Dynamic Modelling and Forecasting of Realized Covariance Matrices*

Roxana Chiriac †
University of Konstanz

Valeri Voev ‡
University of Konstanz,
CoFE

This Version: December 3, 2007

*We would like to thank Ingmar Nolte, Winfried Pohlmeier, Joachim Grammig, participants in the Financial Econometrics Seminar in Riezlern and participants in the Multivariate Volatility Models Conference in Faro for helpful discussions. All remaining errors are ours. The paper was previously circulated under the name "Long Memory Modelling of Realized Covariance Matrices".

†Department of Economics, Box D124, University of Konstanz, 78457 Konstanz, Germany. Phone +49-7531-88-5111, Fax -4450, email: Roxana.Chiriac@uni-konstanz.de. The first author gratefully acknowledges financial support from the German federal state of Baden-Württemberg through a Landesgraduiertenstipendium.

‡Department of Economics, Box D124, University of Konstanz, 78457 Konstanz, Germany. Phone +49-7531-88-2204, Fax -4450, email: Valeri.Voev@uni-konstanz.de. The second author gratefully acknowledges financial support from the German Science Foundation within the project "Price-, Liquidity- and Credit Risks: Measurement and Allocation."

Abstract

This paper proposes a methodology for modelling time series of realized covariance matrices in order to forecast multivariate risks. The model is based on a multivariate, fractionally integrated Autoregressive Moving Average (ARMA) process for the elements of the Cholesky factors of the observed matrix series. This approach allows for joint modelling of the whole covariance matrix and guarantees positive definiteness of the resulting forecasts without imposing parameter restrictions on the model. The model is particularly suited to capture the long memory, typically observed in volatility processes of financial assets. We describe the forecasting procedure and provide an empirical application.

JEL classification: C32, C53, G11

Keywords: Long memory processes, Realized covariance, Forecasting, Fractional integration

1 Introduction

Multivariate volatility modelling is of particular importance in the fields of risk management, portfolio management and asset pricing. Typical approaches employed to model multivariate volatility are the MGARCH models (for a comprehensive review see Bauwens, Laurent, and Rombouts (2006)), stochastic volatility models (reviewed in Asai, McAleer, and Yu (2005)) and multivariate realized covariance models (see Barndorff-Nielsen and Shephard (2004) and Andersen, Bollerslev, Diebold, and Ebens (2001), among others). While the MGARCH and SV approaches model the unobserved daily covariance, the realized covariance models exploit the wealth of information contained in high-frequency data to obtain highly precise estimates of the covariance of the underlying assets at lower (e.g., daily) frequencies. One of the most prominent features of volatility is the presence of long memory, which led, within the GARCH framework, to the development of the integrated GARCH (Engle and Bollerslev (1986)), and the fractionally integrated GARCH (Baillie, Bollerslev, and Mikkelsen (1996)) models. With high frequency data, the long persistence in a univariate series of realized volatilities is portrayed by a slow decay in the autocorrelation functions of the series (see e.g., Andersen and Bollerslev (1997), Andersen, Bollerslev, Diebold, and Ebens (2001)) and is modeled with fractionally integrated ARMA (ARFIMA) processes by Andersen, Bollerslev, Diebold, and Labys (2003), Oomen (2001) and Koopman, Jungbacker, and Hol (2005), among others.

Recently, the literature on MGARCH models is developing towards parsimonious model specifications applicable for higher dimensional problems. At the same time, there is little research on time series models for covariance matrices estimated with high frequency data. The existing literature on dynamic modelling of realized covariance matrices typically concentrates on univariate approaches for a time series of realized volatilities or a single realized covariance (correlation) time series. Andersen, Bollerslev, Diebold, and Ebens (2001) model the series of log-realized volatilities and realized correlations with univariate ARFIMA models, while Corsi (2005) and Corsi and Audrino (2007) apply univariate Heterogenous Autoregressive (HAR) models to capture the high persistency of the series through an autoregressive representation of volatilities/correlations realized over different time horizons. However, the matrix constructed from the correlation forecasts obtained from these univariate models is not guaranteed to be positive definite. In order to produce a forecast of the whole 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. Among the few proposed models for the dynamics of the whole realized covariance matrix are the Wishart Autoregressive (WAR) model of Gourieroux, Jasiak, and Sufana (2004), based on the distribution of the sample variance-covariance matrix, known in the literature as the Wishart distribution, and the model of Bauer and Vorkink (2006), which uses the matrix log transformation to guarantee positive definiteness of the model-implied and the forecasted matrix. In both studies, however, the empirically observed high-persistence of the multivariate volatility process is not taken into account. The model proposed in this paper overcomes the limits of the previous approaches; it describes the joint dynamics of (co)variance time series by modelling the Cholesky decomposition of the realized covariance matrices, which guarantees the positive definiteness of the resulting matrix forecasts, in a framework, which allows for flexible serial dependence patterns, including a slower than exponential decrease in the autocorrelation functions. Furthermore, inclusion of an arbitrary number of explanatory variables is straightforward. The model can be seen as an application and extension of the multivariate ARFIMA model of Sowell (1989) which we estimate by conditional maximum likelihood (ML) based on the work of 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 number of assets. An additional advantage of Beran's (1995) approach is that the process is not restricted to be stationary.[1]

The paper is structured as follows: in Section 2 we present the model theoretically, Section 3 contains the results of the empirical application of the model and the forecasting procedure and Section 4 concludes.

2 The Model

Let Y_t of dimension $n \times n$ be the realized covariance matrix at time t, where n represents the number of assets considered. The Cholesky decomposition of the matrix Y_t is given by the upper triangular matrix P_t, for which $P_t' P_t = Y_t$. Let $X_t = vech(P_t)$ be the

[1]Doornik and Ooms (2004) is a nice review of inference and forecasting for ARFIMA models.

vector obtained by stacking the upper triangular components of the matrix P_t in a vector. In this paper we model the dynamics of the vector X_t of dimension $m \times 1$, where $m = n(n+1)/2$, by using a Vector Autoregressive Fractionally Integrated Moving Average or VARFIMA(p, d, q) model, which is defined below.

Definition 1: The VARFIMA(p, d, q) model for the vector process X_t is defined as

$$\Phi(L)D(L)[X_t - BZ_t] = \Theta(L)\varepsilon_t, \qquad \varepsilon_t \sim N(0, \Sigma_t) \tag{1}$$

where X_t is the vector of dimension $m \times 1$ formed from the elements of the Cholesky decomposition of the realized covariance matrix Y_t. 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 $\Phi_i, i = 1, \ldots, p$ and $\Theta_j, j = 1, \ldots, q$ – 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. Σ_t is the covariance matrix of ε_t. We assume that the roots of $\Phi(L)$ and $\Theta(L)$ lie outside the unit circle.

The model presented here is an extension of the VARFIMA model of Sowell (1989) in which we allow for conditional heteroscedasticity of the error term. An element of the vector X_t, say X_{it}, is stationary if $d_i < 0.5$. Moreover, the whole vector process X_t is stationary if $d_i < 0.5$ for $i = 1, \ldots, m$.

In this paper, we use the following specification of the model:

$$\Phi(L)D(L)[X_t - \mu] = \Theta(L)\varepsilon_t, \qquad \varepsilon_t \sim N(0, \Sigma_t) \tag{2}$$

where μ is a vector of constants of dimension $m \times 1$. As explanatory variables one could consider variables that are documented to have a significant effect on stock market volatility, such as lags of squared daily returns (Black (1976)), functions of trading volume (Lamoureaux and Lastrapes (1990)), corporate bond returns (Schwert (1989)) or short term interest rates (Glosten, Jagannathan, and Runkle (1993)). We allow the matrix Σ_t to be time-varying as there is some evidence that volatility of volatility is time-varying.[2] In this paper, we use the diagonal BEKK$(1, 1, 1)$ specification to

[2]See e.g., Corsi, Kretschmer, Mittnik, and Pigorsch (2005), who show that the residuals of ARFIMA models fitted on univariate time series of realized (co)variances exhibit non–Gaussianity and volatility clustering. Consequently, they extend the ARFIMA framework by including a GARCH component on the volatility of ARFIMA residuals, which substantially improves the goodness-of-fit.

parameterize the time-varying covariance of the VARFIMA(p,d,q) innovation process ε_t:

$$\Sigma_t = C'C + A'\varepsilon_{t-1}\varepsilon'_{t-1}A + B'\Sigma_{t-1}B, \tag{3}$$

where C is an upper $m \times m$ triangular parameter matrix and A and B are $m \times m$ diagonal parameter matrices.

