Modelling and forecasting multivariate realized volatility — **Source link** ↗

Roxana Chiriac, Valeri Voev

Institutions: University of Konstanz, Aarhus University

Published on: 01 Sep 2011 - Journal of Applied Econometrics (John Wiley & Sons, Ltd.)

Topics: Realized variance and Positive definiteness

Related papers:

- A Simple Approximate Long-Memory Model of Realized Volatility

- Modeling and forecasting realized volatility

- Forecasting multivariate realized stock market volatility

- Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models

- The Wishart Autoregressive process of multivariate stochastic volatility

MODELLING AND FORECASTING MULTIVARIATE REALIZED VOLATILITY

ROXANA CHIRIAC[a#] AND VALERI VOEV[b]*

[a] *Department of Economics, University of Konstanz, Germany*
[b] *CREATES, Aarhus University, Denmark*

SUMMARY

This paper proposes a methodology for dynamic modelling and forecasting of realized covariance matrices based on fractionally integrated processes. The approach allows for flexible dependence patterns and automatically guarantees positive definiteness of the forecast. We provide an empirical application of the model, which shows that it outperforms other approaches in the extant literature, both in terms of statistical precision as well as in terms of providing a superior mean-variance trade-off in a classical investment decision setting. Copyright © 2010 John Wiley & Sons, Ltd.

1. INTRODUCTION

Multivariate volatility modelling is of particular importance in the areas of risk management, portfolio management and asset pricing. Typical econometric approaches include multivariate GARCH models (for a comprehensive review see Bauwens *et al*., 2006), stochastic volatility models (reviewed in Asai *et al*., 2006) and, more recently, realized covariance measures (see, for example, Barndorff-Nielsen and Shephard, 2004; Andersen *et al*., 2001). While in the GARCH and stochastic volatility framework the volatility process is latent, the realized covariance methods employ high-frequency data to enable precise estimation of the daily covariance of the underlying assets, thus making it effectively observable.

A prominent feature of volatility is its strong persistence, which motivated the development of the integrated GARCH (Engle and Bollerslev, 1986), the fractionally integrated GARCH (Baillie *et al*., 1996) and the linear ARCH (Robinson, 1991; Giraitis *et al*., 2000) models. Realized volatility series tend to exhibit a slow decay in the autocorrelation function (see, for example, Andersen and Bollerslev, 1997; Andersen *et al*., 2001), and are modelled by means of fractionally integrated ARMA (ARFIMA) processes by Andersen *et al*. (2003), Oomen (2001) and Koopman *et al*. (2005), among others.

Recently, the literature on multivariate GARCH models has been advancing towards flexible model specifications, applicable to a large number of assets. Yet there is little research on time series models for realized covariance matrices. The existing literature has typically focused on univariate analysis of realized volatilities or single realized covariance (correlation) series. Andersen *et al*. (2003) model log-realized volatilities and realized correlations with univariate ARFIMA models,

---

* Correspondence to: Valeri Voev, CREATES, School of Economics and Management, University of Aarhus, Building 1326, Batholins Alle 10, DK-8000 Aarhus C, Denmark. E-mail: vvoev@creates.au.dk
# Since the on-line publication of this article, Roxana Chiriac has changed her name into Roxana Halbleib

while Corsi (2009) and Corsi and Audrino (2007) develop heterogeneous autoregressive (HAR) models to capture the strong persistence through a hierarchical autoregressive structure. A problem which arises in this context is that the matrix constructed from the variance and correlation forecasts obtained from disjoint models is not guaranteed to be positive definite. In order to obtain a forecast of the entire covariance matrix, Voev (2007) proposes a methodology in which the univariate variance and covariance forecasts can be combined to produce a positive definite matrix forecast. A drawback of this approach is that the dynamic linkages among the variance and covariance series (e.g., volatility spillovers) is neglected. The Wishart autoregressive (WAR) model of Gourieroux *et al*. (2009), and the model of Bauer and Vorkink (2007), who employ the matrix log transformation to guarantee positive definiteness of the forecast, are among the few proposed approaches for the dynamics of the whole realized covariance matrix. The standard WAR model, however, is incapable of producing long memory type dependence patterns and is built on latent processes, whose interpretation is difficult and which makes the introduction of exogenous forecasting variables problematic. The study of Bauer and Vorkink (2007) differs from ours in that its primary focus is to investigate the forecasting power of various predictive variables, such as past returns, risk-free interest rate and dividend yield, while our main contribution is to improve upon the ability to characterize the dynamic aspects of volatility and to comprehensively analyse the resulting forecasting implications.

The approach developed in this paper involves the following three steps: first, decomposing the series of covariance matrices into Cholesky factors; second, forecasting the Cholesky series with a suitable time series model; and third, reconstructing the matrix forecast. The positivity of the forecast is thus ensured by 'squaring' of the Cholesky factors, which can be modelled without imposing parameter restrictions. A further advantage of the methodology is that the inclusion of exogenous predictive variables is, at least conceptually, straightforward. The idea of modelling the Cholesky factorization of a volatility matrix is not new. Tsay (2002) discusses its use as a re-parameterization of the latent covariance matrix in a traditional multivariate GARCH framework, while in Gallant and Tauchen (2001) a Cholesky-GARCH type of model is used in the context of efficient method of moments estimation. Interestingly, Pourahmadi (1999) suggests modelling the Cholesky factors of the inverse of the covariance matrix, which can be very appealing in cases where the inverse is the object of direct interest as, for example, in the solution to a minimum-variance portfolio problem. More recently, the idea of modelling the Cholesky factorization of the realized covariance matrix that we advocate here has been put forward, although not implemented empirically, by Andersen *et al*. (2003).

The degree of parameterization (flexibility) of the time series model should be guided by the dimension of the matrix, as well as by the application we have in mind; do we aim at a good in-sample fit, or are we more interested in out-of-sample forecasting? In this paper, our interest is in the latter and hence we tend to favour very parsimonious specifications. The model is based on fractionally integrated processes and can be seen as an application of the multivariate ARFIMA model of Sowell (1989). Estimation is carried out using the conditional maximum likelihood (ML) method developed in Beran (1995). The conditional approach is preferred over the exact ML methods proposed in the univariate case by Sowell (1992) and An and Bloomfield (1993), since the exact ML approach requires the inversion of a $Tn \times Tn$ matrix, where T is the sample size and n is the dimension of the process. For a review of inference of and forecasting with ARFIMA models, we direct the reader to Doornik and Ooms (2004).

To assess the merits of our model in practice, we undertake a comprehensive out-of-sample forecasting study using recent data, partially covering the ongoing financial crisis. In the analysis,

we consider one-step (daily), five-step (weekly) and 10-step (biweekly) horizons, using direct and iterative forecasts from a range of models using both high-frequency and daily data. An issue we need to address in the context of volatility forecast evaluation is that *ex post* measures of volatility are subject to estimation error. We use results from Hansen and Lunde (2006) and Patton (2009), who derive the necessary and sufficient conditions a loss function should satisfy, in order for the ranking of the models to be robust to noise in the *ex post* volatility proxy.

A further problem which arises is how to determine the best-performing model(s) in a model-rich environment. Pairwise comparisons of loss functions can be misleading, unless we consider some sort of Bonferroni bound correction, and would involve a geometrically increasing number of tests as the number of models increases. Fortunately, Hansen *et al*. (2009) have developed a methodology—the model confidence set (MCS) approach—which allows selecting a set of models containing the best one with a certain level of confidence, naturally adapting to the number of models and thus requiring only one test. Using the root mean squared error (RMSE) criterion, we show that the forecasts based on the fractionally integrated model proposed in this paper have the smallest error for all forecasting horizons. Applying the MCS approach reveals that our model is often the only one selected by the procedure. To get a feeling of what this improved statistical performance implies from a practitioner's point of view, we analyse the performance of mean–variance efficient portfolios and document that our approach leads to a superior mean–variance trade-off. Similar studies have been carried out by Fleming *et al*. (2003) and Liu (2009), but for a more restricted set of models and with a different evaluation methodology.

The paper is structured as follows: Sections 2 and 3 describe the conditional covariance models and the forecasting procedures, Section 4 reports estimation and forecasting results and Section 5 concludes.

2. DYNAMIC CONDITIONAL COVARIANCE MODELS

Let r_t be a vector of daily log returns of dimension $n \times 1$, where n represents the number of assets considered. The process r_t can be written as

$$r_t = E[r_t|\mathcal{F}_{t-1}] + \varepsilon_t \tag{1}$$

where $\mathcal{F}_{t-1}$ is the information set consisting of all relevant information up to and including $t-1$. We assume that the innovation term can be expressed as $\varepsilon_t = H_t^{1/2} z_t$, where H_t is a symmetric positive definite matrix of dimension $n \times n$, $H_t^{1/2}$ is its Cholesky decomposition and z_t is an $n \times 1$ vector assumed to be i.i.d. with $E[z_t] = 0$ and $V[z_t] = I_n$. In the GARCH framework, the latent conditional covariance matrix H_t is specified parametrically, while more recently Barndorff-Nielsen and Shephard (2004) and Andersen *et al*. (2001) proposed the realized covariance matrix Y_t as a non-parametric estimator of H_t.[1] Using infill asymptotics, Barndorff-Nielsen and Shephard (2004) show that Y_t converges in probability to H_t under very general assumptions on the underlying price process and is thus a consistent, asymptotically error-free estimator of H_t. In the following section, we introduce our modelling framework for the matrix-valued process Y_t, while in the subsequent sections we briefly review alternative methods proposed in the literature.

---

[1] The realized covariance matrix is the sum of the products of high-frequency (e.g., 5-minute) returns within a given day t. We elaborate on the computation of Y_t in the empirical section.

2.1. VARFIMA(p, d, q) Model

Consider the Cholesky decomposition of Y_t, given by the upper triangular matrix P_t, such that $P'_t P_t = Y_t$. Since Y_t is symmetric and positive definite, whenever the number of intraday returns is greater than n, the elements of P_t are all real (see, for example, Golub and van Loan, 1996). Let $X_t = vech(P_t)$ be the $m \times 1$ vector obtained by stacking the upper triangular components of P_t, where $m = n(n+1)/2$. We propose to model X_t as a vector autoregressive fractionally integrated moving average (VARFIMA(p,d,q)) process, defined below:

Definition 1. *The VARFIMA(p,d,q) process is defined as*

$$\Phi(L)D(L)[X_t - BZ_t] = \Theta(L)\varepsilon_t, \qquad \varepsilon_t \sim N(0, \Sigma) \qquad (2)$$

where Z_t is a vector of exogenous variables of dimension $k \times 1$, B is a matrix of coefficients of dimension $m \times k$, $\Phi(L) = I_m - \Phi_1 L - \Phi_2 L^2 - \ldots - \Phi_p L^p$, $\Theta(L) = I_m - \Theta_1 L - \Theta_2 L^2 - \ldots - \Theta_q L^q$ are matrix lag polynomials with Φ_i, $i = 1, \ldots, p$ and Θ_j, $j = 1, \ldots, q$—the AR- and MA-coefficient matrices, and $D(L) = diag\{(1-L)^{d_1}, \ldots, (1-L)^{d_m}\}$, where $d_1, \ldots, d_m$ are the degrees of fractional integration of each of the m elements of the vector X_t. We assume that the roots of $\Phi(L)$ and $\Theta(L)$ lie outside the unit circle.

