Erschienen in: Journal of Financial Econometrics ; 14 (2016), 2. - S. 383-417
https://dx.doi.org/10.1093/jjfinec/nbu031

Forecasting Covariance Matrices: A Mixed Approach

Roxana Halbleib[1] and Valeri Voev[2]

[1]University of Konstanz, WIN Koleg Zukunftskolleg, Graduate School of Decision Sciences and [2]Aarhus University, CREATES and Siemens Wind Power A/S

Address correspondence to Roxana Halbleib, Department of Economics, University of Konstanz, Univer sitaetsstrasse 10, PO Box 124, 78464, Konstanz, Germany, or e mail: roxana.halbleib@uni konstanz.de.

Abstract

In this article, we introduce a new method of forecasting large-dimensional covariance matrices by exploiting the theoretical and empirical potential of mixing forecasts derived from different information sets. The main theoretical contribution of the article is to find the conditions under which a mixed approach (MA) provides a smaller mean squared forecast error than a standard one. The conditions are general and do not rely on distributional assumptions of the forecasting errors or on any particular model specification. The empirical contribution of the article regards a comprehensive comparative exercise of the new approach against standard ones when forecasting the covariance matrix of a portfolio of thirty stocks. The implemented MA uses volatility forecasts computed from high-frequency-based models and correlation forecasts using realized-volatility-adjusted dynamic conditional correlation models. The MA always outperforms the standard methods computed from daily returns and performs equally well to the ones using high-frequency-based specifications, however at a lower computational cost.

Key words: high-frequency data, multivariate volatility, realized (co)variance, volatility forecasting

JEL classifications: C32, C53

Volatility modeling and forecasting have been of key interest in financial econometrics since the seminal contributions of Engle (1982) and Bollerslev (1986). Recently, research devel opments in the field have been refueled by the availability of high frequency (HF) financial data for various financial instruments. HF data have proven quite useful in forecasting future volatility through the realized volatility measure. Currently, there are a number of methods, mostly univariate, which propose dynamic models for realized volatility time ser ies, or alternatively, ways to integrate realized volatility measures into standard Generalized Autoregressive Conditional Heteroskedastic (GARCH) specifications. Hansen and Lunde (2011) provide a review of this growing literature.

In this article, we propose a new method of forecasting large dimensional covariance matrices by mixing HF volatility specifications with low frequency (LF) or daily correlation models. The main theoretical contribution of the article is the derivation, in a model free en vironment, of the statistical conditions and market environments under which such a mixed forecast outperforms a standard one. Thus, the generality of our results allows a practitioner to verify and decide in advance whether the current market conditions and/or the properties of the correlation and volatility forecasts, he/she has at hand, are favorable for one approach or the other. Moreover, our derivations help in understanding the roles that individual cor relation and volatility forecasts play in providing accurate covariance matrix forecasts sub ject to current market conditions: high versus low volatility/correlation states.

Throughout the article, we make use of the volatility correlation decomposition of the covariance matrix, which enables modeling of large dimensions.[1] Apart from this, no par ticular model specification or distributional assumptions are made. The loss function we consider is the mean squared error (MSE), which satisfies the conditions stated in Patton (2011), Patton and Sheppard (2009), and Laurent et al. (2013) of being robust to the noise in the univariate and multivariate variance proxies. We express the MSE as a function of the *ex post* integrated correlation and volatility over the forecasting period, which allows us to analyze the models' performance under various market conditions.[2] We show that the mixed approach (MA) can lead to significant gains in comparison to standard approaches (SAs) under general assumptions and at a low computational cost.

Although the pure daily models for modeling and forecasting covariance matrices are very popular (see Bauwens et al. 2006 for a comprehensive survey), they suffer from the curse of dimensionality, need to impose heavy parameter restrictions to ensure the positive definiteness of the covariance forecasts and treat the volatility and correlation process as la tent. Alternatively, HF data contains information that allows for almost error free measure ment of *ex post* volatility, based on the estimation of the quadratic variation of the price process, thus making it effectively observable. Early studies in the area (see e.g., Andersen et al. 2001, among others) have recognized that market microstructure effects can distort estimation at high frequencies and have proposed a sparse sampling approach, in which the available data are sampled every 5 , 10 or 15 minutes to mitigate the impact of market microstructure noise. More recently, techniques have been developed to use data more efficiently by designing estimators that are noise robust (see e.g., Barndorff Nielsen et al. 2008, 2009 and Jacod et al. 2009, among others). Most of these approaches are applicable to univariate series (i.e., for volatility) rather than to covariance estimation. Although multivariate extensions of the above mentioned approaches do exist (see e.g., Barndorff Nielsen et al. 2011, among others), they suffer from limitations, especially when applied to many assets. In most empirical work, realized covariance estimation is still carried out using the sparse sampling approach.[3] In general, it can be stated that the covariance/correlation

1 The term "large dimension" is used generically here and not in the sense of the number of assets going to infinity.

2 This differs from the concept of conditional predictive ability analyzed in Giacomini and White (2006). Our loss function is unconditional with respect to past information. Given that we derive the results in a model free environment, we do not explicitly define the conditioning set and, thus, we cannot examine conditional loss functions. With a particular model specification, conditional loss functions can be examined.

3 See Hautsch et al. (2012) and Lunde et al. (2011) for recent contributions on the estimation of large dimensional realized covariance matrices.

estimation using HF data is much more challenging than the volatility estimation due to issues of nonsynchronicity of the raw multivariate series and parameter proliferation. Furthermore, model specifications for realized covariance/correlation matrices have only recently gained more attention (see e.g., Gouriéroux et al. 2009; Bauer and Vorkink 2011; Chiriac and Voev 2011; Jin and Maheu 2013; Noureldin et al. 2012; Bauwens et al. 2013; and Hautsch et al., 2011) and there is still a lot of empirical work needed for these models to gain broader recognition.

Our answer to these challenges is to "mix" the accuracy of volatility models based on HF data with sparsely parameterized correlation models based on daily returns. In comparison to the existing HF based approaches, our method only requires the estimation of volatility series, rather than of covariance matrices, which, as discussed above, is more problematic, especially for large dimensions.

The main empirical contribution of the article is the implementation of the MA and the empirical exercise itself, which is a complex study of its performance in comparison to a series of standard LF and HF based alternatives. We examine the validity of the theoretical conditions in an application to 30 highly liquid stocks traded on the NYSE. The MA that we implement, uses volatility forecasts based on dynamic models for univariate series of daily realized volatilities, which can be estimated by any of the above mentioned techniques. The innovative part of the implementation concerns the correlation matrix forecast, which is conceptually identical to the dynamic conditional correlation (DCC) specification of Engle (2002), but with the important difference that we standardize (de volatilize) daily returns using realized volatilities rather than GARCH volatilities. Thus, one would expect forecast improvements over similar covariance matrix specifications using LF data (such as the DCC model of Engle 2002 and Tse and Tsui 2002) due to improvements in volatility forecasting and less noisy standardized residuals as an input to the correlation model. Andersen et al. (2001) find that the normal distribution provides a "strikingly" close match to the density of returns scaled by realized volatilities. The results of our empirical exercise show that the mixed specification outperforms the SAs based on daily returns and performs as well as a much more computationally intensive model based on HF data.

Our approach mixes forecasts and not frequencies as in the Mixed Data Sampling (MIDAS) model introduced by Ghysels et al. (2005) and Ghysels et al. (2006) (see Ghysels and Valkanov 2011, for an overview of MIDAS) and the High Frequency Data Based Projection Driven (HYBRID) GARCH model introduced by Chen et al. (2014) with extensions in Chen et al. (2011) and Chen and Ghysels (2011). In our approach, the input variables are volatility and correlation forecasts "independently" derived from HF and LF based models, whereas in MIDAS and HYBRID GARCH, the input variables are intra day returns sampled at various frequencies. As mentioned above, in the empirical application, we implement the mixed approach with input volatility forecasts derived from dynamic models on daily series of realized volatilities, such as the Autoregressive Fractional Integrated Moving Average (ARFIMA) and the Heterogenous Autoregressive (HAR) model of Corsi (2009). The information sets are given by series of daily measures computed from HF data, such as the series of daily realized volatilities. Alternatively, for the volatility input, one might consider forecasts based on MIDAS or HYBRID GARCH, where the HF information enters directly into the volatility prediction equation.

To the best of our knowledge, the papers of Bannouh et al. (2012) and Colacito et al. (2011) are the only other studies that consider mixed type covariance models. However, the differences between these studies and our are so stark that it suffices to mention only a few points of departure. First, the model of Bannouh et al. (2012) uses a factor structure in which the factor covariance matrix is estimated with HF data and the loadings on the factors are estimated with daily returns. Our approach does not assume a factor structure of the covariance matrix, although it clearly does not exclude that one exists. Second, the authors focus on the issue of the estimation of large dimensional covariance matrices, ra ther than on forecasting. Differently from our approach, that mixes volatility forecasts stemming from intra daily data with correlation forecasts originating from daily returns, the DCC MIDAS model of Colacito et al. (2011) mixes short run components computed from daily returns with long run components computed from monthly or quarterly data for both volatility and correlation forecasts. In fact, the only thing that the three papers have in common is that they use data of different frequencies, but in very different implementations.

The remainder of the article is structured as follows: Section 1 introduces the MA and presents the theoretical results, Section 2 contains the empirical study, and Section 3 con cludes. The proofs to the propositions in Section 1 can be found in Appendix A. Appendix B contains tables and graphs.

1 Theory

1.1 General Settings

Let r_t be a vector of daily log returns of dimension n the number of assets considered. In this article, we are interested in computing one step ahead covariance matrix forecasts of r_t condi tional on the information set available up to time t (a discussion of multi step ahead forecasts will follow later). More precisely, we aim at providing conditional one step ahead forecasts for Σ_{t+1}, which is the integrated covariance matrix of the (Itô semi martingale) price process from t to $t+1$. To this end, we employ the following decomposition of the covariance matrix:

$$\Sigma_t \quad D_t R_t D_t, \tag{1}$$

where D_t is a diagonal matrix given by the conditional standard deviations of each asset and R_t is the correlation matrix. This decomposition has been used in Engle (2002) and Tse and Tsui (2002) in the DCC framework. Thus, the conditional forecasts of Σ_t from the con ditional forecasts of D_t and R_t can be obtained as follows:

$$\hat{\Sigma}_{t+1|t} \quad \hat{D}_{t+1|t}\hat{R}_{t+1|t}\hat{D}_{t+1|t}, \tag{2}$$

where $\hat{D}_{t+1|t}$ and $\hat{R}_{t+1|t}$ are one step ahead forecasts of D_{t+1} and R_{t+1} conditional on the information set up to time t. In the next section, we will differentiate between volatility and correlation specific information sets. From Equation (2) one can easily derive the condi tional one step ahead variance and covariance forecasts as follows:

$$\hat{\sigma}_{ii,t+1|t} \quad \hat{d}^2_{i,t+1|t}, \qquad \forall i \quad 1,\ldots,n \tag{3}$$

$$\hat{\sigma}_{ij,t+1|t} \quad \hat{d}_{i,t+1|t}\hat{\rho}_{ij,t+1|t}\hat{d}_{j,t+1|t}, \qquad \forall i \;/\; j, \quad i,j \quad 1,\ldots,n \tag{4}$$

where $\hat{\sigma}_{ii,t+1|t}$ and $\hat{d}_{i,t+1|t}$ are the i th diagonal elements of $\hat{\Sigma}_{t+1|t}$ and $\hat{D}_{t+1|t}$, and $\hat{\sigma}_{ij,t+1|t}$ and $\hat{\rho}_{ij,t+1|t}$ are the ij th off diagonal elements of $\hat{\Sigma}_{t+1|t}$ and $\hat{R}_{t+1|t}$.

1.2 The Mixed Approach

In the following, we introduce the MA that differentiates between the information sets for volatility and correlation forecasting. Based on the decomposition from Equation (2), the MA forecast of Σ_t is given by:

$$\hat{\Sigma}^{MA}_{t+1|t} \quad \hat{D}^{\mathcal{F}^D}_{t+1|t}\hat{R}^{\mathcal{F}^R}_{t+1|t}\hat{D}^{\mathcal{F}^D}_{t+1|t}, \tag{5}$$

where $\hat{D}^{\mathcal{F}^D}_{t+1|t} \equiv E[D_{t+1}|\mathcal{F}^D_t]$, $\hat{R}^{\mathcal{F}^R}_{t+1|t} \equiv E[R_{t+1}|\mathcal{F}^R_t]$ and $\mathcal{F}^D_t$ and $\mathcal{F}^R_t$ are volatility and correlation specific information sets up to time t. For example, $\mathcal{F}^D_t$ may contain series of daily realized volatilities (as in the HAR model of Corsi 2009) or HF returns (as in the MIDAS model of Ghysels et al. 2005 and Ghysels et al. 2006); $\mathcal{F}^R_t$ may contain daily returns. However, at this point the choice of the information sets is arbitrary, as long as $\mathcal{F}^D_t \ / \ \mathcal{F}^R_t$. Below, we will compare the performance of the MA against that of the SA for which both volatility and correlation forecasts are conditional on the same information set. To focus solely on comparing the impact of choosing a mixture of information sets rather than a single one, we choose to define the SA based on the same decomposition as given in Equation (2). More precisely, we specify:

$$\hat{\Sigma}^{SA}_{t+1|t} \quad \hat{D}^{\mathcal{F}}_{t+1|t}\hat{R}^{\mathcal{F}}_{t+1|t}\hat{D}^{\mathcal{F}}_{t+1|t}. \tag{6}$$

Although the dynamic models for D_t and R_t in general differ, $\mathcal{F}$ indicates that both volatility and correlation forecasts use the same information set, namely $\mathcal{F}_t$ (e.g., daily returns or daily series of realized measures computed from HF data).

Before turning to the formal comparison of the two approaches, we provide some intuition on why we believe that the MA might be a valuable alternative to the SA. The MA allows for volatility and correlation forecasts to be "independently" derived based on different (specific) information sets and may, thus, provide smaller (co)variance forecast errors, as explained in Section 1.3. Furthermore, the MA can allow to exploit the advantages of HF based models for forecasting volatilities and those of LF based models for forecasting correlations[4]: compared to LF data through the GARCH type of models, the HF data has proven to be extremely useful in the *ex post* measurement and forecasting of daily volatility by means of realized measures or MIDAS frameworks. Nevertheless, multivariate approaches based on HF data (e.g., realized covariance matrices) are not as well developed and suffer from difficulties associated with nonsynchronous trading and dimensionality. Alternatively, models using LF data such as the DCC approach of Engle (2002) have already gained a broad recognition among practitioners and scientists, due to their good forecasting performance and ease of implementation. Although this particular implementation of the MA is an obvious choice, in the following analysis we do not limit our theoretical derivations to it, but instead keep the exposition general. Thus, we derive and discuss the

4 From here on, whenever we refer to HF based information sets, we mean information sets that entail the HF itself such as in the MIDAS model of Ghysels et al. (2005) and Ghysels et al. (2006) or daily realized measures, such as daily realized volatilities or correlations computed from HF data. Please note the difference between the terms "HF information set" and "HF based information set". In this case, an "HF information set" would be the information set that exclusively contains HF data, as in the MIDAS specification.

conditions under which, a general MA as defined in Equation (5) provides smaller MSE than a general SA as defined in Equation (6). These derivations allow a practitioner to eas ily verify whether a particular model satisfies the conditions for providing a superior fore cast in comparison to another (potentially simpler) specification, subject to the current market conditions (e.g., high volatility) and to the statistical properties of the volatility and correlation forecasts.

