UNIVERSITY OF TÜBINGEN

SCHOOL OF BUSINESS AND ECONOMICS

FACULTY OF ECONOMETRICS, STATISTICS AND EMPIRICAL ECONOMICS

Predicting Multivariate Realized Return Volatility using High-Frequency Stock Price Data

Term paper for the
Master Seminar on Econometrics
Summer Semester 2014

Supervisor(s): Tobias Langen (Dipl.-Volkswirt), Dr. Thomas Dimpfl

Author: Andreas Lay
Degree program: MSc Quantitative Economics
Matriculation no.: 3447330

Address: Bismarckstraße 38/1
72072 Tübingen
Phone: +49(0)1756165666
Email address: andreas.lay@student.uni-tuebingen.de

Date of submission: July 4, 2014

Contents

List of Figures

List of Tables

Abstract

In this paper, we estimate three models to forecast multivariate realized volatility: A VARFIMA$(1, \delta, 1)$ model, a Heterogeneous Autoregressive (HAR) model and a Generalized Orthogonal GARCH (GO-GARCH) model. The models are used to forecast the realized volatility-covariation matrix of two stocks, General Electric (GE) and International Business Machines (IBM), in an out-of-sample period for a daily, weekly and biweekly forecasting horizon. Our results suggest that the VARFIMA and HAR models using high-frequency intradaily data outperform the GO-GARCH model based on daily returns over all forecasting horizons. The HAR appears to perform slightly better than the VARFIMA at the weekly and biweekly forecasting horizon, however the performance difference between the two models turns out to be low.

1. Introduction

The risk of any investment is fully characterized by the return distribution of this investment and if investment returns are normally distributed, the distribution is given when its first two moments, i.e. expected value and variance, are given. Although stock return distributions appear to be fat-tailed and leptokurtic, thus contradicting the idea of normally distributed returns, economic theories as the modern portfolio theory require stock returns to be normally distributed for investors to choose portfolio weights solely based on the vector of (conditional) expected returns and the (conditional) return variance-covariance matrix.[1] Moreover, in option pricing the popular Black-Scholes formula requires the return volatility of the underlying as an input parameter. Even if we reject the working hypothesis of normally distributed returns, return volatility remains an important measure to assess the risk of an investment even though it may not capture all characteristics of the underlying distribution – and it can be attempted to forecast these return volatilities by means of stochastic volatility models. The aim of this paper is to assess the goodness of stochastic volatility models in out-of-sample forecasting of multivariate realized stock return volatility over a daily, weekly (five days) and biweekly (ten days) forecasting horizon. To do so, three stochastic volatility models are estimated and their forecasting performance is compared: A VARFIMA model, a Heterogeneous Autoregressive (HAR) model and a Generalized Orthogonal GARCH (GO-GARCH) model. Whereas the former make use of intradaily high-frequency stock price data, the latter one exclusively relies on daily stock returns. Similar analyses have been conducted by Andersen, Bollerslev, Diebold, and Labys (2003) and Chiriac and Voev (2011): The former fit among others a bivariate VAR on series of fractionally differenced realized exchange rate volatilities and find that their so-called *VAR-RV* outperforms multivariate GARCH (MGARCH) models in out-of-sample forecasting on the daily and

[1] As an alternative to the assumption of normally distributed returns, a quadratic utility function can be assumed for investors to choose portfolio weights accordingly.

biweekly forecasting horizon. Chiriac and Voev (2011) forecast multivariate realized stock return volatility using a VARFIMA, a HAR as well as multivariate GARCH models and are able to confirm the finding for the daily, weekly and biweekly forecasting horizon that the high-frequency data based VARFIMA and HAR models outperform MGARCH models.

The fractional difference parameters in the VARFIMA model can either be estimated in a two-step procedure where the fractional difference parameters are estimated in a first step and a VARMA is then fitted to the fractionally differenced data in a second step to determine the VAR and VMA parameters, or all parameters can be estimated simultaneously. Chiriac and Voev (2011) apply an approximate maximum likelihood method as suggested by Beran (1995) to determine all model parameters simultaneously whereas Andersen, Bollerslev, Diebold, and Labys (2003) use a two-step procedure based on a first-step log-periodogram regression as discussed by Robinson (1995). Although the two-step procedure has the drawback of resulting in inefficient parameter estimates for the VAR and VMA parameters, the procedure remains feasible even for high dimensions and has the advantage of being straightforward to implement.

This paper proceeds as follows: Section 2 introduces basic notation and definitions concerning returns and realized volatility before turning to the setup of the models. The theoretical framework for each model does not only comprise the model setup but also the estimation methods which are applied in section 3. The VARFIMA model is introduced in section 2.2, the HAR model in section 2.3 and the GO-GARCH model in section 2.4. Section 3 starts with describing the data used for our analysis and then turns to the parameter estimation results obtained by applying the procedures given in the theoretical section. The results of the out-of-sample forecasting using these parameter estimates are evaluated in section 3.3. Finally, we summarize and assess our results in section 4.

2. Theoretical Framework

2.1. Preliminaries

Throughout this paper, matrices will be denoted by bold uppercase letters, vectors by bold lowercase letters and scalars as well as sets by upper- or lowercase letters. Assume that each trading day $t, t = 1, 2, \ldots, T$ we observe m prices for each stock. Then the *ith* intraday return for stock n at day t is computed as $r_{tn,i} = \ln(p_{tn,i}) - \ln(p_{tn,i-1}), i = 1, 2, \ldots, m-1$.[2] Given that we attempt to model and predict the variance-covariance matrix of a given number of stocks, we need to make sure that our predictions yield positive-definite matrices. Let N denote the number of stocks we consider and $\boldsymbol{r}_{t,i}$ the $N \times 1$ vector of returns at observation i of day $t, t = 1, 2, \ldots, T$, then the realized volatility-covariation matrix of dimension $N \times N$ for day t is

[2]Note that this definition of intradaily returns implies that overnight returns are not considered in this context.

defined as

$$\boldsymbol{H}_t \equiv \sum_{i=1}^{m-1} \boldsymbol{r}_{t,i}\boldsymbol{r}'_{t,i} \tag{2.1.1}$$

where $m-1$ is the number of intradaily returns observed within each day. Forecasting the process realized volatility-covariation matrix, we need to ensure that our forecast yields a positive-definite matrix. To achieve this goal, this paper adopts the approach of Chiriac and Voev (2011) who model the Cholesky factors of matrix $\boldsymbol{H}_t$ instead of its elements directly: Factorize $\boldsymbol{H}_t$ as $\boldsymbol{H}_t = \boldsymbol{X}_t'\boldsymbol{X}_t$ and collect the $K = N(N+1)/2$ Cholesky factors in a vector $\boldsymbol{x}_t$, i.e. $\boldsymbol{x}_t = \mathrm{vech}(\boldsymbol{X}_t)$. Decomposing $\boldsymbol{H}_t$ for each $t, t = 1, 2, \dots, T$ yields a realization of the vector stochastic process $\boldsymbol{x}_t$. As Chiriac and Voev (2011) note, the elements of the realized volatility-covariation matrix can be reconstructed from the Cholesky factors $x_{t,1}, \dots, x_{t,K}$ by

$$h_{t,ij} = \sum_{k=1+\frac{i(i-1)}{2}}^{\frac{i(i+1)}{2}} x_{t,k}x_{t,k+\frac{j(j-1)}{2}-\frac{i(i-1)}{2}} \tag{2.1.2}$$

where $h_{t,ij}$ is the i,j-element of $\boldsymbol{H}_t$, that is it denotes the realized covariation between variable i and j. This "reverse Cholesky" ensures that the matrix constructed from the $x_{t,k}$ is positive definite.