The model presented here has been studied by Sowell (1989), who shows that X_t is stationary if $d_i < 0.5$ for $i = 1, \ldots, m$. In equation (2), one could include in Z_t variables that are documented to have an effect on stock market volatility, such as functions of trading volume (Lamoureaux and Lastrapes, 1990), corporate bond returns (Schwert, 1989) or short-term interest rates (Glosten *et al.*, 1993). Since a single exogenous regressor implies m additional parameters in the model, in order to keep the model tractable, restrictions might need to be imposed by structuring the matrix B appropriately. Whether the response of the volatility/covariance series to exogenous variables can be modelled satisfactorily for large n remains ultimately an empirical question.

The motivation for modelling the Cholesky factors, rather than the elements of Y_t directly, is that by doing so, we do not need to impose parameter restrictions on the model. While in sample, all elements of X_t corresponding to the main diagonal of P_t are positive, certain parameterizations of the model could lead to negative values out-of-sample. This, however, is unproblematic as any (invertible) upper triangular matrix constructed from the elements of X_t provides a positive definite matrix through the 'reverse' of the Cholesky transformation:

$$Y_{ij,t} = \sum_{l=1+\frac{i(i-1)}{2}}^{\frac{i(i+1)}{2}} X_{l,t} X_{l+\frac{j(j-1)}{2}-\frac{i(i-1)}{2},t}, \qquad i, j = 1, \ldots, n, \ j \geq i \qquad (3)$$

where $X_{l,t}$ is the lth element of X_t. This property has also motivated Tsay (2002) to use the Cholesky decomposition as a GARCH model re-parameterization. We note here that an unbiased prediction of X_t will, in general, not lead to an unbiased forecast of Y_t, an issue which we return to further below. In terms of estimation, we face the problem that the parameters of the unrestricted VARFIMA models are not identified, due to the non-uniqueness of VARMA models, discussed in Lütkepohl (2005). In our paper, we consider the so-called final equations form, which provides a unique parameterization and is defined below.

Definition 2. *The n-dimensional VARMA(p,q) representation $\Phi(L)Y_t = \Theta(L)\varepsilon_t$ is said to be in final equations form if $\Theta_0 = I_n$ and $\Phi(L) = 1 - \phi_1 L - \ldots - \phi_p L^p$ is a scalar operator with $\phi_p \neq 0$.*

Following this definition, we estimate the model in final equation form, restricting the AR polynomial to be a scalar polynomial. The assumption of normally distributed error terms gives rise to a Gaussian likelihood function, which, maximized under certain regularity conditions (see Gourieroux and Monfort, 1995), and the assumption that the conditional mean function is well specified, provides consistent estimates of the model parameters. Doornik and Ooms (2004) discuss various ways of estimating ARFIMA models, including maximum likelihood (ML) as well as nonlinear least squares (NLS) methods. In this paper, we opt for the approximate ML approach in the spirit of Beran (1995), which is also applicable for non-stationary processes with $d > 0.5$. The estimation effectively minimizes the sum of squared residuals, circumventing the estimation of Σ, which is irrelevant for the purposes of constructing a point forecast.

Table I summarizes the total number of parameters for a general VARFIMA (p,d,q) model in final equation form, as well as for two restricted model specifications considered in this paper. The workhorse in our empirical study will be a $(1, d, 1)$ specification with a scalar Θ:

$$(1 - \phi L)D(L)[X_t - c] = (1 - \theta L)\varepsilon_t, \qquad \varepsilon_t \sim N(0, \Sigma) \qquad (4)$$

where c is an $m \times 1$ vector of constants and $D(L) = \text{diag}\{(1 - L)^{d_1}, \ldots, (1 - L)^{d_m}\}$ (Model 1) or $D(L) = (1 - L)^d I_m$ (Model 2). We will exclusively rely on Model 2 for forecasting, and we only estimate Model 1 in order to test the null hypothesis $d_1 = \ldots = d_m$, which cannot be rejected for our dataset at the 5% level. The main reason to consider these very restricted specifications is the often observed empirical result that overparameterized models lead to poor out-of-sample forecasts. Hansen (2009) provides a theoretical treatment of the problem and shows that models which are selected on the basis of their in-sample fit are most likely to be the worst models out-of-sample. A further advantage of parsimony, is that it eases estimation and allows more freedom in terms of increasing the dimension n.[2] Regarding the mean vector c, it can be estimated in a first step as the sample mean of X_t which leaves only three parameters for estimation in the second step.[3] To

Table I. Number of parameters for the general VARFIMA(p, d, q) model and its restricted specifications considered in this paper. Model 1 is a VARFIMA(1, d, 1) with scalar Φ and Θ; Model 2 further restricts $d_1 = d_2 = \ldots = d_m$. In all specifications $k = 1$, as we only estimate a constant

	Dimension	Number of parameters	Model 1	Model 2
$\Phi(L)$	1×1	p	1	1
$D(L)$	$m \times m$	m	m	1
B	$m \times k$	km	m	m
$\Theta(L)$	$m \times m$	qm^2	1	1
Total number of parameters		$qm^2 + (k+1)m + p$	$2m + 2$	$m + 3$

[2] Interestingly, in the new RiskMetrics methodology, Zumbach (2007) finds strong empirical support for long memory type of autocorrelation decay of financial volatility, and concludes that 'one decay factor can be used to model all time series'.
[3] This is related to correlation targeting in DCC models in which the unconditional correlation matrix is set equal to the sample correlation matrix

account for the 'generated regressor' problem in the second step of the estimation, we calculate bootstrapped standard errors. The number of second-step parameters in Model 2 is independent of n and therefore the model can, at least in principle, be applied to very large dimensions, say $n > 100$. The Cholesky decomposition is not the only way of transforming the covariance matrix to guarantee positivity. Bauer and Vorkink (2007) use the matrix log transformation and model the dynamics of the so-called log volatilities, defined as $a_t = vech(A_t)$, where A_t is given by $A_t = B_t \log(G_t) B_t'$ and B_t and G_t result from the spectral decomposition $Y_t = B_t G_t B_t'$. In order to assess whether the matrix decomposition method makes a difference empirically, we also apply the VARFIMA model to the log-volatility series a_t.

2.2. Heterogeneous Autoregressive (HAR) Model

An interesting alternative to fractionally integrated models for persistent processes is proposed by Corsi (2009), who introduces a mixed-frequency hierarchical AR model, where the daily volatility is a function of lagged daily, weekly and monthly volatility. For the sake of parsimony, we consider the following HAR specification for X_t (and a_t):

$$X_{t+1,d} = c^{(d)} + \beta^{(d)} X_t^{(d)} + \beta^{(w)} X_t^{(w)} + \beta^{(bw)} X_t^{(bw)} + \beta^{(m)} X_t^{(m)} + \omega_{t+1,d} \tag{5}$$

where d stands for the daily, w for the weekly (5 days), bw for the biweekly (10 days) and m for the monthly (20 days) frequency, $c^{(d)}$ is an $m \times 1$ parameter vector and the β's are scalar parameters, which can easily be estimated by OLS. The regressors $X_t^{(\cdot)}$ are averages of past values of X_t scaled to match the frequency of the left-hand-side (LHS) variable; e.g., in the equation above $X_t^{(bw)} = \frac{1}{10} \sum_{i=0}^{9} X_{t-i}$. While in Corsi (2009) the biweekly frequency is not considered, we include it since we are interested in 5- and 10-day-ahead forecasts.

2.3. Wishart Autoregressive (WAR) Model

Gourieroux *et al.* (2009) introduce the Wishart autoregressive model based on the distribution of the sample variance–covariance matrix, known as the Wishart distribution. Let $x_{k,t}$ with $k = 1, \ldots, K$, be independent Gaussian VAR(1) processes of dimension n:

$$x_{k,t} = M x_{k,t-1} + \varepsilon_{k,t}, \qquad \varepsilon_{k,t} \text{ i.i.d. } N(0, \Omega) \tag{6}$$

Gourieroux *et al.* (2009) define the process given by

$$Y_t = \sum_{k=1}^{K} x_{k,t} x_{k,t}' \tag{7}$$

to be a WAR process of order 1 and dimension n, with K degrees of freedom denoted by $W_n(K, M, \Omega)$. Substituting equation (6) into equation (7), Y_t can be written as

$$Y_t = M Y_{t-1} M' + K\Omega + \eta_t \tag{8}$$

where η_t is a heteroscedastic error term with conditional mean zero. The process Y_t has a non-degenerate Wishart distribution as long as $K \geq n$. In practice, K needs to be estimated and Chiriac

(2007) and Bonato (2009) show that empirically one often obtains $K < n$ which can be attributed to the strong time variation and the presence of extreme values in the covariance matrix series. We estimate a diagonal WAR(1) specification in order to keep it tractable. Bonato *et al.* (2009) suggest combining the WAR model with HAR dynamics on the latent processes $x_{k,t}$ and use blocking of the parameter matrices to reduce parameter proliferation. Following their idea, we implement the following diagonal WAR-HAR specification:

$$Y_{t+1,d} = M^{(d)}Y_t^{(d)}M^{(d)'} + M^{(w)}Y_t^{(w)}M^{(w)'} + M^{(bw)}Y_t^{(bw)}M^{(bw)'} + M^{(m)}Y_t^{(m)}M^{(m)'} + K\Omega + \eta_t$$

where the $Y_t^{(\cdot)}$'s are realized covariance matrices aggregated at the corresponding frequency and scaled to match the frequency of the LHS variable, and the M's are diagonal $n \times n$ parameter matrices.

The list of models is completed by the following two GARCH-class models, based on daily data.

2.4. (Fractionally Integrated) Dynamic Conditional Correlation Model

We assume that the conditional mean of daily returns is constant, $E[r_t|\mathcal{F}_{t-1}] = \mu$ (see equation (1)), and estimate the models on the demeaned series of daily returns. Engle (2002) proposed a multivariate GARCH model with univariate GARCH(1,1) conditional variances, $h_{ii,t}$, and dynamic conditional correlations (DCC):

$$H_t = D_t R_t D_t \tag{9}$$

where $D_t = \text{diag}(h_{11,t}^{1/2} \ldots h_{nn,t}^{1/2})$ and $h_{ii,t} = w_i + \alpha_i \varepsilon_{i,t-1}^2 + \beta_i h_{ii,t-1}$ with $w_i, \alpha_i, \beta_i \geq 0$ and $\alpha_i + \beta_i < 1$, $\forall i = 1, \ldots, n$. The correlation matrix is expressed as

$$R_t = (\text{diag}(Q_t))^{-\frac{1}{2}} Q_t (\text{diag}(Q_t))^{-\frac{1}{2}} \tag{10}$$

$$Q_t = (1 - \theta_1 - \theta_2)\overline{Q} + \theta_1 u_{t-1}u_{t-1}' + \theta_2 Q_{t-1}$$

where u_t is the vector of de-volatilized residuals with elements

$$u_{i,t} = \frac{\varepsilon_{i,t}}{\sqrt{h_{ii,t}}}, \quad i = 1, \ldots, n$$

and $\overline{Q}$ is the unconditional covariance of u_t. In the fractionally integrated version of the DCC (FIDCC), we model the volatilities as FIGARCH(1, d, 0) processes (see Baillie *et al.*, 1996):

$$(1 - L)^{d_i}\varepsilon_{i,t}^2 = \omega_i + v_{i,t} - \beta_i v_{i,t-1} \tag{11}$$

where $0 < d_i < 1$, $v_{i,t} = \varepsilon_{i,t}^2 - h_{ii,t}$, d_i and β_i are positive, and $\beta_i \leq d_i$, $\forall i = 1, \ldots, n$. The FIGARCH(1, d, 0) model can be expressed as an ARCH(∞):

$$h_{ii,t} = \frac{\omega_i}{1 - \beta_i} + \delta_i(L)\varepsilon_{i,t}^2 = \frac{\omega_i}{1 - \beta_i} + \sum_{j=1}^{\infty} \delta_{i,j}\varepsilon_{i,t-j}^2 \tag{12}$$

where $\delta_i(L) = 1 - (1 - \beta_i L)^{-1}(1 - L)^{d_i} = \sum_{j=1}^{\infty} \delta_{i,j} L^j$. R_t remains as above. Again, to take account of the multistep estimation, we compute bootstrapped standard errors.