1.3 MSE Comparisons

Because we focus on one step ahead forecasts, we are able to simplify the notation in the following manner: $\hat{\sigma}_{ij} \equiv \hat{\sigma}_{ij,t+1|t}$, $\hat{d}_i \equiv \hat{d}_{i,t+1|t}$ and $\hat{\rho}_{ij} \equiv \hat{\rho}_{ij,t+1|t}$ for all $i, j \quad 1, \ldots, n$. We will use the representations:

$$\hat{\sigma}_{ij} \quad \sigma_{ij} + \varepsilon_{\sigma_{ij}}, \qquad \forall i, j \quad 1, \ldots, n \tag{7}$$

$$\hat{d}_i \quad d_i + \varepsilon_{d_i}, \qquad \forall i \quad 1, \ldots, n \tag{8}$$

$$\hat{\rho}_{ij} \quad \rho_{ij} + \varepsilon_{\rho_{ij}}, \qquad \forall i \ / \ j, \quad i, j \quad 1, \ldots, n \tag{9}$$

where the ε's represent forecast errors and σ_{ij}, d_i and ρ_{ij} are the true values of the variables at time $t+1$. Based on this notation, we can rewrite Equations (3) and (4) as follows:

$$\hat{\sigma}_{ii} \quad (d_i + \varepsilon_{d_i})^2 \quad d_i^2 + 2d_i\varepsilon_{d_i} + \varepsilon_{d_i}^2 \equiv \sigma_{ii} + \varepsilon_{\sigma_{ii}} \qquad \forall i \quad 1, \ldots, n \tag{10}$$

$$\begin{aligned} \hat{\sigma}_{ij} \quad & (d_i + \varepsilon_{d_i})(\rho_{ij} + \varepsilon_{\rho_{ij}})(d_j + \varepsilon_{d_j}) \\ & d_i\rho_{ij}d_j + d_i\rho_{ij}\varepsilon_{d_j} + d_j\rho_{ij}\varepsilon_{d_i} + d_id_j\varepsilon_{\rho_{ij}} + d_j\varepsilon_{d_i}\varepsilon_{\rho_{ij}} + d_i\varepsilon_{d_j}\varepsilon_{\rho_{ij}} + \rho_{ij}\varepsilon_{d_i}\varepsilon_{d_j} + \varepsilon_{d_i}\varepsilon_{\rho_{ij}}\varepsilon_{d_j} \\ \equiv \sigma_{ij} + \varepsilon_{\sigma_{ij}} & \qquad \forall i \ / \ j, \quad i, j \quad 1, \ldots, n, \end{aligned} \tag{11}$$

with $\varepsilon_{\sigma_{ii}} \equiv 2d_i\varepsilon_{d_i} + \varepsilon_{d_i}^2$ and $\varepsilon_{\sigma_{ij}} \equiv d_i\rho_{ij}\varepsilon_{d_j} + d_j\rho_{ij}\varepsilon_{d_i} + d_id_j\varepsilon_{\rho_{ij}} + d_j\varepsilon_{d_i}\varepsilon_{\rho_{ij}} + d_i\varepsilon_{d_j}\varepsilon_{\rho_{ij}} + \rho_{ij}\varepsilon_{d_i}\varepsilon_{d_j} + \varepsilon_{d_i}\varepsilon_{\rho_{ij}}\varepsilon_{d_j}$. Note that the complexity of the error term $\varepsilon_{\sigma_{ij}}$ (and to a smaller extent that of $\varepsilon_{\sigma_{ii}}$) arises due to the fact that the covariance (variance) forecast is a nonlinear transformation of the correlation and volatility forecasts. The error term, thus, accounts for the noise in these original forecasts and the effect of the transformation. The unbiasedness of correl ation and volatility forecasts is not sufficient to guarantee the unbiasedness of the covari ance unless we also ensure that the volatility and correlation forecasts are independent or at least uncorrelated.

In the analysis below, we compare the MA and the SA on the basis of their MSEs mak ing use of the following decomposition:

$$MSE(\hat{\sigma}_{ij}) \quad E[\varepsilon_{\sigma_{ij}}]^2 + V[\varepsilon_{\sigma_{ij}}]. \tag{12}$$

The mean and variance of $\varepsilon_{\sigma_{ij}}$ are derived in Appendix A. The expressions in Equations (10) and (11) allow to analyze the MSEs as a function of d_i, d_j, and ρ_{ij}, which in turn makes it possible to draw conclusions about the models' performance under different volatility/correlation regimes. We are now in a position to derive the conditions under which the fore cast MSE of the MA, $\hat{\sigma}_{ij}^{MA}$, is smaller than the forecast MSE of the SA, $\hat{\sigma}_{ij}^{SA}$ for each $i, j \quad 1, \ldots, n$. Note that we first examine an element wise MSE. If the element wise domin ance of a forecast holds, dominance in terms of matrix loss follows automatically. However, in practical applications, it may be the case that the MA dominates the SA for some elements of the covariance matrix, while the SA dominates for others. Therefore, to provide a complete picture, we also examine a matrix loss function, such as the squared Frobenius norm of the matrix forecast errors $\varepsilon_{\Sigma} \equiv \hat{\Sigma} - \Sigma$, where $\hat{\Sigma} = \hat{\Sigma}_{t+1|t}$ and Σ is the true covariance matrix. This loss function, which is a member of the family of quadratic loss functions as suggested by Laurent et al. (2013), is consistent with respect to the choice of the covariance matrix proxy. The Frobenius norm of a quadratic matrix A of dimension $n \times n$ is defined to be: $\|A\| = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}^2}$, where a_{ij} is the ij th element of the matrix. Thus, the Frobenius norm of a covariance matrix, which is a symmetric matrix, assigns double weight to the covariance (off-diagonal) elements in comparison to the variance (diagonal) elements. The MSE of $\hat{\Sigma}$ based on the Frobenius norm thus becomes

$$MSE(\hat{\Sigma}) = \sum_{i=1}^{n} \sum_{j=1}^{n} MSE(\hat{\sigma}_{ij}) = \sum_{i=1}^{n} MSE(\hat{\sigma}_{ii}) + 2 \sum_{i=1}^{n-1} \sum_{j>i}^{n} MSE(\hat{\sigma}_{ij}).$$

A conceptually different possibility would be to consider some sort of an economic loss function, such as hedging error variance, portfolio variance, etc. An analysis of such loss functions is outside of the scope of this article, but it can be considered in a separate study.

Proposition 1: Variance elements MSE comparison

If for a given $i \in \{1, 2, \ldots, n\}$, $E[(\varepsilon_{d_i}^{X})^4] < \infty$, where $X \in \{\mathcal{F}^{\mathcal{D}}, \mathcal{F}\}$, and $\mathcal{F}^{\mathcal{D}} \neq \mathcal{F}$, then the difference between the MSE of the MA variance forecast $\hat{\sigma}_{ii}^{MA}$ and that of the SA variance forecast $\hat{\sigma}_{ii}^{SA}$ is given by

$$MSE(\hat{\sigma}_{ii}^{MA}) - MSE(\hat{\sigma}_{ii}^{SA}) = A_{1,i} d_i^2 + A_{2,i} d_i + A_{3,i} \equiv A(d_i) \tag{13}$$

where $A_{1,i} = 4E[(\varepsilon_{d_i}^{\mathcal{F}^{\mathcal{D}}})^2] - E[(\varepsilon_{d_i}^{\mathcal{F}})^2]$, $A_{2,i} = 4E[(\varepsilon_{d_i}^{\mathcal{F}^{\mathcal{D}}})^3] - E[(\varepsilon_{d_i}^{\mathcal{F}})^3]$ and $A_{3,i} = E[(\varepsilon_{d_i}^{\mathcal{F}^{\mathcal{D}}})^4] - E[(\varepsilon_{d_i}^{\mathcal{F}})^4]$. A set of *sufficient (but not necessary) conditions* for $MSE(\hat{\sigma}_{ii}^{MA}) \leq MSE(\hat{\sigma}_{ii}^{SA})$ is that $A_{1,i} \leq 0$, $A_{2,i} \leq 0$ and $A_{3,i} \leq 0$. The minimal sufficient conditions (i.e., the necessary and sufficient conditions) for the inequality to hold are clearly weaker and are provided in Appendix A.

In Proposition 1 (as well as for the subsequent results), apart from the existence of certain moments, we do not make any distribution, unbiasedness and such other assumptions on the forecast (errors). This is theoretically appealing, as it allows us to specialize the expressions should we desire to make such assumptions (e.g., that errors have a zero expectation and/or are normally distributed). Furthermore, models of volatility forecasting often employ specifications regarding a transformation of volatility (e.g., log realized volatility). Strictly speaking, naive forecasts via re-transforming the forecast are biased, although they might be empirically superior to bias-corrected forecasts (see e.g., Halbleib and Voev 2011). Also GARCH models typically model the variance, but there are specifications for the volatility (in its narrow definition as standard deviation) as well. Generally speaking it is difficult to take a stand on the unbiasedness of variance/volatility forecasts and, therefore, we make no assumption in this regard.

Our findings indicate that it is sufficient for the MA to outperform the SA if the second, third, and fourth (uncentered) moments of the MA volatility errors are smaller than their SA counterparts. The fact that the variance is the square of volatility results in the MSE of variance forecasts (the second moment of variance forecast errors) being expressed as a linear combination between the second, third, and fourth moments of the volatility forecast errors. The second moment in Equation (13), namely $E[(\varepsilon_{d_i})^2]$, is in fact the MSE of the con ditional volatility forecasts.

If $\mathcal{F}^{\mathcal{D}}$ is a HF based information set and $\mathcal{F}$ is a LF based set and, since there is evidence (e.g., Chiriac and Voev 2011. among others) that HF data driven forecasts have a smaller MSE than their LF counterparts, one would expect $E[(\varepsilon_{d_i}^{\mathcal{F}^{\mathcal{D}}})^2] \quad E[(\varepsilon_{d_i}^{\mathcal{F}})^2] \leq 0$ to hold. The conditions on the third and fourth moments of volatility forecast errors refer to their skew ness and kurtosis. Because there is no theoretical reason why a certain model should pro vide skewed forecast errors, we expect that $E[(\varepsilon_{d_i}^{\mathcal{F}^{\mathcal{D}}})^3] \quad E[(\varepsilon_{d_i}^{\mathcal{F}})^3] \quad 0$ and, thus, $E[(\varepsilon_{d_i}^{\mathcal{F}^{\mathcal{D}}})^3] \quad E[(\varepsilon_{d_i}^{\mathcal{F}})^3] \quad 0$. Regarding the fourth moment, one can expect HF based methods to provide, in addition to smaller MSE forecast errors, less extreme forecast errors than LF driven methods: i.e., $E[(\varepsilon_{d_i}^{\mathcal{F}^{\mathcal{D}}})^4] \quad E[(\varepsilon_{d_i}^{\mathcal{F}})^4] \leq 0$.

If $\mathcal{F}^{\mathcal{D}}$ and $\mathcal{F}$ are both HF based information sets, but of different frequencies (e.g., $\mathcal{F}^{\mathcal{D}}$ contains series of daily realized volatilities computed from five minute returns and $\mathcal{F}$ contains series of daily realized volatilities computed from one minute returns), then the validity of the conditions stated above must be empirically verified. Although, theoretic ally, increasing the frequency leads to more efficient unbiased realized volatility estimates and, thus, to smaller MSEs, in practice this effect might be destroyed by the market micro structure noise, as our empirical exercise reveals. Moreover, the MSE of variance forecasts depends on additional properties of the volatility forecast errors, such as skewness and kur tosis. So far, there is no evidence regarding how these measures change when the frequency is increased. Therefore, in the empirical application in Section 2, we focus on this issue in more detail.

In the following proposition, we derive and compare the MSEs of the covariance elem ents for the MA and the SA. As before, we make minimal assumptions on the moments and dependence structure of the forecast errors.

Proposition 2: Covariance elements MSE comparison

If, for a given $i \; / \; j$, $i, j \in \{1, 2, \ldots, n\}$ and $k, l, m \quad 0, 1, 2$, it holds that

i. $E[(\varepsilon_{d_i}^{\mathcal{F}^{\mathcal{D}}})^k (\varepsilon_{d_j}^{\mathcal{F}^{\mathcal{D}}})^l (\varepsilon_{\rho_{ij}}^{\mathcal{F}^{\mathcal{R}}})^m] < \infty$

ii. $E[(\varepsilon_{d_i}^{\mathcal{F}})^k (\varepsilon_{d_j}^{\mathcal{F}})^l (\varepsilon_{\rho_{ij}}^{\mathcal{F}})^m] < \infty$

then it follows that

$$MSE(\hat{\sigma}_{ij}^{MA}) \quad MSE(\hat{\sigma}_{ij}^{SA}) \quad G(d_i, d_j, \rho_{ij}), \tag{14}$$

where $G(d_i, d_j, \rho_{ij})$ is a fourth order polynomial in d_i, d_j, and ρ_{ij} as given in Equation (30) in Appendix A. If $\rho_{ij} \geq 0$, *sufficient (but not necessary)* conditions for $MSE(\hat{\sigma}_{ij}^{MA}) \leq MSE(\hat{\sigma}_{ij}^{SA})$ are that all parameters of $G(d_i, d_j, \rho_{ij})$ are nonpositive (see Proof A.2 in Appendix A).

In Proposition 2, the MSEs of the covariance forecasts are expressed as functions of the true *ex post* values of the volatility and correlation. Given the minimal assumptions we make, these expressions are rather lengthy and defy straightforward interpretation. Nevertheless, under the empirically relevant assumption that $\rho_{ij} \geq 0$, some intuition regarding the sufficient conditions stated above can be provided.[5] For the MA to outperform the SA, it is sufficient but not necessary, that:

1. second moments of volatility forecast errors are smaller for the MA than for the SA ($G_{1,i} \leq 0$ and $G_{1,j} \leq 0$);
2. second moments of correlation forecast errors are smaller for the MA than for the SA ($G_{2,ij} \leq 0$); and
3. cross moments up to order six of forecast errors are smaller for the MA than for the SA ($G_{3,ij}, \ldots, G_{22,ij} \leq 0$).

Similar to the variance case, the fact that the covariance is the product between two volatilities and correlation automatically results in the MSE of covariance forecasts being a linear combination of the volatility and correlation forecast MSEs and of further cross moments of the forecast errors. Therefore,

1. The first condition is intuitive, since it refers to the MSE of volatility forecasts:
   a. If $\mathcal{F}^{\mathcal{D}}$ $\mathcal{F}$, then $G_{1,i}$ 0 and $G_{1,j}$ 0 and the condition is automatically fulfilled. This is the case when both the MA and the SA use the same volatility forecasts.
   b. If $\mathcal{F}^{\mathcal{D}}$ / $\mathcal{F}$, then the condition is likely to hold, as discussed in Proposition 1. However, when the SA is computed from an HF based information set of a different frequency than that used in the MA, the validity of the condition must be empirically verified due to the market microstructure noise.
2. The second condition is also intuitive, since it refers to the MSE of the correlation forecasts:
   a. If $\mathcal{F}^{\mathcal{R}}$ $\mathcal{F}$, then $G_{2,ij}$ 0 and the condition is automatically fulfilled. This is the case when both the MA and the SA use the same correlation forecasts.
   b. If $\mathcal{F}^{\mathcal{R}}$ / $\mathcal{F}$, then the condition is likely to hold when the MA correlation forecasts are HF data driven and the SA correlation forecasts stem from models for daily returns (i.e., HF based correlation forecasts have a smaller MSE than their LF based counterparts). However, this scenario is not of practical interest, as the aim of the MA is to avoid the difficulties associated with the estimation and forecasting of correlation matrices computed from HF data. Using LF data driven correlation forecasts makes the MA more tractable than a SA built solely from HF based forecasts. However, in this case, the condition is likely to fail empirically (at least in small dimensions, where the HF model is more feasible). Theoretically, the performance of the MA may be improved by using correlation forecasts from LF based models, such as the DCC of Engle (2002) that use daily returns standardized by HF volatility estimates (realized volatilities), which are more accurate than their LF counterparts. This is confirmed by the empirical results presented in Section 2.
3. The third condition is less intuitive to interpret. Five of G's parameters refer to differences in the cross moments between volatility forecast errors, whereas the other sixteen

5 Note that this assumption is only needed for the ease of interpreting the sufficient conditions. In our empirical exercise, the negative correlation days are about 3% of the total sample. Moreover, the positive correlation assumption is particularly relevant for periods of high volatility, due to the empirically observed volatility in correlation effect (high volatility induces high correlation) typical of stock markets.

refer to cross moments between volatility and correlation forecast errors. If the volatility forecasts are conditional on the same information sets, we can expect a large dependency in their error, regardless of the forecast method. However, in addition to smaller MSEs, HF data may also provide weaker forecast error dependencies than their LF counterparts. Therefore, an MA using HF based volatility forecasts may have some MSE advantage over an SA computed from daily returns. This is confirmed by the empirical results presented in Section 2. The nonpositivity conditions on the cross moments of volatility and correlation forecast errors are expected to hold given that the volatility and correlation forecasts are derived from different information sets and, thus, have less dependent errors than if they are conditional on the same information sets.

