH models. *Journal of Economic Surveys 26*(4), 736–751.

Caporin, M., E. Rossi, and P. Santucci de Magistris (2017). Chasing volatility: A persistent multiplicative error model with jumps. *Journal of Econometrics 198*(1), 122–145.

Chen, X., E. Ghysels, and F. Wang (2011). HYBRID GARCH models and intra-daily return periodicity. *Journal of Time Series Econometrics 3*(1), 1–28.

Chen, X., E. Ghysels, and F. Wang (2015). HYBRID-GARCH: A generic class of models for volatility predictions using high frequency data. *Statistica Sinica 25*(2), 759–786.

Clements, M. P. and A. B. Galvão (2008). Macroeconomic forecasting with mixed-frequency data. *Journal of Business & Economic Statistics 26*(4), 546–554.

Clements, M. P. and A. B. Galvão (2009). Forecasting US output growth using leading indicators: An appraisal using MIDAS models. *Journal of Applied Econometrics 24*(7), 1187–1206.

Colacito, R., R. F. Engle, and E. Ghysels (2011). A component model for dynamic correlations. *Journal of Econometrics 164*(1), 45–59.

De Lira Salvatierra, I. and A. J. Patton (2015). Dynamic copula models and high frequency data. *Journal of Empirical Finance 30*, 120–135.

Engle, R. F. (2002a). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics 20*(3), 339–350.

Engle, R. F. (2002b). New frontiers for ARCH models. *Journal of Applied Econometrics 17*(5), 425–446.

Engle, R. F. and G. M. Gallo (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics 131*(1), 3–27.

Engle, R. F., E. Ghysels, and B. Sohn (2013). Stock market volatility and macroeconomic fundamentals. *Review of Economics and Statistics 95*(3), 776–797.

Engle, R. F. and K. F. Kroner (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory 11*(1), 122–150.

Engle, R. F. and G. Lee (1999). A long-run and short-run component model of stock return volatility. In R. F. Engle and H. White (Eds.), *Causality, Cointegration, and Forecasting: A Festschrift in Honour of Clive W.J. Granger*, pp. 475–497. Oxford University Press.

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*(3), 1187–1222.

Foroni, C., M. Marcellino, and C. Schumacher (2015). Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials. *Journal of the Royal Statistical Society: Series A 178*(1), 57–82.

Forsberg, L. and E. Ghysels (2007). Why do absolute returns predict volatility so well? *Journal of Financial Econometrics 5*(1), 31–67.

Francq, C., L. Horvath, and J.-M. Zakoïan (2011). Merits and drawbacks of variance targeting in GARCH models. *Journal of Financial Econometrics 9*(4), 619–656.

Gagliardini, P., E. Ghysels, and M. Rubin (2017). Indirect inference estimation of mixed frequency stochastic volatility state space models using MIDAS regressions and ARCH models. *Journal of Financial Econometrics 15*(4), 509–560.

Ghysels, E. (2016). Macroeconomics and the reality of mixed frequency data. *Journal of Econometrics 193*(2), 294–314.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2005). There is a risk-return trade-off after all. *Journal of Financial Economics 76*(3), 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*(1–2), 59–95.

Ghysels, E. and A. Sinko (2011). Volatility forecasting and microstructure noise. *Journal of Econometrics 160*(1), 257–271.

Ghysels, E., A. Sinko, and R. Valkanov (2007). MIDAS regressions: Further results and new directions. *Econometric Reviews 26*(1), 53–90.

Ghysels, E. and B. Sohn (2009). Which power variation predicts volatility well? *Journal of Empirical Finance 16*(4), 686–700.

Ghysels, E. and R. Valkanov (2012). Forecasting volatility with midas. In L. Bauwens, C. Hafner, and S. Laurent (Eds.), *Handbook of Volatility Models and Their Applications*, pp. 383–401. Wiley.

Giacomini, R. and H. White (2006). Tests of conditional predictive ability. *Econometrica 74*(6), 1545–1578.