As the Cholesky factors of a positive definite matrix depend on the ordering of the assets, which is usually arbitrary in the realized covariance matrix, for the purpose of forecasting a possible strategy could be to estimate the model for all orderings and to average over the forecasts resulting from the different orderings.

The assumption of normally distributed error terms gives rise to a Gaussian likelihood function, which, maximized under some regularity conditions (see Gourieroux and Monfort (1995)) and the assumption that the conditional mean function is well specified, provides consistent estimates of the parameters of the VARFIMA model defined above. These estimates are known as quasi maximum likelihood (QML) estimates and their consistency is assured as long as the assumed (pseudo true) density function is a member of the linear exponential density family, such as the Gaussian density function. Although the diagonal elements of the Cholesky decomposition are by construction positive, the assumption of normal distribution of the corresponding error terms in Equation (2) is not problematic. The positive definiteness condition for the realized covariance matrix based on forecasted Cholesky factors, does not imply positivity restrictions on the elements of the predicted X_{t+s}, for some $s > 0$. Any (invertible) upper triangular matrix constructed from the elements of the forecasted X_{t+s} vector provides a positive definite matrix of predicted realized covariances.[3] More formally, the reverse transformation from X_t to Y_t is given by

$$Y_t = upmat(xpnd(X_t))'upmat(xpnd(X_t)),$$

where the *xpnd* operator is the inverse of the *vech* operator and the *upmat* operator creates an upper triangular matrix. This transformation provides the main advantage of our model specification: it guarantees the positive definiteness and symmetry of the realized covariance matrix without imposing any restrictions on the parameters.

[3]The Cholesky decomposition can be extended to positive semi-definite matrices, however, in such cases the Cholesky factorization is not unique for a given ordering.

In terms of estimation, we face the problem, that the parameters of the unrestricted VARFIMA models are not identified, which results from the non-uniqueness of VARMA models, discussed at length in Lütkepohl (2005). The problem in the multivariate case is even more severe than in the univariate case, in which root cancelation in the AR and MA-polynomials can occur. In the multivariate case, even after assuming that the AR and the MA polynomials have no common roots, one can still factor out infinitely many times a so-called unimodular lag operator without changing the structure of the process.[4] Lütkepohl (2005) discusses two forms of a general VARMA model which are unique representations of a given VARMA process: final equations form and echelon form. In our paper we consider the final equations form, for which we provide a definition 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 equations form, restricting the AR polynomial to be a scalar polynomial. Apart from guaranteeing uniqueness of the representation, this approach leads to a reduction of the parameters to be estimated. Table 1 gives the total number of parameters for a general VARFIMA(p, d, q) model in final equations form, as well as for certain restricted model specifications considered in this paper.

For our proposes, we employ the model in Equation (2) with AR and MA polynomials of order one and a constant μ of dimension $m \times 1$ (Model 1). Given that $m = \frac{n(n+1)}{2}$, where n represents the number of stocks considered in the application, Model 1 has a total number of $\frac{(n^2+n+2)^2}{4}$ parameters. In order to reduce the number of parameters, we assume in a restricted version of the model that all Cholesky decomposition series are fractionally integrated with the same degree of integration $d = d_1 = \ldots = d_m$, and, consequently, $D(L) = (1 - L)^d I_m$ (Model 2). Further reduction of the number of parameters is achieved by imposing diagonality on the parameter matrix Θ (Model 3).

Since the model in Equation (2) is applied to a transformation of the realized covariance matrix, namely the vector series of Cholesky factors, the estimated parameters

[4] A unimodular lag operator is an operator whose determinant is a non-zero constant, i.e., the determinant does not involve powers of L.

Table 1: Number of parameters for the general VARFIMA(p,d,q) model and its specifications, considered in this paper.

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

Note: $^*k = 1$ (constant). Model 1 is unrestricted VARFIMA$(1,d,1)$, Model 2 is restricted VARFIMA$(1,d,1)$ with $d_1 = \ldots = d_m$, and Model 3 is restricted diagonal VARFIMA$(1,d,1)$, where in addition to the restriction in Model 2, Θ is diagonal.

of the model are not directly interpretable. However, one can derive the dynamic linkages among the variance and covariance series as a function of the estimated parameters; the predicted values of elements of the covariance matrix $E_t[Y_{t+s}]$ are nonlinear functions of the elements of the forecast $E_t[X_{t+s}] \equiv E\left[X_{t+s}|\mathcal{F}_t\right]$, where $\mathcal{F}_t$ is the information set at time t, and, consequently, they are nonlinear functions of the estimated parameter vector and the elements of the vector X_t. We can write the ij-th element of the predicted covariance matrix as (see Appendix A for derivation):

$$E_t[Y_{ij,t+s}] = \sum_{l=1+\frac{i(i-1)}{2}}^{\frac{i(i+1)}{2}} E_t\left[X_{l,t+s}\right] E_t\left[X_{l+\frac{j(j-1)}{2}-\frac{i(i-1)}{2},t+s}\right] \equiv G_{i,j,s}(X_t,\hat{\theta}), \tag{4}$$

where $l,i,j = 1,\ldots,m$, $j \geq i$ and $E_t\left[X_{l,t+s}\right]$ is the l-th element of the vector $E_t\left[X_{t+s}\right]$. $G_{i,j,s}(\cdot)$ is a scalar function of the elements of X_t and $\hat{\theta}$, corresponding to the ij-th element of the matrix $E_t[Y_{t+s}]$, where $\hat{\theta}$ is the vector of all estimated parameters. For example, the impact of covariance $Y_{ij,t}$ on the predicted variance $E_t[Y_{ii,t+s}]$ can be computed as follows:

$$\frac{\partial E_t[Y_{ii,t+s}]}{\partial Y_{ij,t}} = \frac{\partial G_{i,i,s}}{\partial G_{i,j,0}} = \sum_{r=1}^{m} \frac{\partial G_{i,i,s}}{\partial X_r}\frac{\partial X_r}{\partial G_{i,j,0}} = F(X_t,\hat{\theta}),$$

where $G_{i,j,0} \equiv Y_{ij,t} = \sum_{l=1+\frac{i(i-1)}{2}}^{\frac{i(i+1)}{2}} X_{l,t}X_{l+\frac{j(j-1)}{2}-\frac{i(i-1)}{2},t}$, $j \geq i$ and $X_{l,t}$ is the l-th element of the vector X_t, and $F(\cdot)$ is a nonlinear scalar function.

3 Empirical Application

We estimate the model specifications described in Table 1 for $n = 2$ assets for three different combinations of stocks traded at the New York Stock Exchange (NYSE). For the estimation we use a multivariate extension of the conditional maximum likelihood approach of Beran (1995).

3.1 Data

The data we use is taken from the NYSE Trade and Quotations (TAQ) database and consists of tick-by-tick bid and ask quotes sampled from 9:45 until 16:00 over the period January 1, 2001 to June 30, 2006 (1381 trading days).[5] For the current analysis, we select the following four stocks: Home Depot Inc. (HD), Hewlett-Packard (HWP), International Business Machines (IBM) and JPMorgan Chase & Co (JPM). All stocks trade on the NYSE and are highly liquid, which motivated the choice. Before proceeding to construct the equally-spaced returns, we clean the data from possible erroneous entries (quotes canceled or reported much later than they occurred) and compute midquotes. In order to obtain a regularly spaced sequence of midquotes, we use the previous-tick interpolation method, described in Dacorogna, Gençay, Müller, Olsen, and Pictet (2001). The mid-quotes are thus sampled at the 5-minute and daily frequencies[6], from which 5-minute and daily log returns are constructed. Thus we obtain 75 intraday observations which are used to compute the realized variance-covariance matrices for each day. Table B.1 in Appendix B reports summary statistics of both intraday 5-minute and daily returns. We observe typical stylized facts such as overkurtosis and tendency for negative skewness of intradaily and daily returns (across all four stocks, the average kurtosis of 5-minute return series is about 305.8, while of daily returns is about 11.7).