3. FORECASTING

We focus on 1-day, 5-day and 10-day horizons; that is, we consider fairly short-term forecasts. As a notational convention, we use hats to denote estimated parameters and quantities, as well as forecasts. For example, the forecast of the conditional variance of the s-day-ahead return r_{t+s} will be denoted by $\hat{H}_{t+s} \equiv E_t[H_{t+s}]$, where the conditional expectation is given by the particular model at hand. The information set at time t will typically include high-frequency information in the form of history of the process Y_t. An exception is the (FI)DCC model, which uses only the history of daily returns r_t. In most cases what we need, however, is not the variance of r_{t+s}, but rather the variance of the s-day return $r_{t:t+s} = r_{t+1} + \ldots + r_{t+s}$. An important assumption we make is that the returns are conditionally uncorrelated so that $V_t[r_{t:t+s}] = \sum_{i=1}^{s} H_{t+i}$. This assumption is theoretically appealing, empirically reasonable at the daily frequency, and not uncommon in the literature; see, for example, Ghysels *et al.* (2009), who address the issue of longer-horizon volatility forecasting based on daily data.

We take the perspective of an agent who is only interested in revising his forecast every s days. This necessitates working with non-overlapping s-day returns and prevents having to deal with overlap-induced autocorrelation in the returns. In order to carry out an out-of-sample forecast evaluation, we split the whole sample of length T into an in-sample $t = 1, \ldots, T'$ and an out-of-sample $t = T' + 1, \ldots, T$. The aim is to forecast $H_{t:t+s} \equiv \sum_{i=1}^{s} H_{t+i}$, for $t = T', T' + s, \ldots, T' + s\left(\lfloor \frac{T-T'}{s} \rfloor - 1\right)$. One possible strategy is to consider the iterative (cumulative) forecast $\hat{H}_{t:t+s}^{\text{iter}} \equiv \sum_{i=1}^{s} \hat{H}_{t+i}$. Alternatively, we can use the aggregate series $Y_{\tau_s} = \sum_{i=0}^{s-1} Y_{t-i}$ and $r_{\tau_s} = \sum_{i=0}^{s-1} r_{t-i}$, $t = s + (T' \div s), 2s + (T' \div s), \ldots, T' - s, T'$, where $\div$ denotes the modulo operator, and make a direct one-step-ahead forecast $\hat{H}_{t:t+s}^{\text{dir}} \equiv E_{\tau_s}[H_{\tau_s+1}]$, $\tau_s = T', \ldots, \lfloor \frac{T-T'}{s} \rfloor$.[4] Ghysels *et al.* (2009) find that direct approaches (e.g., a GARCH model on monthly return observations used to produce a 1-month-ahead forecast) are in general inferior to iterated and mixed-data approaches. Thus, the data loss associated with sparser sampling is more harmful compared to the effect of potential accumulation of forecast errors associated with the iterated and mixed-data approaches.

We need to emphasize here that the distinction between direct and iterative forecasts in our application does not involve quite the same trade-off as in Ghysels *et al.* (2009). In particular, in our direct approach for the models based on high-frequency data, we use cumulated realized covariance matrices; i.e., a direct model for the 10-day covariance is based on the time series of 10-day cumulated realized covariance matrices. Thus, while there is data resolution loss associated with aggregating the matrices, our direct approach still indirectly uses high-frequency data.

---

[4] Some clarification is due here: in the direct approach we start at $s + (T' \div s)$ since we want the in-sample to end on exactly the same day as in the iterative approach. In our dataset we have a total of 2156 daily observations and our in-sample period contains the first 1508, leaving 648 days out-of-sample. This means that if we want to aggregate data at a 5-day frequency, we need the last 5-day period to end at day 1508. This is achieved by having the first 5-day period to start at $t = 1508 \div 5 + 1 = 4$ and end at $t = 1508 \div 5 + 5 = 8$; i.e., we drop the first $1508 \div 5 = 3$ observations. The out-of-sample period contains $\lfloor \frac{648}{5} \rfloor = 129$ non-overlapping 5-day periods; i.e., we drop the last $648 - \lfloor \frac{648}{5} \rfloor \times 5 = 3$ observations.

3.1. Forecasting with the VARFIMA Model

In this section, we discuss forecasting with the VARFIMA model for the Cholesky factors X_t. The procedure can be implemented in exactly the same way if we wish to use the log volatilities a_t instead. For ease of exposition, and since the exogenous regressors are by assumption predetermined, we neglect the term BZ_t in equation (2).[5] For each $j = 1, \ldots, m$, the fractionally differenced series $(1 - L)^{d_j} X_{j,t}$ is given by

$$(1 - L)^{d_j} X_{j,t} = \sum_{h=0}^{\infty} \lambda_{j,h} X_{j,t-h} = X_{j,t} + \sum_{h=1}^{\infty} \lambda_{j,h} X_{j,t-h} \tag{13}$$

where $\lambda_{j,0} = 1$ and $\lambda_{j,h} = \prod_{0 < r \le h} \dfrac{r - 1 - d_j}{r}$, $h = 1, 2, \ldots$. We can then rewrite equation (2) as

$$\Phi(L)\Lambda(L)X_t = \Theta(L)\varepsilon_t \tag{14}$$

where $\Lambda(L) = I_m + \sum_{h=1}^{\infty} \lambda_h L^h$ and $\lambda_h = \text{diag}\{\lambda_{1,h}, \ldots, \lambda_{m,h}\}$. From equation (14), the VMA(∞) and VAR(∞) representations are given by

$$X_t = \Phi(L)^{-1}\Lambda(L)^{-1}\Theta(L)\varepsilon_t = \sum_{i=0}^{\infty} \Psi_i \varepsilon_{t-i}$$

$$\varepsilon_t = \Phi(L)\Lambda(L)\Theta(L)^{-1}X_t = \sum_{i=0}^{\infty} \Xi_i X_{t-i}$$

where $\Psi_0 = \Xi_0 = I_m$, and the optimal predictor of X_t is

$$\hat{X}_{t+s} \equiv E_t[X_{t+s}] = \sum_{i=s}^{\infty} \Xi_i X_{t+s-i} = \sum_{i=0}^{\infty} \Xi_{s+i} X_{t-i}$$

To obtain multi-step-ahead forecasts from the VARFIMA model, we truncate the VAR(∞) representation at t lags, and use standard results for AR models (see Hamilton, 1994, pp. 80–81). The forecast is unbiased (i.e., the forecast errors have zero mean) and since ε_t is assumed to be normally distributed, the forecast errors are also normally distributed as

$$u_{t,t+s} \equiv X_{t+s} - \hat{X}_{t+s} \sim N(0, \Sigma_s), \quad \text{where}$$

$$\Sigma_s = E[(X_{t+s} - \hat{X}_{t+s})(X_{t+s} - \hat{X}_{t+s})'] = E[u_{t,t+s}u'_{t,t+s}] = \sum_{i=0}^{s-1} \Psi_i \Sigma \Psi'_i$$

Having obtained $\hat{X}_{t+s}$, we construct the forecast $\hat{Y}_{t+s}$ by applying the transformation in equation (3).[6]

---

[5] One can think of the procedure describing how to forecast $\tilde{X}_t = X_t - BZ_t$, from which the forecast of X_{t+s} is formed as $\hat{\tilde{X}}_{t+s} + \hat{B}_s Z_t$, where $\hat{B}_s$ is a function of the estimated parameter matrix $\hat{B}$ and the horizon s. In our case, we simply need to add s times the estimated mean $\hat{c}$ to the forecast.

[6] We have argued that Y_t is almost an error-free measure of the latent H_t, which motivates the whole literature on high-frequency volatility estimation, allowing us to equate $\hat{Y}_{t+s}$ to $\hat{H}_{t+s}$. Of course, $Y_{t+s} \neq H_{t+s}$ even *ex post*, implying that the quality of the forecast does not fully depend on the dynamic specification for Y_t, but also on the quality of the realized covariance estimator. It is beyond the scope of this paper to address the latter issue; the development of multivariate volatility measures using high-frequency data is currently a very active area of research. In this paper, we use an estimator which has been shown to be reliable and much more precise than any estimator based on daily data (see, for example, Barndorff-Nielsen and Shephard, 2004).

Let us now look at the forecast errors for the individual elements of Y_{t+s}, $e_{ij,t+s} = \hat{Y}_{ij,t+s} - Y_{ij,t+s}$. Since $\hat{Y}_{ij,t+s}$ is a quadratic transformation of $\hat{X}_{t+s}$, the mean of $e_{ij,t+s}$ is generally no longer zero, and depends on the variance Σ_s of the forecast error $u_{t,t+s}$. Thus, each component of the covariance matrix forecast, $\hat{Y}_{ij,t+s}$, is biased by $E_t[e_{ij,t+s}] \equiv \sigma^*_{s,ij} \neq 0$, where $\sigma^*_{s,ij}$ can be obtained from the elements of the matrix Σ_s using the following formula (cf. equation (3)):

$$\sigma^*_{s,ij} = \sum_{l=1+\frac{i(i-1)}{2}}^{\frac{i(i+1)}{2}} \sigma_{s(l,l+\frac{j(j-1)}{2}-\frac{i(i-1)}{2})} \tag{15}$$

where $j \geq i$, $i = 1, \ldots, n$ and $\sigma_{s(u,v)}$ is the (u, v)-element of Σ_s. While this expression is only valid if ε_t is homoscedastic, the formula for the heteroscedastic case is readily available since the expression for Σ_s is known in that case to be $\Sigma_{t,s} = \sum_{i=0}^{s-1} \Psi_i \Sigma_{t-i} \Psi'_i$.[7] The forecasts based on the matrix log decomposition will also be biased, and a bias correction strategy is proposed in Bauer and Vorkink (2007). Appendix A.3 contains a discussion of bias correction in the light of our empirical study below.