By summarizing the discussion, we can say that the advantages of the MA in comparison to the SA result from its flexibility in mixing accurate volatility forecasts with less computationally intensive correlation forecasts and from a lower degree of dependency between the volatility and correlation forecast errors. In particular, we expect a MA using HF based volatility forecasts and LF based correlation forecasts to outperform or to perform just as well as the SA. Even though the second condition may not be fulfilled when the SA is purely HF based, it is important to emphasize that the conditions stated above are not necessary. We can still find that $MSE(\hat{\sigma}_{ij}^{MA}) \leq MSE(\hat{\sigma}_{ij}^{SA})$ even when some of the conditions are violated. Moreover, the weight of this condition in the set of all conditions is comparatively small (1 in 22). Thus, in practice, the MA can still provide better or equally good forecasts in comparison to the SA as long as the effect of smaller HF based correlation forecast errors is weaker than the effect of reducing the dependency between the volatility and correlation forecast errors. The empirical results presented in Section 2 prove this. The performances of the MA and the purely HF data driven SA are statistically equivalent (the MSEs are not significantly different) and the MA always outperforms a SA built solely on LF based forecasts.

Proposition 3: Matrix MSE comparison

If for all $i \in \{1, 2, \ldots, n\}$, $E[(\varepsilon_{d_i}^{X})^4] < \infty$, where $X \in \{\mathcal{F}^{\mathcal{D}}, \mathcal{F}\}$ and if for all $j > i$, $i, j \in \{1, 2, \ldots, n\}$ and $k, l, m \quad 0, 1, 2$, it holds that

i. $E[(\varepsilon_{d_i}^{\mathcal{F}^{\mathcal{D}}})^k (\varepsilon_{d_j}^{\mathcal{F}^{\mathcal{D}}})^l (\varepsilon_{\rho_{ij}}^{\mathcal{F}^{\mathcal{R}}})^m] < \infty$

ii. $E[(\varepsilon_{d_i}^{\mathcal{F}})^k (\varepsilon_{d_j}^{\mathcal{F}})^l (\varepsilon_{\rho_{ij}}^{\mathcal{F}})^m] < \infty$

then it follows that

$$MSE(\hat{\sigma}_{ij}^{MA}) \quad MSE(\hat{\sigma}_{ij}^{SA}) \quad \sum_{i \quad 1}^{n} A(d_i) + 2 \sum_{i \quad 1}^{n-1} \sum_{j>i}^{n} G(d_i, d_j, \rho_{ij}) \equiv H(d_i, d_j, \rho_{ij}), \tag{15}$$

where A and G are defined in Propositions 1 and 2. If $\rho_{ij} \geq 0$, *sufficient (but not necessary)* conditions for $MSE(\hat{\sigma}_{ij}^{MA}) \leq MSE(\hat{\sigma}_{ij}^{SA})$ are that all parameters of $H(d_i, d_j, \rho_{ij})$ for all $i \quad 1, \ldots n$ and $j \quad 2, \ldots n$ with $j > i$ are nonpositive.

The proof of Proposition 3 follows directly from the proofs of Propositions 1 and 2. If the MA and the SA use the same volatility forecasts, then Equation (15) reduces to

$2\sum_{i\ 1}^{n-1}\sum_{j>i}^{n} G(d_i, d_j, \rho_{ij})$ with $G_{1,\cdot}\quad G_{5,ij}\quad G_{6,ij}\quad G_{10,ij}\quad G_{14,ij}\quad 0$. If the MA and the SA use the same correlation forecasts, then $G_{2,ij}\quad 0$ in Equation (15).

The sufficient conditions of Proposition 3 are quite restrictive, because the nonpositivity must hold for all assets and pairs of assets. The fact that the difference between the matrix MSEs is a sum over all assets and pairs allows more flexible conditions than in the element wise comparison. In practical applications, it may be the case that for a certain pair some of the G parameters are positive and, thus, the element wise MSE difference becomes positive. However, given the summation over all pairs, this effect may be eliminated by a "strong" negativity among the rest of the pairs. A MA combining HF data driven volatility forecasts with LF based correlation forecasts can benefit most from this summation effect. Thus, in comparison to pure LF methods, the large gain in MSE of volatility forecasts based on HF information sets ($A_{1,i} \leq 0$) outweighs potentially positive G parameters. In comparison to pure HF methods, the loss in MSE of LF based correlation forecasts ($G_{2,ij} > 0$) may be neutralized by the negativity of the G parameters that account for the dependency between volatility and correlation forecasts. In the empirical exercise, we will report results for both the element wise and matrix MSE comparisons.

The usefulness of Propositions 2 and 3 is that they give us the exact form of the element wise and matrix MSE differences as a function of the variables d_i, d_j, and ρ_{ij}. Since the parameters of the polynomial G are easily estimated from data (simply as the sample counterparts of the population moments), we can use their results to analyze how the relative model performance reacts in various volatility/correlation environments. To approach this issue more formally, we compute the partial derivatives of $G(d_i, d_j, \rho_{ij})$ and $H(d_i, d_j, \rho_{ij})$ with respect to their arguments (the derivatives with respect to d_i and d_j are symmetric), which are reported in Appendix A. To give an intuitive description of these derivatives, one should remember that it is a widely observed empirical fact that high volatility induces high correlations (i.e., the volatility in correlation effect) and, thus, when d_i is large, ρ_{ij} is also likely to be large. Our theoretical expressions confirm this empirical fact by the large proportion of common G's in the expressions (31), (32), (33), and (34). Thus, if in general the conditions of Propositions 2 and 3 are fulfilled, then we expect that during high volatility/correlation states of the market, the MA is not outperformed by the SA. As a matter of fact, one can observe that the parameters describing the differences in the moments and cross moments of volatility forecast errors are weighed by a factor of 2, implying that using accurate volatility forecasts may be particularly beneficial during times of high volatility and correlation.

As mentioned above, our focus in this article is on providing conditional one period ahead covariance matrix forecasts. Within our framework, a multi step forecast can easily be obtained by mixing forecasts of volatility and correlation as in Equation (2) as follows:

$$\hat{\Sigma}_{t:t+h|t}\quad \hat{D}_{t:t+h|t}\hat{R}_{t:t+h|t}\hat{D}_{t:t+h|t}. \tag{16}$$

However, the theoretical properties derived above do not directly apply. If the multi step forecasts are obtained from a "direct" procedure, by aggregating data over h periods, then the theoretical results derived so far will hold as long as the finite moment assumptions from the propositions stated above hold for the aggregated forecast errors. However, if the volatility and correlation multi step forecasts result from an iterated procedure, in which the h step ahead forecasts are computed by aggregating h one period forecasts, to derive the theoretical results, one must account for the joint distribution of the h one step forecast errors. To illustrate this issue, we provide an example with $h \quad 2$ (cf. Equation (3) above):

$$\hat{\sigma}_{ii,t:t+2|t} \quad \hat{d}^{2}_{i,t:t+2|t} \tag{17}$$

$$\begin{aligned} \hat{d}_{i,t:t+2|t} \quad & \hat{d}_{i,t:t+1|t} + \hat{d}_{i,t+1:t+2|t} \quad d_{i,t:t+1} + \varepsilon_{d_i,t:t+1|t} + d_{i,t+1:t+2} + \varepsilon_{d_i,t+1:t+2|t} \\ & d_{i,t:t+2} + \underbrace{\varepsilon_{d_i,t:t+1|t} + \varepsilon_{d_i,t+1:t+2|t}}_{\varepsilon_{d_i,t:t+2|t}} . \end{aligned} \tag{18}$$

We must now account for the properties of the aggregated forecast error, $\varepsilon_{d_i,t:t+2|t}$, which imply conditions on the time variation in the moments and cross moments of the daily forecast errors as well as on their serial correlation and cross correlation. We refrain from pursuing this analysis in the current study and leave it for future research.

2 Empirical Application

In this section, we present the results of the MA and the SA presented in Section 1. Patton (2011) and Laurent et al. (2013) provide empirical evidence that, although the ranking among forecasting models based on MSE is insensitive to the proxy choice, a less noisy proxy, such as realized (co)variances computed from data sampled on frequencies from 5 minutes to 20 minutes, allows more efficient discrimination between models. Here, to pro vide a complete picture of our results, we consider as proxies for the true covariance matrix a series of estimates the cross product (square) of daily returns as a LF based proxy and realized covariances computed from 1 , 5 , 15 to 30 minute returns as HF based proxies. To be consistent with the theoretical framework and the results described in Section 1, we compare the forecast performance of the MA to that of the SA solely by means of their MSEs.

2.1 Data

Our data consists of tick by tick transaction prices from the Trade and Quotations (TAQ) database sampled from 9:30 until 16:00 for the period November 1, 2001 to November 27, 2013 ($T \quad 3012$ trading days).[6] For the current analysis, we select a total of 30 liquid assets[7] that were in the past or are currently constituents of the Dow Jones Industrial Average (DJIA) and are traded on the New York Stock Exchange. We employ the previous tick interpolation method, described in Dacorogna et al. (2001) and obtain 390, 78, 26,

6 We are grateful to Sebastian Bayer for preparing the data.

7 The stocks are: Alcoa Inc. (AA), American Express Company (AXP), Boeing Corporation (BA), Bank of America Corporation (BAC), Citigroup Inc. (C), Caterpillar Inc. (CAT), Chevron Corporation (CVX), Dupont (DD), Walt Disney Company (DIS), General Electric Company (GE), The Goldman Sachs Group, Inc. (GS), Home Depot Inc. (HD), Honeywell International Inc. (HON), Hewlett Packard Company (HPQ), International Business Machines (IBM), International Paper Company (IP), Johnson & Johnson (JNJ), J.P. Morgan Chase & Company (JPM), Coca Cola Company Kraft Foods Inc. (KO), McDonald's Corporation (MCD), 3M Company (MMM), Altria Group Inc. (MO), Merck & Company Inc. (MRK), Nike, Inc. (NKE), Pfizer Inc. (PFE), Procter & Gamble Company (PG), United Technologies Corporation (UTX), Verizon Communications Inc. (VZ), Wal Mart Stores Inc. (WMT), Exxon Mobil Corporation (XOM).

and 13 intra day returns by sampling every 1 , 5 , 15 and 30 minutes, respectively, as well as one daily return. For the estimation, we multiply the returns by 100, i.e., we consider percentage returns. For each $t \quad 1, \ldots, 3012$, a series of daily realized covariance matrices can be constructed as:

$$RCov_t^{x-min} \quad \sum_{j \quad 1}^{M} r_{j,t} r'_{j,t} \tag{19}$$

where $x \quad \{1, 5, 15, 30\}$ and $M \quad \{390, 78, 26, 13\}$. The x minute returns, $r_{j,t}$, are com puted as

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

where $\Delta \quad 1/M$ and $p_{j\Delta,t}$ is the log transaction price at time $j\Delta$ on Day t. The realized co variance matrices are symmetric by construction and, for $n < M$, are almost certainly posi tive definite (p.d.). In this case, only the $RCov^{1-min}$ and $RCov^{5-min}$ are p.d. almost surely. However, to enhance the precision, we refine the estimator by subsampling. As in Chiriac and Voev (2011), we construct 6, 30, 90, and 180, respectively, regularly Δ spaced subgrids ($\Delta \quad \{60, 300, 900, 1800\}$ seconds) starting at seconds $1, 11, 21, \ldots, y$, where $y \quad 51, 291, 891, 1791$, compute the realized covariance matrix on each subgrid and take the average. The resulting subsampled realized covariance is much more robust to the so called market microstructure noise than the "naive" one from Equation (19), and, given the high liquidity of all the stocks, we are confident that the effect of nonsynchronicity is rather weak at the chosen frequencies. Moreover, the resulting $RCov^{15-min}$ and $RCov^{30-min}$ matri ces also become p.d. Although the "naive" covariance matrices are of ranks 26 and 13, re spectively, through averaging (summing), the rank becomes full after averaging (summing) over two and three matrices, respectively, due to the relation $\text{rank}(A+B) \leq \text{rank}(A) + \text{rank}(B)$ and to the fact that the assets we consider are very liquid (i.e., the intra day returns generally change from one subgrid to the next).

To avoid the noise induced by measuring the overnight volatility as the squared over night return, we use the open to close data and measure only the volatility over the trading window. Consequently, the daily return is also computed over the open to close period.

Table B.1 in Appendix B reports summary statistics for the realized variances and co variances averaged over all 30 stocks and over all 435 pairs of stocks as well as the sum mary statistics of daily returns averaged over all stocks. For both realized variance and covariance measures, we observe typical stylized facts such as extreme right skewness and leptokurtosis (Andersen et al. 2001). The daily returns exhibit skewness close to zero and overkurtosis.

The thirty daily realized variance series are given by the diagonal elements of the real ized covariance matrix defined above. Please note the distinction we make here between realized variance and its square root, for which we use the term "realized volatility" (*RVol*). The series of daily realized correlation matrices *RCorr* are computed from *RCov* in the usual way.

2.2 Forecasting Models

In this section, we elaborate on the implementation of the two forecasting approaches introduced in Section 1. Based on the theoretical results, we opt to implement the MA with volatility forecasts stemming from dynamic models for daily realized volatilities and correlation forecasts derived from dynamic models for daily returns. Clearly, a MA utilizing LF based volatility forecasts and HF based correlation forecasts is also conceivable, but it is not of practical interest. Furthermore, we consider SAs that use either pure HF or LF based volatility and correlation forecasts. The choice of the models implemented here is broad and, generally speaking, motivated by various theoretical and empirical results reported in the related literature on forecasting daily volatilities and correlations based on daily returns and HF data. For both volatility and correlation forecasts we opt not just for one model choice, but rather for a series of models to let the data "tell" which model and mixing strategy are the best and to make our results more robust. As mentioned above, the choice of the models is ultimately an empirical question.

2.2.1 The MA

To obtain the volatility forecasts, $\hat{D}_{t+1|t}$ we opt to follow the main body of literature (e.g., Andersen et al. 2001, among others) in applying autoregressive frameworks that are able to capture the strong persistence (long memory) of daily log *RVol* series evident in the slowly decaying autocorrelation functions (ACFs) in Figure B.1.[8] Thus, for each i $1, \dots,$ 30 on the demeaned log *Rvol* series, we apply the following fractional integrated model specifications: ARFIMA(1,*d*,1), ARFIMA(1,*d*,0), and ARFIMA(0,*d*,0). We estimate all models using the maximum likelihood approach described in Beran (1995). Moreover, as an alternative to the ARFIMA models, we also apply the HAR model of Corsi (2009), namely $HAR(1, 5, 10, 20)$, which is also able to capture the long persistence in the underlying series.[9]

The one day ahead correlation matrix forecast, $\hat{R}_{t+1|t}$ is based on the DCC approach of Engle (2002) estimated on residuals (we assume that the conditional mean of daily returns is constant and estimate the model on the demeaned series of daily returns), standardized by realized volatilities, rather than by GARCH volatilities as in the standard implementation of the DCC. Theoretically, the standardization by *RVol* is likely to improve upon the correlation model as a secondary channel through which HF data can lead to enhancements. We estimate the DCC model using the maximum likelihood approach described in Engle (2002). However, given that there is empirical evidence (Engle et al. 2008) that the maximum likelihood estimator is severely biased when applied to correlation matrices of high dimensions (larger than twenty), we also apply various techniques proposed in the literature to reduce this bias, including the consistent DCC (cDCC) approach of Aielli (2013) together with two shrinkage approaches proposed by Hafner and Reznikova (2012), for which the targets are the identity matrix (cDCC I) and the equicorrelation matrix (cDCC E).

8 Figure B.1 plots the ACF of the diagonal elements of $RCov^{5-min}$. The ACFs for the diagonal elements of the other *RCov* matrices are similar. They can be obtained from the authors upon request.

9 One could also apply other methods, such as MIDAS of Ghysels et al. (2005) and Ghysels et al. (2006) or HYBRID GARCH of Chen et al. (2014) to obtain forecasts of daily variances. Whereas the MIDAS is a pure HFA, the HYBRID GARCH method uses a mixture of HF and LF information.

2.2.2 The SA

The pure LF approach (LFA): The one day ahead volatility forecasts, $\hat{D}_{t+1|t}$, are obtained from univariate GARCH(1,1) models[10] and the one day ahead correlation forecast $\hat{R}_{t+1|t}$ is derived from the DCC approach of Engle (2002) applied to (de volatilized) residuals stand ardized using GARCH's conditional standard deviations. We estimate the DCC model by means of the maximum likelihood approach described in Engle (2002). As above, we imple ment additional enhancements of the DCC model which aim at correcting the bias induced by the application of the maximum likelihood estimation to large dimensional matrices.