For each $t = 1, \ldots, 1381$, we construct the daily realized covariance matrices from the intraday 5-minute returns as follows:

$$Y_t = \sum_{j=1}^{M} r_{j,t} r'_{j,t} \tag{5}$$

[5] Although the NYSE market opens at 9:30, we filter out the quotes recorded in the first 15 minutes in order to eliminate the opening auction effect on the price process.

[6] The daily closing price is determined as the last midquote that took place before or at 16:00, the closing time of each trading day.

where $M = 75$ and $r_{j,t}$ is the $n \times 1$ vector of intraday returns computed as

$$r_{j,t} = p_{j,t} - p_{j-1,t}, \quad j = 1, \ldots, M \tag{6}$$

where $p_{j,t}$ is the j-th log midquote observation in day t. By construction, the realized covariance matrices are symmetric and, for $n < M$, they are guaranteed to be positive definite. Table B.2 in Appendix B reports the average values of the summary statistics for the realized variances and covariances across the 4 stocks considered in the study. As already documented by Andersen, Bollerslev, Diebold, and Ebens (2001), both realized variance and covariance distributions are extremely right skewed and leptokurtic.

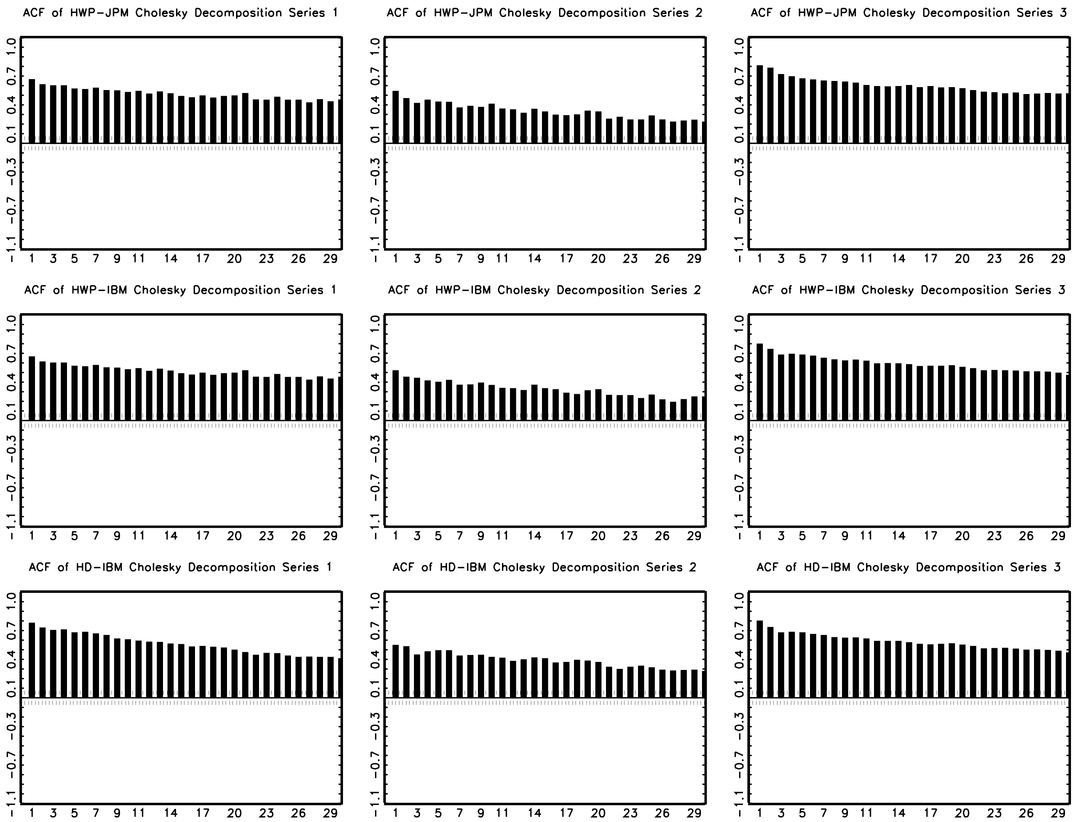
After computing the series of realized covariance matrices, we proceed to construct their Cholesky decompositions. As described in Section 2, the VARFIMA model is estimated for the $m = \frac{n(n+1)}{2}$ non-zero elements in the Cholesky decomposition, which turn out to preserve the long memory properties of the realized (co)variances as documented by e.g., Andersen and Bollerslev (1997) and Andersen, Bollerslev, Diebold, and Ebens (2001). Evidence of this fact is presented in Figure 1: the sample autocorrelations of the Cholesky factors of the realized covariance matrices of the three stock combinations HWP-JPM, HWP-IBM and HD-IBM decay at a slow rate, similar to the autocorrelations of the realized (co)variance series.

3.2 Estimation Results

In Table 2 we report estimation results for the stock combinations HWP-JPM, HWP-IBM and HD-IBM, for which we estimated an unrestricted VARFIMA$(1, d, 1)$ (Model 1). The results are similar across stock pairs; in particular, the long memory is captured in the values of the d-parameters around 0.37 for the diagonal elements of the Cholesky decomposition matrices and 0.29 for the off-diagonal element. The AR parameter is significantly positive (except for the second combination), while the diagonal MA-parameters are significantly negative in the first two cases. Similar results are obtained by Oomen (2001) in the univariate modelling of log realized volatilities.

Figure 2 plots the autocorrelogram of the standardized residual series of the model estimated from the three stock combinations and the values of the multivariate Ljung-Box statistic at lag 30 are included in the last row of Table 2. The residual autocorrelograms do not reveal any systematic autocorrelation left in the residuals, and we presume that the model properly fits the dynamics of the realized covariance Cholesky

Figure 1: Autocorrelation functions of the Cholesky factors of the realized covariance matrices of HWP-JPM, HWP-IBM, and HD-IBM.



factor series. Though one could argue that the values of the multivariate Ljung-Box statistic are high, the reduction is considerable when compared to the values of the statistic of the original series.[7]

To check the robustness of the results with respect to the sampling frequency of the returns used to construct the realized covariance matrices, we carried out the above analysis with realized covariances based on 30-minute returns, which should be much less susceptible to market microstructure noise. The results we obtained are qualitatively similar, so we refrain from reporting them here.[8]

[7]The values of the multivariate Ljung-Box statistic for the raw series X_t at lag 30 is 9803.994, 10747.351, and 11092.249 for HWP-JPM, HWP-IBM, and HD-IBM, respectively.

[8]The results can be obtained on request from the authors.

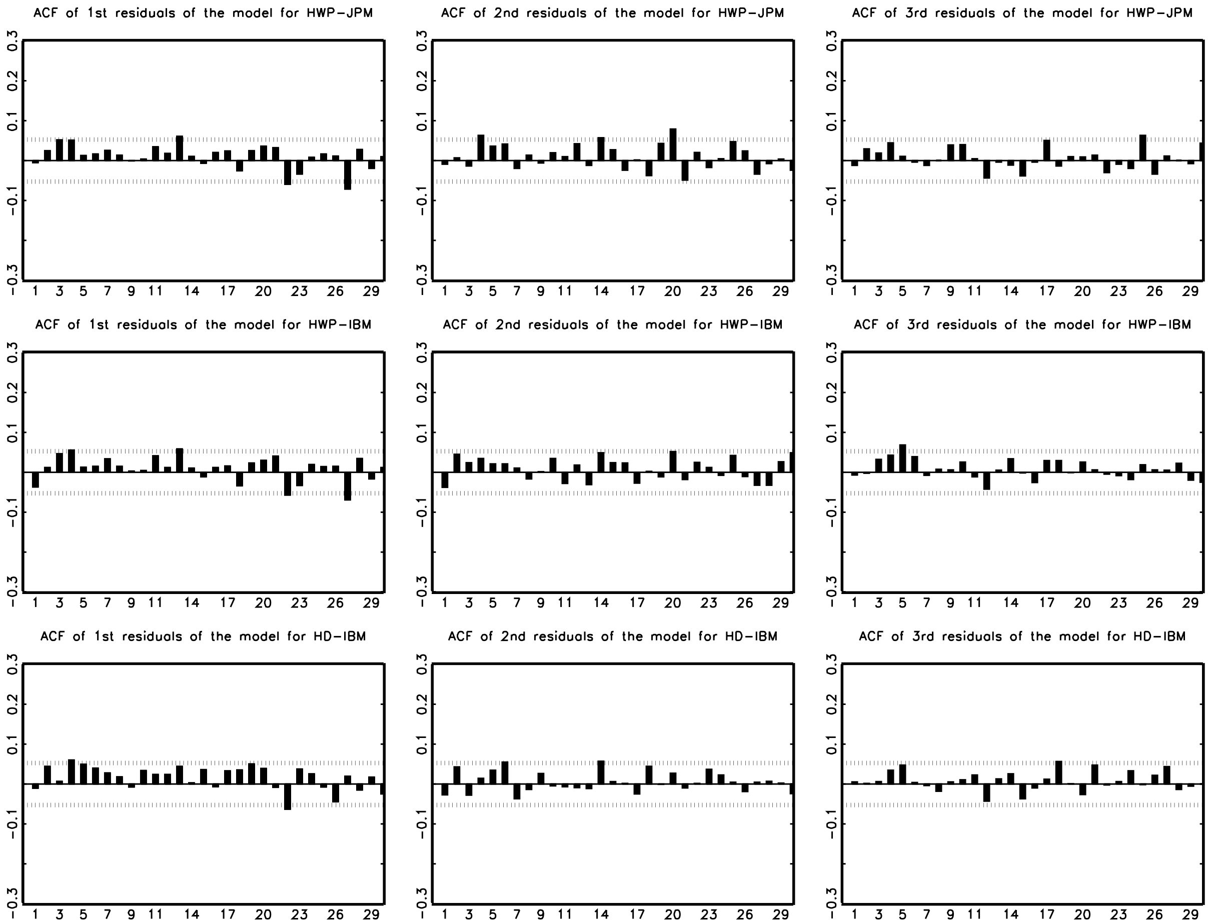
Table 2: Estimation results of the VARFIMA(1,d,1)-diagonal BEKK(1,1,1) model, $n = 2$ (Model 1)