3.2. Forecasting with the HAR Model

Direct forecasting with the HAR model at the desired frequency is easily carried out owing to the hierarchical structure of the model. A 1-day-ahead forecast can be obtained from equation (5), while longer-horizon forecasts are constructed from the corresponding level of the hierarchy; e.g., a 10-day-ahead forecast can be generated by the model[8]

$$X_{t+1,bw} = c^{(bw)} + \beta^{(bw)} X_t^{(bw)} + \beta^{(m)} X_t^{(m)} + \omega_{t+1,bw} \tag{16}$$

We obtain iterative s-step-ahead forecasts for the HAR model by iterating equation (5), using at each step $h = 2, \ldots, s$ the previously obtained forecasts until $h - 1$ in the computation of $X_{t+h-1}^{(w)}$, $X_{t+h-1}^{(bw)}$, and $X_{t+h-1}^{(m)}$.

3.3. Forecasting with the WAR Model

Direct h-day-ahead forecasts with the WAR model are obtained by applying equation (8) at the desired aggregation level. Iterative forecasts are available by cumulating the h-step-ahead forecasts of Y_t for $h = 1, \ldots, s$, given by

$$\hat{Y}_{t+h} = \hat{M} \hat{Y}_{t+h-1} \hat{M}' = \ldots = \underbrace{\hat{M} \cdot \ldots \cdot \hat{M}}_{h} Y_t \underbrace{\hat{M}' \cdot \ldots \cdot \hat{M}'}_{h} \tag{17}$$

Forecasts from the WAR-HAR model are computed in the same way as HAR forecasts.

---

[7] In the last expression, we need to index the $\Sigma_{t,s}$ by t, since with time varying volatility of ε_t the bias correction will be time varying as well.

[8] It should be noted that while theoretically the HAR model has a hierarchical structure, empirically volatility at higher frequencies can influence the volatility at longer horizons. In this study, we keep the hierarchy of the model and relate future volatility to past values of volatility at the same and lower frequencies.

3.4. Forecasting with the (FI)DCC Model

Forecasting with the DCC and FIDCC models involves first projecting the conditional variances based on the univariate GARCH models. The s-step-ahead iterated forecast of $h_{ii,t}$ is based on the following equations for $h = 1, \ldots, s$:

$$\hat{h}_{ii,t+h} = \hat{w}_i + (\hat{\alpha}_i + \hat{\beta}_i)\hat{h}_{ii,t+h-1}, \quad \text{and} \tag{18}$$

$$\hat{h}_{ii,t+h} = \frac{\hat{\omega}_i}{1 - \hat{\beta}_i} + \sum_{j=1}^{\infty} \hat{\delta}_{i,j}\hat{h}_{ii,t+h-j} \tag{19}$$

for GARCH(1,1) and FIGARCH(1, d, 0), truncated at $t + h - 1$ lags, respectively. The direct s-step forecasts are given by the one-step-ahead forecasts available in equations (18) and (19), with data aggregated at the desired frequency. The iterated forecast of Q_t is obtained from cumulating the h-step-ahead forecasts $\hat{Q}_{t+h}$, for $h = 1, \ldots, s$:

$$\hat{Q}_{t+1} = (1 - \hat{\theta}_1 - \theta_2)\hat{\bar{Q}} + \hat{\theta}_1\hat{u}_t\hat{u}_t' + \hat{\theta}_2 Q_t, \quad h = 1 \tag{20}$$

$$\hat{Q}_{t+h} = (1 - \hat{\theta}_1 - \hat{\theta}_2)\hat{\bar{Q}} + (\hat{\theta}_1 + \hat{\theta}_2)\hat{Q}_{t+h-1}, \quad h = 2, \ldots, s$$

The direct forecast is given by the one-step-ahead forecast available in equation (20) with data aggregated at the desired frequency. Given the forecast of the volatility matrix $\hat{D}_{t+s}$ and the matrix $\hat{Q}_{t+s}$, we compute $\hat{H}_{t+s}$ using equations (10) and (9).

4. EMPIRICAL APPLICATION

In this section, we present estimation and forecasting results for the eight approaches presented in Section 2: VARFIMA dynamics of the Cholesky factors X_t (VARFIMA-Cholesky) and log-volatilities a_t (VARFIMA-Log), HAR dynamics of X_t (HAR-Cholesky) and a_t (HAR-Log), a diagonal WAR specification, a diagonal WAR-HAR specification, DCC and FIDCC. The models vary with respect to the frequency of the data they use ((FI)DCC vs. all other models), the way the covariance matrix is handled (Cholesky decomposition, log transformation, WAR), and the type of dynamic structure (long memory or not). Having variability in these essential aspects of the modelling approaches allows us to clearly identify the potential sources of forecast accuracy.[9]

To measure the statistical precision of the forecasts we employ the RMSE criterion, which satisfies the conditions in Patton (2009). In order to have some idea of the economic benefits associated with the accuracy of the volatility forecasts, we also evaluate the mean–variance profile of portfolio optimization strategies. To test the models' performance against each other, we rely on the Model Confidence Set (MCS) methodology, which selects the set of models containing the best one with a given level of confidence. In terms of implementation, we use the Ox package MulCom v1.00 provided by the authors.

---

[9] Given that Bauer and Vorkink (2007) have shown that macroeconomic variables have predictive power for the volatility of market portfolios, we acknowledge that it would be interesting to include such variables in our study, but refrain from doing so, as it is not the main focus of the paper and would constitute a new dimension of the study.

4.1. Data

The data consist of tick-by-tick bid and ask quotes from the NYSE Trade and Quotations (TAQ) database sampled from $9:30$ until $16:00$ for the period $01.01.2000–30.07.2008$ ($T = 2156$ trading days).[10] For the current analysis, we select the following six highly liquid stocks: American Express Inc. (AXP), Citigroup (C), General Electric (GE), Home Depot Inc. (HD), International Business Machines (IBM) and JPMorgan Chase & Co. (JPM). We employ the previous-tick interpolation method, described in Dacorogna *et al*. (2001) and obtain 78 intraday returns by sampling every 5 minutes. Table A.1 in Appendix A.2 reports summary statistics for the 5-minute and daily return series.

For each $t = 1, \ldots, 2156$, a series of daily realized covariance matrices can be constructed as

$$Y_t = \sum\nolimits_{j=1}^{M} r_{j,t} r'_{j,t} \tag{21}$$

where $M = 78$. The 5-minute returns, $r_{j,t}$, are computed as

$$r_{j,t} = p_{j\Delta,t} - p_{(j-1)\Delta,t}, \quad j = 1, \ldots, M$$

where $\Delta = 1/M$ and $p_{j\Delta,t}$ is the log midquote price at time $j\Delta$ in day t. The realized covariance matrices are symmetric by construction and, for $n < M$, positive definite almost surely. Since by sampling sparsely we disregard a lot of data, we refine the estimator by subsampling. With $\Delta = 300$ seconds, we construct 30 Δ-spaced subgrids starting at seconds 1, 11, 21, $\ldots$, 291, compute the realized covariance matrix on each subgrid and take the average. The resulting subsampled realized covariance is much more robust to the so-called market microstructure noise than the simple 5-minute based one. Given the high liquidity of all the stocks and the very recent sample, we are confident that the effect of non-synchronicity is rather mild at the chosen frequency. In order to avoid the noise induced by measuring the overnight volatility as the squared overnight return, we assume that our investor opens a position at the beginning of the trading day and closes at the end of the trading day, which implies that all multivariate volatility models are applied to open-to-close data and measure the volatility over the trading session. Table A.2 in Appendix A.2 reports summary statistics of realized variances and covariances of the six stocks considered in the study. As already documented in Andersen *et al*. (2001), both realized variance and covariance distributions are extremely right skewed and leptokurtic.

Analyzing the series of Cholesky factors, we find that they inherit the long-memory property of realized (co)variances, discussed in Andersen and Bollerslev (1997) and Andersen *et al*. (2001). The sample autocorrelations of the elements of X_t decay at a slow rate, similar to the autocorrelations of the realized (co)variance series. To get an initial idea of the degree of fractional integration, we run OLS regressions of log-autocorrelations on log-lags (see Beran, 1998, pp. 89–92) and obtain an estimate of 0.40 on average across the m series.

4.2. Estimation Results

Before turning to the forecasting evaluation, we briefly discuss the estimation results of the VARFIMA-Cholesky model on the full sample of data, reported in Table A.3 in Appendix A.2,

---

[10] We are grateful to Asger Lunde for providing us with the data.

which also contains the results for the remaining models.[11] The estimation is carried out on centred data by subtracting the sample mean of the series under consideration. As a consequence, in order to assure a correct inference of the second-step model parameters, their standard errors should account for the pre-estimation of the mean, and therefore we report bootstrapped standard errors based on the method developed in Politis and Romano (1994) and Politis *et al.* (1999) for dependent and cross-correlated time series. For the specification with unrestricted integration parameters (Model 1), all d parameters are significant at the 5% level and of similiar magnitude, indicating that the series are possibly integrated of the same order. To verify this, we test the null hypothesis $H_0 : d_1 = \ldots = d_m$ by means of a likelihood ratio test, obtaining a p-value of 0.054, based on the asymptotic χ^2 distribution with 20 degrees of freedom. In the restricted model (Model 2) the degree of fractional integration is estimated at 0.44, a value commonly obtained with such series. The autocorrelogram of the model residuals, as well as the multivariate Ljung–Box portmanteau test, suggest that in-sample the model can be fine-tuned by allowing for a richer structure. Given the strong theoretical arguments in Hansen (2009), as well as the empirically documented poor out-of-sample performance of overparameterized models, we opt against this.

To complete the in-sample investigation of our model, we report the marginal effects among the variance and covariance series in Table A.7.[12] Most values in the table are as expected positive, which can be related to volatility spillover effects across stocks. It is interesting to note that increasing volatility leads to increased correlation and vice versa, which has adverse consequences for portfolio selection, limiting the benefits of diversification in volatile markets. Evidence of positive linkage between volatility and correlation across stocks has been documented by Andersen *et al.* (2001), while Ang and Bekaert (2002) and Ang and Chen (2002) find the same phenomenon at the market level.

4.3. Forecasting Results and Evaluation

We split the whole sample of data into an in-sample from 01.01.2000 to 31.12.2005 ($T' = 1508$ days) and an out-of-sample from 01.01.2006 to 30.07.2008 (648 days, or 129 (64) complete 5 (10)-day periods). The forecasts are carried out in a recursive manner; i.e., at each step the models are re-estimated with all of the available data. As described in Section 2, we consider 1-day, 5-day and 10-day horizons, using direct and iterated methods. We do not bias correct our forecasts, due to theoretical caveats and since we do not find empirical justification to do so (detailed argumentation can be found in Appendix A.3).

