CoFE discussion papers ; 07/01

Dynamic Modelling of Large Dimensional Covariance Matrices

Valeri Voev*
University of Konstanz, CoFE

1 February 2007

Abstract

Modelling and forecasting the covariance of financial return series has always been a challenge due to the so-called "curse of dimensionality". This paper proposes a methodology that is applicable in large dimensional cases and is based on a time series of realized covariance matrices. Some solutions are also presented to the problem of non-positive definite forecasts. This methodology is then compared to some traditional models on the basis of its forecasting performance employing Diebold-Mariano tests. We show that our approach is better suited to capture the dynamic features of volatilities and covolatilities compared to the sample covariance based models.

1 Introduction

Modelling and forecasting the variances and covariances of returns of financial assets is crucial for financial management and portfolio selection and re-balancing. Recently this branch of the econometric literature has grown at a very fast pace. One of the simplest methods used is the sample covariance matrix. A stylized fact, however, is that there is a serial dependence in the second moments of returns. Thus, more sophisticated models had to be developed which incorporate this property, as well as other well-known features of financial return distributions such as leptokurtosis or the so-called "leverage effect". This led to the development of the univariate GARCH processes and their extension - the multivariate GARCH (MGARCH) models (for a comprehensive review see Bauwens, Laurent, and Rombouts (2006)), which include also the modelling of covariances. One of the most severe drawbacks of the MGARCH models, however, is the difficulty of handling dimensions higher than 4 or 5 (or with very restrictive assumptions). Another more practically oriented field of research deals with the problem of how to reduce the noise inherent in simpler covariance estimators such as the sample covariance matrix. Techniques have been developed to "shrink" the sample covariance (SC) matrix, thereby reducing its extreme values in order to mitigate the effect of the so-called error maximization noted by Michaud (1989). One of the shrinkage estimators used

*Financial support from the German Science Foundation, Research Group "Preis-, Liquiditäts- und Kreditrisiken: Messung und Verteilung" is gratefully acknowledged. I would like to thank Winfried Pohlmeier, Michael Lechner, Jens Jackwerth and Günter Franke for helpful comments. Address for correspondence: PBox D-124, University of Konstanz, 78457 Konstanz; Tel. 0049 7531 88 2204; E-mail: valeri.voev@uni-konstanz.de

among practitioners is the Black-Litterman model (Black and Litterman (1992)). This model uses a prior which reflects an investor's beliefs about securities returns and combines it with implied equilibrium expected returns to obtain a posterior distribution, whose variance is a combination of the covariance matrix of implied returns and the confidence of the investor's views (which reflect the prior covariance). Further, Ledoit and Wolf (2003) and Ledoit and Wolf (2004) use shrinkage methods to combine a SC matrix with a more structured estimator (e.g a matrix with equal pairwise correlations, or a factor estimator). The idea is to combine an asymptotically unbiased estimator having a large variance with a biased estimator, which is considerably less noisy. So the shrinkage actually amounts to optimizing in terms of the well-known trade-off between bias and variance.

Recently, with the availability of high-quality transaction databases, the technique of realized variance and covariance (RC) gained popularity. A very comprehensive treatment of volatility modelling with focus on forecasting appears in Andersen, Bollerslev, Christoffersen, and Diebold (2006). Andersen, Bollerslev, Diebold, and Ebens (2001), among others, have shown that there is a long-range persistence (long memory) in daily realized volatilities, which allows one to obtain good forecasts by means of fractionally integrated ARMA processes. At the monthly level, we find that the autocorrelations decline quite quickly to zero, which led us to choose standard ARMA models for fitting and forecasting.

The aim of this paper is to compare the forecasting performance of a set of models, which are suitable to handle large dimensional covariance matrices. Letting H denote the set of considered models, we have $H = \{s, ss, rm, rc, src, drc, dsrc\}$, where the first two models are based on the sample covariance matrix, the third model is a RiskMetrics™ exponentially weighted moving average (EWMA) estimator developed by J.P. Morgan (1996), the fourth and the fifth represent simple forecasts based on the realized and on the shrunk realized covariance matrix, and the last two models employ dynamic modelling of the RC and shrunk RC, respectively. We judge the performance of the models by looking at their ability to forecast individual variance and covariance series by employing a battery of Diebold-Mariano (Diebold and Mariano (1995)) tests. Of course, if we have good forecasts for the individual series, then the whole covariance matrix will also be well forecast. The practical relevance of a good forecast can be seen by considering an investor who faces an optimization problem to determine the weights of some portfolio constituents. One of the crucial inputs in this problem is a forecast of future movements and co-movements in asset returns. Our contribution is to propose a methodology which improves upon the sample covariance estimator and is easy to implement even for very large portfolios. We show that in some sense these models are more flexible than the MGARCH models, although this comes at the expense of some complications.

The remainder of the paper is organized as follows: Section 2 sets up the notation and describes the forecasting models, Section 3 presents the data set used to compare the forecasting performance of the models, Section 4 discusses the results on the forecast evaluation and Section 5 concludes the paper.

2 Forecasting models

In this section we describe each of the covariance forecasting models. First, we introduce some notation and description of the forecasting methodology. We concentrate on one-step

ahead forecasts of covariance matrices of N stocks, and consider the monthly frequency. The information is updated every period and a new forecast is formed. Thus, each new forecast incorporates the newest information which has become available. Such a strategy might describe an active long-run investor, who revises and rebalances her portfolio every month. Let the multivariate price process be defined as $\mathbf{P} = \{\mathbf{P}_t(\omega), t \in (-\infty, \infty), \omega \in \Omega\}$, where Ω is an outcome space.[1] The portfolio is set up at $t = 0$ and updated at each $t = 1, 2, \ldots, \bar{T}$, where $\bar{T}$ is the end of the investment period. The frequency of the observations in our application is daily, which we refer to as intra-periods. In this setup, we can formally define the information set at each time $t \geq 0$ as a filtration $\mathcal{F}_t = \sigma(\mathbf{P}_s(\omega), s \in \mathcal{T})$ generated by $\mathbf{P}$, with $\mathcal{T} = \{s : s = -L + \frac{j}{M}, j = 0, 1, \ldots, (L+t)M\}$, M – the number of intra-periods within each period[2] and L – the number of periods, for which price data is available, before the investment period. It is important to note that not all information is considered in the forecasts based on the sample covariance matrix. For these models only the lower frequency monthly sampling is needed. Furthermore, we define the monthly returns as $\mathbf{r}_t = \ln(\mathbf{P}_t) - \ln(\mathbf{P}_{t-1})$, where $\mathbf{P}_t$ is the realization of the price process at time t, and the j^{th} intra-period return by $\mathbf{r}_{t+\frac{j}{M}} = \ln\left(\mathbf{P}_{t+\frac{j}{M}}\right) - \ln\left(\mathbf{P}_{t+\frac{j-1}{M}}\right)$. The realized covariance at time $t+1$ is given by:

$$\mathbf{\Sigma}_{t+1}^{RC} = \sum_{j=1}^{M} \mathbf{r}_{t+\frac{j}{M}} \mathbf{r}'_{t+\frac{j}{M}}. \tag{1}$$