Parameter	HWP-JPM	HWP-IBM	HD-IBM
d_1	0.3743 (0.0000)	0.3745 (0.0000)	0.3889 (0.0000)
d_2	0.2876 (0.0000)	0.3016 (0.0000)	0.2966 (0.0000)
d_3	0.3899 (0.0000)	0.4097 (0.0000)	0.3940 (0.0000)
ϕ_1	0.5019 (0.0011)	0 .5594 (0.1341)	0.5156 (0.0006)
θ_{11}	-0.5943 (0.0000)	-0.6390 (0.0489)	-0.6086 (0.0000)
θ_{12}	0.0291 (0.5960)	0.0889 (0.3213)	-0.0143 (0.7599)
θ_{13}	0.0933 (0.0178)	0.1241 (0.0710)	0.2000 (0.0000)
θ_{21}	0.0157 (0.3524)	0.0153 (0.2166)	0.0110 (0.6167)
θ_{22}	-0.6372 (0.0000)	-0.6643 (0.0310)	-0.6248 (0.0000)
θ_{23}	0.1236 (0.0001)	0.0865 (0.0146)	0.1197 (0.0009)
θ_{31}	0.0225 (0.2563)	0.0196 (0.2789)	0.0319 (0.1404)
θ_{32}	0.0198 (0.6390)	0.0201 (0.5165)	0.0184 (0.6340)
θ_{33}	-0.5004 (0.0002)	-0.5956 (0.0592)	-0.5664 (0.0000)
μ_1	0.0060 (0.0011)	0.0067 (0.0002)	0.0051 (0.0011)
μ_2	0.0011 (0.0338)	0.0016 (0.0010)	0.0016 (0.0007)
μ_3	0.0012 (0.4475)	0.0039 (0.0031)	0.0038 (0.0005)
C_{11}	0.0002 (0.0000)	0.0002 (0.0000)	0.0007 (0.0059)
C_{12}	0.0001 (0.0527)	0.0002 (0.0000)	0.0001 (0.0078)
C_{13}	-0.0002 (0.0091)	0.0000 (0.5954)	-0.0002 (0.0125)
C_{22}	0.0005 (0.0000)	0.0001 (0.0018)	0.0001 (0.0088)
C_{23}	-0.0001 (0.7252)	-0.0002 (0.0001)	0.0000 (0.3293)
C_{33}	0.0000 (0.9929)	0.0000 (0.9800)	-0.0002 (0.0001)
A_{11}	0.0755 (0.0066)	-0.0552 (0.1204)	0.3328 (0.0000)
A_{22}	0.2257 (0.0000)	-0.2417 (0.0000)	0.2165 (0.0002)
A_{33}	0.3667 (0.0000)	-0.1449 (0.0000)	0.1510 (0.0000)
B_{11}	0.9951 (0.0000)	0.9966 (0.0000)	0.9275 (0.0000)
B_{22}	0.9725 (0.0000)	0.9641 (0.0000)	0.9682 (0.0000)
B_{33}	0.9216 (0.0000)	0.9833 (0.0000)	0.9829 (0.0000)
LB(30) of $\hat{\varepsilon}_t$	335.116	299.920	300.892
AIC	-31.130	-32.177	-33.122

Note: p-values based on QML standard errors are reported in parenthesis.

Given that the estimated degrees of fractional integration, presented in Table 2, seem to be similar not only across stock pairs but also within each stock combination, we estimate, on the same set of data, a restricted VARFIMA$(1, d, 1)$-diagonal BEKK$(1, 1, 1)$ specification (Model 2) where $D(L) = (1 - L)^d I_m$ (all Cholesky decomposition series

Figure 2: Autocorrelation functions of the standardized residual series of the unrestricted VARFIMA(1,d,1)-diagonal BEKK(1,1,1) model (Model 1) for HWP-JPM, HWP-IBM, and HD-IBM.



are assumed to have the same degree of fractional integration, d). The results of this estimation are given in Table 3. The AIC criteria improve in two out of three cases, while the multivariate Ljung-Box statistics do not change considerably, which implies that the in-sample fit is not significantly improved by allowing for different degrees of fractional integration. A further meaningful restriction, motivated by the observation that the off-diagonal Θ coefficients are generally not significant, is to consider a diagonal specification (Model 3), for which we report the results in Table 4.

The diagonal model has values for the AIC close to the values for Model 2 and com-

Table 3: Estimation results of the restricted VARFIMA(1,d,1)-diagonal BEKK(1,1,1) model, $n = 2$, $d = d_1 = \ldots = d_m$ (Model 2)

Parameter	HWP-JPM	HWP-IBM	HD-IBM
d	0.3650 (0.0000)	0.3789 (0.0000)	0.3771 (0.0000)
ϕ_1	0.4520 (0.0004)	0.4927 (0.0784)	0.4536 (0.0004)
θ_{11}	-0.5382 (0.0000)	-0.5796 (0.0172)	-0.5324 (0.0000)
θ_{12}	0.0076 (0.8760)	0.0477 (0.6163)	-0.0265 (0.6049)
θ_{13}	0.1058 (0.0090)	0.1415 (0.0576)	0.2112 (0.0000)
θ_{21}	0.0116 (0.4705)	0.0144 (0.1723)	0.0131 (0.5153)
θ_{22}	-0.6638 (0.0000)	-0.6812 (0.0017)	-0.6581 (0.0000)
θ_{23}	0.1136 (0.0002)	0.0830 (0.0123)	0.1182 (0.0003)
θ_{31}	0.0249 (0.2349)	0.0252 (0.1180)	0.0393 (0.0960)
θ_{32}	0.0342 (0.3912)	0.0207 (0.5597)	0.0276 (0.4658)
θ_{33}	-0.4298 (0.0003)	-0.5036 (0.0478)	-0.4918 (0.0000)
μ_1	0.0064 (0.0002)	0.0067 (0.0006)	0.0058 (0.0001)
μ_2	0.0004 (0.6719)	0.0009 (0.1908)	0.0009 (0.2100)
μ_3	0.0023 (0.1059)	0.0046 (0.0000)	0.0044 (0.0000)
C_{11}	0.0002 (0.0000)	0.0002 (0.0000)	0.0007 (0.0020)
C_{12}	0.0001 (0.0867)	0.0002 (0.0000)	0.0001 (0.0057)
C_{13}	-0.0002 (0.0134)	0.0000 (0.6148)	-0.0002 (0.0123)
C_{22}	0.0005 (0.0000)	0.0001 (0.0013)	0.0001 (0.0030)
C_{23}	0.0000 (0.8605)	-0.0002 (0.0001)	-0.0001 (0.2621)
C_{33}	0.0000 (0.9952)	0.0000 (1.0000)	-0.0002 (0.0000)
A_{11}	0.0768 (0.0058)	0.0537 (0.1366)	0.3266 (0.0000)
A_{22}	0.2329 (0.0001)	0.2462 (0.0000)	0.2275 (0.0003)
A_{33}	0.3626 (0.0000)	0.1417 (0.0000)	0.1481 (0.0000)
B_{11}	0.9950 (0.0000)	0.9967 (0.0000)	0.9300 (0.0000)
B_{22}	0.9712 (0.0000)	0.9631 (0.0000)	0.9653 (0.0000)
B_{33}	0.9235 (0.0000)	0.9838 (0.0000)	0.9832 (0.0000)
LB(30) of $\hat{\varepsilon}_t$	342.465	305.702	299.515
AIC	-31.129	-32.190	-33.147