Statistical Evaluation
In this section, we use Y_{t+s} as a proxy for the unobservable H_{t+s} and compare the models' out-of-sample forecast RMSE, based on the Frobenius norm[13] of the forecast error

$$e_{t,t+s} \equiv Y_{t:t+s} - \hat{H}_{t:t+s} \tag{22}$$

where $\hat{H}_{t:t+s}$ is either a direct or an iterated forecast, $Y_{t:t+s} = \sum_{i=1}^{s} Y_{t+i}$, and $s = 1, 5, 10$. Referring to Table II, the VARFIMA-Cholesky specification has the smallest RMSE among all models,

---

[11] Tables A.4, A.5 and A.6 provide the estimation results of the HAR-Cholesky, HAR-Log, diagonal WAR, diagonal WAR-HAR, DCC and FIDCC models, respectively.
[12] For a derivation of the marginal effects in the VARFIMA-Cholesky model, see Appendix A.1.
[13] The Frobenius norm of a real $m \times n$ matrix A is defined as $||A|| = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2$.

Table II. RMSE based on the Frobenius norm of the forecasting error (equation (22)). For 5- and 10-step-ahead forecasts, the statistic is standardized by dividing by the number of steps to make the results comparable

Model	1 day	Iterated		Direct	
		5 days	10 days	5 days	10 days
VARFIMA-Cholesky	3.897[a]	3.388[a]	3.515[a]	3.540[a]	3.716[a]
VARFIMA-Log	3.937[a]	3.498	3.610[a]	3.525[a]	3.700[a]
HAR-Cholesky	3.940	3.459[a]	3.628	3.652[a]	3.919
HAR-Log	3.943	3.492	3.627	3.585[a]	3.871
Diagonal WAR	4.990	6.198	7.055	4.673	4.608
Diagonal WAR-HAR	4.598	4.995	5.752	4.489	4.659
DCC	5.195	4.727	4.851	5.252	4.945
FIDCC	5.613	4.613	4.767	5.435	5.224

[a] Model belongs to the 5% MCS of Hansen *et al.* (2009).

and is therefore necessarily included in the MCS, regardless of the horizon. The VARFIMA-Log follows closely and is in the 5% MCS four out of five times, indicating that the choice of transformation of the realized covariance matrices plays hardly any role. This is a positive result, suggesting that what is important for forecasting is the specification of the dynamics, rather than the particular implementation.[14] Given this evidence, we drop the VARFIMA-Log and HAR-Log specifications in our further analysis. Interestingly, the error, standardized by the forecasting horizon, is smaller for the 5- and 10-day horizon compared to the 1-day horizon for the majority of the models. This implies that multi-period volatility forecasting is in some sense more precise than short-term prediction, supporting the results in Ghysels *et al.* (2009) for models using daily data. Comparatively, the HAR specification shows a very good forecasting ability, but at the longer, 10-day horizon, is significantly outperformed by the VARFIMA model. In general, the iterated forecasts are better than the direct ones, except for the WAR models. This lends support to the fact that using the data at a higher resolution is beneficial and does not lead to over-accumulation of forecasting error, indirectly suggesting that in most cases we have well-specified dynamics.

Economic Evaluation

In order to assess the economic value of the volatility forecasts, we construct portfolios which maximize the utility of a risk-averse investor. If the utility function is second-degree polynomial or logarithmic and/or the return distribution is completely characterized by its first two moments (as, for example, the normal distribution), the portfolio optimization reduces to finding the asset weights which minimize the portfolio volatility, fixing a given expected return, or maximize the portfolio return, targeting a certain volatility (Markowitz, 1952).

We assume that the investor minimizes portfolio volatility subject to an annualized expected return μ_p. In this context, the optimal portfolio is given by the solution to the following quadratic

---

[14] As a further robustness check of this result, we estimate the model on the Cholesky factors of the covariance matrix with inverted ordering of the stocks. The results do not change qualitatively and actually remain even quantitatively almost the same as in the original implementation: the RMSEs are 3.898, 3.390, 3.510, 3.537 and 3.718, corresponding to the columns of the table.

problem:

$$\min_{w_{t+s|t}} \quad w'_{t+s|t}\hat{H}_{t:t+s}w_{t+s|t} \quad \text{s.t.} \quad w'_{t+s|t}E_t[r_{t:t+s}] = \frac{s\mu_p}{250} \quad \text{and} \quad w'_{t+s|t}\iota = 1$$

where $w_{t+s|t}$ is the $n \times 1$ vector of portfolio weights chosen at t and held until $t + s$, ι is an $n \times 1$ vector of ones, and $\frac{s\mu_p}{250}$ is the target return scaled to the investment horizon. To assess the ability of the models to correctly predict the conditional covariance matrix, we compare the *ex post* realizations of the conditional portfolio mean and standard deviation; that is, given a set of weights at time t, $w_{t+s|t}$, we compute $r^p_{t:t+s} = w'_{t+s|t}r_{t:t+s}$ and $\sigma^p_{t:t+s} = \sqrt{w'_{t+s|t}Y_{t:t+s}w_{t+s|t}}$, for $t = T', T' + s, \ldots, T' + s\left(\lfloor\frac{T-T'}{s}\rfloor - 1\right)$. The optimization problem described above can be solved for various levels of the target portfolio return μ_p, thus obtaining an efficiency frontier, characterizing the best mean–variance trade-off achievable by using a particular forecasting model. To have some sort of ideal case scenario, we also construct the efficiency frontier using the 'oracle' forecast $\hat{H}_{t:t+s} = Y_{t:t+s}$.

The results are illustrated in Figure 1 for $s = 1$.[15] Obviously, the 'oracle' forecast leads to a by far superior mean–variance trade-off, which is a testimony to the precision of Y_t as an
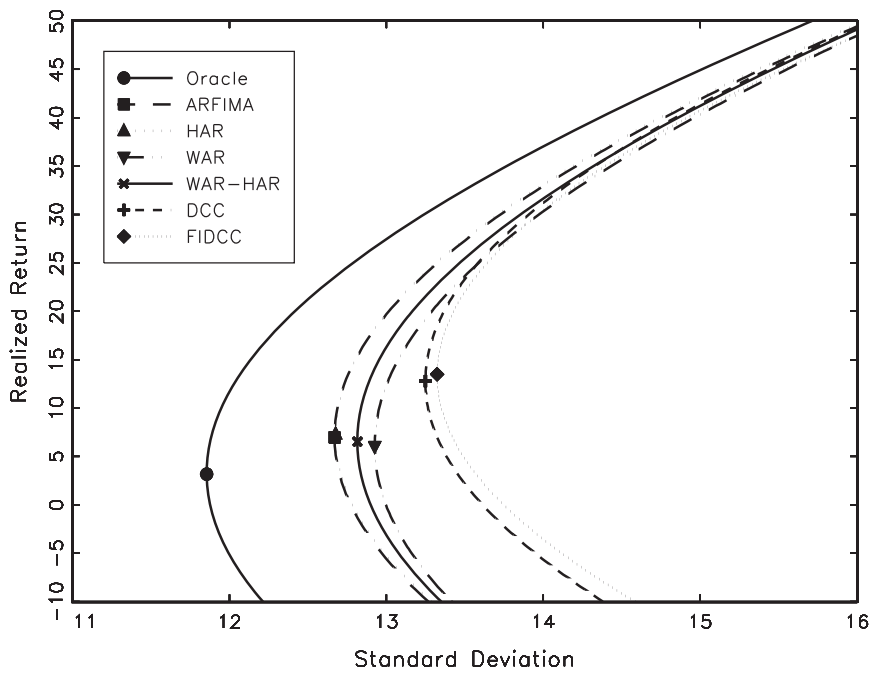

Figure 1. Mean–variance plots for the *ex post* realized conditional mean (on the *y*-axis in %, annualized) against realized conditional standard deviation (on the *x*-axis in %, annualized). The global minimum variance portfolio is symbolized as follows: circle (oracle), square (VARFIMA), triangle (HAR), a '×' (diagonal WAR-HAR), inverted triangle (diagonal WAR), cross (DCC), and diamond (FIDCC). All plots are averages across the 648 out-of-sample periods (days)

---

[15] Graphs for the iterated and direct 5- and 10-day-ahead forecasts are available upon request from the authors.

estimator of H_t. From the six models at our disposal, the VARFIMA and HAR-based plots almost overlap and offer seemingly significant improvements over the remaining four models. By fixing the level of expected return μ_p, the difference of these curves can be statistically tested on a point-by-point basis, i.e., by comparing the series $\sigma^p_{t:t+s}$ for a particular value of μ_p. A natural point to be considered is the level of return corresponding to the global minimum variance portfolio (GMVP), accentuated in Figure 1 by a particular symbol for each model.

While this return level is unknown a priori, *ex post* there necessarily exists a value of μ_p that has led to the smallest $\sigma^p_{t:t+s}$. In Table III, we report the average of the realized conditional standard deviation of the GMVP and use again the MCS methodology to select the set of models which contains the one with the smallest standard deviation at the 5% confidence level. Pertaining to the model ranking, the results are in line with the comparison based on the forecast RMSE. Contrary to the purely statistically motivated loss function, however, the direct models seem to perform a bit better according to the economic evaluation criteria at least for the high-frequency models. Furthermore, the standardized loss, defined as the annualized portfolio volatility, increases with the forecast horizon. These differences suggest that the two loss functions are indeed different in nature, and nevertheless tend to suggest the same models as the best performing ones. We interpret this as a strong evidence in favour of the methodology developed in this paper.

Recently, Liu (2009) analyzed the benefits from high-frequency data in the context of minimizing the variance of the error of a portfolio of the 30 DJIA stocks, tracking the S&P 500 index. His findings suggest that if the portfolio is rebalanced daily, models using high-frequency data outperform models with daily data, while if the rebalancing is done on a monthly basis, models using daily data can potentially perform equally well.

He does not consider long-memory models, however, which might explain the similar performance of daily and high-frequency models at the longer horizons. Our results support his findings in the sense that high-frequency data are definitely beneficial also at horizons of a week or two, with the difference that the performance gap remains fairly constant across forecasting horizons. All in all, we are optimistic that the parsimony of the proposed model will make it viable also in problems of larger cross-sectional dimension, while its ability to characterize processes with strong persistence will also make it suitable for longer-horizon forecasting.

Table III. Annualized realized conditional standard deviations of the *ex post* global minimum variance portfolio (%). All numbers are averages across the out-of-sample periods

Model	1 day	Iterated		Direct	
		5 days	10 days	5 days	10 days
VARFIMA-Cholesky	12.669[a]	12.931[a]	13.016[a]	12.924[a]	13.001[a]
HAR-Cholesky	12.676	12.934[a]	13.023[a]	12.934	13.018
Diagonal WAR	12.925	13.462	13.786	13.219	13.143
Diagonal WAR-HAR	12.814	13.186	13.398	13.178	13.150
DCC	13.248	13.501	13.588	13.805	15.202
FIDCC	13.323	13.552	13.530	14.134	15.279

[a] Model belongs to the 5% MCS of Hansen *et al.* (2009).

5. CONCLUSION

In this paper, we develop an approach for the dynamics of realized covariance matrices. The model explicitly accounts for the empirically observed long memory of financial volatility and can accommodate exogenous predictive variables. The matrices are decomposed into Cholesky series, modelled as a a multivariate vector fractionally integrated ARMA (VARFIMA) process without imposing restrictions on the admissible parameter space. By subsequent 'squaring' of the forecast, we automatically obtain positive definite covariance forecasts.