2.2. VARFIMA(p, d, q) Model

The general VARFIMA(p, d, q) takes the form

$$\boldsymbol{\Phi}(\mathrm{L})\boldsymbol{D}(\mathrm{L})\left[\boldsymbol{x}_t - \boldsymbol{B}\boldsymbol{z}_t\right] = \boldsymbol{\Theta}(\mathrm{L})\boldsymbol{\epsilon}_t \tag{2.2.1}$$

where $\boldsymbol{x}_t$ is a $K \times 1$ vector of Cholesky factors, $\boldsymbol{z}_t$ an $M \times 1$ vector of exogenous variables, $\boldsymbol{B}$ a $K \times M$ coefficient matrix, $\boldsymbol{\Phi}(\mathrm{L}) \equiv \boldsymbol{I}_K - \boldsymbol{\Phi}_1\mathrm{L} - \boldsymbol{\Phi}_2\mathrm{L}^2 - \dots - \boldsymbol{\Phi}_p\mathrm{L}^p$, $\boldsymbol{\Theta}(\mathrm{L}) \equiv \boldsymbol{I}_K + \boldsymbol{\Theta}_1\mathrm{L} + \boldsymbol{\Theta}_2\mathrm{L}^2 + \dots + \boldsymbol{\Theta}_q\mathrm{L}^q$ matrix lag polynomials of dimension $K \times K$ and $\boldsymbol{D}(\mathrm{L}) \equiv \mathrm{diag}\{(1-\mathrm{L})^{d_1}, \dots, (1-\mathrm{L})^{d_K}\}$ with $d_k \equiv \delta_k + m_k$, $k \in \{1, \dots, K\}$, $\delta_k \in (-\frac{1}{2}, \frac{1}{2})$ and $m_k \in \mathbb{Z}_+$. The parameter m_k is a non-negative integer giving the number variable k has to be differenced to achieve sationarity whereas the parameter δ_k gives the degree of fractional integration of variable k. The innovations $\boldsymbol{\epsilon}_t$ are assumed to be i.i.d. white noise, in particular $\boldsymbol{\epsilon}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$. The multivariate VARFIMA(p, d, q) from equation (2.2.1) is discussed by Sowell (1989) and is stationary for $d_k < 0.5$ for $k = 1, 2, \dots, K$.

In addition the lag polynomials are assumed to be diagonal matrices with identical entries along the main diagonal, i.e.

$$\boldsymbol{\Phi}(\mathrm{L}) = \phi(\mathrm{L})\boldsymbol{I}_K,$$
$$\boldsymbol{\Theta}(\mathrm{L}) = \theta(\mathrm{L})\boldsymbol{I}_K$$

where $\phi(\mathrm{L})$ and $\theta(\mathrm{L})$ are both invertible because we want to model the vector stochastic process of Cholesky factors $\boldsymbol{x}_t$ as a stationary long-memory process. For the same reason the degree of integration m_k is assumed to be zero for each variable. Whereas the restriction on the autoregressive lag polynomial is necessary for identification of model (2.2.1) as pointed out by Lütkepohl (2005), the restriction on the moving average lag polynomial is arbitrary to avoid overparameterization of the model. Contrary to the model estimated by Chiriac and Voev (2011), we allow the fractional difference parameters to vary between variables. Exogenous variables $\boldsymbol{z}_t$ are not included in this analysis but only a vector of constants $\boldsymbol{c}$ – which equals the expected value $\boldsymbol{\mu}$ of $\boldsymbol{x}_t$ in case of stationarity – such that the model we are going to estimate can be written as

$$\phi(\mathrm{L})\boldsymbol{D}(\mathrm{L})\left[\boldsymbol{x}_t - \boldsymbol{\mu}\right] = \theta(\mathrm{L})\boldsymbol{\epsilon}_t. \tag{2.2.2}$$

Note that

$$(1-\mathrm{L})^{\delta_k} = \sum_{j=0}^{\infty} b_j(\delta_k)\mathrm{L}^j \tag{2.2.3}$$

and

$$\begin{aligned} b_j(\delta_k) &= (-1)^j \frac{\Gamma(\delta_k+1)}{\Gamma(j+1)\Gamma(\delta_k-j+1)} \\ &= \prod_{i=1}^{j} \frac{i-1-\delta_k}{i} \end{aligned} \tag{2.2.4}$$

with $b_0(\delta_k) \equiv 1$ and where the second line of (2.2.4) is required for the practical computation of $b_j(\delta_k)$ for large j. For notational convenience, assume for a moment that the fractional difference parameters δ_k are constant over all variables, then the VAR(∞) representation of equation (2.2.2) for $p = q = 1$ – which exists under the assumptions made above – takes the form

$$\boldsymbol{x}_t - \boldsymbol{\mu} = \boldsymbol{\epsilon}_t - \rho_1(\boldsymbol{x}_{t-1} - \boldsymbol{\mu}) - \rho_2(\boldsymbol{x}_{t-2} - \boldsymbol{\mu}) - \dots \tag{2.2.5}$$

with $\rho_j \equiv \sum_{l=0}^{j} b_l(\delta)\alpha_{j-l}$, $\alpha_j \equiv (-1)^j(\phi\theta^{j-1} + \theta^j)$ for $j \geq 1$ and $\alpha_0 \equiv 1$. A derivation of (2.2.5) is given in appendix C.1. In case of varying δ_k, simply calculate the ρ_j for each variable seperately with their corresponding fractional difference parameters.

Estimation

Following Andersen, Bollerslev, Diebold, and Labys (2003), the fractional difference parameters are estimated in a first step by a log-periodogram regression as introduced for the univariate case by Geweke and Porter-Hudak (1983) and generalized to the multivariate case by Robinson (1995) who also provides asymptotic results for the regression estimates. The performed regression takes the form

$$\boldsymbol{y} = \boldsymbol{\Delta}\boldsymbol{\lambda} + \boldsymbol{u} \tag{2.2.6}$$

with

$$\boldsymbol{y} \equiv \begin{bmatrix} y_1 \\ \vdots \\ y_K \end{bmatrix}, \boldsymbol{\Delta} \equiv \begin{bmatrix} c_1 & \delta_1 \\ \vdots & \vdots \\ c_K & \delta_K \end{bmatrix}, \boldsymbol{\lambda} \equiv \begin{bmatrix} 1 \\ -2\log\lambda \end{bmatrix}, \boldsymbol{u} \equiv \begin{bmatrix} u_1 \\ \vdots \\ u_K \end{bmatrix}$$

where $\lambda_l = (l2\pi)/T$ is computed for $l = 1, 2, \ldots, T-1$ and hence represents frequencies between 0 and 2π associated with cycles of T/l. Variable y_k is just the log of the periodogram of variable x_k, I_k, thus $y_k = \log(I_k)$ with

$$I_k(\lambda) = \frac{1}{2\pi T}\left|\sum_{t=1}^{T} x_{kt} e^{it\lambda}\right|^2 .$$

The system of equations (2.2.6) can then be easily estimated by (system) OLS.

In this paper, a VARFIMA$(1, \delta, 1)$ is estimated. Thus, in a second step each time series of Cholesky factors is fractionally differenced using the parameters obtained in the log-periodogram regression and a VARMA$(1, 1)$ is fitted to the fractionally differenced data. The estimated VARMA$(1, 1)$ parameter vector $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\phi}, \hat{\theta})'$ is determined by numerically minimizing the log determinant of the estimated innovation variance-covariance matrix:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}^*} \log\left\{\left|\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\epsilon}_t(\boldsymbol{\theta}^*)\boldsymbol{\epsilon}_t(\boldsymbol{\theta}^*)'\right|\right\}. \tag{2.2.7}$$

Here the $\boldsymbol{\epsilon}_t(\boldsymbol{\theta}^*)$ are the estimated innovation terms that can be recursively calculated from the set of observations of the fractionally differenced series applying parameter vector $\boldsymbol{\theta}^* = (\mu^*, \phi^*, \theta^*)'$. This procedure is equivalent to maximizing the conditional log-likelihood function (Lütkepohl, 2005, p. 469).

2.3. Heterogeneous Autoregressive (HAR) Model

The heterogeneous autoregressive (HAR) model is described by a linear relationship between the dependent variable and its moving time averages:

$$\boldsymbol{x}_{t+1} = \boldsymbol{c}^{(d)} + \beta^{(d)}\boldsymbol{x}_t^{(d)} + \beta^{(w)}\boldsymbol{x}_t^{(w)} + \beta^{(bw)}\boldsymbol{x}_t^{(bw)} + \beta^{(m)}\boldsymbol{x}_t^{(m)} + \boldsymbol{\omega}_{t+1,d} \tag{2.3.1}$$

where $\boldsymbol{c}^{(d)}$ is a $K \times 1$ vector of constants, the $\beta^{(\cdot)}$ are scalar parameters and $\boldsymbol{\omega}_{t+1,d}$ a $K \times 1$ vector of error terms. The independent variables $\boldsymbol{x}_t^{(\cdot)}$ are averages of the m most recent observations of the vector of Cholesky factors $\boldsymbol{x}_t$, with $m^{(d)} = 1$ (daily), $m^{(w)} = 5$ (weekly), $m^{(bw)} = 10$ (biweekly) and $m^{(m)} = 20$ (monthly), i.e.

$$\boldsymbol{x}_t^{(\cdot)} \equiv \frac{1}{m^{(\cdot)}}(\boldsymbol{x}_t + \boldsymbol{x}_{t-1} + \cdots + \boldsymbol{x}_{t-m^{(\cdot)}+1}).$$

The univariate version of (2.3.1) was introduced by Corsi (2009) who models the realized volatility of a stock directly whereas Chiriac and Voev (2011) estimate the multivariate generalization and fit the model to Cholesky factors instead of realized covariations as well. The HAR-RV can be derived under the assumption of heterogeneous market participants with different investment horizons. Corsi (2009) divides investors into three groups: Investors with a short investment horizon, that is daytraders, investors with an intermediate investment horizon who balance their portfolios weekly and long-term investors adjusting their portfolios monthly. Short-term investors' trading decisions are influenced by mid- and long-term volatility, thus long- and mid-term volatility influence short-term volatility caused by short-horizon traders. On the other hand, mid-term investors care about long-term volatility, but not about short-term price movements. Last but not least, long-horizon investors care neither about short- nor mid-term volatility but only about long-term price movements. This results in a hierarchical structure of volatility where a cascade from low- to high-frequency volatility can be observed in the market. As we are interested in daily, weekly and biweekly forecasting, the biweekly average is included addition to the monthly one in (2.3.1). The parameters of our model (2.3.1) are readily estimated by (system) OLS.