Assessing the performance of variance forecasts has been quite problematic, since the true covariance matrix $\mathbf{\Sigma}_t$ is not directly observable. This has long been a hurdle in evaluating GARCH models. Traditionally, the squared daily return was used as a measure of the daily variance. Although this is an unbiased estimator, it has a very large estimation error due to the large idiosyncratic noise component of daily returns. Thus a good model may be evaluated as poor, simply because the target is measured with a large error. In an important paper, Andersen and Bollerslev (1998) showed that GARCH models actually provide good forecasts when the target to which they are compared is estimated more precisely, by means of sum of squared intradaily returns. Since then, it has become a practice to take the realized variance as the relevant measure for comparing forecasting performance. In this spirit, we use the realized monthly covariance in place of the true matrix. Thus we will assess a given forecast $\hat{\mathbf{\Sigma}}_{t+1|t}^{(h)}$, $h \in H$ by its deviation from $\mathbf{\Sigma}_{t+1}^{RC}$.

2.1 A sample covariance forecast

In this section we describe a forecasting strategy based on the sample covariance matrix, which will serve as a benchmark. The sample covariance is a consistent estimator for the true population covariance under weak assumptions. We use a rolling window scheme and define

[1] Of course, in reality the price process could not have started in the infinite past. Since we are interested in when the process became observable, and not in its beginning, we leave the latter unspecified.

[2] This number is not necessarily the same for all periods and should be denoted more precisely by $M(t)$. This is not done in the text to avoid cluttering of the notation.

the forecast as:

$$\hat{\mathbf{\Sigma}}^{(s)}_{t+1|t} = \frac{1}{T} \sum_{s=t-T+1}^{t} (\mathbf{r}_s - \bar{\mathbf{r}}_{t,T})(\mathbf{r}_s - \bar{\mathbf{r}}_{t,T})', \tag{2}$$

where for each t, $\bar{\mathbf{r}}_{t,T}$ is the sample mean of the return vector $\mathbf{r}$ over the last T observations. We will denote the sample covariance matrix at time t by $\mathbf{\Sigma}^{SC}_t$. For T we choose a value of 60, which with monthly data corresponds to a time span of five years. As the near future is of the highest importance in volatility forecasting, this number might seem too large. Too small a number of periods, however, would lead to a large variance of the estimator, therefore other authors (e.g. Ledoit and Wolf (2004)) have also chosen 60 months as a balance between precision and relevance of the data. A problem of this approach, as simple as it is, is that new information is given the same weight as very old information. Another obvious oversimplification is that we do not account for the serial dependence present in the second moments of financial returns.

2.2 A shrinkage sample covariance forecast

In this section we briefly present the shrinkage estimator, proposed by Ledoit and Wolf (2003), in order to give an idea of the shrinkage principle.

The shrinkage estimator of the covariance matrix $\mathbf{\Sigma}_t$ is defined as a weighted linear combination of some shrinkage target F_t and the sample covariance matrix, where the weights are chosen in an optimal way. More formally, the estimator is given by

$$\mathbf{\Sigma}^{SS}_t = \hat{\alpha}^*_t F_t + (1 - \hat{\alpha}^*_t)\mathbf{\Sigma}^{SC}_t, \tag{3}$$

$\hat{\alpha}^*_t \in [0, 1]$ is an estimate of the optimal shrinkage constant α^*_t.

The shrinking intensity is chosen to be optimal with respect to a loss function defined as a quadratic distance between the true and the estimated covariance matrices based on the Frobenius norm. The Frobenius norm of an $N \times N$ symmetric matrix Z with elements $(z_{ij})_{i,j=1,\ldots,N}$ is defined by

$$\|Z\|^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} z^2_{ij}. \tag{4}$$

The quadratic loss function is the Frobenius norm of the difference between $\mathbf{\Sigma}^{SS}_t$ and the true covariance matrix:

$$L(\alpha_t) = \left\| \alpha_t F_t + (1 - \alpha_t)\mathbf{\Sigma}^{SC}_t - \mathbf{\Sigma}_t \right\|^2. \tag{5}$$

The optimal shrinkage constant is defined as the value of α which minimizes the expected

value of the loss function (i.e. the risk) in expression (5):

$$\alpha_t^* = \underset{\alpha_t}{\operatorname{argmin}} \operatorname{E}\left[L(\alpha_t)\right]. \tag{6}$$

For an arbitrary shrinkage target F and a consistent covariance estimator S, Ledoit and Wolf (2003) show that

$$\alpha^* = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\operatorname{Var}\left[s_{ij}\right] - \operatorname{Cov}\left[f_{ij}, s_{ij}\right]\right)}{\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\operatorname{Var}\left[f_{ij} - s_{ij}\right] + (\phi_{ij} - \sigma_{ij})^2\right)}, \tag{7}$$

where f_{ij} is a typical element of the sample shrinkage target, s_{ij} – of the covariance estimator, σ_{ij} – of the true covariance matrix, and ϕ_{ij} – of the population shrinkage target Φ. Further they prove that this optimal value is asymptotically constant over T and can be written as[3]:

$$\kappa_t = \frac{\pi_t - \rho_t}{\nu_t}. \tag{8}$$

In the formula above, π_t is the sum of the asymptotic variances of the entries of the sample covariance matrix scaled by $\sqrt{T}$: $\pi_t = \sum_{i=1}^{N}\sum_{j=1}^{N} \operatorname{AVar}\left[\sqrt{T}s_{ij,t}\right]$, ρ_t is the sum of asymptotic covariances of the elements of the shrinkage target with the elements of the sample covariance matrix scaled by $\sqrt{T}$: $\rho_t = \sum_{i=1}^{N}\sum_{j=1}^{N} \operatorname{ACov}\left[\sqrt{T}f_{ij,t}, \sqrt{T}s_{ij,t}\right]$, and ν_t measures the misspecification of the shrinkage target: $\nu_t = \sum_{i=1}^{N}\sum_{j=1}^{N}(\phi_{ij,t} - \sigma_{ij,t})^2$. Following their formulation and assumptions, $\sum_{i=1}^{N}\sum_{j=1}^{N} \operatorname{Var}\left[\sqrt{T}(f_{ij} - s_{ij})\right]$ converges to a positive limit, and so $\sum_{i=1}^{N}\sum_{j=1}^{N} \operatorname{Var}\left[f_{ij} - s_{ij}\right] = \mathrm{O}(1/T)$. Using this result and the $\sqrt{T}$ convergence in distribution of the elements of the sample covariance matrix, Ledoit and Wolf (2003) show that the optimal shrinkage constant is given by:

$$\alpha_t^* = \frac{1}{T}\frac{\pi_t - \rho_t}{\nu_t} + O\left(\frac{1}{T^2}\right). \tag{9}$$

Since α^* is unobservable, it has to be estimated. Ledoit and Wolf (2004) propose a consistent estimator of α^* for the case where the shrinkage target is a matrix in which all pairwise correlations are equal to the same constant. This constant is the average value of all pairwise correlations from the sample correlation matrix. The covariance matrix resulting from combining this correlation matrix with the sample variances, known as the equicorrelated matrix, is the shrinkage target. The equicorrelated matrix is a sensible shrinkage target as it involves only a small number of free parameters (hence less estimation noise). Thus the elements of the sample covariance matrix, which incorporate a lot of estimation error and hence can take rather extreme values are "shrunk" towards a much less noisy average. Using

[3] In their paper the formula appears without the subscript t. By adding it here we want to emphasize that these variables are changing over time.

the equicorrelated matrix as the shrinkage target F_t in equation (3) the forecast is given by

$$\hat{\mathbf{\Sigma}}_{t+1|t}^{(ss)} = \mathbf{\Sigma}_t^{SS}. \tag{10}$$

2.3 A RiskMetrics™ forecast

The RiskMetrics™ forecasting methodology is a modification of the sample covariance matrix, in which observations which are further in the past are given exponentially smaller weights, determined by a factor λ. For the generic $(i,j), i,j = 1,\ldots,N$ element of the EWMA covariance matrix $\mathbf{\Sigma}_t^{RM}$ we have:

$$\sigma_{ij,t}^{RM} = (1-\lambda)\sum_{s=1}^{t} \lambda^{s-1} \left(r_{i,s} - \bar{r}_i\right)\left(r_{j,s} - \bar{r}_j\right), \tag{11}$$

where $\bar{r}_i = \frac{1}{t}\sum_{s=1}^{t} r_{i,s}$. Again, the forecast is given by:

$$\hat{\mathbf{\Sigma}}_{t+1|t}^{(rm)} = \mathbf{\Sigma}_t^{RM}. \tag{12}$$

Methods to choose the optimal λ are discussed in J.P. Morgan (1996). In this paper we set $\lambda = 0.97$, the value used by J.P. Morgan for monthly (co)volatility forecasts. Note that contrary to the sample covariance matrix, for which we use a rolling window scheme, in the RiskMetrics approach we use at each t all the available observations from the beginning of the observation period up to t. Since in the RiskMetrics approach the weights decrease exponentially, the observations which are further away in the past are given relatively smaller weights and hence do not influence the estimate as much as in the sample covariance matrix.

2.4 A simple realized covariance forecast

The realized covariance estimator was already defined in expression (1). Its univariate and multivariate properties have been studied among others, by Barndorff-Nielsen and Shephard (2004) and by Andersen, Bollerslev, Diebold, and Labys (2003). In the limit, when $M \to \infty$, Barndorff-Nielsen and Shephard (2004) have shown that realized covariance is an error-free measure for the integrated covariation of a very broad class of stochastic volatility models. In the empirical part we compute monthly realized covariance by using daily returns (see also French, Schwert, and Stambaugh (1987)). The simple forecast is defined by:

$$\hat{\mathbf{\Sigma}}_{t+1|t}^{(rc)} = \mathbf{\Sigma}_t^{RC}. \tag{13}$$

Thus an investor who uses this strategy simply computes the realized covariance at the end of each month and then uses it as his best guess about the true covariance matrix of the next month. A nice feature of this method is that it only uses recent information which is of most value for the forecast but imposes a very simple and restrictive time dependence. Practically equation (13) states that all variances and covariances follow a random walk process. However,

as we shall see later, the estimated series of monthly variances and covariances show weak stationarity.

2.5 A shrinkage realized covariance forecast

Although the estimator discussed in the previous section is asymptotically error-free, in practice one cannot record observations continuously. A much more serious problem is the fact that at very high frequencies, the martingale assumption needed for the convergence of the realized covariances to the integrated covariation is no longer satisfied. At trade-by-trade frequencies, market microstructure affects the price process and results in microstructure noise induced autocorrelations in returns and hence biased variance estimates. Methods to account for this bias and correct the estimates have been developed by Hansen and Lunde (2006), Oomen (2005), Aït-Sahalia, Mykland, and Zhang (2005), Bandi and Russell (2005), Zhang, Mykland, and Aït-Sahalia (2005), and Voev and Lunde (2007), among others. At low frequencies the impact of market microstructure noise can be significantly mitigated, but this comes at the price of higher variance of the estimator. Since we are using daily returns, market microstructure is not an issue. Thus we will suggest a possible way to reduce variance. Again as in Section 2.2, we will try to find a compromise between bias and variance applying the shrinkage methodology. The estimator looks very much like the one in expression (3). In this case we have:

$$\mathbf{\Sigma}_t^{SRC} = \hat{\alpha}_t^* F_t + (1 - \hat{\alpha}_t^*)\mathbf{\Sigma}_t^{RC}, \tag{14}$$

where now F_t is the equicorrelated matrix, constructed from the realized covariance matrix $\mathbf{\Sigma}_t^{RC}$ in the same fashion as the equicorrelated matrix constructed from the sample covariance matrix, as explained in Section 2.2. Similarly to the previous section, the forecast is simply