In a forecasting application, we show that the model performs significantly better than other currently available approaches at various prediction horizons, not only in terms of minimizing the RMSE of the forecast, but also in terms of improving the performance of mean–variance efficient portfolios.

We hope that our study opens up a number of further interesting research directions in the context of volatility forecasting. An issue we have not explored in this study is the potential benefits of combining high- and low-frequency data in the spirit of the mixed data-sampling (MIDAS) approach, which Ghysels *et al*. (2009) show to be very promising for longer-horizon forecasting. In their study, however, the data are only available at the daily and lower frequency, and so one would conjecture that higher-frequency data should bring further improvements. An obvious further extension of our work is to examine whether the methods discussed here remain feasible in higher dimensions. Most approaches we advocate have a fixed number of parameters, independent of the number of assets. It is an open empirical question, however, whether such tightly parameterized models will still yield good forecasts in an environment with many more, possibly heterogeneous assets. Last, but not least, we have fully neglected the impact of macroeconomic variables on future volatility, which has been addressed in detail by Bauer and Vorkink (2007), and would arguably be an interesting issue to analyse in our modelling framework.

REFERENCES

An S, Bloomfield P. 1993. *Cox and Reid's modification in regression models with correlated error*. Discussion paper, Department of Statistics, North Carolina State University.

Andersen TG, Bollerslev T. 1997. Heterogeneous information arrivals and return volatility dynamics: uncovering the long-run in high frequency returns. *Journal of Finance* **52**: 975–1005.

Andersen TG, Bollerslev T, Diebold FX, Ebens H. 2001. The distribution of realized stock return volatility. *Journal of Financial Economics* **61**: 43–76.

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

Ang A, Bekaert G. 2002. International asset allocation with regime shifts. *Review of Financial Studies* **15**(4): 1137–1187.

Ang A, Chen J. 2002. Asymmetric correlations of equity portfolios. *Journal of Financial Economics* **63**(3): 443–494.

Asai M, McAleer M, Yu J. 2006. Multivariate stochastic volatility: a review. *Econometric Reviews* **25**: 145–175.

Baillie R, Bollerslev T, Mikkelsen H. 1996. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **74**: 3–30.

Barndorff-Nielsen OE, Shephard N. 2004. Econometric analysis of realised covariation: high frequency based covariance, regression and correlation in financial economics. *Econometrica* **72**: 885–925.

Bauer GH, Vorkink K. 2007. *Multivariate realized stock market volatility*. Working Paper 2007-20, Bank of Canada.

Bauwens L, Laurent S, Rombouts J. 2006. Multivariate GARCH models: a survey. *Journal of Applied Econometrics* **21**: 79–109.

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.

Beran J. 1998. *Statistics for Long-Memory Processes*. CRC Press: Boca Raton, FL.

Bonato M. 2009. *Estimating the degrees of freedom of the realized volatility Wishart autoregressive model*. Working paper.

Bonato M, Caporin M, Ranaldo A. 2009. *Forecasting realized (co)variances with a block Wishart autoregressive model*. SSRN. Available: http://ssrn.com/abstract=1282254 [2 December 2009].

Chiriac R. 2007. *Nonstationary Wishart autoregressive model*. Working paper, CoFE, University of Konstanz.

Corsi F. 2009. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics* **7**: 174–196.

Corsi F, Audrino F. 2007. *Realized correlation tick-by-tick*. Working paper, University of Lugano.

Dacorogna MM, Gençay R, Müller UA, Olsen RB, Pictet OV. 2001. *An Introduction to High-Frequency Finance*. Academic Press: San Diego, CA.

Doornik JA, Ooms M. 2004. Inference and forecasting for AEFIMA models with an application to US and UK inflation. *Studies in Nonlinear Dynamics and Econometrics* **8**: Article 14.

Engle R. 2002. Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroscedasticity models. *Journal of Business and Economic Statistics* **20**: 339–350.

Engle R, Bollerslev T. 1986. Modeling the persistence of conditional variances. *Econometric Reviews* **5**: 1–50.

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

Gallant R, Tauchen G. 2001. *Efficient method of moments*. Working paper, Duke University.

Ghysels E, Rubia A, Valkanov R. 2009. *Multi-period forecasts of volatility: direct, iterated and mixed-data approaches*. EFA 2009 Bergen Meetings Paper.

Giraitis L, Robinson PM, Surgailis D. 2000. A model for long memory conditional heteroskedasticity. *Annals of Applied Probability* **10**: 1002–1024.

Glosten LR, Jagannathan R, Runkle DE. 1993. On the relation between the expected value and the volatility of the nominal excess return of stocks. *Journal of Finance* **48**: 1179–1801.

Golub GH, van Loan CF. 1996. *Matrix Computations*. Johns Hopkins University Press: Baltimore, MD.

Gourieroux C, Monfort A. 1995. *Statistics and Econometric Models*, Vol. 1. Cambridge University Press: Cambridge, UK.

Gourieroux C, Jasiak J, Sufana R. 2009. The Wishart autoregressive process of multivariate stochastic volatility. *Journal of Econometrics* **150**: 167–181.

Hamilton JD. 1994. *Time Series Analysis*. Princeton University Press: Princeton, NJ.

Hansen PR. 2009. *In-sample fit and out-of-sample fit: their joint distribution and its implications for model selection*. Working paper, Stanford University.

Hansen PR, Lunde A. 2006. Consistent ranking of volatility models. *Journal of Econometrics* **131**: 97–121.

Hansen PR, Lunde A, Nason JM. 2009. *The model confidence set*. Working paper, Aarhus University.

Koopman SJ, Jungbacker B, Hol E. 2005. Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements. *Journal of Empirical Finance* **12**: 445–475.

Lamoureaux C, Lastrapes W. 1990. Heteroskedasticity in stock return data: volume versus GARCH effects. *Journal of Finance* **45**(1): 221–229.

Liu Q. 2009. On portfolio optimization: how and when do we benefit from high-frequency data? *Journal of Applied Econometrics* **24**: 560–582.

Lütkepohl H. 2005. *New Introduction to Multiple Time Series Analysis*. Springer: Berlin.

Markowitz H. 1952. Portfolio selection. *Journal of Finance* **7**: 77–91.

Oomen R. 2001. *Using high frequency stock market index data to calculate, model and forecast realized return variance*. Economics Discussion Paper No. 2001/6, European University.

Patton A. 2009. Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics* (forthcoming).

Politis D, Romano J. 1994. Large sample confidence regions based on subsamples under minimal assumptions. *Annals of Statistics* **22**: 2031–2050.

Politis D, Romano J, Wolf M. 1999. *Subsampling*. Springer: New York.

Pourahmadi M. 1999. Joint mean–covariance models with applications to longitudinal data: unconstrained parameterization. *Biometrika* **86**(3): 677–690.

Robinson PM. 1991. Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *Journal of Econometrics* **47**: 67–84.

Schwert GW. 1989. Why does stock market volatility change over time? *Journal of Finance* **44**: 1115–1153.

Sowell F. 1989. *Maximum likelihood estimation of fractionally integrated time series models*. Carnegie Mellon University.

Sowell F. 1992. Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics* **53**: 165–188.

Tsay RS. 2002. *Analysis of Financial Time Series*. Wiley-Interscience: New York.

Voev V. 2007. Dynamic modelling of large-dimensional covariance matrices. In *High Frequency Financial Econometrics: Recent Developments* (1st edn), Bauwens L, Pohlmeier W, Veredas D (eds). Springer: Berlin; 293–312.

Zumbach G. 2007. *The RiskMetrics 2006 methodology*. Technical document, RiskMetrics.

APPENDIX

A.1. Derivation of the Marginal Effects

We denote the estimated parameter vector by $\hat{\vartheta}$ and the history of the process X_t up to time t by X_t. The (i, j)-element of the predicted covariance matrix can be written as (see Equation 3):

$$\hat{Y}_{ij,t+s} = \sum_{l=1+\frac{i(i-1)}{2}}^{\frac{i(i+1)}{2}} E_t\left[X_{l,t+s}X_{l+\frac{j(j-1)}{2}-\frac{i(i-1)}{2},t+s}\right] \equiv G_{i,j,s}(X_t, \hat{\vartheta}), \quad i, j = 1, \ldots, n \quad \text{(A.1)}$$

where $G_{ij,s}(\cdot)$ is a scalar function of X_t and $\hat{\vartheta}$, corresponding to the (i, j)-element of the matrix $\hat{Y}_{t+s}$. For example, the impact of a shock in the covariance $Y_{ij,t}$ on the predicted variance $\hat{Y}_{ii,t+s}$ can be computed as follows:

$$\frac{\partial \hat{Y}_{ii,t+s}}{\partial Y_{ij,t}} = \frac{\partial G_{ii,s}}{\partial G_{ij,0}} = \sum_{r=1}^{m} \frac{\partial G_{ii,s}}{\partial X_{r,t}} \frac{\partial X_{r,t}}{\partial G_{ij,0}} = F_{ii,ij}^{s,t}(X_t, \hat{\vartheta}) \quad \text{(A.2)}$$

where $G_{ij,0} \equiv Y_{ij,t} = \sum_{l=1+\frac{i(i-1)}{2}}^{\frac{i(i+1)}{2}} X_{l,t}X_{l+\frac{j(j-1)}{2}-\frac{i(i-1)}{2},t}$, $j \geq i$ and $F_{ii,ij}^{s,t}(\cdot)$ is a scalar function. In a similar way, one can derive the the impact of the variance $Y_{ii,t}$ on the predicted covariance $\hat{Y}_{ij,t+s}$:

$$\frac{\partial \hat{Y}_{ij,t+s}}{\partial Y_{ii,t}} = \frac{\partial G_{ij,s}}{\partial G_{ii,0}} = \sum_{r=1}^{m} \frac{\partial G_{ij,s}}{\partial X_{r,t}} \frac{\partial X_{r,t}}{\partial G_{ii,0}} = F_{ij,ii}^{s,t}(X_t, \hat{\vartheta}) \quad \text{(A.3)}$$

where $G_{ii,0} \equiv Y_{ii,t} = \sum_{l=1+\frac{i(i-1)}{2}}^{\frac{i(i+1)}{2}} X_{l,t}^2$. We derive the expressions for $G_{i,j,s}$, $G_{i,i,s}$, $F_{ii,ij}^{s,t}$ and $F_{ij,ij}^{s,t}$, for the VARFIMA model with unrestricted Θ and D matrices. An element l, $l = 1, \ldots, m$ of X_t is given by

$$(1 - \phi L)(1 - L)^{d_l}[X_{l,t} - c_l] = \varepsilon_{l,t} - \theta_{ll}\varepsilon_{l,t-1} - \sum_{s=1,s\neq l}^{m} \theta_{ls}\varepsilon_{s,t-1}, \quad \varepsilon_{l,t} \sim N(0, \sigma_{ll}) \quad \text{(A.4)}$$

where σ_{ll} is the (l, l)-element of Σ and θ_{ls} is the (l, s)-element of Θ. Given the representation in Equation (13), we can write Equation (A.4) as follows:

$$(1 - \phi L)\left[X_{l,t} - c_l + \sum_{h=1}^{\infty} \lambda_{l,h}(X_{l,t-h} - c_l)\right] = \varepsilon_{l,t} - \theta_{ll}\varepsilon_{l,t-1} - \sum_{s=1,s\neq l}^{m} \theta_{ls}\varepsilon_{s,t-1},$$

where $\lambda_{l,h} = \prod_{0<r\leq h} \frac{r-1-d_l}{r}$, $h = 1, 2, \ldots$. Thus:

$$X_{l,t} = c_l + (\phi - \lambda_{l,1})(X_{l,t-1} - c_l) + \sum_{h=2}^{\infty} (\phi\lambda_{l,h-1} - \lambda_{l,h})(X_{l,t-h} - c_l) + \varepsilon_{l,t} - \theta_{ll}\varepsilon_{l,t-1}$$

$$- \sum_{s=1,s\neq l}^{m} \theta_{ls}\varepsilon_{s,t-1}$$

From the expression above, we can derive the conditional expectation of $Y_{ij,t+s}$ from equation (A.1) for any $s \geq 1$ and $i, j = 1, \ldots, n$ with $j \geq i$. We focus here on $s = 1$ (generalization

to $s > 1$ is straightforward):

$$
\hat{Y}_{ij,t+1} = \sum_{l=1+\frac{i(i-1)}{2}}^{\frac{i(i+1)}{2}} E_t\left[X_{l,t+1}X_{l+\frac{j(j-1)}{2}-\frac{i(i-1)}{2},t+1}\right] \equiv \sum_{l=1+\frac{i(i-1)}{2}}^{\frac{i(i+1)}{2}} E_t[X_{l,t+1}X_{p,t+1}]
$$

$$
= \sum_{l=1+\frac{i(i-1)}{2}}^{\frac{i(i+1)}{2}} E_t\left[\left(c_l + (\phi - \lambda_{l,1})(X_{l,t} - c_l) + \sum_{h=2}^{\infty}(\phi\lambda_{l,h-1} - \lambda_{l,h})(X_{l,t-h+1} - c_l)\right.\right.
$$

$$
+\varepsilon_{l,t+1} - \theta_{ll}\varepsilon_{l,t} - \sum_{s=1,s\neq l}^{m}\theta_{ls}\varepsilon_{s,t}\bigg)\left(c_p + (\phi - \lambda_{p,1})(X_{p,t} - c_p) + \sum_{h=2}^{\infty}(\phi\lambda_{p,h-1}\right.
$$

$$
\left.\left.-\lambda_{p,h})(X_{p,t-h+1} - c_p) + \varepsilon_{p,t+1} - \theta_{pp}\varepsilon_{p,t} - \sum_{s=1,s\neq p}^{m}\theta_{ps}\varepsilon_{s,t}\right)\right]
$$

$$
\equiv G_{i,j,1}(X_t, \vartheta) \tag{A.5}
$$

where the index p is defined (implicitly as a function of i, j) as $p = l + \dfrac{j(j-1)}{2} - \dfrac{i(i-1)}{2}$ and $\vartheta = (c', \phi, vech(\Theta)', d_1, \ldots, d_m, \ldots, vech(\Sigma))'$. From equation (A.5) we derive the expression for $G_{i,j,1}(X_t, \vartheta)$ to be

$$
G_{i,j,1}(X_t, \vartheta) = \sum_{l=1+\frac{i(i-1)}{2}}^{\frac{i(i+1)}{2}}\left\{\left[c_l + (\phi - \lambda_{l,1})(X_{l,t} - c_l) + \sum_{h=2}^{\infty}(\phi\lambda_{l,h-1} - \lambda_{l,h})(X_{l,t-h+1} - c_l)\right.\right.
$$

$$
+\theta_{ll}\varepsilon_{l,t} + \sum_{s=1,s\neq l}^{m}\theta_{ls}\varepsilon_{s,t}\bigg]\left[c_p + (\phi - \lambda_{p,1})(X_{p,t} - c_p) + \sum_{h=2}^{\infty}\right.
$$

$$
\left.\left.(\phi\lambda_{p,h-1} - \lambda_{p,h})(X_{p,t-h+1} - c_p) + \theta_{pp}\varepsilon_{p,t} + \sum_{s=1,s\neq p}^{m}\theta_{ps}\varepsilon_{s,t}\right] + \Sigma_{l,p}\right\} \tag{A.6}
$$

In a similar manner, we derive $G_{i,i,1}(X_t, \vartheta)$ to be

$$
G_{i,i,1}(X_t, \vartheta) = \sum_{l=1+\frac{i(i-1)}{2}}^{\frac{i(i+1)}{2}}\left\{\left[c_l + (\phi - \lambda_{l,1})(X_{l,t} - c_l) + \sum_{h=2}^{\infty}(\phi\lambda_{l,h-1} - \lambda_{l,h})(X_{l,t-h+1} - c_l)\right.\right.
$$

$$
\left.\left.-\theta_{ll}\varepsilon_{l,t} - \sum_{s=1,s\neq l}^{m}\theta_{ls}\varepsilon_{s,t}\right]^2 + \Sigma_{l,l}\right\} \tag{A.7}
$$

Given that $G_{i,j,0}(X_t, \vartheta) = \sum_{l=1+\frac{i(i-1)}{2}}^{\frac{i(i+1)}{2}} X_{l,t}X_{p,t}$, we can derive $F_{ii,ij}^{1,t}$ and $F_{ij,ii}^{1,t}$ from equations (A.2) and (A.3) for any (i, j) combination. For example, the marginal effect of the volatility $Y_{11,t}$ at time t on the conditional expectation of the covariance $Y_{12,t+1}$ at time $t+1$ is given by

$$
F_{12,11}^{1,t}(X_t, \vartheta) = \frac{\partial E_t[Y_{12,t+1}]}{\partial Y_{11,t}} = \frac{\partial G_{1,2,1}(X_t, \vartheta)}{\partial G_{1,1,0}(X_t, \vartheta)} = \frac{\partial G_{1,2,1}(X_t, \vartheta)}{\partial X_{1,t}^2}
$$

where $G_{1,2,1}(X_t, \hat{\vartheta})$ is obtained from Equation (A.6). Thus:

$$
F_{12,11}^{1,t}(X_t, \vartheta) = \frac{(\phi - \lambda_{1,1} - \theta_{11})A_2 - \theta_{21}A_1}{2X_{1,t}} \tag{A.8}
$$

where

$$A_1 = c_1 + (\phi - \lambda_{1,1})(X_{1,t} - c_1) - \theta_{11}\varepsilon_{1,t} + \sum_{h=2}^{\infty}(\phi\lambda_{1,h-1} - \lambda_{1,h})(X_{1,t-h+1} - c_1) - \sum_{s=2}^{m}\theta_{1s}$$

$$A_2 = c_2 + (\phi - \lambda_{2,1})(X_{2,t} - c_2) - \theta_{22}\varepsilon_{2,t} + \sum_{h=2}^{\infty}(\phi\lambda_{2,h-1} - \lambda_{2,h})(X_{2,t-h+1} - c_2) - \sum_{s=1,s\neq2}^{m}\theta_{2s}$$

Similarly we can derive the marginal effect of the covariance $Y_{12,t}$ on the conditional expectation of the volatility $Y_{11,t+1}$ denoted by $F_{11,12}^{1,t}$:

$$F_{11,12}^{1,t}(X_t, \vartheta) = \frac{\partial E_t[Y_{11,t+1}]}{\partial Y_{12,t}} = \frac{\partial G_{1,1,1}(X_t, \vartheta)}{\partial G_{1,2,0}(X_t, \vartheta)} = \frac{\partial G_{1,1,1}(X_t, \vartheta)}{\partial X_{1,t}X_{2,t}}$$

$$= \frac{\partial G_{1,1,1}(X_t, \vartheta)}{\partial X_{1,t}}\frac{\partial X_{1,t}}{\partial X_{1,t}X_{2,t}} + \frac{\partial G_{1,1,1}(X_t, \vartheta)}{\partial X_{2,t}}\frac{\partial X_{2,t}}{\partial X_{1,t}X_{2,t}}$$

where $G_{1,1,1}(X_t, \vartheta)$ is obtained from equation (A.7). Thus

$$F_{11,12}^{1,t}(X_t, \vartheta) = \frac{2(\phi - \lambda_{1,1} - \theta_{11})A_1}{X_{2,t}} - \frac{2\theta_{12}A_1}{X_{1,t}}$$

where A_1 is defined above.

A.2. Supplementary Tables

Table A.1. Descriptive statistics of the 5-minute and daily returns over the period 01.01.2000–30.07.2008. The means are scaled by 10^4

Stock	Mean	Max	Min	SD	Skewness	Kurtosis
5-minute returns						
AXP	0.000	0.054	−0.034	0.0021	0.353	22.679
C	−0.100	0.060	−0.048	0.0022	0.235	24.834
GE	−0.050	0.039	−0.029	0.0018	0.257	15.937
HD	−0.061	0.033	−0.029	0.0021	0.078	12.680
IBM	0.080	0.074	−0.022	0.0017	0.849	36.369
JPM	0.039	0.056	−0.051	0.0023	0.198	22.785
Daily returns						
AXP	0.004	0.094	−0.093	0.186	−0.049	5.917
C	−7.829	0.159	−0.130	0.179	0.312	10.732
GE	−3.911	0.099	−0.083	0.154	0.213	6.687
HD	−4.770	0.120	−0.086	0.184	0.466	6.369
IBM	6.225	0.123	−0.095	0.153	0.093	7.671
JPM	3.056	0.254	−0.164	0.203	1.258	20.051

Table A.2. Descriptive statistics of realized covariances and variances of the six stocks. The realized variances and covariances are calculated from 5-minute intraday returns (measured in %), as described in the main text

Stock	Mean	Max	Min	SD	Skewness	Kurtosis
Realized variance						
AXP	3.443	57.583	0.073	4.684	4.229	32.783

Table A.2. (*Continued*)

Stock	Mean	Max	Min	SD	Skewness	Kurtosis
C	3.610	119.857	0.107	5.910	7.646	108.488
GE	2.429	51.402	0.103	3.174	4.904	46.966
HD	3.456	51.376	0.165	3.968	3.919	28.013
IBM	2.275	56.909	0.119	3.049	5.684	67.597
JPM	1.009	28.766	−1.108	1.644	6.415	74.871
Realized covariance						
AXP-C	1.587	37.664	−0.548	2.775	5.320	46.132
AXP-GE	1.106	26.317	−1.467	1.845	5.899	58.081
AXP-HD	1.161	27.657	−2.452	1.967	5.327	47.599
AXP-IBM	0.917	23.434	−0.789	1.464	5.647	55.889
AXP-JPM	1.582	40.305	−0.975	2.812	6.098	62.128
C-GE	1.241	41.686	−0.583	2.117	7.019	91.587
C-HD	1.269	27.336	−0.928	2.175	5.019	39.512
C-IBM	1.028	36.726	−3.269	1.737	7.558	109.960
C-JPM	2.003	107.553	−0.472	3.995	11.349	245.086
GE-HD	1.040	26.852	−1.139	1.702	5.903	59.205
GE-IBM	0.901	24.054	−0.329	1.445	5.757	57.766
GE-JPM	1.197	49.23	−0.624	2.136	8.667	144.595
HD-IBM	0.875	18.318	−1.203	1.339	5.213	44.183
HD-JPM	1.233	26.399	−2.567	2.077	5.003	39.756
IBM-JPM	4.017	188.049	0.117	6.924	11.919	262.259