The pure HF approach (HFA): The volatility forecasts are obtained in the same manner as for the MA and the one day ahead correlation forecasts $\hat{R}_{t+1|t}$ are obtained from the fol lowing autoregressive framework:

$$\hat{R}_{t+1|t} \quad \left(1 \quad \sum_{l \;\; 1}^{t} \lambda_l\right) \overline{R}Corr + \sum_{l \;\; 1}^{t} \lambda_l \tilde{R}Corr_{t-l+1}, \tag{20}$$

where $\overline{R}Corr \quad \frac{1}{t}\sum_{i \;\; 1}^{t} RCorr_i$, $\tilde{R}Corr_t \quad RCorr_t \quad \overline{R}Corr$ and λ_l is the sequence of coeffi cients of a pure autoregressive (AR) representation of the following fractional integrated vector ARFIMA (0,d,0) process:

$$D(L)Y_t \quad \zeta_t, \qquad \zeta_t \sim N(0,\Omega), \tag{21}$$

where Y_t is the vector obtained by stacking the lower triangular portion of $\tilde{R}Corr_t$ *without* the main diagonal and $D(L) \quad (1 \quad L)^d I_m$, where m is the number of correlation series $m \quad n(n \quad 1)/2$. We label this model: VARFIMA(0,d,0). Thus, the parameters λ_l are derived from d as follows:

$$\lambda_l \quad (\quad 1)^l \binom{d}{l} \quad \frac{\Gamma(l \quad d)}{\Gamma(l+1)\Gamma(\quad d)} \quad \prod_{0<i\leq l} \frac{i \quad 1 \quad d}{i}, \quad l \quad 1,2,3\ldots \tag{22}$$

The selection of the long memory process is due to the persistence of the 435 series of daily realized correlations as displayed by the ACFs of the first 30 *RCorr* series depicted in Figure B.2.[11]

Due to the large dimension of Y_t ($m \quad 435$), we also consider a simple short memory vector AR process, as it follows:

$$Y_t \quad \psi Y_{t-1} + \xi_t, \qquad \xi_t \sim N(0,\Upsilon), \tag{23}$$

where $\quad 1 < \psi < 1$ to assure the stationarity of the Y_t process. We label this model: VARFIMA(1,0,0). It remains an empirical question whether a fractional autoregressive or a simple autoregressive model is better for forecasting realized correlations. We present em pirical results for both the VARFIMA(0,d,0) and VARFIMA(1,0,0) models. We note that the model specifications in Equations (21) and (23) are clearly not inferior to the DCC model in terms of flexibility.

10 Different specifications of GARCH are possible. However, as Hansen and Lunde (2005) show, the GARCH(1,1) is hard to beat within the GARCH class of models.

11 The ACFs of the other *RCorr* series are similar. They can be obtained from the authors upon request.

To provide further evidence, in addition to the realized measures computed from five minute returns (the frequency used in the volatility forecasts of MA), we also consider measures computed from 1 , 15 and 30 minute intra day returns.

2.3 Forecast Evaluation

We split the entire sample of data into an in sample period from November 1, 2001 to June 6, 2007 (1393 days) and an out of sample period from June 7, 2007 to November 27, 2013 (1609 days). The forecasts are carried out in a recursive manner, that is, at each step the models are re estimated with all of the available data. Given that the results for the proxies $RCov^{15-min}$ and $RCov^{30-min}$ do not significantly differ from those for $RCov^{5-min}$, we re frain from presenting them here. However, they can be obtained from the authors upon request.

The results are reported in Tables B.2 B.5 in Appendix B. Table B.2 reports the average value of the MSEs over all elements of the covariance matrix, where the proxy is the *RCov* computed on one and five minute returns and the cross product of vectors of daily returns ($RCoV^{daily}$), respectively. Tables B.3 and B.4 provide separate statistics for the 30 variance and 435 covariance forecasts, respectively, whereas Table B.5 reports the Frobenius norms for the three proxies. To compare the MSE results, we implement the model confidence set (MCS) of Hansen et al. (2011), which selects the set of models containing the best one with a certain degree of confidence.[12,13]

Overall, five conclusions can be drawn: (i) the 95% MCS appears to be consistent across the proxies built on five minute and daily returns. The instability of the set with $RCOV^{1-min}$ as the proxy may be explained by quality deterioration due to the increased market microstructure noise; (ii) the LFA always provides the largest MSEs, regardless of the proxy choice and it is almost never (with two exceptions in Tables B.4 and B.5) in the 95% MCS; (iii) averaging over all elements of the covariance matrix (465), as well as over the variances[14] (30) and the covariances (435), the 95% MCS includes both MAs and HFAs (using one and five minute intra day returns) regardless of the proxy choice that sta bilizes the rankings (built on five minute and daily returns); (iv) the same result holds for the matrix MSE; and (v) among the models of forecasting volatility, the HAR(1,5,10,20) provides the smallest MSEs, whereas among the models for forecasting correlations, the cDDC E approach is the method with the smallest MSEs for both the MA and the LFA. VARFIMA(0,d,0) is the best choice for the HFA.

The analysis across each element of the covariance matrix confirms the results presented above. Additionally, it provides evidence of the variation of the results across the matrix elements. Thus, although HFAs have the highest probability of being in the 95% MCS, they also exhibit the largest variance: the HFA using one minute data is the most likely

12 We implement the Ox package MulCom v2.00 provided by the authors.

13 For a robustness check, we also implement the Diebold Mariano test with the benchmark as the model with the smallest MSE. These results are similar to those of the MCS and, therefore, we refrain from reporting them here. They can be obtained from the authors upon request.

14 The HF variance forecasts are subject to bias when obtained naively from log volatility forecasts. Based on the results of Chiriac and Voev (2011), Lütkepohl and Xu (2013), and our empirical find ings from a previous version of the article, which show that the bias correction increases the MSE of variance forecasts regardless of the model choice, we opt for no bias correction.

(100%) to be included in the 95% MCS; lowering the frequency diminishes the probability of being part of the 95% MCS up to 13%. Comparably, the variation in the probability of the MA and the LFA being in the 95% MCS is much smaller: for the MA the probability varies between 70% and 100% and for the LFA between 50% and 99%.

Summing up the results, we can state that the MA and the HFAs computed on one and five minute intra day returns provide comparable MSEs, both smaller than the MSE of the LFA. Lowering the frequency worsens the performance of the HFA in comparison to the MA. This result is encouraging, especially when we consider larger dimensions. In addition to its good forecasting power, the MA is much easier to implement, given its straightfor ward correlation specification. The difficulty in the implementation of the HFA increases with the dimension of the system due to issues of nonsynchronicity and parameter prolifer ation. Thus, the results we provide should be viewed in light of the considerable simplicity of the MA in comparison to the HFA.

In the following, we report some empirical results on the necessary and sufficient conditions of Propositions 1, 2, and 3. To this end, we choose the volatility and correlation forecasts that perform best (see conclusions iii and v above) and the $RCOV^{5-min}$ as a proxy.[15]

The sufficient conditions of Proposition 1 are empirically validated when the SA is a pure LFA: the average estimate of the parameters $A_{1,i}$, $A_{2,i}$ and $A_{3,i}$ from Equation (13) are 0.197, 0.168, and 4.565, respectively. The sufficient conditions of nonpositivity of the parameters are completely fulfilled. As expected, on average, HF based forecasts provide gains in the MSEs and in the kurtosis of volatility forecast errors in comparison to LF based forecasts. Although, there is no particular theoretical reason for asymmetric forecast errors, the skewness is also negative.[16] Consequently, the MA provides smaller variance MSEs than the LFA. This result is confirmed by the entries in the tables.

In Figure B.3, we plot the average value of the polynomial in Equation (13) for a wide range of values of d_i, i.e, d_i $[0, 10]$, where the upper limit is chosen to approximately cor respond to an extreme value of 160% annualized volatility, along with the average 95% confidence intervals. To avoid any distributional assumptions on the forecast errors, we opt to approximate the confidence intervals using the bootstrapped percentile intervals as described in Efron and Tibshirani (1993, 170 173). We implement the stationary bootstrap procedure of Politis and Romano (1994b), which accounts for some degree of serial correl ation of the forecast errors. To account for cross correlation among different forecast errors, we bootstrap them in groups.

One can observe that for values of d larger than 1, both the parabola and the confidence bounds lie below the zero line, indicating that on average the MA provides significantly smaller mean squared forecast errors than the LFA, especially during periods of high vola tility. Moreover, one can observe the left asymmetry of the confidence intervals around the

15 Similar results for the other covariance matrix proxies can be obtained from the authors upon request.

16 Negative skewness means that the probability of positive forecast errors is larger than that of negative forecasts errors: that is, the forecasts are larger than the proxy in most cases. This "over forecasting" effect holds for all proxies.

mean value of the estimated polynomial, especially for small values of the true volatility. This suggests that we may have a larger probability of a negative than a positive MSE dif ference for values of d close to zero.

However, when considering the HFA using one minute data, the sufficient conditions of Proposition 1 are only partially empirically validated: the average estimates of the parameters $A_{1,i}$ and $A_{2,i}$ from Equation (13) are negative and equal to 0.006 and 0.34, respectively, and the average estimate of $\hat{A}_{3,i}$ is positive and equal to 1.95. Thus, increasing the frequency does not necessarily improve the MSE of HF based volatility forecasts (average of $\hat{A}_{1,i}$ is negative, but close to zero) due to the market microstructure noise, but it provides less extreme forecast errors (the average of $A_{3,i}$ is positive). The skewness is again negative, indicating an "over forecasting" of the variances. Therefore, an analysis with respect to the magnitude of market volatility, d is necessary to make conclusive statements.

Figure B.4 provides some evidence in this direction. It seems that for low values of d, increasing the frequency does not necessarily improve the variance MSEs, while for large values of d it actually worsens the MSE.

Thus, one may conclude that when the market is calm, both the MA and the HFA pro vide equally good variance forecasts. However, when the market is turbulent, the MA is preferable to the LFA (due to more informative underlying data), but also to the HFA (due to less market microstructure noise).

Next, we empirically verify the conditions of Proposition 2 regarding the MSE inequal ities between the MA and the SA covariance forecasts. On average, the sufficient conditions are validated: the parameters of the polynomial $G(d_i, d_j, \rho_{ij})$ are all negative for all ranking stabilizing proxies. Moreover, the parameters of the comparison between the MA and the HFA are on average negative, but closer to zero than the difference between the MA and the LFA. This indicates that the MA and the HFA perform equally well, while the MA pro vides smaller covariance MSEs than the LFA. These results are confirmed by the entries in Table B.4.

For a detailed analysis with respect to the market conditions of the covariance MSE comparison MA versus LFA and MA versus HFA in Figures B.5 and B.6 we plot the aver age estimate of the polynomial from Proposition 2 along with the average 95% percentile bootstrapped confidence intervals for d_i $[0, 10]$ and d_j $[0, 10]$ and for different values of ρ_{ij}. The choice of ρ_{ij} is motivated by the descriptive statistics of the *RCorr* computed on five minute returns over the whole window: on average, the daily realized correlation among all stocks is around 0.32, the maximum is 0.95 and the minimum is 0.57. As one can see from Figure B.5, the MA provides significantly smaller MSE in particular for large correlations (the surfaces lie below the zero bound) and large volatilities. The fact that the parameters of the polynomials given in Equations (31) and (32) are on average negative is primarily driven by the fact that the parameters describing the differences in the moments and cross moments of volatility forecast errors are in almost all cases negative and far from zero. This indicates that the MA is especially attractive in a highly volatile environment that is also generally characterized by high correlations.

With respect to the choice of ρ_{ij}, almost no difference between the graphs can be observed in Figure B.6. This is due to the fact that the associated parameters in Equations (31) and (32) are quite small: the gains from using HF based correlation forecasts are min imal compared to those using LF based forecasts.

Moreover, one can observe that half of the polynomial and confidence surfaces lie under the zero bound, which indicates better or equal performance of the MA in compari son to the HFA. Thus, when the market faces high volatility (one extreme, of around 160% annually and another one of around 80% annually), the MA provides better forecasts than the HFA. The graph further shows that, in the case of extreme volatility, the HFA out performs the MA. Thus, in such a situation, an investor should think twice before going through the computational burden specific to the HFA for a limited MSE gain in compari son to the MA.

The empirical validation of the conditions of Proposition 3 follows directly from the val idity of the conditions of the first two propositions. However, the reverse of this statement is not necessarily true. Although the conditions of the first two propositions may not hold in all cases, the MA can still provide smaller or equal matrix MSEs than the SA when sum ming over all stocks and pairs of stocks.

Thus, from Table B.5 we observe that the MA outperforms the LFA even when six out of the twenty two of the G parameters are positive for most pairs. This effect is outweighed by the large gain in volatility forecast MSE for all stocks obtained by using HF based fore casts ($\hat{A}_{1,i} < 0$ for each $i \quad 1, \ldots, 30$). Moreover, one can observe that the MA performs just as well as the HFA built on one minute data, although the estimated value of $G_{2,ij}$, which accounts for the difference in the correlation MSEs is positive for almost all pairs of stocks (as expected, HF based correlation forecasts have smaller MSEs than their LF coun terparts). This effect is, however, neutralized by the negativity of the other G parameters that account for dependencies between correlation and volatility forecast errors: $\hat{G}_{19,ij}$, $\hat{G}_{20,ij}$, $\hat{G}_{21,ij}$ and $\hat{G}_{22,ij}$ are negative for 80% of the pairs.

The fact that, we obtain in almost all cases negative G parameters that account for the dependencies between correlation and volatility forecast errors (up to 98% of the pairs) em phasizes the advantage of using the MA over the SA even though the SA may provide better individual volatility or correlation forecasts (as with the pure HFA).

Summing up the empirical results, we can conclude that the implemented MA is a better choice in forecasting covariance matrices than the LFA and the HFA for three reasons: (i) it has the advantage of using HF rather than LF data to estimate and forecast volatilities, which has already widely been recognized in the literature; (ii) it exploits the fact that mix ing forecasts empirically reduces the dependency between volatility and correlation forecast errors; and (iii) although it uses a less accurate correlation forecast (based on the outer product of daily returns, rather than daily realized correlation matrices), it has a substantial computational advantage over the HFA because it avoids the estimation issue and param eter proliferation in modeling the dynamics of large dimensional realized correlation matrices.

3 Conclusion

In this article, we introduce a new method for covariance matrix forecasting by theoretic ally and empirically exploiting the potential of mixing "independently" computed volatility and correlation forecasts. This method is based on a decomposition of the covariance ma trix into a diagonal matrix of volatilities and a correlation matrix. Through this decompos ition it becomes easy to employ HF data in volatility modeling, while still using a standard LF model for correlations in the spirit of the DCC model of Engle (2002).

In the theoretical section of the article, we derive the necessary and sufficient conditions for a MA of this kind to provide smaller mean squared forecast errors relative to a SA. The conditions do not depend on distributional assumptions of the forecast error or on particular model specifications that are easy to empirically check. This allows practi tioners to verify whether a MA is worth considering or not in a particular empirical implementation. We also provide some intuition regarding the results, as well as some general guidance about the environments in which the MA is likely to have the greatest advantages over a standard one, for example, market states with high/low volatility and correlation.

In the empirical section, we show how the theoretical results can be implemented in practice with an application involving thirty highly liquid assets. The analysis re veals that the MA decisively outperforms standard LF models and performs as well as pure HF models, which are much more cumbersome to implement and considerably more computationally costly, especially when the dimension of the covariance matrix becomes very large.

Acknowledgements

The authors would like to thank Luc Bauwens, Robert F. Engle, Christian Gourieroux, Asger Lunde, Andrew Patton, Jeroen Rombouts, Kevin Sheppard, the two anonymous referees, and Eric Ghysels (editor) for the useful comments and suggestions. All errors remain the authors' own. R.H. acknowledges financial support from the European Social Fund and from the Ministry of Science, Research and Arts of Baden Wurttemberg, Germany. V.V. gratefully acknowledges fi nancial support by the Center for Research in Econometric Analysis of Time Series, CREATES, funded by the Danish National Research Foundation.

Appendix A

In this appendix, to keep notational burden at a minimum, whenever we do not index variables by superscripts for models (*MA*, *SA*), it means that the equation holds for any model.