$$\hat{\mathbf{\Sigma}}_{t+1|t}^{(src)} = \mathbf{\Sigma}_t^{SRC}. \tag{15}$$

Since the realized covariance is a consistent estimator, we can still apply formula (7) taking into account the different rate of convergence. In order to compute the estimates for the variances and covariances, we need a theory for the distribution of the realized covariance, which is developed in Barndorff-Nielsen and Shephard (2004), who provide asymptotic distribution results for the realized covariation matrix of continuous stochastic volatility semimartingales ($\mathcal{SVSM}^c$). Assuming that the log price process $\ln \mathbf{P} \in \mathcal{SVSM}^c$, we can decompose it as $\ln \mathbf{P} = a^* + m^*$, where a^* is a process with continuous finite variation paths and m^* is a local martingale. Furthermore, under the condition that m^* is a multivariate stochastic volatility process, it can be defined as $m^*(t) = \int_0^t \Theta(u)\mathrm{d}w(u)$, where Θ is the spot covolatility process and w is a vector standard Brownian motion. Then the spot covariance is defined as:

$$\Sigma(t) = \Theta(t)\Theta(t)', \tag{16}$$

assuming that (for all $t < \infty$)

$$\int_0^t \Sigma_{kl}(u)\mathrm{d}u < \infty, \quad k, l = 1, \ldots, N, \tag{17}$$

where $\Sigma_{kl}(t)$ is the (k,l) element of the $\Sigma(t)$ process. Having laid this notation we will now interpret the "true" covariance matrix as:

$$\mathbf{\Sigma}_{t+1} = \int_t^{t+1} \Sigma(u)\mathrm{d}u. \tag{18}$$

Thus the covariance matrix at time $t+1$ is the increment of the integrated covariance matrix of the continuous local martingale from time t to time $t+1$. The realized covariance as defined in expression (1) consistently estimates $\mathbf{\Sigma}_{t+1}$ as given in equation (18). Furthermore, Barndorff-Nielsen and Shephard (2004) show that under a set of regularity conditions the realized covariation matrix follows asymptotically, as $M \to \infty$, the normal law with $N \times N$ matrix of means $\int_t^{t+1} \Sigma(u)\mathrm{d}u$. The asymptotic covariance of

$$\sqrt{M}\left\{\mathbf{\Sigma}_{t+1}^{RC} - \int_t^{t+1} \Sigma(u)\mathrm{d}u\right\}$$

is Ω_{t+1}, a $N^2 \times N^2$ array with elements

$$\Omega_{t+1} = \left\{\int_t^{t+1} \left\{\Sigma_{kk'}(u)\Sigma_{ll'}(u) + \Sigma_{kl'}(u)\Sigma_{lk'}(u)\right\} \mathrm{d}u\right\}_{k,k',l,l'=1,\ldots,N}.$$

Of course, this matrix is singular due to the equality of the covariances in the integrated covariance matrix. This can easily be avoided by considering only its unique lower triangular elements, but for our purposes it will be more convenient to work with the full matrix. The result above is not useful for inference, since the matrix Ω_{t+1} is not known. Barndorff-Nielsen and Shephard (2004) show that a consistent, positive semi-definite estimator is given by a random $N^2 \times N^2$ matrix:

$$H_{t+1} = \sum_{j=1}^{M} x_{j,t+1}x'_{j,t+1} - \frac{1}{2}\sum_{j=1}^{M-1}\left(x_{j,t+1}x'_{j+1,t+1} + x_{j+1,t+1}x'_{j,t+1}\right), \tag{19}$$

where $x_{j,t+1} = vec\left(\mathbf{r}_{t+\frac{j}{M}}\mathbf{r}'_{t+\frac{j}{M}}\right)$ and the *vec* operator stacks the columns of a matrix into a vector. It holds that $MH_{t+1} \xrightarrow{p} \Omega_{t+1}$ with $M \to \infty$.

With the knowledge of this matrix, we can combine the asymptotic results for the realized covariance, with the result in equation (7) to compute the estimates for π_t, ρ_t and ν_t.

For the equicorrelated matrix F we have that[4] $f_{ij} = \bar{r}\sqrt{\sigma_{ii}^{(RC)}\sigma_{jj}^{(RC)}}$, where $\bar{r}$ is the average value of all pairwise correlations, implied by the realized covariance matrix, and $\sigma_{ij}^{(RC)}$ is the (i,j) element of the realized covariance matrix. Thus Φ, the population equicorrelated matrix, has a typical element $\phi_{ij} = \bar{\varrho}\sqrt{\sigma_{ii}\sigma_{jj}}$, where σ_{ij} is the (i,j) of the true covariance matrix Σ and $\bar{\varrho}$ is the average correlation implied by it. Substituting $\sigma_{ij}^{(RC)}$ for s_{ij} in equation

[4] In the following exposition, the time index is suppressed for notational convenience.

(7) and multiplying by M gives for the optimal shrinkage intensity:

$$M\alpha^* = \frac{\sum_{i=1}^N \sum_{j=1}^N \left(\text{Var}\left[\sqrt{M}\sigma_{ij}^{(RC)}\right] - \text{Cov}\left[\sqrt{M}f_{ij}, \sqrt{M}\sigma_{ij}^{(RC)}\right]\right)}{\sum_{i=1}^N \sum_{j=1}^N \left(\text{Var}\left[f_{ij} - \sigma_{ij}^{(RC)}\right] + (\phi_{ij} - \sigma_{ij})^2\right)}. \tag{20}$$

Note that this equation resembles expression (8). The only difference is the scaling by $\sqrt{M}$ instead of $\sqrt{T}$, which is due to the $\sqrt{M}$ convergence. In this case π_t, the first summand in the numerator, is simply the sum of all diagonal elements of Ω_t. By using the definition of the equicorrelated matrix, it can be shown that the second term, ρ_t, can be written as (suppressing the index t):

$$\rho = \sum_{i=1}^N \text{AVar}\left[\sqrt{M}\sigma_{ii}^{(RC)}\right] + \sum_{i=1}^N \sum_{j=1, j\neq i}^N \text{ACov}\left[\sqrt{M}\bar{r}\sqrt{\sigma_{ii}^{(RC)}\sigma_{jj}^{(RC)}}, \sqrt{M}\sigma_{ij}^{(RC)}\right]. \tag{21}$$

Applying the delta method the second term can be expressed as[5]

$$\frac{\bar{r}}{2}\left(\sqrt{\frac{\sigma_{jj}^{(RC)}}{\sigma_{ii}^{(RC)}}}\text{ACov}\left[\sqrt{M}\sigma_{ii}^{(RC)}, \sqrt{M}\sigma_{ij}^{(RC)}\right] + \sqrt{\frac{\sigma_{ii}^{(RC)}}{\sigma_{jj}^{(RC)}}}\text{ACov}\left[\sqrt{M}\sigma_{jj}^{(RC)}, \sqrt{M}\sigma_{ij}^{(RC)}\right]\right).$$

From this expression we see that ρ also involves summing properly scaled terms of the Ω matrix. In the denominator of equation (20), the first term is of order $\text{O}(1/M)$, and the second one is consistently estimated by $\hat{\nu} = \sum_{i=1}^N \sum_{j=1}^N \left(f_{ij} - \sigma_{ij}^{(RC)}\right)^2$.