2.4. Generalized Orthogonal GARCH Model

MGARCH models just make use of daily returns and no intradaily data. Thus, define the return of stock n from day $t-1$ to day t as $r_{tn} \equiv \ln(p_{t,n}) - \ln(p_{t-1,n})$ where p_{tn} is the closing price of stock n observed at trading day t. Stacking the daily returns for each stock $n, n = 1, 2, \ldots, N$ in a vector yields the $N \times 1$ daily return vector $\boldsymbol{r}_t$. The MGARCH model assumes conditionally normally distributed returns:

$$\boldsymbol{r}_t | \mathscr{F}_{t-1} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_t) \tag{2.4.1}$$

where $\mathscr{F}_{t-1}$ denotes the information set at day $t-1$ and $\boldsymbol{\Sigma}_t$ the conditional variance-covariance matrix at day t. Let the return vector $\boldsymbol{r}_t$ be stationary and hence the unconditional variance-covariance matrix is given by $\boldsymbol{\Sigma} \equiv \mathrm{E}(\boldsymbol{\Sigma}_t)$.

Orthogonal GARCH

The orthogonal GARCH model (O-GARCH) as discussed by Alexander (2001) is based on a principal components analysis of the time series of returns: Recall that any $N \times N$ matrix $\boldsymbol{\Sigma}_t$ with distinct eigenvalues can be decomposed as $\boldsymbol{\Sigma}_t = \boldsymbol{W}_t \boldsymbol{\Lambda}_t \boldsymbol{W}_t$ where $\boldsymbol{\Lambda}_t$ denotes the $N \times N$ diagonal matrix of eigenvalues of $\boldsymbol{\Sigma}_t$ and $\boldsymbol{W}_t$ the $N \times N$ matrix of eigenvectors of $\boldsymbol{\Sigma}_t$ where each column represents an eigenvector (Hamilton, 1994, p. 730). It can be shown that symmetric matrices with distinct eigenvalues have orthogonal eigenvectors and thus matrix $\boldsymbol{W}_t$ is orthogonal, i.e. $\boldsymbol{W}_t' = \boldsymbol{W}_t^{-1}$. Assume $\boldsymbol{W}_t = \boldsymbol{W}$ for $t = 1, 2, \ldots, T$ and define the $N \times 1$ vector of *principal components* or *factors* as $\boldsymbol{p}_t \equiv \boldsymbol{W} \boldsymbol{r}_t$, then the return vector can be rewritten as

$$\boldsymbol{r}_t = \boldsymbol{W}' \boldsymbol{p}_t. \tag{2.4.2}$$

It can be shown that the unconditional factor variance-covariance matrix is given by $\mathrm{Var}(\boldsymbol{p}_t) = \boldsymbol{\Lambda}$, which implies that the factors are unconditionally uncorrelated as the matrix of eigenvalues of $\boldsymbol{\Sigma}$, $\boldsymbol{\Lambda}$, is diagonal. The O-GARCH makes the assumption that factors at time t are also uncorrelated conditioning on the information set $\mathscr{F}_{t-1}$, so $\boldsymbol{\Omega}_t \equiv \text{Var}(\boldsymbol{p}_t|\mathscr{F}_{t-1})$ is diagonal as well. It follows that the conditional return variance-covariance matrix is given by

$$\boldsymbol{\Sigma}_t = \boldsymbol{W}\boldsymbol{\Omega}_t\boldsymbol{W}'. \tag{2.4.3}$$

The O-GARCH now models the diagonal elements of $\boldsymbol{\Omega}_t$ by means of simple univariate GARCH models. In practice, $\boldsymbol{W}$ is estimated by using the eigenvectors of the estimated unconditional variance-covariance matrix $\hat{\boldsymbol{\Sigma}}$ and the time series of principal components is computed. Finally, simple univariate GARCH models are fitted to the principal components, resulting in estimates of $\boldsymbol{\Omega}_t$ and thus $\boldsymbol{\Sigma}_t$ which is necessarily positive definite by construction as can be seen from (2.4.3). To sum up, the O-GARCH enables us to model $K = N(N+1)/2$ variances and covariances of the original process by just fitting univariate GARCH models on $N < K \ \forall \ N > 1$ uncorrelated factors as our data is considered to be a linear combination of these factors which remains feasible even for high dimensions N. However, as van der Weide (2002) points out, the orthogonal matrix $\boldsymbol{W}$ is not identified by the eigenvectors of $\boldsymbol{\Sigma}$ anymore if the eigenvalues of $\boldsymbol{\Sigma}$ are not distinct which may be the case in systems of only weakly correlated variables.

Generalized Orthogonal GARCH

The generalized orthogonal GARCH (GO-GARCH) as introduced by van der Weide (2002) assumes that the $N \times 1$ observed return process $\boldsymbol{r}_t$ can be expressed as a linear combination of N uncorrelated and unobserved economic factors $\boldsymbol{f}_t$:

$$\boldsymbol{r}_t = \boldsymbol{Z}\boldsymbol{f}_t \tag{2.4.4}$$

where the factors $\boldsymbol{f}_t$ are normalized to have an unconditional unit variance-covariance matrix, $\text{Var}(\boldsymbol{f}_t) = \boldsymbol{I}_N$, and the $N \times N$ matrix $\boldsymbol{Z}$ is non-singular. Assuming conditionally normally distributed and uncorrelated factors – $\boldsymbol{f}_t|\mathscr{F}_{t-1} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Omega}_t)$ – gives rise to a Gaussian likelihood function. As in the O-GARCH, the factors follow simple univariate GARCH processes, i.e. $\boldsymbol{\Omega}_t = \text{diag}\{\omega_{1,t}, \ldots, \omega_{N,t}\}$ with

$$\omega_{n,t} = (1 - \alpha_n - \beta_n) + \alpha_n f_{n,t-1}^2 + \beta_n \omega_{n,t-1} \tag{2.4.5}$$

where $f_{n,t-1}$ denotes the *nth* element of $\boldsymbol{f}_{t-1}$ for $n = 1, \ldots, N$. The crucial difference to the standard O-GARCH model is that the linkage $\boldsymbol{Z}$ between the factors and the returns process need *not* be orthogonal. As van der Weide (2002) shows, the map $\boldsymbol{Z}$ can be decomposed as

$$\boldsymbol{Z} = \boldsymbol{W}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{U}_0 \tag{2.4.6}$$

where $\boldsymbol{U}_0$ is an orthogonal matrix with $\det(\boldsymbol{U}_0) = 1$. Again, $\boldsymbol{W}$ and $\boldsymbol{\Lambda}$ are readily estimated by computing the eigenvalues and eigenvectors of the estimated unconditional variance-covariance matrix $\hat{\boldsymbol{\Sigma}}$. For the estimation of $\boldsymbol{U}_0$, van der Weide (2002) uses the fact that $\boldsymbol{U}_0$ can be written as the product of rotation matrices that can be estimated applying maximum likelihood.[3]

[3]The estimation results for the GO-GARCH printed in this paper are computed using the *rmgarch* R-package by Ghalanos (2014).

The log likelihood takes the form

$$\mathscr{L} = -\frac{1}{2}\sum_{t=1}^{T}\left(N\log(2\pi) + \log|\boldsymbol{Z}_\theta \boldsymbol{Z}_\theta'| + \log|\boldsymbol{\Omega}_t| + \boldsymbol{f}_t'\boldsymbol{\Omega}_t^{-1}\boldsymbol{f}_t\right) \tag{2.4.7}$$

where

$$\boldsymbol{Z}_\theta \equiv \begin{bmatrix} 1 & 0 \\ \cos\theta & \sin\theta \end{bmatrix}$$

defines the Euler angles required for the calculation of $\boldsymbol{Z}$ as the product of rotation matrices. Like in the conventional O-GARCH, GO-GARCH variance-covariance matrix forecasts are always positive definite.