Note: p-values based on QML standard errors are reported in parenthesis.

pared to the values for Model 1, in two out of three cases the values have improved. Considering these results we believe that a rather parsimonious model is more reasonable to capture the dynamics of realized covariance series and definitely easier to apply in practice. In terms of interdependencies among the variance and covariance series, it is important to note that the diagonal specification does not exclude the possibility of spillover effects among the series. This is due to the non-linearity of the Cholesky

Table 4: Estimation results of the diagonal VARFIMA(1,d,1)-diagonal BEKK(1,1,1) model, $n = 2$, $d = d_1 = \ldots = d_m$, (Model 3)

Parameter	HWP-JPM	HWP-IBM	HD-IBM
d	0.3650 (0.0000)	0.3789 (0.0000)	0.3771 (0.0000)
ϕ_1	0.3165 (0.0175)	0.3473 (0.1492)	0.1691 (0.3968)
θ_{11}	-0.3978 (0.0017)	-0.4379 (0.0368)	-0.2358 (0.2254)
θ_{22}	-0.4703 (0.0001)	-0.5192 (0.0112)	-0.3620 (0.0564)
θ_{33}	-0.3326 (0.0079)	-0.3899 (0.0756)	-0.2340 (0.2475)
μ_1	0.0067 (0.0001)	0.0069 (0.0002)	0.0058 (0.0001)
μ_2	0.0005 (0.5107)	0.0008 (0.2269)	0.0008 (0.2861)
μ_3	0.0023 (0.0580)	0.0045 (0.0000)	0.0043 (0.0000)
C_{11}	0.0002 (0.0000)	0.0002 (0.0000)	0.0007 (0.0020)
C_{12}	0.0001 (0.0867)	0.0002 (0.0000)	0.0001 (0.0057)
C_{13}	-0.0002 (0.0134)	0.0000 (0.6148)	-0.0002 (0.0123)
C_{22}	0.0005 (0.0000)	0.0001 (0.0013)	0.0001 (0.0030)
C_{23}	0.0000 (0.8605)	-0.0002 (0.0001)	-0.0001 (0.2622)
C_{33}	0.0000 (0.9984)	0.0000 (1.0000)	0.0002 (0.0000)
A_{11}	0.0768 (0.0058)	0.0537 (0.1365)	0.3266 (0.0000)
A_{22}	0.2329 (0.0001)	0.2462 (0.0000)	0.2275 (0.0003)
A_{33}	0.3626 (0.0000)	0.1417 (0.0000)	0.1481 (0.0000)
B_{11}	0.9950 (0.0000)	0.9967 (0.0000)	0.9300 (0.0000)
B_{22}	0.9712 (0.0000)	0.9631 (0.0000)	0.9654 (0.0000)
B_{33}	0.9235 (0.0000)	0.9838 (0.0000)	0.9832 (0.0000)
LB(30) of $\hat{\varepsilon}_t$	435.266	364.185	373.933
AIC	-31.120	-32.188	-33.127

Note: p-values based on QML standard errors are reported in parenthesis.

transformation. Even if the X_t series are estimated independently of each other, the resulting Y_t components are functions of the X_t series (see Equation 4), and therefore related to each other. Furthermore, anticipating the results of the next section, in terms of out-of-sample forecasting ability, a model with fewer parameters and more structure may be a better choice than a model which provides the best in-sample fit.

3.3 Forecasting

The motivation for finding a proper model describing the dynamics of the variance-covariance matrix is based on obtaining good return covariance forecasts and, correspondingly, deriving precise risk measures.

For ease of exposition and since the exogenous regressors in the model in Equation (1) are by assumption predetermined, we neglect the term BZ_t. The fractionally differenced series $D(L)X_t$ follow a stationary VARMA process, and therefore we can obtain forecasting formulas through their infinite Vector Moving Average (VMA(∞)) representations (see e.g., Lütkepohl (2005), pp. 228−230). For each $X_{j,t}$ $j = 1, \dots, m$, the fractionally differenced series $(1-L)^{d_j} X_{j,t}$ is given by:

$$(1-L)^{d_j} X_{j,t} = \sum_{h=0}^{\infty} \delta_{j,h} X_{j,t-h} = X_t + \sum_{h=1}^{\infty} \delta_{j,h} X_{j,t-h}, \tag{7}$$

where $\delta_{j,0} = 1 \quad \forall j = 1, \dots, m$, and $\delta_{j,h} = \prod_{0<r\leq h} \frac{r-1-d_j}{r}, r = 1, 2, \dots$.Therefore, we can rewrite Equation (1) as:

$$\Phi(L)\Lambda(L)X_t = \Theta(L)\varepsilon_t, \tag{8}$$

where $\Lambda(L) = I_m + \sum_{h=1}^{\infty} \Delta_h X_{j,t-h}$ and $\Delta_h = diag\{\delta_{1,h}, \dots.\delta_{m,h}\}$. From Equation (8) we can derive the VMA(∞) representation:

$$X_t = \Phi(L)^{-1}\Lambda(L)^{-1}\Theta(L)\varepsilon_t = \sum_{i=0}^{\infty} \Psi_i \varepsilon_{t-i}, \tag{9}$$

and the optimal predictor of X_t in terms of the VMA(∞) representation is given by:

$$E_t[X_{t+s}] = \sum_{i=s}^{\infty} \Psi_i \varepsilon_{t+s-i} = \sum_{i=0}^{\infty} \Psi_{s+i} \varepsilon_{t-i} \tag{10}$$

The resulting forecast is unbiased, that is, the forecast errors have zero mean, and since the ε_t are assumed to be normally distributed, the forecast errors are also normally distributed as:

$$u_{t,t+s} \equiv X_{t+s} - E_t[X_{t+s}] \sim N(0, \Sigma_{t,s}), \tag{11}$$

where

$$\Sigma_{t,s} = E[(X_{t+s} - E_t[X_{t+s}])(X_{t+s} - E_t[X_{t+s}])'] = E[u_{t,t+s} u'_{t,t+s}] = \sum_{i=0}^{s-1} \Psi_i \Sigma_{t+s-i} \Psi'_i \tag{12}$$

and

$$\Sigma_{t+s-i} = C'C + A'\varepsilon_{t+s-i-1}\varepsilon'_{t+s-i-1}A + B'\Sigma_{t+s-i-1}B.$$

It follows that the forecast errors of the one-step ahead forecast, $u_{t,t+1}$, are normally distributed with zero mean and variance-covariance matrix $\Sigma_{t,1} = \Sigma_t$. As we saw above, for each $j = 1, ..., m$, the $X_{j,t}$ process has an infinite autoregressive representation that can be truncated at, say $h = 1000$ lags for practical purposes.

After forecasting X_t, which is of dimension $m \times 1$, we focus on forecasting the symmetric matrix Y_t (daily volatility matrix) of dimension $n \times n$. The forecast error terms for the $m = \frac{n(n+1)}{2}$ distinct elements of the symmetric matrix Y_{t+s} are defined as $e_{ij,t+s} = E_t[Y_{ij,t+s}] - Y_{ij,t+s}$. Since Y_{t+s} is a quadratic transformation of X_{t+s}, the mean of $e_{ij,t+s}$ is no longer zero, but is a function of the elements of the variance $\Sigma_{t,s}$ of the forecast error of X_{t+s}. Thus, in order to obtain unbiased predictors, the forecasts of each component of the covariance matrix, $E_t[Y_{ij,t+s}]$, should be corrected for the bias given by $E_t[e_{ij,t+s}] \equiv \sigma^*_{t,ij} \neq 0$.