Since we have a consistent estimator for Ω, we can now also estimate π and ρ. In particular, we have

$$\begin{aligned}
\hat{\pi} &= \sum_{i=1}^N \sum_{j=1}^N h_{ij,ij} \\
\hat{\rho} &= \sum_{i=1}^N h_{ii,ii} + \frac{\bar{r}}{2}\sum_{i=1}^N \sum_{j=1}^N \sqrt{\frac{\sigma_{jj}^{(RC)}}{\sigma_{ii}^{(RC)}}}h_{ii,ij} + \sqrt{\frac{\sigma_{ii}^{(RC)}}{\sigma_{jj}^{(RC)}}}h_{jj,ij},
\end{aligned}$$

where $h_{kl,k'l'}$ is the element of H which estimates the corresponding element of Ω. Thus we can estimate κ_t by $\hat{\kappa}_t = \frac{\hat{\pi}_t - \hat{\rho}_t}{\hat{\gamma}_t}$ and the estimator for the optimal shrinkage constant is:

$$\hat{\alpha}_t^* = \max\left\{0, \min\left\{\frac{\hat{\kappa}_t}{M}, 1\right\}\right\}. \tag{22}$$

[5] cf. Ledoit and Wolf (2004)

The estimated optimal shrinkage constants for our dataset range from 0.0205 to 0.2494 with a mean of 0.0562.

2.6 Dynamic realized covariance forecasts

This model is an alternative to the one in Section 2.4. The most popular models for time varying variances and covariances are the GARCH models. The most significant problem of these models is the large number of parameters in large dimensional systems. The recent DCC models of Tse and Tsui (2002) and Engle (2002) propose a way to mitigate this problem by using the restriction that all correlations obey the same dynamics. Recently Gourieroux, Jasiak, and Sufana (2004) have suggested an interesting alternative – the WAR (Wishart autoregressive) model, which has certain advantages over the GARCH models, e.g. smaller number of parameters, easy construction of non-linear forecasts, simple verification of stationarity conditions, etc. Even quite parsimonious models, however, have a number of parameters of the order $N(N+1)/2$. With $N = 15$ this means more than 120 parameters, which would be infeasible for estimation. We therefore suggest a simple approach in which all variance and covariance series are modelled univariately as ARMA processes and individual forecasts are made, which are then combined into a forecast of the whole matrix. This approach can also be extended by including lags of squared returns which can be interpreted as a kind of ARCH-terms. A theoretical drawback of this model, is that such a methodology does not guarantee the positive definiteness of the forecast matrix. It turns out that this problem could be quite severe, especially if we include functions of lagged returns in the specification. Hence we propose two possible solutions. First, if the above mentioned problem occurs relatively rarely, then in these cases we can define the forecast as in Section 2.4, which would ensure that all forecast matrices are positive definite. More precisely, instead of assuming a random walk process for the realized covariance series (as in Section 2.4) we now model each of them as ARMAX$(p, q, 1)$[6] processes as follows:

$$\sigma_{ij,t}^{(RC)} = \omega + \sum_{s=1}^{p} \varphi_s \sigma_{ij,t-s}^{(RC)} + \sum_{u=0}^{q} \theta_u \varepsilon_{ij,t-u} + \alpha r_{i,t-1} r_{j,t-1}, \tag{23}$$

with $\theta_0 = 1$ and $\varepsilon_{ij,t}$ – a Gaussian white noise process. The model easily extends to an ARMAX(p, q, k) specification with k lags of crossproducts. The parameters φ_s, θ_u and α are estimated by maximum likelihood starting at $t = 100$ and the forecasts $\hat{\sigma}_{ij,t+1|t}^{(RC)}$ are collected in a matrix $\mathbf{\Sigma}_{t+1}^{DRC}$. At time $t+1$ the new information is taken into account and the procedure is repeated. The best model for each series is selected by minimizing the Akaike information criterion (AIC).

In this case the forecast is:

$$\hat{\mathbf{\Sigma}}_{t+1|t}^{(drc)} = \begin{cases} \mathbf{\Sigma}_{t+1}^{DRC}, & \text{if } \mathbf{\Sigma}_{t+1}^{DRC} \text{ is positive definite} \\ \mathbf{\Sigma}_{t}^{RC}, & \text{otherwise.} \end{cases} \tag{24}$$

[6] The last parameter shows the number of lags of the X variable.

A more robust solution is to factorize the sequence of realized covariance matrices into their Cholesky decompositions, model the dynamics and forecast the Cholesky series and then reconstruct the variance and covariance forecasts. This ensures the positive definiteness of the resulting forecast. In this case the Cholesky series are modelled like in equation (23), the forecasts are collected in a lower triangular matrix $\mathbf{C}_{t+1}$ and the covariance forecast is given by:

$$\hat{\mathbf{\Sigma}}_{t+1|t}^{(drc-Chol)} = \mathbf{C}_{t+1}\mathbf{C}_{t+1}'. \tag{25}$$

Analogously, we can use these two strategies to model dynamically the series of shrunk variance covariance matrices which defines the forecasts $\mathbf{\Sigma}_{t+1|t}^{(dsrc)}$ and $\mathbf{\Sigma}_{t+1|t}^{(dsrc-Chol)}$.