3. Estimation, Forecasting & Evaluation

3.1. Data

Our dataset consists of intradaily transaction prices for stocks of *General Electric* (GE) and *International Business Machines* (IBM) on a one-minute frequency from 9:30 a.m. to 4:00 p.m. covering a period from January 3*rd* 2011 until June 29*th* 2012 which makes 377 trading days with 391 price observations per trading day. An equally spaced grid of 1-minute-intervals is achieved by a simple previous-tick interpolation. For each trading day, realized volatility-covariation matrices are calculated on a five-minute-interval basis to account for market microstructure noise that biases our estimator. Intradaily returns are computed as the difference in log prices based on five-minute-intervals. In order to use all observations and get a more efficient as well as a robust estimator, the realized volatility-covariation matrix is computed as the the sum of squared intradaily returns and cross-product returns for the five-minute-interval starting at minute $m = 1, 2, 3, 4$ so that we obtain four realized covariation matrices for each day. The realized covariation matrix $\boldsymbol{H}_t$ is then defined to be the average of those four matrices. For each of the 377 trading days, the realized covariation matrix is decomposed into its Cholesky factors so that we obtain three time series of Cholesky factors which are used to fit the VARFIMA and HAR models. The dataset for the GO-GARCH model is obtained by keeping only closing prices of each day and calculating the time series of daily returns.

As we are interested in out-of-sample forecasting, the sample is split into an in-sample period of 252 observations ranging from January 3*rd* 2011 to December 30*th* 2011 and an out-of-sample period of 125 observations from January 3*rd* to June 29*th* 2012. The evolution of realized volatility and covariation for the two stocks over the whole period is plotted in figure 1, the beginning of the out-of-sample period is marked by the dashed vertical line. As can be seen in table B.3 in the appendix, realized volatility and covariation for the two stocks are right-skewed and leptokurtic, i.e. the right tail of the distribution is relatively long. Note that the out-of-sample period appears to be a period with relatively low levels of both volatility and covariation, so our out-of-sample forecasting evaluation will not be able to check how the models perform in times of high volatility as for example observed at the end of the in-sample period. Further descriptive statistics on daily and intradaily returns can be found in table B.2 and B.1 of appendix B.

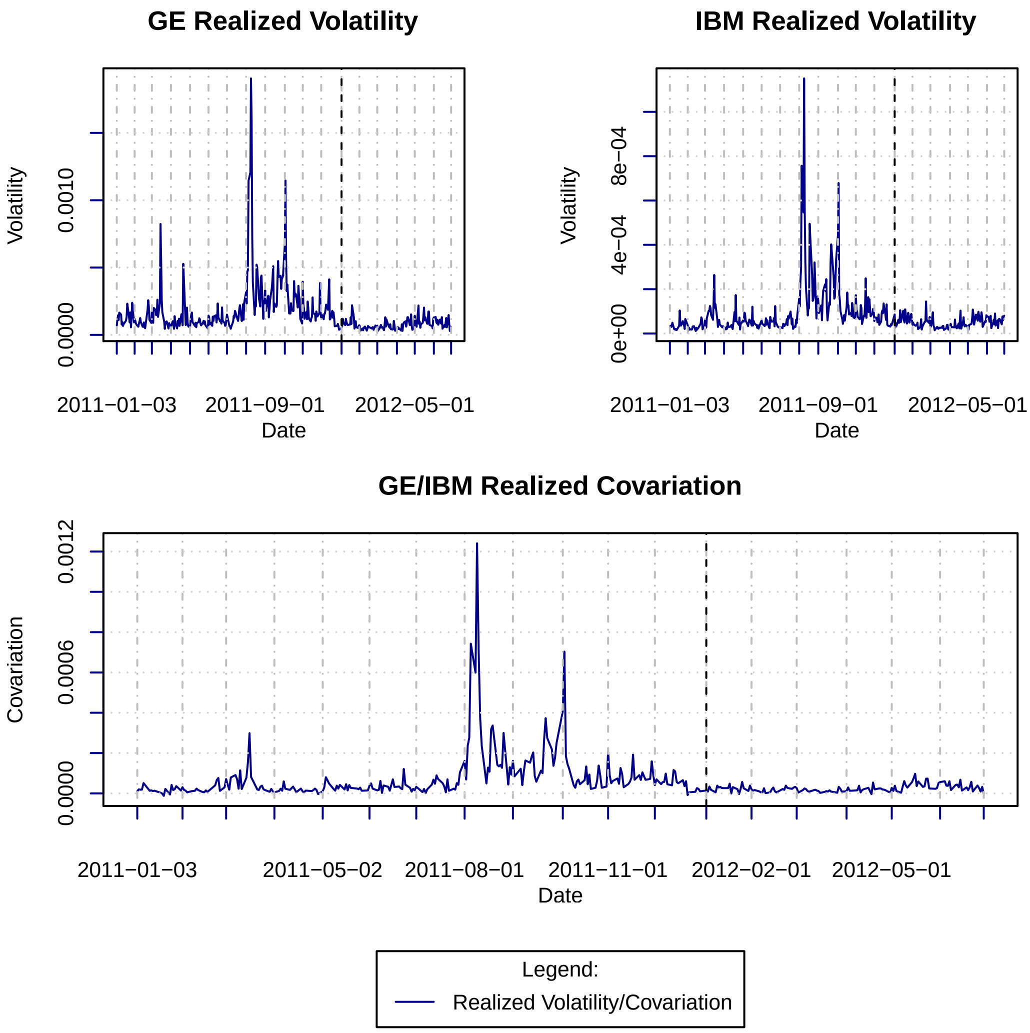

Figure 1: The evolution of realized volatility and covariation from January *3rd* 2011 to June 29*th* 2012 consisting of 377 observations for each stock. The dashed vertical line marks the beginning of the out-of-sample period which covers the latest 125 observations.

3.2. Estimation Results

The parameter estimates for our three models are printed in table 1 for both only the in-sample period as well as the full set of observations, standard errors are given in brackets. Estimation for the VARFIMA$(1, \delta, 1)$ is carried out using the two-step procedure as described in section 2.2. Our estimates for the fractional difference parameters δ_k lie between 0.36 and 0.47 which is consistent with the results reported by Chiriac and Voev (2011) or Andersen, Bollerslev, Diebold, and Labys (2003) which are 0.449 for the stock return volatility series and 0.401 for the exchange rate volatility series. This implies that our series of realized volatilities and covariations indeed exhibit long-memory properties which is inherited by the series of Cholesky factors. Whereas the levels of the fractional difference parameters decrease slightly for each series of Cholesky factors when using the full sample compared to the in-sample estimation, the autoregressive and moving average parameters ϕ and θ increase in absolute value, indicating that the magnitude of the long-memory process is slightly over-estimated using only the in-sample observations.[4] The HAR parameter estimates appear to be robust to changes in sample size which is what we expect given that the series of Cholesky factors exhibit long-memory stationarity and thus a regression of the Cholesky factors on moving factor averages will yield consistent estimates such that we can observe convergence with an increasing sample size. Finally, the factor variances of our GO-GARCH$(1,1)$ model appear to be highly persistent given that the sum of the two univariate GARCH$(1,1)$, α_n and β_n, are close to one for $n = 1, 2$. As any MGARCH model applied to daily return series can be interpreted as fitting a VARMA model on the squared returns (Lütkepohl, 2005, p. 565) and the GO-GARCH regards the returns as generated by a linear combination of uncorrelated factors, highly-persistent factor variances imply high persistence in the series of squared daily returns as well. As Andersen and Bollerslev (1998) note, squared daily returns are unbiased – however extremely noisy – estimators for the daily variance, and therefore our GO-GARCH estimates indicate high persistence in daily return variance. Our standard errors are computed using the stationary bootstrap method as proposed by Politis and Romano (1994). This non-parametric bootstrap method is based on drawing random subsamples of random length from our original sample and combine these subsamples to a new bootstrap sample of length T which equals the length of our original time series. This way we account for the two-step estimation applied in the VARFIMA and GO-GARCH model: Although asymptotic theory for the GO-GARCH is implicitly given as it is just a special case of a BEKK-GARCH, the additional uncertainty concerning the first-step estimation of the eigenvectors and eigenvalues of $\mathbf{\Sigma}$ need to be taken into account. The same is true for the VARFIMA model where fractional difference parameters have been estimated in a first step by a log-periodogram regression and are treated as known in the VARMA$(1,1)$ fitting.