A.1. Preliminaries

We have that:

$$E[\varepsilon_{\sigma_{ii}}] \quad 2d_i E[\varepsilon_{d_i}] + E[\varepsilon_{d_i}^2] \tag{24}$$

$$\begin{aligned} E[\varepsilon_{\sigma_{ij}}] \quad & d_i \rho_{ij} E[\varepsilon_{d_j}] + d_j \rho_{ij} E[\varepsilon_{d_i}] + d_i d_j E[\varepsilon_{\rho_{ij}}] + d_i E[\varepsilon_{d_j} \varepsilon_{\rho_{ij}}] + d_j E[\varepsilon_{d_i} \varepsilon_{\rho_{ij}}] \\ & + \rho_{ij} E[\varepsilon_{d_i} \varepsilon_{d_j}] + E[\varepsilon_{d_i} \varepsilon_{\rho_{ij}} \varepsilon_{d_j}] \end{aligned} \tag{25}$$

$$\begin{aligned} V[\varepsilon_{\sigma_{ii}}] \quad & V[2d_i \varepsilon_{d_i} + \varepsilon_{d_i}^2] \quad 4d_i^2 V[\varepsilon_{d_i}] + V[\varepsilon_{d_i}^2] + 4d_i \mathrm{Cov}[\varepsilon_{d_i}, \varepsilon_{d_i}^2] \\ & 4d_i^2 V[\varepsilon_{d_i}] + (E[\varepsilon_{d_i}^4] \quad E[\varepsilon_{d_i}^2]^2) + 4d_i (E[\varepsilon_{d_i}^3] \quad E[\varepsilon_{d_i}] E[\varepsilon_{d_i}^2]) \end{aligned} \tag{26}$$

$$
\begin{aligned}
V[\varepsilon_{\sigma_{ij}}] \quad & d_i^2\rho_{ij}^2 V[\varepsilon_{d_j}] + d_j^2\rho_{ij}^2 V[\varepsilon_{d_i}] + d_i^2 d_j^2 V[\varepsilon_{\rho_{ij}}] + d_i^2 V[\varepsilon_{d_j}\varepsilon_{\rho_{ij}}] + d_j^2 V[\varepsilon_{d_i}\varepsilon_{\rho_{ij}}] \\
& + \rho_{ij}^2 V[\varepsilon_{d_i}\varepsilon_{d_j}] + V[\varepsilon_{d_i}\varepsilon_{\rho_{ij}}\varepsilon_{d_j}] \\
& + 2d_i d_j \rho_{ij}^2 \mathrm{Cov}[\varepsilon_{d_i}, \varepsilon_{d_j}] + 2d_i^2 d_j \rho_{ij} \mathrm{Cov}[\varepsilon_{d_j}, \varepsilon_{\rho_{ij}}] + 2d_i d_j^2 \rho_{ij} \mathrm{Cov}[\varepsilon_{d_i}, \varepsilon_{\rho_{ij}}] \\
& + 2d_i d_j \rho_{ij} \mathrm{Cov}[\varepsilon_{d_i}, \varepsilon_{d_j}\varepsilon_{\rho_{ij}}] + 2d_j^2 \rho_{ij} \mathrm{Cov}[\varepsilon_{d_i}, \varepsilon_{d_i}\varepsilon_{\rho_{ij}}] + 2d_j \rho_{ij}^2 \mathrm{Cov}[\varepsilon_{d_i}, \varepsilon_{d_i}\varepsilon_{d_j}] \\
& + 2d_i d_j \rho_{ij} \mathrm{Cov}[\varepsilon_{d_j}, \varepsilon_{d_i}\varepsilon_{\rho_{ij}}] + 2d_i^2 \rho_{ij} \mathrm{Cov}[\varepsilon_{d_j}, \varepsilon_{d_j}\varepsilon_{\rho_{ij}}] + 2d_i \rho_{ij}^2 \mathrm{Cov}[\varepsilon_{d_j}, \varepsilon_{d_i}\varepsilon_{d_j}] \\
& + 2d_i^2 d_j \mathrm{Cov}[\varepsilon_{\rho_{ij}}, \varepsilon_{d_j}\varepsilon_{\rho_{ij}}] + 2d_i d_j^2 \mathrm{Cov}[\varepsilon_{\rho_{ij}}, \varepsilon_{d_i}\varepsilon_{\rho_{ij}}] + 2d_i d_j \rho_{ij} \mathrm{Cov}[\varepsilon_{\rho_{ij}}, \varepsilon_{d_i}\varepsilon_{d_j}] \\
& + 2d_i \rho_{ij} \mathrm{Cov}[\varepsilon_{d_j}, \varepsilon_{d_i}\varepsilon_{\rho_{ij}}\varepsilon_{d_j}] + 2d_j \rho_{ij} \mathrm{Cov}[\varepsilon_{d_i}, \varepsilon_{d_i}\varepsilon_{\rho_{ij}}\varepsilon_{d_j}] + 2d_i d_j \mathrm{Cov}[\varepsilon_{\rho_{ij}}, \varepsilon_{d_i}\varepsilon_{\rho_{ij}}\varepsilon_{d_j}] \\
& + 2d_i d_j \mathrm{Cov}[\varepsilon_{d_j}\varepsilon_{\rho_{ij}}, \varepsilon_{d_i}\varepsilon_{\rho_{ij}}] + 2d_i \rho_{ij} \mathrm{Cov}[\varepsilon_{d_j}\varepsilon_{\rho_{ij}}, \varepsilon_{d_i}\varepsilon_{d_j}] + 2d_j \rho_{ij} \mathrm{Cov}[\varepsilon_{d_i}\varepsilon_{\rho_{ij}}, \varepsilon_{d_i}\varepsilon_{d_j}] \\
& + 2d_i \mathrm{Cov}[\varepsilon_{d_j}\varepsilon_{\rho_{ij}}, \varepsilon_{d_i}\varepsilon_{\rho_{ij}}\varepsilon_{d_j}] + 2d_j \mathrm{Cov}[\varepsilon_{d_i}\varepsilon_{\rho_{ij}}, \varepsilon_{d_i}\varepsilon_{\rho_{ij}}\varepsilon_{d_j}] \\
& + 2\rho_{ij} \mathrm{Cov}[\varepsilon_{d_i}\varepsilon_{d_j}, \varepsilon_{d_i}\varepsilon_{\rho_{ij}}\varepsilon_{d_j}].
\end{aligned}
\tag{27}
$$

A.2. Proofs

Proof A.1 (Proposition 1):
We have that

$$E[\varepsilon_{\sigma_{ii}}]^2 \quad 4d_i^2 E[\varepsilon_{d_i}]^2 + E[\varepsilon_{d_i}^2]^2 + 4d_i E[\varepsilon_{d_i}] E[\varepsilon_{d_i}^2]$$

Thus the MSE of $\hat{\sigma}_{ii}$ is given by

$$
\begin{aligned}
MSE(\hat{\sigma}_{ii}) \quad & E[\varepsilon_{\sigma_{ii}}]^2 + V[\varepsilon_{\sigma_{ii}}] \quad 4d_i^2 E[\varepsilon_{d_i}]^2 + E[\varepsilon_{d_i}^2]^2 + 4d_i E[\varepsilon_{d_i}] E[\varepsilon_{d_i}^2] \\
& +4d_i^2 (E[\varepsilon_{d_i}^2] \quad E[\varepsilon_{d_i}]^2) + (E[\varepsilon_{d_i}^4] \quad E[\varepsilon_{d_i}^2]^2) + 4d_i (E[\varepsilon_{d_i}^3] \quad E[\varepsilon_{d_i}] E[\varepsilon_{d_i}^2]) \\
& 4d_i^2 E[\varepsilon_{d_i}^2] + 4d_i E[\varepsilon_{d_i}^3] + E[\varepsilon_{d_i}^4]
\end{aligned}
$$

It follows that

$$
\begin{aligned}
MSE(\hat{\sigma}_{ii}^{MA}) \quad MSE(\hat{\sigma}_{ii}^{SA}) \quad & 4d_i^2 \left(E[(\varepsilon_{d_i}^{\mathcal{F}^{\mathcal{D}}})^2] \quad E[(\varepsilon_{d_i}^{\mathcal{F}})^2] \right) + 4d_i \left(E[(\varepsilon_{d_i}^{\mathcal{F}^{\mathcal{D}}})^3] \quad E[(\varepsilon_{d_i}^{\mathcal{F}})^3] \right) \\
& +(E[(\varepsilon_{d_i}^{\mathcal{F}^{\mathcal{D}}})^4] \quad E[(\varepsilon_{d_i}^{\mathcal{F}})^4]).
\end{aligned}
$$

Let

$$A_{1,i} \equiv E[(\varepsilon_{d_i}^{\mathcal{F}^{\mathcal{D}}})^2] \quad E[(\varepsilon_{d_i}^{\mathcal{F}})^2],\ A_{2,i} \equiv E[(\varepsilon_{d_i}^{\mathcal{F}^{\mathcal{D}}})^3] \quad E[(\varepsilon_{d_i}^{\mathcal{F}})^3],\ A_{3,i} \equiv E[(\varepsilon_{d_i}^{\mathcal{F}^{\mathcal{D}}})^4] \quad E[(\varepsilon_{d_i}^{\mathcal{F}})^4]. \tag{28}$$

Then we have that

$$MSE(\hat{\sigma}_{ii}^{MA}) \quad MSE(\hat{\sigma}_{ii}^{SA}) \quad 4A_{1,i}d_i^2 + 4A_{2,i}d_i + A_{3,i}. \tag{29}$$

Clearly, $E[(\varepsilon_{d_i}^{\mathcal{F}^{\mathcal{D}}})^2] \quad E[(\varepsilon_{d_i}^{\mathcal{F}})^2] \le 0$, $E[(\varepsilon_{d_i}^{\mathcal{F}^{\mathcal{D}}})^3] \quad E[(\varepsilon_{d_i}^{\mathcal{F}})^3] \le 0$ and $E[(\varepsilon_{d_i}^{\mathcal{F}^{\mathcal{D}}})^4] \quad E[(\varepsilon_{d_i}^{\mathcal{F}})^4] \le 0$ are sufficient conditions for $MSE(\hat{\sigma}_{ii}^{MA}) \le MSE(\hat{\sigma}_{ii}^{SA})$ since $d_i \ge 0$. To verify the necessary and sufficient conditions, consider that $4A_{1,i}d_i^2 + 4d_iA_{2,i} + A_{3,i} \le 0$ if and only if

1. $A_{1,i} \quad 0$, and $d_i \le \quad \frac{A_{3\,i}}{4A_{2,i}}$ or
2. $A_{1,i} > 0,\ A_{2,i}^2 \quad A_{1,i}A_{3,i} > 0$ and $d_i \in [d_{1,i}, d_{2,i}]$ where $d_{1/2,ij} \quad \frac{-A_{2,i} \mp \sqrt{A_{2,i}^2 - A_{1,i}A_{3,i}}}{2A_{1,i}}$ or

3. $A_{1,i} < 0$ and if either
   3.1. $A_{2,i}^2 \quad A_{1,i}A_{3,i} \leq 0$ or
   3.2. $A_{2,i}^2 \quad A_{1,i}A_{3,i} > 0$ and $d_i \notin (d_{1,i}, d_{2,i})$ where $d_{1/2,ij} \quad \frac{-A_{2,i} \mp \sqrt{A_{2,i}^2 - A_{1,i}A_{3,i}}}{2A_{1,i}}$.

The following table summarizes the necessary and sufficient conditions for $MSE(\hat{\sigma}_{ii}^{MA}) \leq MSE(\hat{\sigma}_{ii}^{SA})$ taking into account that $d_i \geq 0$:

$A_{1,i}$	$A_{2,i}$	$A_{3,i}$	d_i
≤ 0	≤ 0	≤ 0	$\forall d_i$
> 0	≤ 0	≤ 0	$d_i \leq \frac{A_{2,i} + \sqrt{A_{2,i}^2 \quad A_{1,i}A_{3,i}}}{2A_{1,i}}$
$\leq \frac{A_{2,i}^2}{A_{3,i}}$	> 0	≤ 0	$\forall d_i$
$0 > A_{1,i} > \frac{A_{2,i}^2}{A_{3,i}}$	> 0	≤ 0	$0 \leq d_i \leq \frac{A_{2,i} \quad \sqrt{A_{2,i}^2 \quad A_{1,i}A_{3,i}}}{2A_{1,i}}, \; d_i \geq \frac{A_{2,i} + \sqrt{A_{2,i}^2 \quad A_{1,i}A_{3,i}}}{2A_{1,i}}$
$= 0$	> 0	≤ 0	$0 \leq d_i \leq \quad \frac{A_{3,j}}{4A_{2,i}}$
> 0	> 0	≤ 0	$0 \leq d_i \leq \frac{A_{2,i} + \sqrt{A_{2,i}^2 \quad A_{1,i}A_{3,i}}}{2A_{1,i}}$
≥ 0	≥ 0	$= 0$	$d_i = 0$
< 0	< 0	> 0	$d_i \geq \frac{A_{2,i} + \sqrt{A_{2,i}^2 \quad A_{1,i}A_{3,i}}}{2A_{1,i}}$
$= 0$	< 0	> 0	$d_i \geq \quad \frac{A_{3,i}}{4A_{2,i}}$
$0 < A_{1,i} \leq \frac{A_{2,i}^2}{A_{3,i}}$	< 0	> 0	$\frac{A_{2,i} \quad \sqrt{A_{2,i}^2 \quad A_{1,i}A_{3,i}}}{2A_{1,i}} \leq d_i \leq \frac{A_{2,i} + \sqrt{A_{2,i}^2 \quad A_{1,i}A_{3,i}}}{2A_{1,i}}$
< 0	≥ 0	≥ 0	$d_i \geq \frac{A_{2,i} + \sqrt{A_{2,i}^2 \quad A_{1,i}A_{3,i}}}{2A_{1,i}}$

□

Proof A.2 (Proposition 2)**:**

Squaring the bias term in Equation (25), adding it to the variance expression in Equation (27) and simplifying (alternatively one can obtain the MSE directly as the expectation of the squared forecast error), one obtains

$$
\begin{aligned}
MSE[\hat{\sigma}_{ij}] \quad & d_i^2\rho_{ij}^2 E[\varepsilon_{d_j}^2] + d_j^2\rho_{ij}^2 E[\varepsilon_{d_i}^2] + d_i^2 d_j^2 E[\varepsilon_{\rho_{ij}}^2] + d_i^2 E[\varepsilon_{d_j}^2\varepsilon_{\rho_{ij}}^2] \\
& + d_j^2 E[\varepsilon_{d_i}^2\varepsilon_{\rho_{ij}}^2] + \rho_{ij}^2 E[\varepsilon_{d_i}^2\varepsilon_{d_j}^2] + 2d_i d_j \rho_{ij}^2 E[\varepsilon_{d_j}\varepsilon_{d_i}] + 2d_i^2 d_j \rho_{ij} E[\varepsilon_{d_j}\varepsilon_{\rho_{ij}}] \\
& + 2d_i^2\rho_{ij} E[\varepsilon_{d_j}^2\varepsilon_{\rho_{ij}}] + 6d_i d_j \rho_{ij} E[\varepsilon_{d_i}\varepsilon_{d_j}\varepsilon_{\rho_{ij}}] + 2d_i\rho_{ij}^2 E[\varepsilon_{d_i}\varepsilon_{d_j}^2] + 4d_i\rho_{ij} E[\varepsilon_{d_i}\varepsilon_{d_j}^2\varepsilon_{\rho_{ij}}] \\
& + 2d_i d_j^2 \rho_{ij} E[\varepsilon_{d_i}\varepsilon_{\rho_{ij}}] + 2d_j^2\rho_{ij} E[\varepsilon_{d_i}^2\varepsilon_{\rho_{ij}}] + 2d_j\rho_{ij}^2 E[\varepsilon_{d_i}^2\varepsilon_{d_j}] + 4d_j\rho_{ij} E[\varepsilon_{d_i}^2\varepsilon_{d_j}\varepsilon_{\rho_{ij}}] \\
& + 2d_i^2 d_j E[\varepsilon_{d_j}\varepsilon_{\rho_{ij}}^2] + 2d_i d_j^2 E[\varepsilon_{d_i}\varepsilon_{\rho_{ij}}^2] + 4d_i d_j E[\varepsilon_{d_i}\varepsilon_{d_j}\varepsilon_{\rho_{ij}}^2] + 2d_i E[\varepsilon_{d_i}\varepsilon_{d_j}^2\varepsilon_{\rho_{ij}}^2] \\
& + 2d_j E[\varepsilon_{d_i}^2\varepsilon_{d_j}\varepsilon_{\rho_{ij}}^2] + 2\rho_{ij} E[\varepsilon_{d_i}^2\varepsilon_{d_j}^2\varepsilon_{\rho_{ij}}] + E[\varepsilon_{d_i}^2\varepsilon_{d_j}^2\varepsilon_{\rho_{ij}}^2]
\end{aligned}
$$