3 Data

The data we have used consists of 15 stocks from the current composition of the Dow Jones Industrial Average index from 01.01.1980 to 31.12.2002. The stocks are Alcoa (NYSE ticker symbol: AA), American Express Company (AXP), Boeing Company (BA), Caterpillar Inc. (CAT), Coca-Cola Company (KO), Eastman Kodak (EK), General Electric Company (GE), General Motors Corporation (GM), Hewlett-Packard Company (HPQ), International Business Machines (IBM), McDonald's Corporation (MCD), Philip Morris Companies Incorporated (MO), Procter & Gamble (PG), United Technologies Corporation (UTX) and Walt Disney Company (DIS). The reason that we have considered only 15 stocks is due to fact that the realized covariance matrices are of full rank only if $M > N$, where M is the number of intra-period observations used to construct the realized covariance, in our case number of daily returns used to construct each monthly realized covariance. Usually there are 21 trading days per month, but some months have had fewer trading days (e.g. September 2001). With intradaily data this problem would not be of importance, since then we can easily have hundreds of observations within a day. Such datasets are already common, but they still do not cover large periods of time. Nevertheless, the dynamic properties of daily realized volatilities, covariances and correlations are studied by e.g. Andersen, Bollerslev, Diebold, and Ebens (2001) and Andersen, Bollerslev, Diebold, and Labys (2001). It has been shown that there is a long-range persistence, which allows for construction of good forecasts by means of ARFIMA processes.

All the stocks are traded on the NYSE and we take the daily closing prices and monthly closing prices to construct corresponding returns. The data is adjusted for splits and dividends. We find the typical properties of financial returns: negative skewness (with the exception of PG), leptokurtosis and non-normality. The average (across stocks) mean daily return is 0.05% and the average daily standard deviation is 1.9 %. From the daily data log monthly returns are constructed by using the opening price of the first trading day of the month and the closing price of the last day. These returns are then used to construct rolling window sample covariance matrices, used in the first two forecasting models.

4 Results

In this section we present and discuss the results on the performance of the forecasting models described in Section 2.

In order to asses the forecasting performance, we employ Diebold-Mariano tests for each of the variance and covariance series. Then we measure the deviation of the forecast as a matrix from its target by using again the Frobenius norm, which gives an overall idea of the comparative performance of the models. Of course, if the individual series are well forecast, so will be the matrix. As a target or "true" covariance matrix, we choose the realized covariance matrix. First, we present some graphical results. Out of the total of 120 variance and covariance forecast series, Figure 1 plots 9 representative cases, for the sample covariance and the RiskMetrics™ model, against the realized series. The name, which appears above each block in the figure, represents either a variance series (e.g. EK), or a covariance one (e.g. GE,AA).

Both forecasts are quite close, and as can be seen, they cannot account properly for the variation in the series. As the tests show, however, the Riskmetrics™ fares better and is the best model among the sample based ones. It is already an acknowledged fact that financial returns have the property of volatility clustering. This feature is also clearly evident in the figure, where periods of low and high volatility can be easily distinguished, which suggests that variances and covariances tend to exhibit positive autocorrelation. Figure 2 shows the autocorrelation functions for the same 9 series of realized (co)variances. The figure clearly shows that there is some positive serial dependence, which usually dies out quickly, suggesting stationarity of the series. Stationarity is also confirmed by running Augmented Dickey-Fuller (ADF) tests, which reject the presence of an unit root in all series at the 1% significance level.

The observed dependence patterns suggest the idea of modelling the variance and covariance series as well as their shrunk versions as ARMA processes. This resulted in a few cases in which the matrix forecast was not positive definite (16 out of 176 for the original series and 8 out of 176 for the shrunk series). Thus the forecast in expression (24) seems to be reasonable and as we shall see later, compares well to the sample covariance based models. In a GARCH framework, the conditional variance equation includes not only lags of the variance, but also lags of squared innovations (shocks). When mean returns are themselves unpredictable (the usual approach is to model the mean equation as an ARMA process), the shock is simply the return. This fact led us to include lags of squared returns (for the variance series) and cross-products (for the covariance series) as in the ARMAX$(p, q, 1)$ model in equation (23). This added flexibility, however, comes at the price of a drastic increase of the non-positive definite forecasts (108 and 96 out of 176, respectively). Thus the forecast in equation (24) comes quite close to the simple realized and shrunk realized covariance models in Sections 2.4 and 2.5, respectively. A solution to this issue is to decompose the matrices into their lower triangular Cholesky factors, forecast the Cholesky series, and then reconstruct the matrix. This leads to the forecasting formula in equation (25), which defines the $drc - Chol$ and $dsrc - Chol$ forecasting models for the simple realized and shrunk realized covariance case, respectively. A drawback of this approach is that the Cholesky series do not have an intuitive interpretation. They are simply used as a tool to constrain the forecasts to satisfy the complicated restrictions implied by the positive definiteness requirement. Another drawback is that the Cholesky decomposition involves nonlinear transformations of the original series.

Figure 1: Comparison of the sample covariance based (Sample) and Riskmetrics™ (RM) forecast against the realized covariance (True).

The dashed line is the upper 95% confidence band.

Figure 2: Autocorrelation functions of the realized variance and covariance series.

Thus, if one can adequately forecast the nonlinear transformation, this does not immediately mean that applying the inverse transformation to the forecast will produce a good forecast of the initial series. So there is a trade-off between the possibility of including more information in the forecast and obtaining positive definite matrices on the one hand, and the distortions caused by the non-linearity of the transformation on the other. It turns out that in our case the beneficial effects outweigh the negative ones. Figure 3 shows the $drc - Chol$ and the RiskMetrics™ forecast for the same 9 variance and covariance series. From the figure it is evident that the dynamic forecasts track the true series much closer than the RiskMetrics™ forecasts, especially at the end of the period when the (co)volatilities were more volatile. The $dsrc - Chol$ forecast looks quite similar to the $drc - Chol$ (due to the usually small shrinkage constants), but as we shall see later the forecasts are in fact somewhat better.

Turning to the statistical comparison of the forecasting methods, we first briefly present the Diebold-Mariano testing framework as in Harvey, Leyborne, and Newbold (1997). Suppose a pair of l-step ahead forecasts h_1 and h_2, $h_1, h_2 \in H$ have produced errors (e_{1t}, e_{2t}), $t = 1, \ldots, T$. The null hypothesis of equality of forecasts is based on some function $g(e)$ of the forecast errors and has the form $\mathrm{E}\left[g(e_{1t}) - g(e_{2t})\right] = 0$. Defining the loss differential $d_t = g(e_{1t}) - g(e_{2t})$ and its average $\bar{d} = T^{-1} \sum_{t=1}^{T} d_t$, the authors note that "the series d_t is likely to be autocorrelated. Indeed, for optimal l-steps ahead forecasts, the sequence of forecast errors follows a moving average process of order $(l-1)$. Thus result can be expected to hold approximately for any reasonably well-conceived set of forecasts." Consequently, it can be shown that the variance of $\bar{d}$ is, asymptotically,