Note that the estimates of our log-periodogram regression are not efficient, in particular the ratio of their variance and the variance of exact maximum likelihood estimates will converge to infinity, i.e. the rate of convergence of log-periodogram estimates is lower than the rate of convergence of exact maximum likelihood estimates.

[4]Note that our estimate for θ reported in table 1 is negative whereas the estimate reported by Chiriac and Voev (2011) is positive which is due to the fact that they define the VMA lag polynomial as $\mathbf{\Theta}(\mathrm{L}) = \boldsymbol{I}_K - \mathbf{\Theta}_1\mathrm{L} - \mathbf{\Theta}_2\mathrm{L}^2 - \cdots - \mathbf{\Theta}_q\mathrm{L}^q$.

Table 1: Parameter estimates for the VARFIMA$(1, \delta, 1)$, the HAR and the GO-GARCH$(1, 1)$ models. Parameter estimates are printed for estimations based on the in-sample period (the first 252 trading days) and the full sample (all 377 trading days). The GO-GARCH parameters correspond to the univariate GARCH parameter estimates of the two factors. Standard errors are given in brackets and are determined using stationary bootstrap.

Parameter	Estimates	
	In-sample $(T = 252)$	*Full sample* $(T = 377)$
VARFIMA$(1, \delta, 1)$		
ϕ	0.6911	0.8033
	(0.4793)	(0.4467)
θ	-0.6823	-0.7737
	(0.4390)	(0.4100)
δ_1	0.4741	0.4344
	(0.0889)	(0.0772)
δ_2	0.4357	0.4130
	(0.0783)	(0.0777)
δ_3	0.4152	0.3560
	(0.0738)	(0.0690)
HAR		
$c^{(d)}$	0.0006	0.0004
	(0.0003)	(0.0001)
$\beta^{(d)}$	0.6861	0.6252
	(0.1509)	(0.1526)
$\beta^{(w)}$	0.0052	0.0656
	(0.1441)	(0.1347)
$\beta^{(bw)}$	0.1489	0.1347
	(0.1939)	(0.0759)
$\beta^{(m)}$	0.1015	0.1245
	(0.1206)	(0.0567)
GO-GARCH		
α_{11}	0.0751	0.0806
	(0.0629)	(0.0616)
β_{11}	0.9019	0.8958
	(0.1160)	(0.1052)
α_{21}	0.0000	0.0000
	(0.0750)	(0.0514)
β_{21}	0.9990	0.9990
	(0.1231)	(0.1148)

3.3. Forecasting

For forecasting the realized volatility and covariation in our out-of-sample period, three forecasting horizons are chosen: A daily, weekly (five days) and biweekly (ten days) forecasting period. Forecasts using the VARFIMA model were conducted using the VAR(∞) representation as given by (2.2.5) (with variable-specific δ_k due to different degrees of fractional integration) truncated at the first observation. Let $\boldsymbol{P}_j \equiv \operatorname{diag}\{\rho_{j1}, \ldots, \rho_{jK}\}$ with $\rho_{jk} \equiv \sum_{l=0}^{j} b_l(\delta_k)\alpha_{j-l}$ and α_j as defined in section 2.2. Then the one-step ahead prediction of $\boldsymbol{x}_{t+1}$ at time t, denoted by $\hat{\boldsymbol{x}}_{t+1}$, is given by the expected value of $\boldsymbol{x}_{t+1}$ conditional on information given at time t:

$$\begin{aligned} \hat{\boldsymbol{x}}_{t+1} &\equiv \mathrm{E}(\boldsymbol{x}_{t+1} | \mathscr{F}_t) \\ &= \boldsymbol{\mu} - \boldsymbol{P}_1(\boldsymbol{x}_t - \boldsymbol{\mu}) - \boldsymbol{P}_2(\boldsymbol{x}_{t-1} - \boldsymbol{\mu}) - \cdots - \boldsymbol{P}_{t-1}(\boldsymbol{x}_1 - \boldsymbol{\mu}) \end{aligned} \tag{3.3.1}$$

A feasible forecast follows after substituting $\boldsymbol{\mu}$ and $\boldsymbol{P}_j$ by their estimated values, where $\boldsymbol{\mu}$ is estimated by its sample mean based on all observations available at time t. Multi-step forecasts $\hat{\boldsymbol{x}}_{t+s}, s > 1$ in the VARFIMA framework are obtained by computing a one-step ahead forecast as previously described which is then subsequently treated as an actual observation in (3.3.1). One-step ahead forecasts in HAR are readily obtained from (2.3.1), multi-step forecasts are computed following the same iteration logic as for the VARFIMA model by replacing the values required to compute $\boldsymbol{x}_{t+s}$, that is $\boldsymbol{x}_{t+s-1}^{(\cdot)}$, with their forecasted values obtained in the previous step. Forecasts for the GO-GARCH as a special of a BEKK-GARCH model as discussed by Lütkepohl (2005) follow from the VARMA representation of the squared- and cross-product daily returns.

In the VARFIMA and HAR model, the Cholesky factor forecasts are then transformed back into realized volatilities and covariations using equation (2.1.2) which simplifies in our case with $N = 2$ to

$$\begin{aligned} \hat{h}_{t+s,11} &= \hat{x}_{t+s,1}^2, \\ \hat{h}_{t+s,12} &= \hat{x}_{t+s,1}\hat{x}_{t+s,2}, \\ \hat{h}_{t+s,22} &= \hat{x}_{t+s,2}^2 + \hat{x}_{t+s,3}^2 \end{aligned}$$

where $\hat{h}_{t+s,ij}$ denotes the i,j-element of $\hat{\boldsymbol{H}}_{t+s}$ – the s-step ahead realized volatility matrix forecast – and $\hat{x}_{t+s,k}$ denotes the s-step ahead forecast of the kth Cholesky factor. Although (3.3.1) provides unbiased forecasts for the Cholesky factors, a problem arises from the nonlinear transformation of the Cholesky factors back into realized volatilities: The elements of $\hat{\boldsymbol{H}}_{t+s}$ will generally be biased forecasts for the elements of $\boldsymbol{H}_{t+s}$. Chiriac and Voev (2011) deal with this problem and discuss a bias correction based on the theoretical size of the bias. They refuse this approach as the forecast bias depends on the unknown, model-dependent size of the variance of the forecast error. Comparing the ratio of volatility forecasts to their actual ex-post realizations suggests that the bias arising from the nonlinear transformation is negligible, thus we dispense with bias correction as well.