The bias correction for $E_t[Y_{ij,t+s}]$ is obtained from the elements of the matrix $\Sigma_{t,s}$ using the following formula:

$$\sigma^*_{t,ij} = \sum_{l=1+\frac{i(i-1)}{2}}^{\frac{i(i+1)}{2}} \sigma_{t,s(l,l+\frac{j(j-1)}{2}-\frac{i(i-1)}{2})}, \tag{13}$$

where $j \geq i$, $i = 1, ..., n$ and $\sigma_{t,s(v,u)}$ is the v, u-th element of the matrix $\Sigma_{t,s}$. This expression is derived in Appendix A. It is important to note that while the bias correction may theoretically lead to non-positive definite forecasts, in the application we present below all forecasts remain positive definite. As the correction is based on estimated parameters, however, it also introduces more estimation noise to the forecasts. Since it is not clear whether this effect dominates the effect of the reduced bias, we consider forecasts with and without bias correction.

We illustrate the forecasting performance of the three specifications of the VARFIMA$(1, d, 1)$ - diagonal BEKK$(1, 1, 1)$ model by considering the stock combination HWP-JPM. To this end, we divide the overall sample of 1381 daily observations in two subsets: an in-sample period on which we estimate the model, and an out-of-sample period which serves to evaluate the forecasting performance. The in-sample period contains initially the first 1181 observations. In each forecasting step, the in-sample period is increased by one observation, the models are re-estimated for different orderings of the stocks in the realized covariances matrices, and a new one-step ahead forecast is made by averaging over the forecasts of the Cholesky factors from the different orderings. This procedure is carried out 200 times, and as a result we obtain a total of 200 one-step ahead forecasts. To evaluate the forecasting performance of the three model specifica-

Table 5: R^2 statistics of the Mincer-Zarnovitz regressions (Equation (14)) for one-step ahead forecasts.

		Bias Correction			No Bias Correction		
		Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
1^{st} Ordering	Y_{11}	0.180	0.183	0.196	0.180	0.182	0.196
	Y_{12}	0.167	0.167	0.177	0.170	0.170	0.179
	Y_{22}	0.255	0.256	0.258	0.258	0.260	0.260
2^{nd} Ordering	Y_{11}	0.162	0.190	0.198	0.155	0.180	0.189
	Y_{12}	0.182	0.184	0.185	0.179	0.181	0.182
	Y_{22}	0.268	0.270	0.270	0.268	0.269	0.269
Average	Y_{11}	0.176	0.190	0.199	0.170	0.183	0.193
	Y_{12}	0.176	0.177	0.182	0.176	0.177	0.182
	Y_{22}	0.263	0.264	0.265	0.264	0.265	0.265

tions we employ Mincer-Zarnovitz type regressions, which are used for evaluation of volatility forecasts by e.g., Andersen and Bollerslev (1998) and Andersen, Bollerslev, and Meddahi (2005). The regression is given by:

$$Y_{ij,t+s} = \alpha_{ij} + \beta_{ij} E_t[Y_{ij,t+s}] + \nu_{ij,t+s}, \quad i,j = 1,\ldots,n, \quad t = 1181,\ldots,1380, \qquad (14)$$

where $\nu_{ij,t+s}$ is the error term of the regression.

The corresponding values of the R^2 statistics for the three model forecasts based on the different orderings and the average forecast are contained in Table 5. The corresponding α and β coefficients are reported in Table B.3 in Appendix B. From the entries of Table 5 it is evident that, in all cases, the R^2 measure based on forecasts from Model 3 is higher or equal to the R^2 values for the other model specifications. This suggests that imposing the restriction of equal degrees of fractional integration for the three considered series and diagonal MA parameter matrix improves the forecasting ability of the proposed model. Regarding the magnitude of the R^2 statistic, similar results have been obtained by Andersen, Bollerslev, Diebold, and Labys (2003) for the out-of-sample one-step-ahead exchange rates univariate volatility forecast based on an ARFIMA$(1, 0.4, 0)$ specification. The R^2's they report range between 20% and 25%. It should be also noted that exchange rates have much lower volatility (and variation in volatility) compared to stocks, and hence we can a priori expect higher predictive power of the model for exchange rate volatility than for stock volatility. Since the realized covariance is subject to estimation noise (due to market microstructure noise, and, additionally for the covariance, due to non-synchronicity), the R^2's understate the true predictive power of the models. Based on the entries of Table B.3, we can not reject the hypothesis that $\alpha_{ij} = 0$ in all cases for Model 3 without bias correction, and in most of the cases for the other models. The 95% confidence intervals for $\beta_i j$ contain the value 1 in all cases.

The entries of Table 5 suggest that unbiased predictors are not necessarily the optimal choice when forecasting the daily covariance; the forecasts with no bias correction are, on average, closer to the true realizations than the bias-corrected ones. Therefore, we can conclude that the restricted model specifications without bias correction for the forecast generally provide, within our framework, the best forecasts of positive definite and symmetric variance-covariance matrices.

4 Conclusion

In this paper, we present an approach for modelling the dynamics of realized covariance matrices. The model we propose explicitly accounts for the empirically observed long memory of financial volatility and also allows for the inclusion of exogenous or endogenous variables (e.g., lagged or contemporaneous traded volume) which have been found to influence volatility. The main feature of our specification is the decomposition of the realized covariance matrices, which are positive definite by construction, into their Cholesky factors. The dynamics of the elements of the Cholesky decompositions are modeled with a multivariate vector fractionally integrated ARMA (VARFIMA) model without imposing restrictions on the admissible parameter space. Within this framework, the forecasts which we obtain for the variance-covariance matrix by "squaring" the forecasted Cholesky elements are positive definite by construction.

We present estimation results and carry out a forecasting exercise within a small empirical application. A problem which arises is that the non-linear transformation of the covariance matrices leads to a non-zero expectation of the forecasting error, resulting from the "squaring" of the predicted Cholesky factors. To account for this bias, we derive the necessary correction to ensure the unbiasedness of the covariance forecasts. The forecasting evaluation, however, shows that a diagonal model specification with equality restriction on the degrees of fractional integration of the Cholesky series as well as no bias correction of the predictions, provides generally the best forecasts for the daily covariance matrices.

As further research, it would be interesting to carry out a more comprehensive empirical analysis as well as to compare the forecasting performance of the proposed specification against multivariate GARCH specifications and the existing multivariate approaches by Gourieroux, Jasiak, and Sufana (2004) and Bauer and Vorkink (2006).

References

An, S., and P. Bloomfield (1993): “Cox and Reid’s Modification in Regression Models with Correlated Error,” Discussion Paper, Department of Statistics, North Carolina State University.

Andersen, T. G., and T. Bollerslev (1997): “Heterogeneous Information Arrivals and Return Volatility Dynamics: Uncovering the Long-Run in High Frequency Returns,” *Journal of Finance*, 52, 975–1005.

——— (1998): “Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts,” *International Economic Review*, 39, 885–905.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens (2001): “The Distribution of Stock Return Volatility,” *Journal of Financial Economics*, 61, 43–76.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003): “Modeling and Forecasting Realized Volatility,” *Econometrica*, 71, 579–625.

Andersen, T. G., T. Bollerslev, and N. Meddahi (2005): “Correcting the Errors: Volatility Forecast Evaluation Using High-Frequency Data and Realized Volatilities,” *Econometrica*, 73, 279–296.

Asai, M., M. McAleer, and J. Yu (2005): “Multivariate Stochastic Volatility,” Working Paper, Tokyo Metropolitan University.

Baillie, R., T. Bollerslev, and H. Mikkelsen (1996): “Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity,” *Journal of Econometrics*, 74, 3–30.

Barndorff-Nielsen, O. E., and N. Shephard (2004): “Econometric Analysis of Realised Covariation: High Frequency Based Covariance, Regression and Correlation in Financial Economics,” *Econometrica*, 72, 885–925.

Bauer, G. H., and K. Vorkink (2006): “Multivariate Realized Volatility,” Working Paper, Bank of Canada.

Bauwens, L., S. Laurent, and J. Rombouts (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.