$$\mathrm{Var}\left[\bar{d}\right] \approx T^{-1} \left[\gamma_0 + 2 \sum_{k=1}^{l-1} \gamma_k\right], \tag{26}$$

where γ_k is the k^{th} autocovariance of d_t. The Diebold-Mariano test statistic is:

$$S_1 = \left[\widehat{\mathrm{Var}}\left[\bar{d}\right]\right]^{-1/2} \bar{d}, \tag{27}$$

where $\widehat{\mathrm{Var}}\left[\bar{d}\right]$ is obtained from equation (26) by substituting for γ_0 and γ_k the sample variance and autocovariances of d_t, respectively. Tests are then based on the asymptotic normality of the test statistic. Noting that we only consider 1-step ahead forecasts in this paper, the series d_t should not be autocorrelated. As already noted above, this is expected to hold for any *reasonably* constructed forecasts. Actually, however, the sample based forecasts are not really *reasonable* in the sense that they do not account for the serial dependence of the process they are supposed to forecast. Thus, the degree of autocorrelation in the d_t series, when either h_1 or h_2 is a sample based forecast, will correspond to the degree of dependence in the series to be forecast. For this reason, ignoring autocovariances in the construction of the Diebold-Mariano tests will lead to an error in the test statistic. To correct for this we include in $\widehat{\mathrm{Var}}\left[\bar{d}\right]$ the first k significant autocorrelations for each of the 120 series.

Table 1 summarizes the results of the Diebold-Mariano tests carried out pairwise between all models for all 120 series. The first entry in each cell of the table shows the number of series (out of 120) for which the model in the corresponding column outperforms the model

Figure 3: Comparison of the Riskmetrics™ forecast (RM) and the dynamic realized covariance forecast based on Cholesky series (DRC-Chol) against the realized covariance (True).

Table 1: Results from the Diebold-Mariano tests.

	s	*ss*	*rm*	*rc*	*src*	*drc*	*dsrc*	*drc*− *Chol*	*dsrc*− *Chol*
s	-	**85/28**	106/50	14/1	16/1	47/20	89/37	93/49	100/55
ss	**20/0**	-	106/47	14/1	16/1	47/20	89/37	92/49	100/55
rm	14/0	14/0	-	7/1	11/1	37/7	73/29	85/33	89/37
rc	106/60	106/61	113/69	-	**105/86**	119/59	120/88	115/80	117/88
src	104/55	104/56	109/69	**0/0**	-	119/50	120/86	114/77	117/85
drc	73/12	73/12	83/26	1/0	1/0	-	**104/31**	98/47	103/58
dscr	31/3	31/3	47/8	0/0	0/0	**1/0**	-	69/28	83/35
drc (Chol)	27/8	28/8	35/10	5/1	6/1	22/7	51/12	-	91/19
dsrc (Chol)	20/7	20/7	31/8	3/1	3/1	17/6	37/11	29/3	-

Note: Due to the definition of the shrinkage target, the first numbers in the pairs highlighted in bold do not sum up to 120, since the variance series are unchanged in their respective "shrunk" versions. Thus, in these cases there are only 105 series forecasts to be compared.

in the corresponding row. The second entry corresponds to the number of significant outperformances according to the Diebold-Mariano tests at the 5% significance level. Hence, the table is in a sense symmetric, as the number of times model h_1 outperforms model h_2 plus the number of times model h_2 outperforms model h_1 (given by the first number in each cell) sum up to 120 – the total number of series. This is not the case, only for the pairs highlighted in bold, because the 15 variance series are unchanged in their respective "shrunk" versions.[7] Thus, in these cases there are only 105 covariance series forecasts to be compared.

At first glance one can notice that the worst performing models are the *rc* and *src* models. Among the sample based forecasts the RiskMetrics™ is the one which delivers the best performances. The comparison between the sample and the shrinkage sample forecasts shows that shrinking has indeed improved upon the sample covariance matrix. This holds also for the realized covariance matrix. Here, the result is reinforced by the fact that shrinking also increases the probability of obtaining a positive definite forecast. In fact, the quite poor performance of the *drc* model is not due to the poor forecasting of the series themselves, but due to the large error, introduced by taking the previous realized covariance matrix, in case of a non-positive definite forecast (see equation (24)). Even though this only happens in 16 out of 176 cases, it is enough to distort the forecast considerably. The main result of this paper, however, arises from the comparison of the dynamic models with the sample based ones, which can be drawn by considering the last three columns of the table. For most of the series the dynamic models provide better forecasts, which results in smaller errors in the covariance matrix forecasts, as will be shown later. Despite the fact that the number of significant outperformances is not strikingly high (due to the small number of periods for evaluation), it is still clear that the dynamic models outperform decisively even the best model among the sample based ones. Furthermore, as noted earlier, the forecasts using the Cholesky decomposition appear to be better compared to those which model the variance and covariance series directly. This result comes mainly as a consequence of the considerable explanatory power of the lagged shocks in addition to the lagged (co)variances, which could not have been utilized had not we assured the positive definiteness of the forecasts.

[7] By shrinking towards the equicorrelated matrix, the variances do not change.

Table 2: Root mean squared prediction errors based on the Frobenius norm.

RMSPE^{s}	0.06021
RMSPE^{ss}	0.06016
RMSPE^{rm}	0.05887
RMSPE^{rc}	0.06835
RMSPE^{src}	0.06766
RMSPE^{drc}	0.06004
RMSPE^{dsrc}	0.05749
$\text{RMSPE}^{src-Chol}$	0.05854
$\text{RMSPE}^{dsrc-Chol}$	0.05799

In order to understand better the benefits from modelling the variance and covariance series dynamically, we shall consider an alternative (but closely related) measure of forecasting error. In section 2.2 it was shown how the Frobenuis norm can be used as a measure of distance between two matrices. Here we will utilize this concept again by considering the following definition of the forecast error in terms of a matrix forecast:

$$e_t^{(h)} = \left\| \hat{\mathbf{\Sigma}}_{t|t-1}^{(h)} - \mathbf{\Sigma}_t^{RC} \right\|^2, \quad h \in H. \tag{28}$$

The root mean squared prediction errors (RMSPE) are collected in Table 2. The ranking of the models according to this table is quite similar to the one following from Table 1. The only difference is that now the $dsrc$ model appears to be somewhat better than the $dsrc - Chol$, which is most probably due to chance, since as we saw earlier the latter model forecasts most of the series better. As a conclusion, we can state again that in general, the dynamic models outperform the sample covariance based ones.