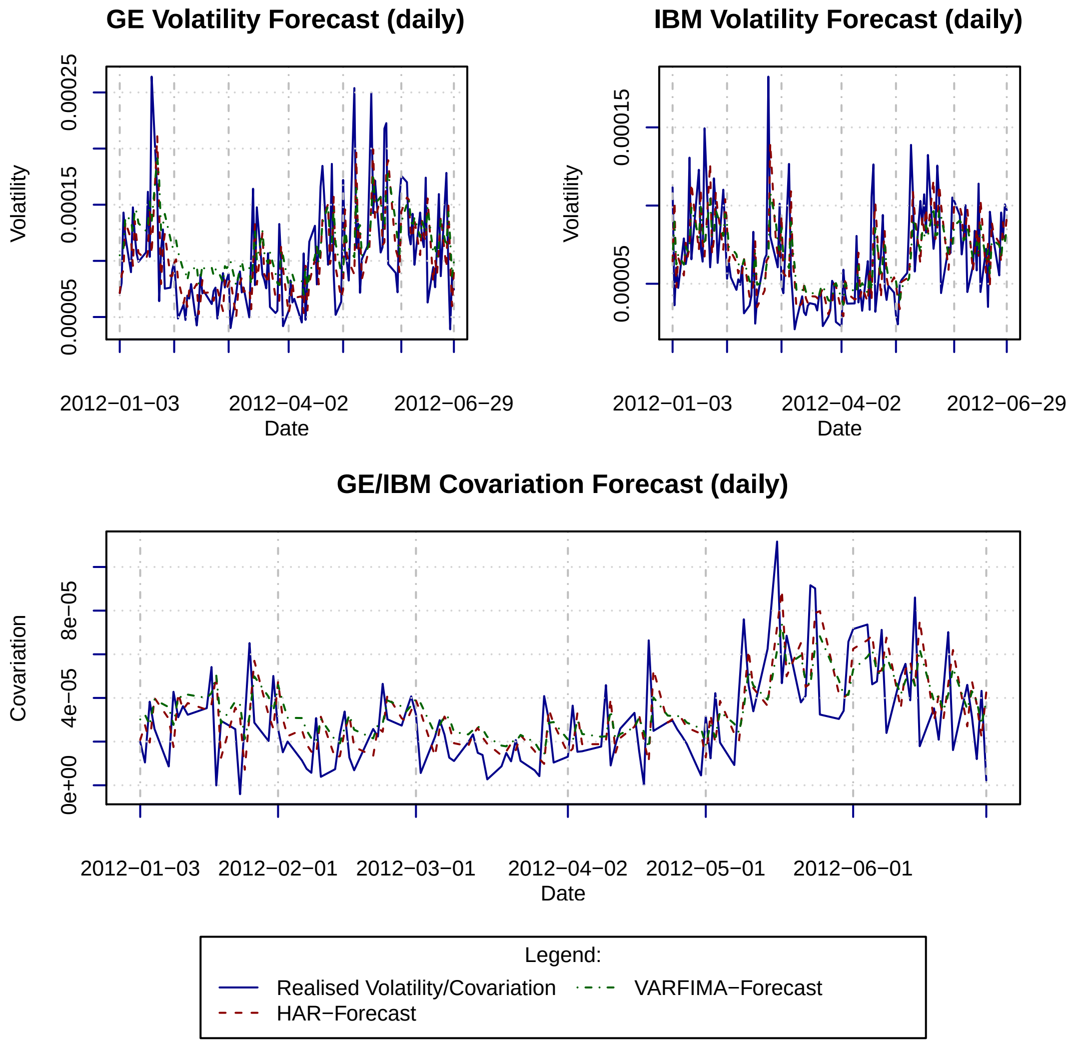

Figure 2: Daily forecasts of realized volatility and covariation for the GE and IBM stocks using the VARFIMA and the HAR model. At each day in the above illustrated out-of-sample period, parameters are re-estimated using all data available at this day. Forecasts are then carried out using these parameters.

Having computed multi-step forecasts, forecasts over a period of h days given information at day t are simply obtained by summing up the individual s-step forecasts:

$$\hat{\boldsymbol{H}}_{t:t+h} \equiv \sum_{s=1}^{h} \hat{\boldsymbol{H}}_{t+s} \tag{3.3.2}$$

The resulting HAR and VARFIMA forecasts for the GE and IBM realized volatility as well as their realized covariation are plotted in figure 2 for the daily forecast horizon ($h = 1$) and in figure A.1 of appendix A for the weekly ($h = 5$) and biweekly ($h = 10$) horizon.

Forecast Evaluation

We evaluate our forecasts by performing Mincer-Zarnowitz regressions, a method for example used by Andersen, Bollerslev, Diebold, and Labys (2003) to analyze the performance of different high- and low-frequency based models for forecasting return variances and covariances of exchange rate returns. The regression takes the following form

$$h_{ij,t:t+h} = b_0 + b_1 \hat{h}_{ij,t:t+h} + u_{ij,t:t+h} \tag{3.3.3}$$

where $h_{ij,t:t+h}$ denotes the i,j-element of the h-day realized volatility matrix $\boldsymbol{H}_{t:t+h}$ and $\hat{h}_{ij,t:t+h}$ its forecast, the i,j-element of $\hat{\boldsymbol{H}}_{t:t+h}$. This regression is conducted for each forecasting horizon and each of our models for the three covolatilities, resulting in 27 regressions overall. These regressions' R^2 are then used to evaluate the models' forecasting performance. The estimation results are printed in table 2: For the daily out-of-sample forecasting period, the VARFIMA performs best at forecasting the IBM stock's realized volatility and the GE-IBM realized covariation whereas the HAR is best at forecasting the GE stock's realized volatility. Nonetheless, the two models' R^2 for the daily horizon are quite close. The GO-GARCH is clearly the worst and only explains a significant share of the GE realized volatility variance where it comes with an R^2 of 0.1452 close to the HAR's performance which exhibits an R^2 of 0.1509. The picture of the weekly forecasts is similar to that of the daily ones: The VARFIMA performs best at forecasting IBM realized volatility and the GE-IBM realized covariation whereas HAR is best at forecasting GE realized volatility. Although the GO-GARCH performs better in terms of higher R^2 at the weekly horizon, it can still just explain significant shares of the GE realized volatility. The picture changes for the biweekly forecasting horizon: Comparing the VARFIMA and the HAR, the HAR's R^2 are higher for each realized covariation than the VARFIMA's. Moreover, the GO-GARCH now outperforms the other two models at the forecasting of the GE realized volatility.

All in all, the GO-GARCH performs relatively good at forecasting the GE stock's realized volatility over all forecasting horizons, but can hardly explain the variance in the remaining two realized covariations. This result is consistent with the results of Andersen, Bollerslev, Diebold, and Labys (2003) who fit univariate GARCH models to the DM/\$, ¥/\$ and the ¥/DM return series and find relatively high R^2 for the ¥/\$ and ¥/DM series whereas the R^2 for the DM/\$ series is relatively low over all forecasting horizons. Furthermore, we observe that the R^2 are highest for the weekly forecasting horizon implying that medium-horizon forecasts are actually more exact than short-term (i.e. daily) forecasts.

Another way of evaluating our forecast is by means of a loss-function as e.g. used by Chiriac and Voev (2011) who opt for the Root Mean Squared Error (RMSE) based on the Frobenius norm of the matrix of forecast errors

$$\boldsymbol{E}_{t:t+h} \equiv \boldsymbol{H}_{t:t+h} - \hat{\boldsymbol{H}}_{t:t+h} \tag{3.3.4}$$

and thus

$$\text{RMSE} \equiv \sqrt{\sum_{\tau}(e^2_{11,\tau:\tau+h-1} + e^2_{22,\tau:\tau+h-1} + 2e^2_{12,\tau:\tau+h-1})} \tag{3.3.5}$$

where $e_{ij,\tau:\tau+h-1}$ denotes the i,j-element of $\boldsymbol{E}_{\tau:\tau+h-1}$, $\tau \equiv p(h) + (p(h) - 1)(h-1)$, $p(h) = 1, 2, \ldots, \lfloor T^*/h \rfloor$ and T^* is the size of the out-of-sample period. For example, as in our case we have $T^* = 125$, thus for a forecasting period of $h = 10$ days there are $\lfloor T^*/h \rfloor = 12$ out-of-sample periods and hence $\tau = 1, 11, \ldots, 111$. Note that even ex-post, the true variance of h-day returns $\boldsymbol{\Sigma}_{t:t+h}$, $\boldsymbol{\Sigma}_{t:t+h} \equiv \text{Var}(\sum_{s=1}^{h} \boldsymbol{r}_{t+s})$, is not known and thus the ex-post realized volatility $\boldsymbol{H}_{t:t+h}$ is just a proxy for the true variance-covariance matrix. However, as pointed out by Andersen and Bollerslev (1998), $\boldsymbol{\Sigma}_{t:t+h}$ can in principle be approximated arbitrarily close by the realized volatility matrix $\boldsymbol{H}_{t:t+h}$ by increasing the intraday return frequency. Hence, it is appropriate to evaluate the GO-GARCH using realized volatilities instead of squared and cross-product returns as these measures are just less efficient estimates of the true return volatility. The RMSE's are given in table 3: For the daily forecasting horizon, the VARFIMA performs best whereas for the weekly and biweekly horizon, the HAR outperforms the other models. The high forecasting errors of the GO-GARCH models can be explained by looking at the squared returns: As mentioned above, MGARCH models have a VARMA representation and can thus be interpreted as fitting a line through squared and cross-product daily returns. Comparing table B.3 and B.4 in appendix B shows that the full-sample average realized volatility for the GE stock is 0.0163%, for the IBM stock 0.0084% and their average realized covariation is 0.0062% whereas the full-sample average squared daily returns for the GE stock is 0.0296%, for the IBM stock 0.0165% and their average cross-product return is 0.0141%. So it appears that volatility estimates based on daily returns are strongly upward biased and therefore high RMSE's are not surprising for the GO-GARCH model. Nonetheless the average squared returns and cross-products will asymptotically converge to the true unconditional daily volatility and results may at least improve for the GO-GARCH in larger samples than the one used in this paper. Our results concerning the GO-GARCH model are consistent with the findings of Chiriac and Voev (2011) that MGARCH models are outperformed by models using high-frequency data. On the other hand, they find that the VARFIMA is best at forecasting over the 1-, 5- and 10-day forecasting horizon whereas in our case, the VARFIMA is best at the daily forecasting horizon and the HAR performs best at the weekly and biweekly forecasting horizon. Note, however, that the weekly (biweekly) horizon only makes use of 25 (12) forecasts whereas Chiriac and Voev (2011) use 129 (64) forecasts, thus different results for the weekly and biweekly horizons may be attributed to the differences in sample sizes.