Given this general expression for the MSE, we obtain that

$$
\begin{aligned}
MSE(\hat{\sigma}_{ij}^{MA}) \quad MSE(\hat{\sigma}_{ij}^{SA}) \quad & G_{1,j}d_i^2\rho_{ij}^2 + G_{1,i}d_j^2\rho_{ij}^2 + G_{2,ij}d_i^2d_j^2 + G_{3,ij}d_i^2 + G_{4,ij}d_j^2 \\
& + G_{5,ij}\rho_{ij}^2 + G_{6,ij}d_id_j\rho_{ij}^2 + G_{7,ij}d_i^2d_j\rho_{ij} + G_{8,ij}d_i^2\rho_{ij} + G_{9,ij}d_id_j\rho_{ij} \\
& + G_{10,ij}d_i\rho_{ij}^2 + G_{11,ij}d_i\rho_{ij} + G_{12,ij}d_id_j^2\rho_{ij} + G_{13,ij}d_j^2\rho_{ij} + G_{14,ij}d_j\rho_{ij}^2 \\
& + G_{15,ij}d_j\rho_{ij} + G_{16,ij}d_i^2d_j + G_{17,ij}d_id_j^2 + G_{18,ij}d_id_j + G_{19,ij}d_i + G_{20,ij}d_j \\
& + G_{21,ij}\rho_{ij} + G_{22,ij} \equiv G(d_i, d_j, \rho_{ij}),
\end{aligned}
\tag{30}
$$

where

$$
\begin{aligned}
& G_{1,i} \quad E[(\varepsilon_{d_j}^{\mathcal{F}^{\mathcal{D}}})^2] \quad E[(\varepsilon_{d_j}^{\mathcal{F}})^2] \\
& G_{2,ij} \quad E[(\varepsilon_{\rho_{ij}}^{\mathcal{F}^{\mathcal{R}}})^2] \quad E[(\varepsilon_{\rho_{ij}}^{\mathcal{F}})^2] \\
& G_{3,ij} \quad E[(\varepsilon_{d_j}^{\mathcal{F}^{\mathcal{D}}}\varepsilon_{\rho_{ij}}^{\mathcal{F}^{\mathcal{R}}})^2] \quad E[(\varepsilon_{d_j}^{\mathcal{F}}\varepsilon_{\rho_{ij}}^{\mathcal{F}})^2] \\
& G_{4,ij} \quad E[(\varepsilon_{d_i}^{\mathcal{F}^{\mathcal{D}}}\varepsilon_{\rho_{ij}}^{\mathcal{F}^{\mathcal{R}}})^2] \quad E[(\varepsilon_{d_i}^{\mathcal{F}}\varepsilon_{\rho_{ij}}^{\mathcal{F}})^2] \\
& G_{5,ij} \quad E[(\varepsilon_{d_i}^{\mathcal{F}^{\mathcal{D}}}\varepsilon_{d_j}^{\mathcal{F}^{\mathcal{D}}})^2] \quad E[(\varepsilon_{d_i}^{\mathcal{F}}\varepsilon_{d_j}^{\mathcal{F}})^2] \\
& G_{6,ij} \quad 2(E[\varepsilon_{d_j}^{\mathcal{F}^{\mathcal{D}}}\varepsilon_{d_i}^{\mathcal{F}^{\mathcal{D}}}] \quad E[\varepsilon_{d_j}^{\mathcal{F}}\varepsilon_{d_i}^{\mathcal{F}}]) \\
& G_{7,ij} \quad 2(E[\varepsilon_{d_j}^{\mathcal{F}^{\mathcal{D}}}\varepsilon_{\rho_{ij}}^{\mathcal{F}^{\mathcal{R}}}] \quad E[\varepsilon_{d_j}^{\mathcal{F}}\varepsilon_{\rho_{ij}}^{\mathcal{F}}]) \\
& G_{8,ij} \quad 2(E[(\varepsilon_{d_j}^{\mathcal{F}^{\mathcal{D}}})^2\varepsilon_{\rho_{ij}}^{\mathcal{F}^{\mathcal{R}}}] \quad E[(\varepsilon_{d_j}^{\mathcal{F}})^2\varepsilon_{\rho_{ij}}^{\mathcal{F}}]) \\
& G_{9,ij} \quad 6(E[\varepsilon_{d_i}^{\mathcal{F}^{\mathcal{D}}}\varepsilon_{d_j}^{\mathcal{F}^{\mathcal{D}}}\varepsilon_{\rho_{ij}}^{\mathcal{F}^{\mathcal{R}}}] \quad E[\varepsilon_{d_i}^{\mathcal{F}}\varepsilon_{d_j}^{\mathcal{F}}\varepsilon_{\rho_{ij}}^{\mathcal{F}}]) \\
& G_{10,ij} \quad 2(E[\varepsilon_{d_i}^{\mathcal{F}^{\mathcal{D}}}(\varepsilon_{d_j}^{\mathcal{F}^{\mathcal{D}}})^2] \quad E[\varepsilon_{d_i}^{\mathcal{F}}(\varepsilon_{d_j}^{\mathcal{F}})^2]) \\
& G_{11,ij} \quad 4(E[\varepsilon_{d_i}^{\mathcal{F}^{\mathcal{D}}}(\varepsilon_{d_j}^{\mathcal{F}^{\mathcal{D}}})^2\varepsilon_{\rho_{ij}}^{\mathcal{F}^{\mathcal{R}}}] \quad E[\varepsilon_{d_i}^{\mathcal{F}}(\varepsilon_{d_j}^{\mathcal{F}})^2\varepsilon_{\rho_{ij}}^{\mathcal{F}}]) \\
& G_{12,ij} \quad 2(E[\varepsilon_{d_i}^{\mathcal{F}^{\mathcal{D}}}\varepsilon_{\rho_{ij}}^{\mathcal{F}^{\mathcal{R}}}] \quad E[\varepsilon_{d_i}^{\mathcal{F}}\varepsilon_{\rho_{ij}}^{\mathcal{F}}]) \\
& G_{13,ij} \quad 2(E[(\varepsilon_{d_i}^{\mathcal{F}^{\mathcal{D}}})^2\varepsilon_{\rho_{ij}}^{\mathcal{F}^{\mathcal{R}}}] \quad E[(\varepsilon_{d_i}^{\mathcal{F}})^2\varepsilon_{\rho_{ij}}^{\mathcal{F}}]) \\
& G_{14,ij} \quad 2(E[(\varepsilon_{d_i}^{\mathcal{F}^{\mathcal{D}}})^2\varepsilon_{d_j}^{\mathcal{F}^{\mathcal{D}}}] \quad E[(\varepsilon_{d_i}^{\mathcal{F}})^2\varepsilon_{d_j}^{\mathcal{F}}]) \\
& G_{15,ij} \quad 4(E[(\varepsilon_{d_i}^{\mathcal{F}^{\mathcal{D}}})^2\varepsilon_{d_j}^{\mathcal{F}^{\mathcal{D}}}\varepsilon_{\rho_{ij}}^{\mathcal{F}^{\mathcal{R}}}] \quad E[(\varepsilon_{d_i}^{\mathcal{F}})^2\varepsilon_{d_j}^{\mathcal{F}}\varepsilon_{\rho_{ij}}^{\mathcal{F}}]) \\
& G_{16,ij} \quad 2(E[\varepsilon_{d_j}^{\mathcal{F}^{\mathcal{D}}}(\varepsilon_{\rho_{ij}}^{\mathcal{F}^{\mathcal{R}}})^2] \quad E[\varepsilon_{d_j}^{\mathcal{F}}(\varepsilon_{\rho_{ij}}^{\mathcal{F}})^2]) \\
& G_{17,ij} \quad 2(E[\varepsilon_{d_i}^{\mathcal{F}^{\mathcal{D}}}(\varepsilon_{\rho_{ij}}^{\mathcal{F}^{\mathcal{R}}})^2] \quad E[\varepsilon_{d_i}^{\mathcal{F}}(\varepsilon_{\rho_{ij}}^{\mathcal{F}})^2]) \\
& G_{18,ij} \quad 4(E[\varepsilon_{d_i}^{\mathcal{F}^{\mathcal{D}}}\varepsilon_{d_j}^{\mathcal{F}^{\mathcal{D}}}(\varepsilon_{\rho_{ij}}^{\mathcal{F}^{\mathcal{R}}})^2] \quad E[\varepsilon_{d_i}^{\mathcal{F}}\varepsilon_{d_j}^{\mathcal{F}}(\varepsilon_{\rho_{ij}}^{\mathcal{F}})^2]) \\
& G_{19,ij} \quad 2(E[\varepsilon_{d_i}^{\mathcal{F}^{\mathcal{D}}}(\varepsilon_{d_j}^{\mathcal{F}^{\mathcal{D}}}\varepsilon_{\rho_{ij}}^{\mathcal{F}^{\mathcal{R}}})^2] \quad E[\varepsilon_{d_i}^{\mathcal{F}}(\varepsilon_{d_j}^{\mathcal{F}}\varepsilon_{\rho_{ij}}^{\mathcal{F}})^2]) \\
& G_{20,ij} \quad 2(E[(\varepsilon_{d_i}^{\mathcal{F}^{\mathcal{D}}})^2\varepsilon_{d_j}^{\mathcal{F}^{\mathcal{D}}}(\varepsilon_{\rho_{ij}}^{\mathcal{F}^{\mathcal{R}}})^2] \quad E[(\varepsilon_{d_i}^{\mathcal{F}})^2\varepsilon_{d_j}^{\mathcal{F}}(\varepsilon_{\rho_{ij}}^{\mathcal{F}})^2]) \\
& G_{21,ij} \quad 2(E[(\varepsilon_{d_i}^{\mathcal{F}^{\mathcal{D}}}\varepsilon_{d_j}^{\mathcal{F}^{\mathcal{D}}})^2\varepsilon_{\rho_{ij}}^{\mathcal{F}^{\mathcal{R}}}] \quad E[(\varepsilon_{d_i}^{\mathcal{F}}\varepsilon_{d_j}^{\mathcal{F}})^2\varepsilon_{\rho_{ij}}^{\mathcal{F}}]) \\
& G_{22,ij} \quad E[(\varepsilon_{d_i}^{\mathcal{F}^{\mathcal{D}}}\varepsilon_{d_j}^{\mathcal{F}^{\mathcal{D}}}\varepsilon_{\rho_{ij}}^{\mathcal{F}^{\mathcal{R}}})^2] \quad E[(\varepsilon_{d_i}^{\mathcal{F}}\varepsilon_{d_j}^{\mathcal{F}}\varepsilon_{\rho_{ij}}^{\mathcal{F}})^2]
\end{aligned}
$$

Partial Derivatives of $G(d_i, d_j, \rho_{ij})$ and $H(d_i, d_j, \rho_{ij})$ wrt d_i and ρ_{ij}

$$\frac{\partial G(d_i, d_j, \rho_{ij})}{\partial d_i} \quad 2G_{1,j}d_i\rho_{ij}^2 + 2G_{2,ij}d_id_j^2 + 2G_{3,ij}d_i + G_{6,ij}d_j\rho_{ij}^2 + 2G_{7,ij}d_id_j\rho_{ij} + 2G_{8,ij}d_i\rho_{ij} + G_{9,ij}d_j\rho_{ij} + G_{10,ij}\rho_{ij}^2 + G_{11,ij}\rho_{ij} + G_{12,ij}d_j^2\rho_{ij} + 2G_{16,ij}d_id_j + G_{17,ij}d_j^2 + G_{18,ij}d_j + G_{19,ij} \tag{31}$$

$$\frac{\partial G(d_i, d_j, \rho_{ij})}{\partial \rho_{ij}} \quad 2G_{1,j}d_i^2\rho_{ij} + 2G_{2,i}d_j^2\rho_{ij} + 2G_{5,ij}\rho_{ij} + 2G_{6,ij}d_id_j\rho_{ij} + G_{7,ij}d_i^2d_j + G_{8,ij}d_i^2 + G_{9,ij}d_id_j + 2G_{10,ij}d_i\rho_{ij} + G_{11,ij}d_i + G_{12,ij}d_id_j^2 + G_{13,ij}d_j^2 + 2G_{14,ij}d_j\rho_{ij} + G_{15,ij}d_j + G_{21,ij} \tag{32}$$

$$\frac{\partial H(d_i, d_j, \rho_{ij})}{\partial d_i} \quad \sum_{i\ 1}^{n}\frac{\partial A(d_i)}{\partial d_i} + 2\sum_{i\ 1}^{n-1}\sum_{j>i}^{n}\frac{\partial G(d_i, d_j, \rho_{ij})}{\partial d_i} \quad 2A_{1,i}d_i + A_{2,i} + 2\sum_{i\ 1}^{n-1}\sum_{j>i}^{n}\frac{\partial G(d_i, d_j, \rho_{ij})}{\partial d_i} \tag{33}$$

$$\frac{\partial H(d_i, d_j, \rho_{ij})}{\partial \rho_{ij}} \quad 2\frac{\partial G(d_i, d_j, \rho_{ij})}{\partial \rho_{ij}} \tag{34}$$

where the expressions for A's and G's are given above.

Appendix B

Table B.1 Averaged descriptive statistics of daily realized variances, daily realized covariances computed from intra day returns sampled at 1 , 5 , 15 and 30 minutes and daily returns over the period November 1, 2001 to November 27, 2013 (3012 trading days)

Series	Mean	Max	Min	Std. dev.	Skewness	Kurtosis
Frequency: 1 minute						
Daily realized variances	3.211	199.445	0.223	7.085	13.340	343.218
Daily realized covariances	0.952	76.677	0.619	2.616	14.008	341.805
Frequency: 5 minutes						
Daily realized variances	2.916	188.278	0.156	6.774	13.397	335.571
Daily realized covariances	1.052	85.187	1.211	2.961	14.290	352.007
Frequency: 15 minutes						
Daily realized variances	2.607	158.252	0.099	6.035	11.160	232.017
Daily realized covariances	1.014	69.118	2.193	2.785	11.664	230.285
Frequency: 30 minutes						
Daily realized variances	2.415	158.879	0.064	5.881	11.256	232.136
Daily realized covariances	0.985	70.175	2.890	2.828	11.904	240.720
Frequency: daily						
Daily returns	0.000	12.048	12.104	1.603	0.034	11.069

The realized variance and covariance series are scaled by 10^4, while the daily returns series are scaled by 10^2.

Table B.2 Average values of the MSEs over all elements (465) of the covariance matrix forecasts

Proxy	Forecasting approach	Model for forecasting standard deviations		Model for forecasting correlations		
				cDCC	cDCC-E	cDCC-i
$RCOV^{1\ min}$	MA	ARFIMA(0,d,0)		13.34	13.27	13.34
		ARFIMA(1,d,0)		13.31	13.24	13.31
		ARFIMA(1,d,1)		12.86	12.81	12.87
		HAR(1,5,10,20)		12.67	12.63	12.68
	LFA	GARCH(1,1)		15.54	15.37	15.51
				VARFIMA (1,0,0)		VARFIMA (0,d,0)
	HFA	1-min	ARFIMA(0,d,0)	12.47		12.09
			ARFIMA(1,d,0)	12.41		12.03
			ARFIMA(1,d,1)	11.81		**11.49**
			HAR(1,5,10,20)	11.67		**11.37**
		5-min	ARFIMA(0,d,0)	13.09		12.60
			ARFIMA(1,d,0)	13.06		12.57
			ARFIMA(1,d,1)	12.55		12.11
			HAR(1,5,10,20)	12.32		11.92
		15-min	ARFIMA(0,d,0)	14.32		13.71
			ARFIMA(1,d,0)	14.32		13.71
			ARFIMA(1,d,1)	14.07		13.47
			HAR(1,5,10,20)	13.52		12.96
		30-min	ARFIMA(0,d,0)	15.54		14.89
			ARFIMA(1,d,0)	16.19		15.59
			ARFIMA(1,d,1)	15.39		14.74
			HAR(1,5,10,20)	14.55		13.92
				cDCC	cDCC-E	cDCC-i
$RCOV^{5\ min}$	MA	ARFIMA(0,d,0)		15.53	15.45	15.54
		ARFIMA(1,d,0)		15.51	**15.42**	15.51
		ARFIMA(1,d,1)		**15.07**	**14.99**	**15.08**
		HAR(1,5,10,20)		**14.89**	**14.82**	**14.90**
	LFA	GARCH(1,1)		17.84	17.70	17.83
				VARFIMA (1,0,0)		VARFIMA (0,d,0)
	HFA	1-min	ARFIMA(0,d,0)	**15.06**		**14.58**
			ARFIMA(1,d,0)	**15.02**		**14.54**
			ARFIMA(1,d,1)	**14.48**		**14.05**
			HAR(1,5,10,20)	**14.36**		**13.95**
		5-min	ARFIMA(0,d,0)	15.38		**14.78**
			ARFIMA(1,d,0)	15.36		**14.76**
			ARFIMA(1,d,1)	**14.87**		**14.30**
			HAR(1,5,10,20)	**14.65**		**14.12**
		15-min	ARFIMA(0,d,0)	16.55		15.83
			ARFIMA(1,d,0)	16.55		15.83
			ARFIMA(1,d,1)	16.32		15.60
			HAR(1,5,10,20)	15.76		**15.07**
		30-min	ARFIMA(0,d,0)	17.76		17.02
			ARFIMA(1,d,0)	18.45		17.77
			ARFIMA(1,d,1)	17.62		16.87
			HAR(1,5,10,20)	16.78		16.04
				cDCC	cDCC-E	cDCC-i
$RCOV^{daily}$	MA	ARFIMA(0,d,0)		65.62	65.48	65.63
		ARFIMA(1,d,0)		65.56	**65.42**	65.57
		ARFIMA(1,d,1)		**64.90**	**64.76**	**64.90**
		HAR(1,5,10,20)		**64.49**	**64.35**	**64.50**
	LFA	GARCH(1,1)		66.94	66.90	66.95

(continued)

Table B.2 (continued)

Proxy	Forecasting approach	Model for forecasting standard deviations		Model for forecasting correlations	
				VARFIMA (1,0,0)	VARFIMA (0,d,0)
	HFA	1-min	ARFIMA(0,d,0)	65.78	**65.14**
			ARFIMA(1,d,0)	65.69	**65.05**
			ARFIMA(1,d,1)	**64.93**	**64.33**
			HAR(1,5,10,20)	**64.62**	**64.02**
		5-min	ARFIMA(0,d,0)	65.93	**65.21**
			ARFIMA(1,d,0)	65.87	**65.16**
			ARFIMA(1,d,1)	65.20	**64.51**
			HAR(1,5,10,20)	**64.78**	**64.10**
		15-min	ARFIMA(0,d,0)	67.37	66.55
			ARFIMA(1,d,0)	67.37	66.55
			ARFIMA(1,d,1)	67.03	66.19
			HAR(1,5,10,20)	66.33	65.52
		30-min	ARFIMA(0,d,0)	68.84	68.02
			ARFIMA(1,d,0)	69.46	68.70
			ARFIMA(1,d,1)	68.65	67.81
			HAR(1,5,10,20)	67.65	66.81

The bold entries are in the 95% MCS.