Black, F. (1976): "Studies in Stock Price Volatility," in *Proceedings of the 1976 Business Meeting of the Business and Economic Statistics Section, American Statitistical Association*, pp. 177–181.

Corsi, F. (2005): "Measuring and Modelling Realized Volatility: from Tick-by-tick to Long Memory," Univerity of Lugano, Mimeo.

Corsi, F., and F. Audrino (2007): "Realized Correlation Tick-By-Tick," Working Paper, University of St. Gallen.

Corsi, F., U. Kretschmer, S. Mittnik, and C. Pigorsch (2005): "Volatility of Realized Volatility," Working Paper, Center for Financial Studies, University of Frankfurt.

Dacorogna, M. M., R. Gençay, U. A. Müller, R. B. Olsen, and O. V. Pictet (2001): *An Introduction to High-Frequency Finance*. San Diego Academic Press.

Doornik, J. A., and M. Ooms (2004): "Inference and Forecasting for ARFIMA Models with an Application to US and UK Inflation," *Studies in Nonlinear Dynamics & Econometrics*, 8, Article 14.

Engle, R., and T. Bollerslev (1986): "Modeling the Persistence of Conditional Variances," *Econometric Reviews*, 5, 1–50.

Glosten, L. R., R. Jagannathan, and D. E. Runkle (1993): "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return of Stocks," *Journal of Finance*, 48, 1179–1801.

Gourieroux, C., J. Jasiak, and R. Sufana (2004): "The Wishart Autoregressive Process of Multivariate Stochastic Volatility," Working Paper, University of Toronto.

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

Koopman, S. J., B. Jungbacker, and E. Hol (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., and W. Lastrapes (1990): "Heteroskedasticity in Stock Return Data: Volume versus GARCH effects," *Journal of Finance*, 45.

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

Oomen, R. (2001): "Using High Frequency Stock Market Index Data to Calculate, Model and Forecast Realized Return Variance," European University, Economics Discussion Paper No. 2001/6.

Schwert, G. W. (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.

——— (1992): "Maximum Likelihood Estimation of Stationary Univariate Fractionally Integrated Time Series Models," *Journal of Econometrics*, 53, 165–188.

Voev, V. (2007): "Dynamic Modelling of Large-Dimensional Covariance Matrices," in *Recent Developments in High Frequency Financial Econometrics*, ed. by L. Bauwens, W. Pohlmeier, and D. Veredas. Springer, Berlin.

A Appendix Proofs

Derivation of $Y_{ij,t}$ from its Cholesky decomposition.

Let us consider a time series of positive definite and symmetric matrices of realized variance-covariance of dimension 3×3, $Y_t = \begin{pmatrix} Y_{11,t} & Y_{12,t} & Y_{13,t} \\ Y_{12,t} & Y_{22,t} & Y_{23,t} \\ Y_{13,t} & Y_{23,t} & Y_{33,t} \end{pmatrix}$.
Its Cholesky decomposition matrix (dimension 3×3) is given by:

$$\tilde{X}_t = \begin{pmatrix} X_{1,t} & 0 & 0 \\ X_{2,t} & X_{3,t} & 0 \\ X_{4,t} & X_{5,t} & X_{6,t} \end{pmatrix}' \tag{A.1}$$

and the vector formed by the non-zero Cholesky decomposition elements is given by:

$$X_t = (X_{1,t}, X_{2,t}, X_{3,t}, X_{4,t}, X_{5,t}, X_{6,t})' \tag{A.2}$$

Reconstructing the matrix Y_t starting from its Cholesky decomposition, we get:

$$\begin{aligned} \tilde{X}_t{}' \tilde{X}_t &= \begin{pmatrix} X_{1,t} & 0 & 0 \\ X_{2,} & X_{3,t} & 0 \\ X_{4,t} & X_{5,t} & X_{6,t} \end{pmatrix} \begin{pmatrix} X_{1,t} & X_{2,t} & X_{4,t} \\ 0 & X_{3,t} & X_{5,t} \\ 0 & 0 & X_{6,t} \end{pmatrix} \\ &= \begin{pmatrix} X_{1,t}^2 & X_{1,t}X_{2,t} & X_{1,t}X_{4,t} \\ X_{1,t}X_{2,t} & X_{2,t}^2 + X_{3,t}^2 & X_{2,t}X_{4,t} + X_{3,t}X_{5,t} \\ X_{1,t}X_{4,t} & X_{2,t}X_{4,t} + X_{3,t}X_{5,t} & X_{4,t}^2 + X_{5,t}^2 + X_{6,t}^2 \end{pmatrix} \\ &= \begin{pmatrix} Y_{11,t} & Y_{12,t} & Y_{13,t} \\ Y_{12,t} & Y_{22,t} & Y_{23,t} \\ Y_{13,t} & Y_{23,t} & Y_{33,t} \end{pmatrix} \end{aligned} \tag{A.3}$$

Therefore each element of matrix Y_t can be written as a function of its Cholesky factors as it follows:

$$Y_{ij,t} = \sum_{l=1+\frac{i(i-1)}{2}}^{\frac{i(i+1)}{2}} X_{l,t} X_{l+\frac{j(j-1)}{2}-\frac{i(i-1)}{2},t} \tag{A.4}$$

where $X_{l,t}$ is the $l-$element of the vector X_t.

Derivation of the forecasting bias in Equation (13).

Let us consider a time series of positive definite and symmetric matrixes of realized variance-covariance of dimension 3×3. The Cholesky decomposition and its vector form are given in Equations (A.1) and (A.2).
The vector of s-step ahead forecast error Equation (1) is given by:

$$u_{t+s} = (u_{1,t+s}, u_{2,t+s}, u_{3,t+s}, u_{4,t+s}, u_{5,t+s}, u_{6,t+s})' \tag{A.5}$$

and has the variance covariance matrix:

$$\Sigma_s = E[u_{t+s}u'_{t+s}] = \begin{pmatrix} \sigma_{s(1,1)} & \sigma_{s(1,2)} & \sigma_{s(1,3)} & \sigma_{s(1,4)} & \sigma_{s(1,5)} & \sigma_{s(1,6)} \\ \sigma_{s(1,2)} & \sigma_{s(2,2)} & \sigma_{s(2,3)} & \sigma_{s(2,4)} & \sigma_{s(2,5)} & \sigma_{s(2,6)} \\ \sigma_{s(1,3)} & \sigma_{s(2,3)} & \sigma_{s(3,3)} & \sigma_{s(3,4)} & \sigma_{s(3,5)} & \sigma_{s(3,6)} \\ \sigma_{s(1,4)} & \sigma_{s(2,4)} & \sigma_{s(3,4)} & \sigma_{s(4,4)} & \sigma_{s(4,5)} & \sigma_{s(4,6)} \\ \sigma_{s(1,5)} & \sigma_{s(2,5)} & \sigma_{s(3,5)} & \sigma_{s(4,5)} & \sigma_{s(5,5)} & \sigma_{s(5,6)} \\ \sigma_{s(1,6)} & \sigma_{s(2,6)} & \sigma_{s(3,6)} & \sigma_{s(4,6)} & \sigma_{s(5,6)} & \sigma_{s(6,6)} \end{pmatrix} \tag{A.6}$$

The vector of s-step ahead forecast error of Y_t is given by:

$$e_{t+s} = \begin{pmatrix} e_{11,t+s} & e_{12,t+s} & e_{13,t+s} \\ e_{12,t+s} & e_{22,t+s} & e_{23,t+s} \\ e_{13,t+s} & e_{23,t+s} & e_{33,t+s} \end{pmatrix} = \begin{pmatrix} u_{1,t+s} & 0 & 0 \\ u_{2,t+s} & u_{3,t+s} & 0 \\ u_{4,t+s} & u_{5,t+s} & u_{6,t+s} \end{pmatrix} \begin{pmatrix} u_{1,t+s} & u_{2,t+s} & u_{4,t+s} \\ 0 & u_{3,t+s} & u_{5,t+s} \\ 0 & 0 & u_{6,t+s} \end{pmatrix}$$

$$= \begin{pmatrix} u^2_{1,t+s} & u_{1,t+s}u_{2,t+s} & u_{1,t+s}u_{4,t+s} \\ u_{1,t+s}u_{2,t+s} & u^2_{2,t+s} + u^2_{3,t+s} & u_{2,t+s}u_{4,t+s} + u_{3,t+s}u_{5,t+s} \\ u_{1,t+s}u_{4,t+s} & u_{2,t+s}u_{4,t+s} + u_{3,t+s}u_{5,t+s} & u^2_{4,t+s} + u^2_{5,t+s} + u^2_{6,t+s} \end{pmatrix} \tag{A.7}$$