Table 2: Mincer-Zarnowitz regressions based on the out-of-sample predictions for the two realized volatilities (RV) and the realized covariation (RCV) between the two stocks. For each of the three forecasting horizons, the ex-post realized volatilities and covariation were projected on the VARFIMA-, HAR- and GO-GARCH forecasts. The estimated parameters for this linear projection as well as the resulting R^2 are printed in the table below, standard errors are given in brackets. The regression for the daily forecasting horizon makes use of $N = 125$ observations, i.e. each single day of the out-of-sample period, the regression for the weekly horizon uses $N = 125/5 = 25$ forecasts and the regression for the biweekly horizon uses $N = \lfloor 125/10 \rfloor = 12$ forecasts.

	Daily ($N = 125$)			*Weekly* ($N = 25$)			*Biweekly* ($N = 12$)		
	b_0	b_1	R^2	b_0	b_1	R^2	b_0	b_1	R^2
GE-RV									
VARFIMA	0.0000	0.6985	0.1475	0.0000	0.8069	0.2683	0.0008	0.1827	0.0148
	(0.0000)	(0.1514)		(0.0002)	(0.2779)		(0.0006)	(0.4717)	
HAR	0.0001	0.5312	0.1509	0.0002	0.5839	0.2784	0.0003	0.7093	0.0155
	(0.0000)	(0.1136)		(0.0001)	(0.196)		(0.0003)	(0.3192)	
GO-GARCH	0.0000	0.2872	0.1452	0.0003	0.2341	0.2294	0.0000	0.2872	0.1452
	(0.0000)	(0.06283)		(0.0001)	(0.08945)		(0.0000)	(0.0628)	
IBM-RV									
VARFIMA	0.0000	0.7224	0.1579	0.0001	0.7778	0.2405	0.0003	0.5562	0.1788
	(0.0000)	(0.1504)		(0.0001)	(0.2882)		(0.0003)	(0.3769)	
HAR	0.0000	0.4650	0.1227	0.0000	0.5155	0.2007	0.0002	0.6975	0.2123
	(0.0000)	(0.1121)		(0.0001)	(0.2145)		(0.0002)	(0.2580)	
GO-GARCH	0.0000	0.2876	0.0374	-0.0002	0.2630	0.0715	0.0004	0.2225	0.0704
	(0.0000)	(0.1315)		(0.0001)	(0.1976)		(0.0004)	(0.2558)	
GE-IBM-RCV									
VARFIMA	0.0000	0.8254	0.2205	0.0000	0.9654	0.3437	0.0001	0.5715	0.1188
	(0.0000)	(0.1399)		(0.0001)	(0.2782)		(0.0002)	(0.4921)	
HAR	0.0000	0.5901	0.2062	0.0001	0.5761	0.2956	0.0000	0.8007	0.2511
	(0.0000)	(0.1044)		(0.0000)	(0.1854)		(0.0000)	(0.2587)	
GO-GARCH	0.0000	0.1177	0.0459	0.0001	0.1308	0.1048	0.0003	0.0653	0.0220
	(0.0000)	(0.0484)		(0.0000)	(0.0797)		(0.0001)	(0.1378)	

Table 3: RMSE's based on the Frobenius norm of the forecast errors. RMSE's were computed for each model for the daily, weekly and biweekly forecasting horizon.

	Daily[(a)]	*Weekly*[(a)]	*Biweekly*[(a)]
VARFIMA	0.6734185	1.057364	1.794726
HAR	0.7046182	1.002439	1.651637
GO-GARCH	1.801261	4.132416	5.850871

(a) All values are scaled by factor 100.

4. Conclusion

We have estimated a VARFIMA$(1, \delta, 1)$, an HAR as well as a GO-GARCH model and seen that – using Mincer-Zarnowitz regressions and the RMSE criterion for forecast evaluation – our high-frequency data based models outperform the GO-GARCH model in out-of-sampling forecasting over a daily, weekly and biweekly forecasting horizon. For the daily and weekly forecasting horizon, both the HAR and VARFIMA can explain significant shares of the variance in any of the three realized volatility/covariation series where the forecasting performance is best at the weekly forecasting horizon. Both the Mincer-Zarnowitz regressions as well as the RMSE criterion suggest that the VARFIMA is best at forecasting at the daily horizon, for the weekly forecasting horizon however, the VARFIMA appears to perform best according to the Mincer-Zarnowitz regressions whereas applying the RMSE criterion, the HAR model outperforms the VARFIMA model. Given that the differences RMSE's and R^2 between the HAR and the VARFIMA model are low for the daily and weekly forecasts (e.g. a maximum of 5 percentage points difference for the R^2), it is generally not clear which of the two models to prefer. Although the HAR outperforms the VARFIMA in case of the biweekly forecasts, this result needs to be interpreted carefully as there are only 12 biweekly out-of-sample forecasting periods included in the forecast evaluation.

A possible problem not addressed in this paper is the *curse of dimensionality* when a high number of stocks is included in the analysis. The results of Chiriac and Voev (2011) suggest that the analysis remains feasible for six stocks, estimation and forecasting using even more stocks remains a question for further research. Furthermore, our analysis is restricted to volatility forecasting relying exclusively on stock price data, the general form of the VARFIMA, however, allows to include other economic variables that may have an effect on stock return volatility. Including such factors may improve the forecasting performance of our VARFIMA model and remains another problem to be addressed by future research.

References

ALEXANDER, C. (2001): "A Primer on the Orthogonal GARCH Model," *ISMA Centre, The Business School for Financial Markets, University of Reading*.

ANDERSEN, T. G., AND T. BOLLERSLEV (1998): "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts," *International Economic Review*, 39, pp. 885–905.

ANDERSEN, T. G., T. BOLLERSLEV, F. X. DIEBOLD, AND P. LABYS (2003): "Modeling and Forecasting Realized Volatility," *Econometrica*, 71(2), 579–625.

BERAN, J. (1995): "Maximum Likelihood Estimation of the Differencing Parameter for Invertible Short and Long Memory Autoregressive Integrated Moving Average Models," *Journal of the Royal Statistical Society*, 57, 659–672.

CHIRIAC, R., AND V. VOEV (2011): "Modelling and forecasting multivariate realized volatility," *Journal of Applied Econometrics*, 26(6), 922–947.

CORSI, F. (2009): "A Simple Approximate Long-Memory Model of Realized Volatility," *Journal of Financial Econometrics*, 7(2), 174–196.

GEWEKE, J., AND S. PORTER-HUDAK (1983): "The Estimation and Application of Long Memory Time Series Model," *Journal of Time Series Analysis*, 4, 221–238.

GHALANOS, A. (2014): *rmgarch: Multivariate GARCH models.*R package version 1.2-6.

HAMILTON, J. D. (1994): *Time Series Analysis*. Princeton University Press.

LÜTKEPOHL, H. (2005): *New Introduction to Multiple Time Series Analysis*. Springer.

POLITIS, D. N., AND J. P. ROMANO (1994): "The Stationary Bootstrap," *Journal of the American Statistical Association*, 89, 1303–1313.

ROBINSON, P. M. (1995): "Log-Periodogram Regression of Time Series with Long Range Dependence," *The Annals of Statistics*, 23(3), 1048–1072.

SOWELL, F. (1989): "Maximum Likelihood Estimation of Fractionally Integrated Time Series Models," *Carnegie Mellon University*.

VAN DER WEIDE, R. (2002): "GO-GARCH: a multivariate generalized orthogonal GARCH model," *Journal of Applied Econometrics*, 17(5), 549–564.