Table B.3 Average values of the MSEs over all variances (30) of the covariance matrix forecasts

Proxy	Forecasting approach	Model for forecasting standard deviations		
$RCOV^{1\ min}$	MA	ARFIMA(0,d,0)		84.82
		ARFIMA(1,d,0)		84.55
		ARFIMA(1,d,1)		80.17
		HAR(1,5,10,20)		79.41
	LFA	GARCH(1,1)		97.88
	HFA	1-min	ARFIMA(0,d,0)	78.38
			ARFIMA(1,d,0)	77.81
			ARFIMA(1,d,1)	**73.37**
			HAR(1,5,10,20)	**73.13**
		15-min	ARFIMA(0,d,0)	92.73
			ARFIMA(1,d,0)	92.73
			ARFIMA(1,d,1)	91.07
			HAR(1,5,10,20)	88.03
		30-min	ARFIMA(0,d,0)	101.35
			ARFIMA(1,d,0)	104.53
			ARFIMA(1,d,1)	101.72
			HAR(1,5,10,20)	96.07
$RCOV^{5\ min}$	MA	ARFIMA(0,d,0)		**77.59**
		ARFIMA(1,d,0)		**77.43**
		ARFIMA(1,d,1)		**74.38**
		HAR(1,5,10,20)		**74.37**
	LFA	GARCH(1,1)		95.87
	HFA	1-min	ARFIMA(0,d,0)	**74.35**
			ARFIMA(1,d,0)	**74.06**
			ARFIMA(1,d,1)	**71.97**
			HAR(1,5,10,20)	**72.47**
		15-min	ARFIMA(0,d,0)	83.23
			ARFIMA(1,d,0)	83.23
			ARFIMA(1,d,1)	**82.06**
			HAR(1,5,10,20)	**80.05**

(continued)

Table B.3 (continued)

Proxy	Forecasting approach	Model for forecasting standard deviations		
		30-min	ARFIMA(0,d,0)	90.22
			ARFIMA(1,d,0)	93.06
			ARFIMA(1,d,1)	90.69
			HAR(1,5,10,20)	86.33
$RCOV^{daily}$	MA	ARFIMA(0,d,0)		**227.15**
		ARFIMA(1,d,0)		**226.71**
		ARFIMA(1,d,1)		**222.08**
		HAR(1,5,10,20)		**220.82**
	LFA	GARCH(1,1)		239.66
	HFA	1-min	ARFIMA(0,d,0)	**223.14**
			ARFIMA(1,d,0)	**222.50**
			ARFIMA(1,d,1)	**218.69**
			HAR(1,5,10,20)	**217.60**
		15-min	ARFIMA(0,d,0)	234.58
			ARFIMA(1,d,0)	234.58
			ARFIMA(1,d,1)	232.80
			HAR(1,5,10,20)	**230.07**
		30-min	ARFIMA(0,d,0)	243.17
			ARFIMA(1,d,0)	245.65
			ARFIMA(1,d,1)	243.81
			HAR(1,5,10,20)	238.45

The bold entries are in the 95% MCS.

Table B.4 Average values of the MSEs over all covariances (435) of the covariance matrix forecasts

Proxy	Forecasting approach	Model for forecasting standard deviations		Model for forecasting correlations		
				cDCC	cDCC-E	cDCC-i
$RCOV^{1\ min}$	MA	ARFIMA(0,d,0)		8.41	8.34	8.41
		ARFIMA(1,d,0)		8.39	8.32	8.40
		ARFIMA(1,d,1)		8.22	8.17	8.22
		HAR(1,5,10,20)		8.07	8.02	8.07
	LFA	GARCH(1,1)		9.86	9.68	9.83
				VARFIMA (1,0,0)		VARFIMA (0,d,0)
	HFA	1-min	ARFIMA(0,d,0)	7.92		**7.52**
			ARFIMA(1,d,0)	7.90		**7.49**
			ARFIMA(1,d,1)	**7.56**		**7.22**
			HAR(1,5,10,20)	**7.43**		**7.11**
		5-min	ARFIMA(0,d,0)	8.14		**7.62**
			ARFIMA(1,d,0)	8.12		**7.61**
			ARFIMA(1,d,1)	7.88		**7.42**
			HAR(1,5,10,20)	**7.69**		**7.26**
		15-min	ARFIMA(0,d,0)	8.91		8.27
			ARFIMA(1,d,0)	8.91		**8.27**
			ARFIMA(1,d,1)	8.76		8.12
			HAR(1,5,10,20)	8.38		7.78
		30-min	ARFIMA(0,d,0)	9.62		8.93
			ARFIMA(1,d,0)	10.10		9.46
			ARFIMA(1,d,1)	9.44		8.75
			HAR(1,5,10,20)	8.92		8.26

(continued)

Table B.4 (continued)

Proxy	Forecasting approach	Model for forecasting standard deviations		Model for forecasting correlations		
				cDCC	cDCC-E	cDCC-i
		ARFIMA(0,d,0)		**11.25**	**11.16**	**11.26**
	MA	ARFIMA(1,d,0)		**11.24**	**11.14**	**11.24**
		ARFIMA(1,d,1)		**10.98**	**10.90**	**10.99**
		HAR(1,5,10,20)		**10.79**	**10.71**	**10.79**
	LFA	GARCH(1,1)		12.46	12.31	12.44
				VARFIMA (1,0,0)		VARFIMA (0,d,0)
		1-min	ARFIMA(0,d,0)	**10.98**		**10.46**
			ARFIMA(1,d,0)	**10.95**		**10.43**
			ARFIMA(1,d,1)	**10.52**		**10.05**
			HAR(1,5,10,20)	**10.35**		**9.91**
$RCOV^{5\ min}$		5-min	ARFIMA(0,d,0)	**11.09**		**10.45**
			ARFIMA(1,d,0)	**11.07**		**10.43**
			ARFIMA(1,d,1)	**10.76**		**10.16**
			HAR(1,5,10,20)	**10.53**		**9.96**
	HFA	15-min	ARFIMA(0,d,0)	11.95		**11.18**
			ARFIMA(1,d,0)	11.95		**11.18**
			ARFIMA(1,d,1)	11.78		**11.02**
			HAR(1,5,10,20)	11.33		**10.59**
		30-min	ARFIMA(0,d,0)	12.76		11.97
			ARFIMA(1,d,0)	13.31		12.58
			ARFIMA(1,d,1)	12.58		11.78
			HAR(1,5,10,20)	11.98		**11.19**
				cDCC	cDCC-E	cDCC-i
		ARFIMA(0,d,0)		54.48	**54.33**	54.49
	MA	ARFIMA(1,d,0)		**54.44**	**54.29**	**54.45**
		ARFIMA(1,d,1)		**54.05**	**53.91**	**54.06**
		HAR(1,5,10,20)		**53.71**	**53.56**	**53.72**
	LFA	GARCH(1,1)		**55.03**	**54.99**	**55.04**
				VARFIMA (1,0,0)		VARFIMA (0,d,0)
		1-min	ARFIMA(0,d,0)	54.92		**54.24**
			ARFIMA(1,d,0)	54.87		**54.19**
			ARFIMA(1,d,1)	**54.33**		**53.68**
			HAR(1,5,10,20)	**54.07**		**53.43**
$RCOV^{daily}$		5-min	ARFIMA(0,d,0)	54.82		**54.05**
			ARFIMA(1,d,0)	54.78		**54.01**
			ARFIMA(1,d,1)	54.38		**53.65**
			HAR(1,5,10,20)	**54.02**		**53.29**
	HFA	15-min	ARFIMA(0,d,0)	55.84		54.96
			ARFIMA(1,d,0)	55.84		54.96
			ARFIMA(1,d,1)	55.59		54.70
			HAR(1,5,10,20)	55.04		**54.17**
		30-min	ARFIMA(0,d,0)	56.82		55.95
			ARFIMA(1,d,0)	57.31		56.49
			ARFIMA(1,d,1)	56.57		55.67
			HAR(1,5,10,20)	55.87		54.97

The bold entries are in the 95% MCS.

Table B.5 Average values of matrix MSEs of the covariance matrix forecasts

Proxy	Forecasting approach	Model for forecasting standard deviations		Model for forecasting correlations		
				cDCC	cDCC-E	cDCC-i
$RCOV^{1\ min}$	MA	ARFIMA(0,d,0)		9858.78	9797.40	9863.15
		ARFIMA(1,d,0)		9839.52	9778.38	9843.88
		ARFIMA(1,d,1)		9555.99	9508.89	9559.83
		HAR(1,5,10,20)		9403.75	9359.38	9407.12
	LFA	GARCH(1,1)		11511.23	11355.01	11489.68
				VARFIMA (1,0,0)		VARFIMA (0,d,0)
	HFA	1-min	ARFIMA(0,d,0)	9245.75		8890.35
			ARFIMA(1,d,0)	9207.88		8854.31
			ARFIMA(1,d,1)	8780.52		**8481.38**
			HAR(1,5,10,20)	**8655.26**		**8381.36**
		5-min	ARFIMA(0,d,0)	9626.44		9176.46
			ARFIMA(1,d,0)	9605.18		9156.26
			ARFIMA(1,d,1)	9264.22		8856.24
			HAR(1,5,10,20)	9074.03		8701.41
		15-min	ARFIMA(0,d,0)	10537.77		9972.62
			ARFIMA(1,d,0)	10537.94		9972.79
			ARFIMA(1,d,1)	10353.91		9797.57
			HAR(1,5,10,20)	9929.80		9408.55
		30-min	ARFIMA(0,d,0)	11407.73		10809.69
			ARFIMA(1,d,0)	11920.71		11366.67
			ARFIMA(1,d,1)	11260.42		10660.20
			HAR(1,5,10,20)	10646.18		10065.38
				cDCC	cDCC-E	cDCC-i
$RCOV^{5\ min}$	MA	ARFIMA(0,d,0)		**12119.10**	**12036.88**	**12124.70**
		ARFIMA(1,d,0)		**12100.25**	**12018.20**	**12105.85**
		ARFIMA(1,d,1)		**11783.31**	**11713.38**	**11788.49**
		HAR(1,5,10,20)		**11618.04**	**11550.88**	**11622.76**
	LFA	GARCH(1,1)		13717.14	13582.26	13702.01
				VARFIMA (1,0,0)		VARFIMA (0,d,0)
	HFA	1-min	ARFIMA(0,d,0)	**11778.97**		**11328.99**
			ARFIMA(1,d,0)	**11744.85**		**11296.00**
			ARFIMA(1,d,1)	**11310.82**		**10905.03**
			HAR(1,5,10,20)	**11180.25**		**10794.96**
		5-min	ARFIMA(0,d,0)	**11978.27**		**11419.31**
			ARFIMA(1,d,0)	**11957.63**		**11399.41**
			ARFIMA(1,d,1)	**11594.63**		**11069.12**
			HAR(1,5,10,20)	**11392.85**		**10896.31**
		15-min	ARFIMA(0,d,0)	**12895.22**		**12226.78**
			ARFIMA(1,d,0)	**12895.38**		**12226.94**
			ARFIMA(1,d,1)	**12712.65**		**12050.53**
			HAR(1,5,10,20)	**12256.30**		**11617.91**
		30-min	ARFIMA(0,d,0)	13806.51		13118.17
			ARFIMA(1,d,0)	14371.34		13738.26
			ARFIMA(1,d,1)	13662.85		12970.18
			HAR(1,5,10,20)	13013.00		12327.65
				cDCC	cDCC-E	cDCC-i
$RCOV^{daily}$	MA	ARFIMA(0,d,0)		54210.60	54079.69	54218.47
		ARFIMA(1,d,0)		54167.84	**54036.43**	54175.72
		ARFIMA(1,d,1)		**53689.95**	**53567.09**	**53697.46**
		HAR(1,5,10,20)		**53351.47**	**53224.22**	**53358.77**
	LFA	GARCH(1,1)		55069.07	**55028.84**	55071.84
				VARFIMA (1,0,0)		VARFIMA (0,d,0)
	HFA	1-min	ARFIMA(0,d,0)	54478.78		**53885.07**
			ARFIMA(1,d,0)	54414.72		**53821.86**

(continued)

Table B.5 (continued)

Proxy	Forecasting approach	Model for forecasting standard deviations		Model for forecasting correlations	
			ARFIMA(1,d,1)	**53825.32**	**53265.22**
			HAR(1,5,10,20)	**53566.68**	**53010.51**
		5-min	ARFIMA(0,d,0)	54503.75	**53834.48**
			ARFIMA(1,d,0)	54461.79	**53793.13**
			ARFIMA(1,d,1)	53974.97	**53335.67**
			HAR(1,5,10,20)	**53619.78**	**52990.66**
		15-min	ARFIMA(0,d,0)	55620.98	54852.79
			ARFIMA(1,d,0)	55620.95	54852.74
			ARFIMA(1,d,1)	55351.18	54576.94
			HAR(1,5,10,20)	54789.09	**54027.02**
		30-min	ARFIMA(0,d,0)	56729.93	55968.09
			ARFIMA(1,d,0)	57230.45	56519.40
			ARFIMA(1,d,1)	56529.22	55746.17
			HAR(1,5,10,20)	55761.13	54978.48

The bold entries are in the 95% MCS.

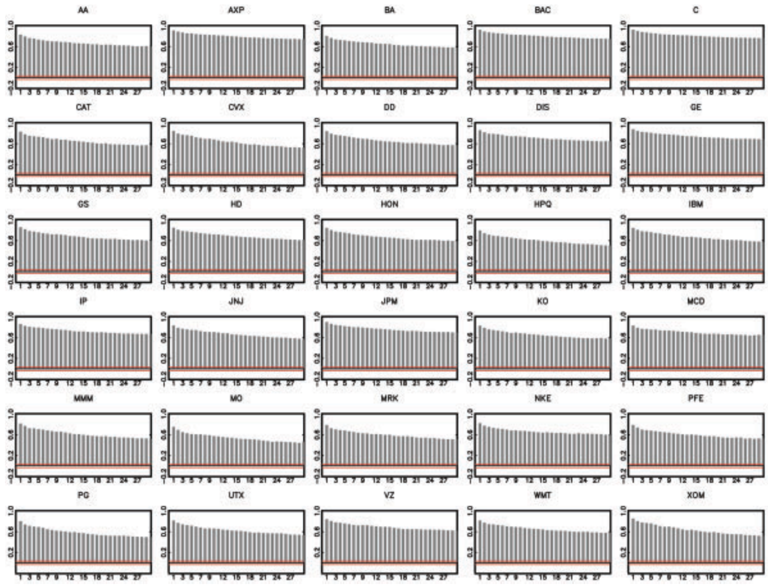

Figure B.1 ACF of daily $RCov^{5\ min}$ from November 1, 2001 to November 27, 2013 (3012 trading days). The dotted lines represent the 95% confidence intervals. On X axis we plot lags from 1 to 30.