Table A.3. Parameter estimates and bootstrapped standard errors (in parentheses) for the VARFIMA(1, d, 1) model in equation (4) based on Cholesky factors X_t and log-volatilities a_t. The first two columns correspond to the two specifications (Model 1 and 2, respectively) of the VARFIMA-Chol, while the third column corresponds to the VARFIMA-Log

Parameter		Variable		
		X_t (Model 1)	X_t (Model 2)	a_t (Model 2)
AR	ϕ	0.4842 (0.3867)	ϕ 0.5625 (0.3105)	0.5273 (0.3366)
MA	θ	0.5138 (0.3432)	θ 0.5054 (0.3173)	0.4693 (0.3457)
Fractional integration	d_1	0.4888 (0.0286)	d 0.449 (0.0340)	0.475 (0.0340)
	d_2	0.4712 (0.0432)		
	d_3	0.5275 (0.0439)		
	d_4	0.3811 (0.0288)		
	d_5	0.3600 (0.0261)		
	d_6	0.4721 (0.0317)		
	d_7	0.3947 (0.0261)		
	d_8	0.3027 (0.0227)		
	d_9	0.3249 (0.0264)		
	d_{10}	0.4736 (0.0281)		
	d_{11}	0.3297 (0.0254)		
	d_{12}	0.2970 (0.0273)		
	d_{13}	0.2814 (0.0225)		
	d_{14}	0.1568 (0.0210)		
	d_{15}	0.4934 (0.0278)		
	d_{16}	0.4500 (0.0466)		
	d_{17}	0.3753 (0.0318)		
	d_{18}	0.2277 (0.0233)		
	d_{19}	0.1565 (0.0215)		
	d_{20}	0.1603 (0.0211)		
	d_{21}	0.5326 (0.0550)		

Table A.4. Parameter estimates and bootstrapped standard errors (in parentheses) for the HAR-Cholesky (1st column) and HAR-Log (2nd column) models

Parameter	Variable	
	X_t	a_t
$\beta^{(d)}$	0.322	0.322
	(0.055)	(0.015)
$\beta^{(w)}$	0.262	0.323
	(0.043)	(0.033)
$\beta^{(bw)}$	0.179	0.128
	(0.055)	(0.028)
$\beta^{(m)}$	0.181	0.194
	(0.072)	(0.047)

Table A.5. Parameter estimates and bootstrapped standard errors (in parentheses) for the diagonal WAR and the diagonal WAR-HAR models

Parameter	Diagonal WAR					
M	0.817	0.837	0.801	0.841	0.783	0.826
	(0.019)	(0.015)	(0.027)	(0.022)	(0.029)	(0.013)

Parameter	Diagonal WAR-HAR					
$M^{(d)}$	0.752	0.778	0.736	0.795	0.706	0.785
	(0.052)	(0.035)	(0.036)	(0.071)	(0.065)	(0.029)
$M^{(w)}$	0.079	0.090	0.067	0.060	0.046	0.164
	(0.105)	(0.103)	(0.097)	(0.072)	(0.075)	(0.123)
$M^{(bw)}$	0.419	0.415	0.424	0.359	0.449	0.352
	(0.137)	(0.129)	(0.134)	(0.113)	(0.129)	(0.117)
$M^{(m)}$	−0.029	−0.021	0.003	−0.012	0.051	−0.006
	(0.011)	(0.006)	(0.001)	(0.004)	(0.0139)	(0.001)

Table A.6. Parameter estimates and bootstrapped standard errors (in parentheses) for the DCC and FIDCC models

Parameter	Variable					
	AXP	C	GE	HD	IBM	JPM
DCC						
w_i	0.009	0.011	0.003	0.016	0.008	0.012
	(0.005)	(0.008)	(0.003)	(0.008)	(0.005)	(0.008)
α_i	0.071	0.085	0.039	0.049	0.057	0.092
	(0.017)	(0.030)	(0.013)	(0.011)	(0.019)	(0.028)
β_i	0.929	0.916	0.960	0.946	0.939	0.912
	(0.016)	(0.029)	(0.013)	(0.011)	(0.020)	(0.025)
	θ_1	0.013		θ_2	0.970	
		(0.002)			(0.005)	
FIDCC						
w_i	0.059	0.084	0.065	0.186	0.061	0.077
	(0.024)	(0.031)	(0.028)	(0.055)	(0.021)	(0.037)

Table A.6. (*Continued*)

Parameter	Variable					
	AXP	C	GE	HD	IBM	JPM
β_i	0.462	0.318	0.298	0.351	0.448	0.386
	(0.008)	(0.008)	(0.006)	(0.013)	(0.006)	(0.003)
d_i	0.499	0.467	0.411	0.413	0.499	0.499
	(0.009)	(0.055)	(0.006)	(0.018)	(0.109)	(0.014)
θ_1		0.013		θ_2		0.966
		(0.002)				(0.009)

Table A.7. Estimated marginal effects (equations (A.2) and (A.3)) and bootstrapped standard errors (in parentheses) from the VARFIMA-Chol model. The table is read by substituting the dot ($\cdot$) in the formula in the first column with the corresponding variable from columns 2–7. The realized covariances are computed with returns in %

Partial effect	Variable					
	AXP	C	GE	HD	IBM	JPM
$\frac{\partial V\ AXP_{t+1}}{\partial COV\ AXP,\ \cdot_t}$	0.3572	0.6692	2.3606	1.2048	4.7499	0.7870
	(0.0416)	(0.0780)	(0.2752)	(0.1404)	(0.5538)	(0.0917)
$\frac{\partial COV\ AXP,\ \cdot_{t+1}}{\partial V\ AXP_t}$	0.3572	0.1861	0.0482	0.0967	0.0253	0.1578
	(0.0416)	(0.0222)	(0.0078)	(0.0135)	(0.0046)	(0.0191)
$\frac{\partial V\ C_{t+1}}{\partial COV\ C,\ \cdot_t}$	0.7444	0.7130	5.9233	3.9596	30.1783	2.7510
	(0.0888)	(0.0829)	(0.6874)	(0.4581)	(3.4788)	(0.3181)
$\frac{\partial COV\ C,\ \cdot_{t+1}}{\partial V\ C_t}$	0.1673	0.7130	0.0736	0.1329	0.0293	0.2192
	(0.0195)	(0.0829)	(0.0142)	(0.0194)	(0.0057)	(0.0257)
$\frac{\partial V\ GE_{t+1}}{\partial COV\ GE,\ \cdot_t}$	0.1928	0.2946	0.9512	2.5289	8.6499	4.7101
	(0.0314)	(0.0570)	(0.1534)	(0.3441)	(1.5233)	(0.5795)
$\frac{\partial COV\ GE,\ \cdot_{t+1}}{\partial V\ GE_t}$	0.5901	1.4808	0.9512	0.5919	0.1458	0.8791
	(0.0688)	(0.1718)	(0.1534)	(0.0853)	(0.0265)	(0.1043)
$\frac{\partial V\ HD_{t+1}}{\partial COV\ HD,\ \cdot_t}$	0.3870	0.5316	2.3678	1.3482	23.1274	5.2299
	(0.0543)	(0.0777)	(0.3415)	(0.1787)	(2.9623)	(0.6526)
$\frac{\partial COV\ HD,\ \cdot_{t+1}}{\partial V\ HD_t}$	0.3012	1.0139	0.6322	1.3482	0.1648	0.6085
	(0.0351)	(0.1210)	(0.0860)	(0.1787)	(0.0264)	(0.0819)
$\frac{\partial V\ IBM_{t+1}}{\partial COV\ IBM,\ \cdot_t}$	0.1013	0.1174	0.5835	0.6592	1.9578	−2.7829
	(0.0184)	(0.0228)	(0.1062)	(0.1058)	(0.2936)	(0.3591)
$\frac{\partial COV\ IBM,\ \cdot_{t+1}}{\partial V\ IBM_t}$	1.1874	7.7684	2.1624	9.1196	1.9578	4.2042
	(0.1384)	(0.9271)	(0.3808)	(1.1279)	(0.2936)	(0.5465)
$\frac{\partial V\ JPM_{t+1}}{\partial COV\ JPM,\ \cdot_t}$	0.6315	0.8771	3.5164	2.4340	16.8169	1.7010
	(0.0765)	(0.1029)	(0.4174)	(0.3276)	(2.1862)	(0.3623)
$\frac{\partial COV\ JPM,\ \cdot_{t+1}}{\partial V\ JPM_t}$	0.1967	0.7049	1.1775	1.9826	−0.6957	1.7010
	(0.0229)	(0.0841)	(0.1448)	(0.2412)	(0.0897)	(0.3623)

A.3. Bias Correction

Forecasting a nonlinear transformation of Y_t, such as X_t or a_t, and then inverting the transformation to obtain a prediction of Y_t, leads to biases. For our model, the theoretical bias correction was derived in equation (15), while Bauer and Vorkink (2007) derive the bias arising from using the matrix log transformation. They note that the theoretical result is only valid under restrictive assumptions, and therefore suggest using a data-driven procedure in empirical work. Our theoretical formula depends crucially on the second moment of the forecasting error, which is model-dependent and has to be estimated in practice. Consequently, we also argue that in practice a data-driven bias correction would arguably be more appropriate. In their study, Bauer and Vorkink (2007) suggest scaling $\hat{Y}_t$ to match the level of Y_t, which, however, is unavailable if we want to correct a forecast.

One could still consider the following procedure: construct the series $\zeta_{i,u} = \dfrac{\sqrt{Y_{ii,u}}}{\sqrt{\hat{Y}_{ii,u}}}$, $u = 1, \ldots, t$, $i = 1, \ldots, n$ and scale the series $\sqrt{\hat{Y}_{ii,t+s}}$ by the mean $\frac{1}{J}\sum_{j=0}^{J-1}\zeta_{i,t-j}$ (or the median of $\zeta_{i,u}$, for $u = t - J + 1, \ldots, t$) so that all predicted volatilities match the level of the average (median) observed volatility, with J controlling for the trade-off between the bias (favouring small J) and the precision (favouring large J) of the correction. Given the ambiguity related to the choice of J, we would only consider this correction if the biases are large. To investigate whether this is the case, we construct the mean and the median of $\zeta_{i,t}$ for $t = T', \ldots, T$ and $J = t$. Both statistics are slightly below unity, both for the forecasts based on X_t and a_t, implying that the model fit is somewhat upward based, as the theory suggests. For the X_t-based forecast, the largest deviation from unity over all periods and series is 0.93 (0.95) for the mean (median) of $\zeta_{i,t}$, while for the a_t series the statistics are even closer to one at 0.98 (0.99). Summarizing the evidence, we hardly find any empirical justification for bias correction. Given that the Cholesky-based forecasts seem more prone to bias compared to the log-volatility-based predictions, refraining from bias correction can only distort the evaluation to the disadvantage of the Cholesky factor-based model.