A. Supplementary Figures

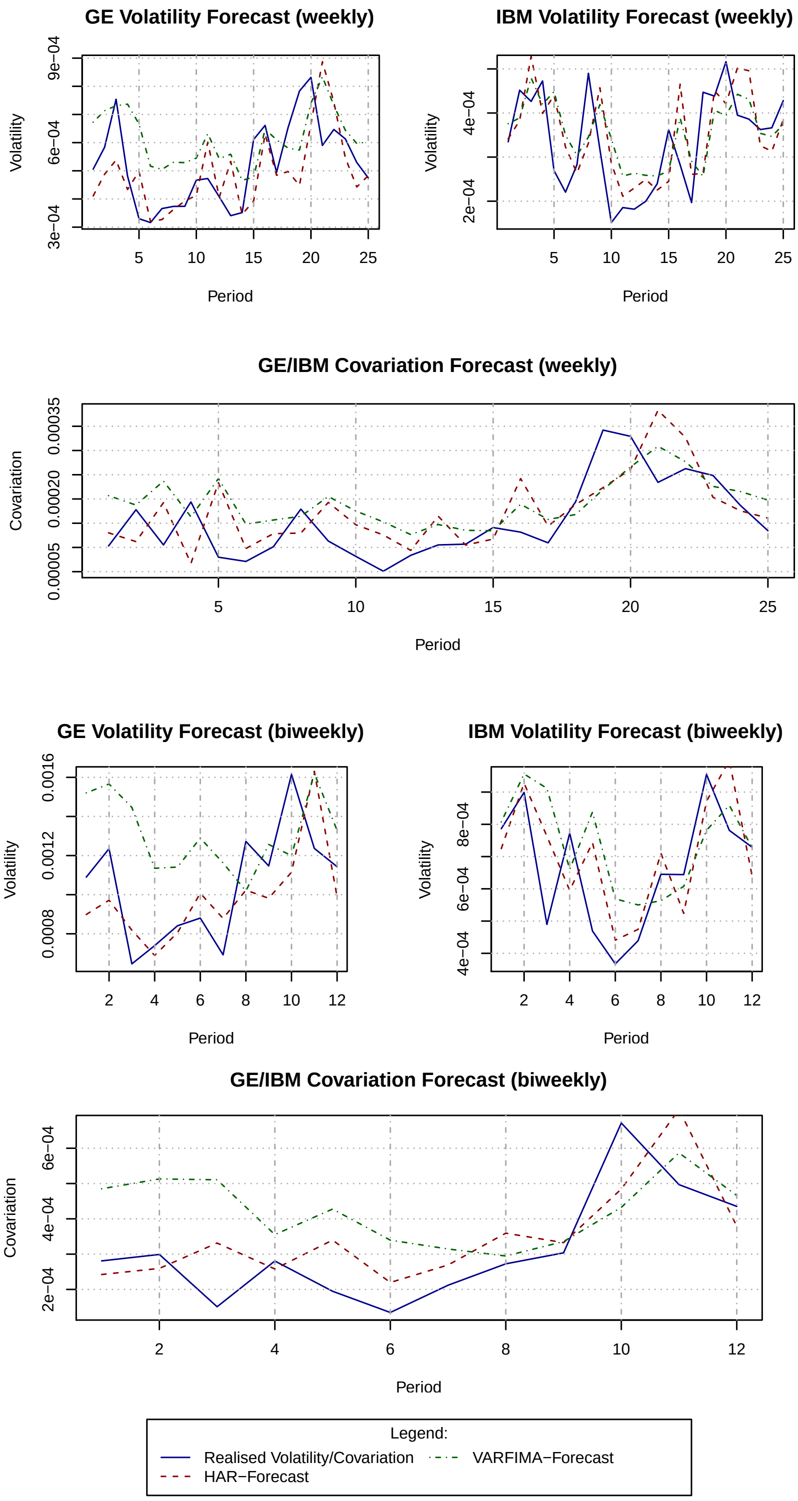

Figure A.1: Weekly and biweekly forecasts of realized volatility and covariation for the GE and IBM stocks.

B. Supplementary Tables

Table B.1: Descriptive statistics on intradaily returns, overnight returns are exluded from the sample.

Stock	Min	Max	Mean	SD	Skewness	Kurtosis
Intradaily Returns						
GE	-2.3080%	3.6690%	-0.0001%	0.1665%	0.2647	17.8555
IBM	-1.5430%	1.3410%	0.0014%	0.1207%	-0.0077	11.8468

Table B.2: Descriptive statistics on daily returns for the GE and IBM stocks, based on 376 daily return observations.

Stock	Min	Max	Mean	SD	Skewness	Kurtosis
Daily Returns						
GE	-6.7286%	6.8631%	0.0347%	1.7217%	-0.1827	4.9217
IBM	-5.2901%	5.3548%	0.0750%	1.2852%	-0.2766	5.7277

Table B.3: Descriptive statistics on realized volatility for the GE and IBM stocks. The statistics are based on $T = 377$ observations, i.e. our full sample.

Stock	Min	Max	Mean	SD	Skewness	Kurtosis
Realized Volatility						
GE	0.0029%	0.1905%	0.0163%	0.0189%	4.8640	35.3210
IBM	0.0011%	0.1151%	0.0084%	0.0105%	5.0870	39.3690
Realized Covariation						
GE/IBM	-0.0013%	0.1241%	0.0062%	0.0110%	5.7020	47.5700

Table B.4: Descriptive statistics on squared and cross-product daily returns for the GE and IBM stocks, based on 376 daily return observations.

Stock	Min	Max	Mean	SD	Skewness	Kurtosis
Squared Returns						
GE	0.0000%	0.4710%	0.0296%	0.0585%	4.5151	27.8088
IBM	0.0000%	0.2867%	0.0165%	0.0357%	4.2879	25.3950
Cross-Product						
GE/IBM	-0.1340%	0.2967%	0.0141%	0.0373%	3.8370	25.8503

C. Derivations

C.1. VAR(∞) Representation of VARFIMA($1, \delta, 1$)

First of all, define $\boldsymbol{x}_t^* \equiv \boldsymbol{x}_t - \boldsymbol{\mu}$, assume $\delta_k = \delta \ \forall \ k \in \{1, 2, \dots, K\}$, then write model (2.2.2) as

$$(1 - \phi \mathrm{L})(1 + \theta \mathrm{L})^{-1}(1 - \mathrm{L})^{\delta} \boldsymbol{x}_t^* = \boldsymbol{\epsilon}_t \tag{C.1}$$

and express $(1 - \phi \mathrm{L})(1 + \theta \mathrm{L})^{-1}$ as

$$\begin{aligned}
(1 - \phi \mathrm{L})(1 + \theta \mathrm{L})^{-1} &= (1 - \phi \mathrm{L})(1 + (-\theta \mathrm{L}) + (-\theta \mathrm{L})^2 + \dots) \\
&= 1 + (-\phi - \theta)\mathrm{L} + (\phi\theta + \theta^2)\mathrm{L}^2 + (-\phi\theta^2 - \theta^3)\mathrm{L}^3 + \dots \\
&= \sum_{j=0}^{\infty} \alpha_j L^j
\end{aligned} \tag{C.2}$$

with $\alpha_j \equiv (-1)^j(\phi\theta^{j-1} + \theta^j)$ for $j \geq 1$ and $\alpha_0 \equiv 1$. Next, use the definition of $(1 - \mathrm{L})^{\delta}$ to write

$$\begin{aligned}
(1 - \phi \mathrm{L})(1 + \theta \mathrm{L})^{-1}(1 - \mathrm{L})^{\delta} &= \left(\sum_{j=0}^{\infty} \alpha_j L^j\right)\left(\sum_{j=0}^{\infty} b_j(\delta) \mathrm{L}^j\right) \\
&= (1 + \alpha_1 \mathrm{L} + \alpha_2 \mathrm{L}^2 + \dots)(1 + b_1(\delta)\mathrm{L} + b_2(\delta)\mathrm{L}^2 + \dots) \\
&= 1 + \left[b_1(\delta) + \alpha_1\right]\mathrm{L} + \left[b_2(\delta) + b_1(\delta)\alpha_1 + \alpha_2\right]\mathrm{L}^2 + \\
&\quad \left[b_3(\delta) + b_2(\delta)\alpha_1 + b_1(\delta)\alpha_2 + \alpha_3\right]\mathrm{L}^3 + \dots \\
&= \sum_{j=0}^{\infty} \rho_j \mathrm{L}^j
\end{aligned} \tag{C.3}$$

where $\rho_j \equiv \sum_{l=0}^{j} b_l(\delta)\alpha_{j-l}$. Plugging (C.3) into (C.1) yields the VAR(∞) representation of the system.

D. Affidavit (Eidesstattliche Versicherung)

I affirm that this paper was written by myself without any unauthorised third-party support. All used references and resources are clearly indicated. All quotes and citations are properly referenced. This paper was never presented in the past in the same or similar form to any examination board.
I agree that my paper may be subject to electronic plagiarism check.

Ich versichere, dass ich die vorliegende Arbeit ohne Hilfe Dritter und ohne Benutzung anderer als der angegebenen Quellen und Hilfsmittel angefertigt und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Ich bin damit einverstanden, dass meine Arbeit zum Zwecke eines Plagiatsabgleichs in elektronischer Form anonymisiert versendet und gespeichert werden kann.

Place, Date/Ort, Datum	Signature/Unterschrift