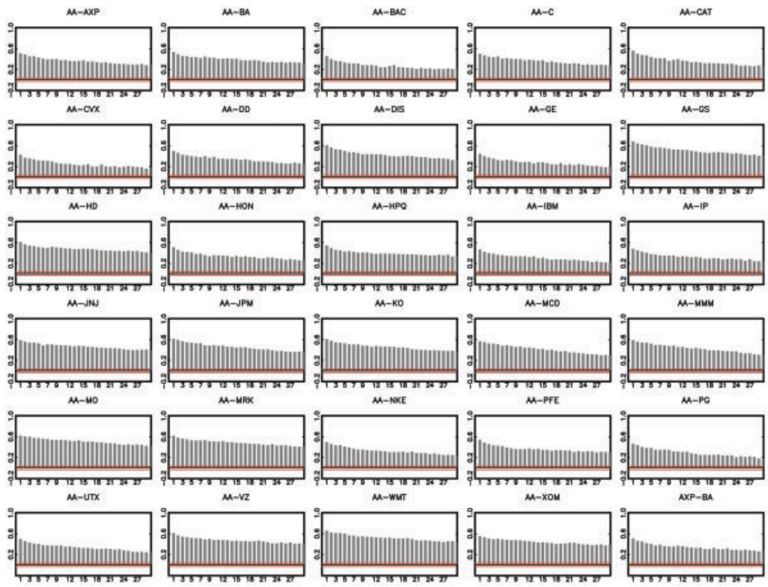

Figure B.2 ACF of the first thirty daily *RCorr* computed from $RCov^{5\ min}$ from November 1, 2001 to November 27, 2013 (3012 trading days). The dotted lines represent the 95% confidence intervals. On X axis we plot lags from 1 to 30.

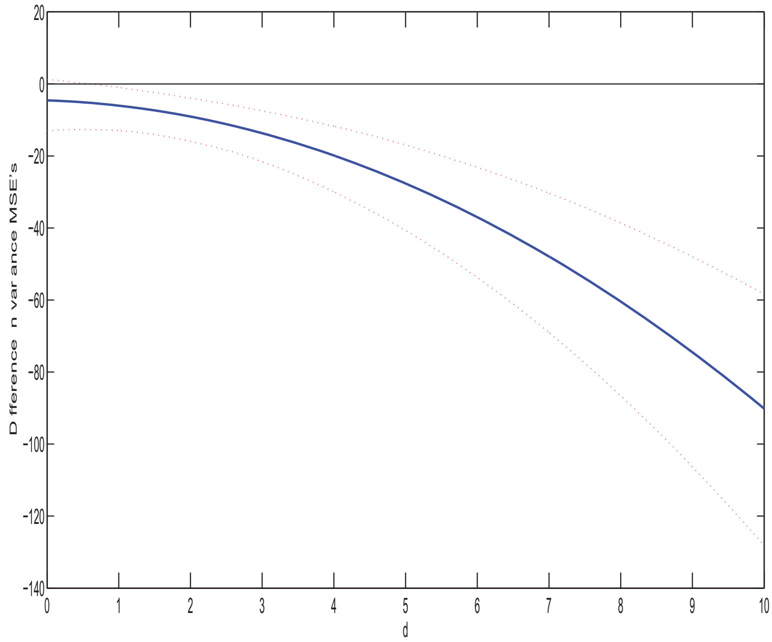

Figure B.3 Average value of $\widehat{MSE}(\widehat{\sigma}_{ii}^{MA}) \quad \widehat{MSE}(\widehat{\sigma}_{ii}^{LFA}) = 4\widehat{A}_{1,i}d_i^2 + 4\widehat{A}_{2,i}d_i + \widehat{A}_{3,i}$. The average value of the polynomial is given by the thick solid line. The dotted lines represent the 95% confidence intervals and the simple solid line represents the zero bound. On X axis we plot the true standard deviation, d_i that ranges from 0 to 10 and generically given by the symbol d.

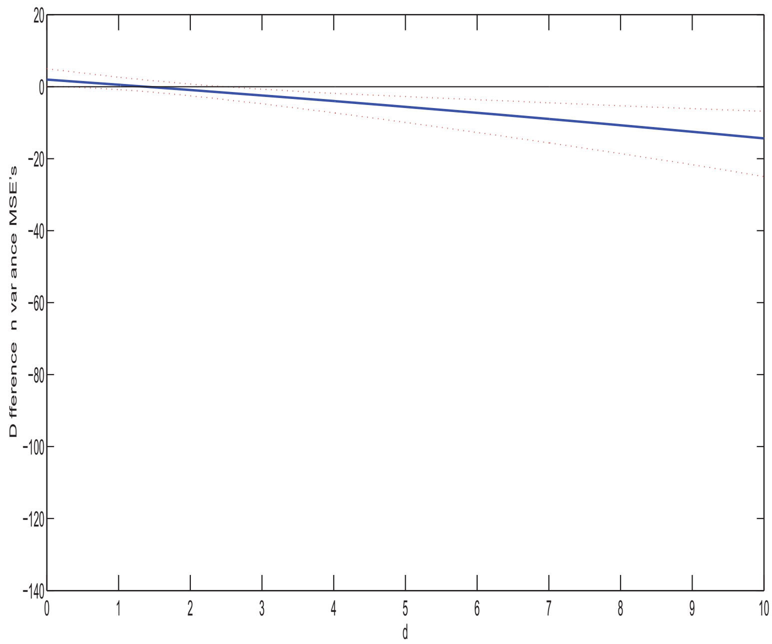

Figure B.4 Average value of $\widehat{MSE}(\hat{\sigma}_{ii}^{MA})$ $\widehat{MSE}(\hat{\sigma}_{ii}^{HFA}) = 4\hat{A}_{1,i}d_i^2 + 4\hat{A}_{2,i}d_i + \hat{A}_{3,i}$. The average value of the polynomial is given by the thick solid line. The dotted lines represent the 95% confidence intervals and the simple solid line represents the zero bound. On X axis we plot the true standard deviation, d_i that ranges from 0 to 10 and generically given by the symbol d.

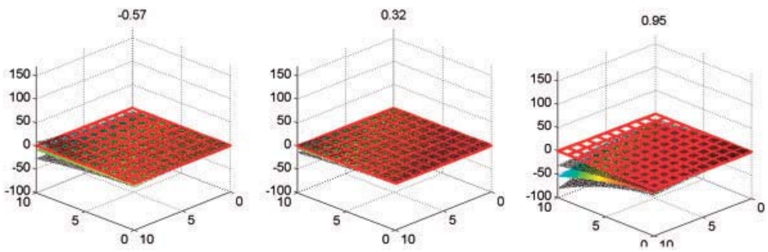

Figure B.5 Average value of $\widehat{MSE}(\hat{\sigma}_{ij}^{MA})$ $\widehat{MSE}(\hat{\sigma}_{ij}^{LFA})$. The average value of the polynomial is given by the rough surface. The dotted surfaces represent the 95% confidence intervals and the surface with rectangular pattern represents the zero bound. On X axis and Y axis we plot the true standard devi ations, d_i and d_j, respectively, that range from 0 to 10. On Z axis we plot the differences in MSE's. The left graph is computed for $\rho_{ij} = 0.57$, the middle graph for $\rho_{ij} = 0.32$ and the right graph for $\rho_{ij} = 0.95$.

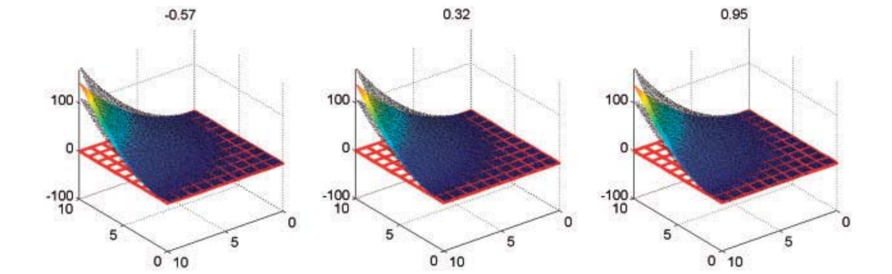

Figure B.6 Average value of $\widehat{MSE}(\hat{\sigma}_{ij}^{MA})$ $\widehat{MSE}(\hat{\sigma}_{ij}^{HFA})$, where HFA is built on one minute data. The average value of the polynomial is given by the rough surface. The dotted surfaces represent the 95% confidence intervals and the surface with rectangular pattern represents the zero bound. On *X* axis and *Y* axis we plot the true standard deviations, d_i and d_j, respectively, that range from 0 to 10. On *Z* axis we plot the differences in MSE's. The left graph is computed for ρ_{ij}= 0.57, the middle graph for ρ_{ij}= 0.32 and the right graph for ρ_{ij}= 0.95.

References

Aielli, G. P. 2013. Dynamic Conditional Correlations: On Properties and Estimation. *Journal of Business & Economic Statistics* 31: 282 299.

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 Realized Exchange Rate Volatility. *Journal of the American Statistical Association* 96: 42 55.

Bannouh, K., M. Martens, R. Oomen, and D. van Dijk. 2012. Realized Mixed Frequency Factor Models for Vast Dimensional Covariance Estimation. Discussion Paper ERS 2012 017 F&A, Erasmus University Rotterdam, ERIM Report Series Research in Management. http://hdl.handle.net/1765/37470.

Barndorff Nielsen, O. E., P. Hansen, A. Lunde, and N. Shephard. 2008. Designing Realized Kernels to Measure the Ex Post Variation of Equity Prices in the Presence of Noise. *Econometrica* 76: 1481 1536.

Barndorff Nielsen, O. E., P. Hansen, A. Lunde, and N. Shephard. 2009. Realized Kernels in Practice: Trades and Quotes. *Econometrics Journal* 12: C1 C32.

Barndorff Nielsen, O. E., P. Hansen, A. Lunde, and N. Shephard. 2011. Multivariate Realised Kernels: Consistent Positive Semi Definite Estimators of the Covariation of Equity Prices with Noise and Non Synchronous Trading. *Journal of Econometrics* 162: 149 169.

Bauer, G. H. and K. Vorkink. 2011. Forecasting Multivariate Realized Stock Market Volatility. *Journal of Econometrics* 160: 93 101.

Bauwens, L., S. Laurent, and J. Rombouts. 2006. Multivariate GARCH Models: A Survey. *Journal of Applied Econometrics* 21: 79 109.

Bauwens, L., G. Storti, and F. Violante. 2013. "Dynamic Conditional Correlation Models for Realized Covariance Matrices". ECORE Discussion Paper 2012/60. www.ecore.be/DPs/dp 1359373419.pdf.

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.

Bollerslev, T. 1986. Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* 31: 307 327.

Chen, X. and E. Ghysels. 2011. News Good or Bad and its Impact on Volatility Predictions over Multiple Horizons. *The Review of Financial Studies* 24: 46 81.

Chen, X., E. Ghysels, and F. Wang. 2011. Hybrid GARCH Models and Intra Daily Return Periodicity. *Journal of Time Series Econometrics* 3: 1 28.

Chen, X., E. Ghysels, and F. Wang. 2014. “Hybrid GARCH: A Generic Class of Models for Volatility Predictions Using Mixed Frequency Data”. Statistica Sinica Preprint No SS 12 283R3. http://www.stat.sinica.edu.tw/statictica.

Chiriac, R. and V. Voev. 2011. Modelling and Forecasting Multivariate Realized Volatility. *Journal of Applied Econometrics* 26: 922 947.

Colacito, R., R. Engle, and E. Ghysels. 2011. A Component Model for Dynamic Correlations. *Journal of Econometrics* 164: 45 59.

Corsi, F. 2009. A Simple Approximate Long Memory Model of Realized Volatility. *Journal of Financial Econometrics* 7: 174 196.

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

Efron, B. and R. Tibshirani. 1993. *An Introduction to the Bootstrap*. New York: Chapman and Hall.

Engle, R. 2002. Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroscedasticity Models. *Journal of Business and Economic Statistics* 20: 339 350.

Engle, R. F. 1982. Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation. *Econometrica* 50: 987 1007.

Engle, R., N. Shephard, and K. Sheppard. 2008. “Fitting Vast Dimensional Time Varying Covariance Models”. NYU Working Paper No. FIN 08 009. www.economics.px.ac.uk/materi als/working papers/paper403.pdf.

Ghysels, E., P. Santa Clara, and R. Valkanov. 2005. There is a Risk Return Tradeoff After All. *Journal of Financial Economics* 76: 509 548.

Ghysels, E., P. Santa Clara, and R. Valkanov. 2006. Predicting Volatility: Getting the Most Out of Return Data Sampled at Different Frequencies. *Journal of Econometrics* 131: 59 95.

Ghysels, E. and R. Valkanov. 2011. “Forecasting Volatility with Midas”. In L. Bauwens, C. Haffner and S. Laurent (eds.), *Volatility Models and their Applications*. New York: Wiley.

Giacomini, R. and H. White. 2006. Tests of Conditional Predictive Ability. *Econometrica* 74: 1545 1578.

Gouriéroux, C., J. Jasiak, and R. Sufana. 2009. The Wishart Autoregressive Process of Multivariate Stochastic Volatility. *Journal of Econometrics* 150: 167 181.

Hafner, C. and O. Reznikova. 2012. On the Estimation of Dynamic Conditional Correlation Models. *Computational Statistics and Data Analysis* 56: 3533 3545.

Halbleib, R. and V. Voev. 2011. Forecasting Multivariate Volatility Using the VARFIMA Model on Realized Covariance Cholesky Factors. *Journal of Economics and Statistics (Jahrbücher für Nationalökonomie und Statistik)* 231: 134 152.

Hansen, P. and A. Lunde. 2011. “Forecasting Volatility Using High Frequency Data”. In M. P. Clements and D. F. Hendry (eds.), *Oxford Handbook of Economic Forecasting*. Oxford, Oxford University Press, pp. 525 556.

Hansen, P. R. and A. Lunde. 2005. A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)? *Journal of Applied Econometrics* 20: 873 889.

Hansen, P. R., A. Lunde, and J. M. Nason. 2011. The Model Confidence Set. *Econometrica* 79: 453 497.

Hautsch, N., L. M. Kyj, and P. Malec. 2011. The Merit of High Frequency Data in Portfolio Allocation. SFB649 Discussion Paper. SFB649DP2011 059, Humboldt University, Berlin.

Hautsch, N., L. M. Kyj, and R. Oomen. 2012. A Blocking and Regularization Approach to High Dimensional Realized Covariance Estimation. *Journal of Applied Econometrics* 27: 625 645.

Jacod, J., Y. Li, P. A. Mykland, M. Podolskij, and M. Vetter. 2009. Microstructure Noise in the Continuous Case: The Pre Averaging Approach. *Stochastic Processes and their Applications* 119: 2249 2276.

Jin, X. and J. M. Maheu. 2013. Modelling Realized Covariances and Returns. *Journal of Financial Econometrics* 11: 335 369.

Laurent, S., J. Rombouts, and F. Violante. 2013. On Loss Functions and Ranking Forecasting Performances of Multivariate Volatility Models. *Journal of Econometrics* 173: 1 10.

Lunde, A., N. Shephard, and K. Sheppard. 2011. Econometric Analysis of Vast Covariance Matrices Using Composite Realized Kernels. Working Paper. www.kevinsheppard.com/Working Papers.

Lutkepohl, H. and F. Xu. 2013. The Role of the Log Transformation in Forecasting Economic Variables. *Empirical Economics* 42: 619 638.

Noureldin, D., N. Shephard, and K. Sheppard. 2012. Multivariate High Frequency Based Volatility (HEAVY) Models. *Journal of Applied Econometrics* 27: 907 933.

Patton, A. 2011. Volatility Forecast Comparison Using Imperfect Volatility Proxies. *Journal of Econometrics* 160: 246 256.

Patton, A. and K. Sheppard. 2009. "Evaluating Volatility and Correlation Forecasts". In T. Andersen, R. Davis, J. P. Kreiss, and T. Mikosch (eds.), *Handbook of Financial Time Series*. Heidelberg, Spinger Verlag Berlin Heidelberg.

Politis, D. and J. Romano. 1994b. The Stationary Bootstrap. *Journal of the American Statistical Assosiation* 89: 1303 1313.

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.