As can be seen from Equation (A.7), $\sigma^*_{t,ij} \equiv E_t[e_{ij,t+s}] \neq 0$, and therefore the s-step ahead forecast of $Y_{ij,t}$ has to be corrected element by element by $\sigma^*_{t,ij}$ according to the following expression:

$$\sigma^*_{t,ij} = E[e_{ij,t+s}] = \sum_{l=1+\frac{i(i-1)}{2}}^{\frac{i(i+1)}{2}} E[u_{l,t+s}u_{l+\frac{j(j-1)}{2}-\frac{i(i-1)}{2},t+s}] = \sum_{l=1+\frac{i(i-1)}{2}}^{\frac{i(i+1)}{2}} \sigma_{t,s\left(l,l+\frac{j(j-1)}{2}-\frac{i(i-1)}{2}\right)}$$

where $j \geq i$, $i = 1, 2, 3$.
The forecasting procedure described above is similarly derived for any $n \geq 1$ number of stocks.

B Appendix Tables

Table B.1: Summary statistics of 5-minute and daily stock returns

Return		HWP	JPM	HD	IBM
5-min	Mean	-1.6×10^{-7}	-3.7×10^{-7}	-2.4×10^{-6}	-1.2×10^{-6}
	Max.	0.1122	0.0807	0.1085	0.1082
	Min.	-0.1590	-0.1197	-0.1271	-0.1059
	Std.	0.0032	0.0025	0.0024	0.0020
	Skew.	-0.8580	-0.9577	-2.1449	1.5326
	Kurt.	239.3	132.6	256.1	283.7
Daily	Mean	3.5×10^{-5}	-2.8×10^{-5}	-17.4×10^{-5}	-7.2×10^{-5}
	Max.	0.1595	0.1577	0.1188	0.1169
	Min.	-0.2064	-0.1998	-0.1516	-0.1116
	Std.	0.0267	0.0218	0.0210	0.0176
	Skew.	-0.0054	0.0745	-0.2271	0.4579
	Kurt.	10.7	13.4	9.2	10.2

Note: This table reports descriptive statistics of the 5-minute and daily returns for the stocks HWP, JPM HD and IBM over the period from 1^{st} January 2001 to 30^{th} June 2006.

Table B.2: Summary statistics of realized variances and realized covariances of the stocks HWP, JPM HD and IBM

	Realized variances	Realized covariances
Mean	3.03×10^{-3}	1.03×10^{-3}
Maximum	0.01546	0.00404
Minimum	1.17×10^{-5}	82×10^{-5}
Std. dev	0.00063	0.00021
Skewness	13.1	8.6
Kurtosis	297.3	153.1

Note: This table reports the average values of descriptive statistics of the realized covariances and variances across all four stocks. The realized variances and covariances are calculated from 5-minute intraday returns, as described in the main text.

Table B.3: Estimated parameters of the Mincer-Zarnovitz regression, one-step ahead forecasts

		α_{ij}					
		Bias Correction			No Bias Correction		
		Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
1^{st} Ordering	Y_{11}	2.53×10^{-5}	2.39×10^{-5}	2.00×10^{-5}	1.78×10^{-5}	1.64×10^{-5}	1.21×10^{-5}
		(0.201)	(0.227)	(0.307)	(0.392)	(0.432)	(0.557)
	Y_{12}	4.81×10^{-6}	5.35×10^{-6}	3.20×10^{-6}	3.69×10^{-6}	4.32×10^{-6}	2.19×10^{-6}
		(0.355)	(0.295)	(0.543)	(0.487)	(0.407)	(0.683)
	Y_{22}	1.25×10^{-5}	1.25×10^{-5}	3.25×10^{-6}	9.25×10^{-6}	9.31×10^{-6}	1.78×10^{-6}
		(0.195)	(0.192)	(0.759)	(0.351)	(0.345)	(0.868)
2^{nd} Ordering	Y_{11}	1.81×10^{-5}	1.60×10^{-5}	1.54×10^{-5}	1.05×10^{-5}	8.82×10^{-6}	8.24×10^{-6}
		(0.038)	(0.073)	(0.086)	(0.270)	(0.364)	(0.399)
	Y_{12}	1.22×10^{-5}	1.18×10^{-5}	1.04×10^{-5}	6.01×10^{-6}	5.63×10^{-6}	3.60×10^{-6}
		(0.002)	(0.003)	(0.013)	(0.214)	(0.248)	(0.482)
	Y_{22}	5.58×10^{-5}	5.73×10^{-5}	5.07×10^{-5}	3.72×10^{-5}	3.72×10^{-5}	2.98×10^{-5}
		(0.000)	(0.000)	(0.001)	(0.058)	(0.040)	(0.110)
Average	Y_{11}	3.80×10^{-5}	4.00×10^{-5}	3.53×10^{-5}	2.56×10^{-5}	2.64×10^{-5}	2.13×10^{-5}
		(0.037)	(0.020)	(0.042)	(0.209)	(0.174)	(0.277)
	Y_{12}	8.59×10^{-6}	8.57×10^{-6}	6.90×10^{-6}	4.71×10^{-6}	4.79×10^{-6}	2.79×10^{-6}
		(0.058)	(0.058)	(0.140)	(0.353)	(0.341)	(0.594)
	Y_{22}	1.52×10^{-5}	1.41×10^{-5}	9.90×10^{-6}	9.76×10^{-6}	8.91×10^{-6}	5.11×10^{-6}
		(0.095)	(0.127)	(0.307)	(0.316)	(0.363)	(0.616)
		β_{ij}					
		Bias Correction			No Bias Correction		
		Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
1^{st} Ordering	Y_{11}	0.998	1.008	1.049	0.989	0.998	1.039
		[0.702, 1.294]	[0.711, 1.304]	[0.753, 1.345]	[0.695, 1.283]	[0.704, 1.293]	[0.746, 1.333]
	Y_{12}	1.025	1.012	1.120	1.013	0.997	1.098
		[0.706, 1.345]	[0.697, 1.326]	[0.783, 1.456]	[0.701, 1.325]	[0.690, 1.303]	[0.771, 1.425]
	Y_{22}	1.158	1.158	1.334	1.061	1.061	1.186
		[0.882, 1.434]	[0.884, 1.433]	[1.019, 1.649]	[0.811, 1.312]	[0.811, 1.310]	[0.907, 1.465]
2^{nd} Ordering	Y_{11}	1.030	1.056	1.052	0.992	1.013	1.009
		[0.793, 1.266]	[0.814, 1.298]	[0.810, 1.293]	[0.764, 1.220]	[0.781, 1.246]	[0.778, 1.241]
	Y_{12}	0.876	0.903	0.967	0.891	0.920	0.990
		[0.618, 1.135]	[0.638, 1.168]	[0.685, 1.250]	[0.625, 1.156]	[0.647, 1.193]	[0.698, 1.282]
	Y_{22}	0.892	0.857	0.903	0.812	0.792	0.835
		[0.610, 1.174]	[0.610, 1.103]	[0.650, 1.156]	[0.548, 1.075]	[0.557, 1.027]	[0.593, 1.076]
Average	Y_{11}	0.971	0.943	0.981	0.913	0.895	0.931
		[0.679, 1.264]	[0.672, 1.214]	[0.707, 1.254]	[0.632, 1.193]	[0.632, 1.158]	[0.665, 1.196]
	Y_{12}	0.953	0.962	1.044	0.955	0.963	1.046
		[0.667, 1.240]	[0.673, 1.250]	[0.736, 1.352]	[0.667, 1.242]	[0.674, 1.253]	[0.737, 1.355]
	Y_{22}	1.093	1.109	1.179	1.027	1.038	1.093
		[0.838, 1.348]	[0.851, 1.366]	[0.905, 1.452]	[0.789, 1.266]	[0.798, 1.279]	[0.839, 1.346]

Note: For estimated α_{ij}, the p-values are reported in round parenthesis. For estimated β_{ij}, the 95% confidence intervals are reported in the squared parenthesis