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

Guérin, P. and M. Marcellino (2013). Markov-switching MIDAS models. *Journal of Business & Economic Statistics 31*(1), 45–56.

Hafner, C. M. and O. Linton (2010). Efficient estimation of a multivariate multiplicative volatility model. *Journal of Econometrics 159*(1), 55–73.

Hansen, P. R., Z. Huang, and H. H. Shek (2012). Realized GARCH: A joint model for returns and realized measures of volatility. *Journal of Applied Econometrics 27*(6), 877–906.

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*(4), 525–554.

Hansen, P. R. and A. Lunde (2006). Consistent ranking of volatility models. *Journal of Econometrics 131*(1–2), 97–121.

Hansen, P. R., A. Lunde, and J. M. Nason (2011). The model confidence set. *Econometrica 79*(2), 453–497.

Hansen, P. R., A. Lunde, and V. Voev (2014). Realized beta GARCH: A multivariate GARCH model with realized measures of volatility. *Journal of Applied Econometrics 29*(5), 774–799.

Harvey, A. C. (2013). *Dynamic models for volatility and heavy tails: With applications to financial and economic time series*, Volume 52 of *Econometric Society Monographs*. Cambridge University Press.

Kuzin, V., M. Marcellino, and C. Schumacher (2011). MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the euro area. *International Journal of Forecasting 27*(2), 529–542.

Laurent, S., J. V. Rombouts, and F. Violante (2012). On the forecasting accuracy of multivariate GARCH models. *Journal of Applied Econometrics 27*(6), 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), 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*(5), 447–464.

Linton, O. B. and J. Wu (2017). A coupled component GARCH model for intraday and overnight volatility. USC-INET Research Paper No. 18-01.

Liu, L. Y., A. J. Patton, and K. Sheppard (2015). Does anything beat 5-minute RV? a comparison of realized measures across multiple asset classes. *Journal of Econometrics 187*(1), 293–311.

Marcellino, M., M. Porqueddu, and F. Venditti (2016). Short-term GDP forecasting with a mixed-frequency dynamic factor model with stochastic volatility. *Journal of Business & Economic Statistics 34*(1), 118–127.

Marcellino, M. and C. Schumacher (2010). Factor MIDAS for nowcasting and forecasting with ragged-edge data: A model comparison for German GDP. *Oxford Bulletin of Economics and Statistics 72*(4), 518–550.

Marcellino, M. and V. Sivec (2016). Monetary, fiscal and oil shocks: Evidence based on mixed frequency structural FAVARs. *Journal of Econometrics 193*(2), 335–348.

Martens, M. (2002). Measuring and forecasting S&P 500 index-futures volatility using high-frequency data. *Journal of Futures Markets 22*(6), 497–518.

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. In R. F. Engle and D. McFadden (Eds.), *Handbook of Econometrics*, Volume 4, Chapter 36, pp. 2111–2245. Elsevier.

Noureldin, D., N. Shephard, and K. Sheppard (2012). Multivariate high-frequency-based volatility (heavy) models. *Journal of Applied Econometrics 27*(6), 907–933.

Oh, D. H. and A. J. Patton (2016). High-dimensional copula-based distributions with mixed frequency data. *Journal of Econometrics 193*(2), 349–366.

Pedersen, R. S. and A. Rahbek (2014). Multivariate variance targeting in the BEKK-GARCH model. *The Econometrics Journal 17*(1), 24–55.

Pettenuzzo, D., A. Timmermann, and R. Valkanov (2016). A MIDAS approach to modeling first and second moment dynamics. *Journal of Econometrics 193*(2), 315–334.

Rodríguez-Poo, J. M. and O. Linton (2001). Nonparametric factor analysis of residual time series. *Test 10*(1), 161–182.

Schorfheide, F. and D. Song (2015). Real-time forecasting with a mixed-frequency VAR. *Journal of Business & Economic Statistics 33*(3), 366–380.

Shephard, N. and K. Sheppard (2010). Realising the future: forecasting with high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics 25*(2), 197–231.