5 Conclusion

Volatility forecasting is crucial for portfolio management, option pricing and other fields of financial economics. Starting with Engle (1982) a new class of econometric models was developed to account for the typical characteristics of financial returns volatility. This class of models grew rapidly and numerous extensions were proposed. In the late 1980's these models were extended to handle not only volatilities, but also covariance matrices. The main practical problem of these models is the large number of parameters to be estimated, if one decides to include more than a few assets in the specification. Partial solutions to this "curse of dimensionality" were proposed, which imposes restrictions on the system dynamics. Still, modelling and forecasting return covariance matrices remains a challenge. This paper proposes a methodology which is more flexible than the traditional sample covariance based models and at the same time is capable of handling a large number of assets. Although conceptually this methodology is more elaborate than the above mentioned traditional models, it is easily applicable in practice and actually requires shorter historical samples, but with a higher frequency. The gains come from the fact that with high-frequency observations, the latent volatility comes close to being observable. This enables the construction of realized variance and covariance series, which can be modelled and forecast on the basis of their dynamic properties. Additionally, we show that shrinking, which has been shown to improve upon the sample covariance

matrix, can also be helpful in reducing the error in the realized covariance matrices. A practical drawback which appears in this framework is that the so constructed forecasts are not always positive definite. One possible solution to this is to use the Cholesky decomposition as a method of incorporating the positive definiteness requirement in the forecast.

The paper shows that on the monthly frequency, this approach produces better forecasts based on results from Diebold-Mariano tests. The possible gains from a better forecast are, e.g., construction of mean-variance efficient portfolios. Providing a more accurate forecast of future asset comovements will result in better balanced portfolios. These gains will be most probably higher and more pronounced if intradaily returns are used for the construction of daily realized covariance matrices, which remains a possible avenue for further research. It has been shown (e.g. by Andersen, Bollerslev, Diebold, and Ebens (2001)) that realized daily volatilities and correlations exhibit high persistence. Since by incorporating intra-daily information these realized measures are also quite precise, this serial dependence can be exploited for volatility forecasting. A possible extension of the methodological framework suggested in the paper could be modelling the realized series in a vector ARMA system, in order to analyze volatility spillovers across stocks, industries or markets, which however would again involve a large number of parameters.

A closely related area of research is concerned with the methods for evaluation of covariance matrix forecasts. In this paper we have used purely statistical evaluation tools based on a symmetric loss function. An asymmetric measure in this case may have more economic meaning, since it is quite plausible to assume that if a portfolio variance has been overestimated, the consequences are less adverse than if it has been underestimated. In a multivariate context Byström (2002) uses as an evaluation measure of forecasting performance the profits generated by a simulated trading of portfolio of rainbow options. The prices of such options depend on the correlation between the underlying assets. Thus the agents who forecast the correlations more precisely should have higher profits on average.

Further, the models presented in this paper can be extended by introducing the possibility of asymmetric reaction of (co)volatilities to previous shocks (leverage). This can be achieved by introducing some kind of asymmetry in equation (23), e.g., by including products of absolute shocks or products of indicator functions for positivity of the shocks.

References

Aït-Sahalia, Y., P. A. Mykland, and L. Zhang (2005): "How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise," *Review of Financial Studies*, 18(2), 351–416.

Andersen, T., T. Bollerslev, P. F. Christoffersen, and F. X. Diebold (2006): "Volatility Forecasting," in *Handbook of Economic Forecasting*, ed. by G. Elliott, C. Granger, and A. Timmermann, chap. 15. Elsevier, 1 edn.

Andersen, T. G., and T. Bollerslev (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 (2001): "The Distribution of Exchange Rate Volatility," *Journal of the American Statistical Association*, 96, 42–55.

——— (2003): "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579–625.

Bandi, F. M., and J. R. Russell (2005): "Microstructure Noise, Realized Volatility, and Optimal Sampling," Working paper, Graduate School of Business, The University of Chicago.

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.

Bauwens, L., S. Laurent, and J. Rombouts (2006): "Multivariate GARCH Models: a Survey," *Journal of Applied Econometrics*, 21, 79–109.

Black, F., and R. Litterman (1992): "Global Portfolio Optimization," *Financial Analysts Journal*, 48(5), 28–43.

Byström, H. (2002): "Using Simulated Currency Rainbow Options to Evaluate Covariance Matrix Forecasts," *Journal of International Financial Markets, Institutions and Money*, 12, 216–230.

Diebold, F. X., and R. S. Mariano (1995): "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13(3), 253–263.

Engle, R. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1007.

——— (2002): "Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroscedasticity Models," *Journal of Business and Economic Statistics*, 20, 339–350.

French, K. R., G. W. Schwert, and R. F. Stambaugh (1987): "Expected Stock Returns and Volatility," *Journal of Financial Economics*, 19, 3–29.

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

Hansen, P. R., and A. Lunde (2006): "Realized Variance and Market Microstructure Noise," *Journal of Business and Economic Statistics*, 24, 127–218.

Harvey, D., S. Leyborne, and P. Newbold (1997): "Testing the Equality of Prediction Mean Squared Errors," *International Journal of Forecasting*, 13, 281–291.

Ledoit, O., and M. Wolf (2003): "Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection," *Journal of Empirical Finance*, 10(5), 603–621.

——— (2004): "Honey, I Shrunk the Sample Covariance Matrix," *Journal of Portfolio Management*, 31, 110–119.

Michaud, R. O. (1989): "The Markowitz Optimization Enigma: Is 'Optimized' Optimal?," *Financial Analysts Journal*, 45(1), 31–42.

Oomen, R. C. A. (2005): "Properties of Bias-Corrected Realized Variance under Alternative Sampling Schemes," *Journal of Financial Econometrics*, 3, 555–577.

Tse, Y., and A. Tsui (2002): "A Multivariate Generalized Auto-Regressive Conditional Heteroscedasticity Model with Time-Varying Correlations," *Journal of Business and Economic Statistics*, 20, 351–362.

Voev, V., and A. Lunde (2007): "Integrated Covariance Estimation using High-frequency Data in the Presence of Noise," *Journal of Financial Econometrics*, 5, 68–104.

Zhang, L., P. A. Mykland, and Y. Aït-Sahalia (2005): "A tale of two time scales: Determining integrated volatility with noisy high frequency data," *Journal of the American Statistical Association*, 100, 1394–1411.