

Herwartz, Helmut; Golosnoy, Vasyl

**Working Paper**

## Semiparametric Approaches to the Prediction of Conditional Correlation Matrices in Finance

Economics Working Paper, No. 2007-23

**Provided in Cooperation with:**

Christian-Albrechts-University of Kiel, Department of Economics

*Suggested Citation:* Herwartz, Helmut; Golosnoy, Vasyl (2007) : Semiparametric Approaches to the Prediction of Conditional Correlation Matrices in Finance, Economics Working Paper, No. 2007-23, Kiel University, Department of Economics, Kiel

This Version is available at:

<http://hdl.handle.net/10419/22039>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Semiparametric Approaches to the Prediction of Conditional Correlation Matrices in Finance

by Vasyl Golosnoy and Helmut Herwartz

C | A | U

Christian-Albrechts-Universität Kiel

Department of Economics

*Economics Working Paper*

*No 2007-23*



# Semiparametric Approaches to the Prediction of Conditional Correlation Matrices in Finance

Vasyl Golosnoy <sup>\*</sup>      Helmut Herwartz <sup>†</sup>

30th October 2006

## Abstract

We consider the problem of ex-ante forecasting conditional correlation patterns using ultra high frequency data. Flexible semiparametric predictors referring to the class of dynamic panel and dynamic factor models are adopted for daily forecasts. The parsimonious set up of our approach allows to forecast correlations exploiting both estimated realized correlation matrices and exogenous factors. The Fisher- $z$  transformation guarantees robustness of correlation estimators under elliptically constrained departures from normality. For the purpose of performance comparison we contrast our methodology with prominent parametric and nonparametric alternatives to correlation modeling. Based on economic performance criteria, we distinguish dynamic factor models as having the highest predictive content.

**JEL Classification:** C13, C14, C53, G12

**Keywords:** Correlation forecasting; Epps effect; Fourier method; Dynamic panel model; Dynamic factor model

---

<sup>\*</sup>Institut für Statistik und Ökonometrie, Christian Albrechts–Universität zu Kiel, Olshausenstr. 40, D–24098 Kiel, Germany [vgolosnoy@stat-econ.uni-kiel.de](mailto:vgolosnoy@stat-econ.uni-kiel.de)

<sup>†</sup>Institut für Statistik und Ökonometrie, Christian Albrechts–Universität zu Kiel, Olshausenstr. 40, D–24098 Kiel, Germany [Herwartz@stat-econ.uni-kiel.de](mailto:Herwartz@stat-econ.uni-kiel.de) (corresponding author). We thank Albrecht Mengel for computational assistance in data handling.

# 1 Introduction

Forecasting conditional covariances in high dimensional systems of speculative returns is required in numerous economic and financial applications as, for instance, portfolio optimization, dynamic hedging or Value-at-Risk evaluation. Adopting a parametric framework for covariance modeling, a multivariate GARCH model say, easily gets intractable for practical purposes since the parameter space of such models is quite large even for small systems of vector returns. The latter caveat applies explicitly for model versions like the BEKK of Engle and Kroner (1995), half-vec, or diagonal model of Bollerslev et al. (1988). The class of correlation models, see Bollerslev (1990), Engle (2002), Tse and Tsui (2002), proceeds from the separation of dynamic variance and correlation features with the former modeled via univariate GARCH specifications. Having only a few parameters governing news response and persistence of correlations, the dynamic conditional correlation model (DCC) offers a parsimonious strategy to model and predict conditional covariances. For a recent overview of multivariate volatility and correlation models the reader may consult Bauwens et al. (2006).

Though providing a feasible approach to conditional risk evaluation in high dimensional systems the DCC model involves various sources of potential misspecification. In the first place the low dimensional parameter space implies strong homogeneity restrictions placed on persistence features of all correlations characterizing a system of returns. Moreover, as any parametric model the postulated dynamic relationship may hold only over a local time history. When applied to longer series of vector returns, parametric specifications are likely to suffer from structural variation of the underlying true data generating process. Thirdly, parametric volatility and correlation models are typically set out to exploit sample information at some medium observation frequency, the daily say. Alternatively, covariance matrices may be estimated directly from cross products of vector returns observed at (ultra) high frequencies. Adopting the concept of *realized volatility* e.g. Andersen et al. (2001a,b) illustrate that daily volatility can be estimated accurately by summing squared returns measured at sufficiently high frequencies. Owing to the expansion of the information base employed for modeling dynamic correlations it is intuitive to expect high frequency models to improve the forecasting performance offered by medium frequency approaches.

Highlighted by seminal contributions of Andersen et al. (2003), Barndorff-Nielsen and Shephard (2002, 2004, 2006) realized (co)variance methods have seen a vast development of both underlying statistical theory and empirical contributions to (co)variance modeling and forecasting. In comparison with the body of literature concerned with the prediction of realized variances, see e.g. Andersen and Bollerslev (1998), Andersen et al. (2003, 2004), Poon and Granger (2003), Ghysels et al. (2006), Gospodinov et al. (2006), the empirical literature on forecasting conditional covariances is up to now rather scarce. Borrowing from the class of parametric correlation models the task of covariance forecasting can be separated into volatility and correlation forecasting. The aim of this paper is to elaborate efficient forecasting methods for high-dimensional correlation matrices using information from intraday ultra high frequency observations.

Owing to lead-lag relationships between assets the estimation of realized correlation suffers from the Epps (1979) effect describing that empirical correlations decrease when the underlying sampling frequency increases. Additionally, non-synchronicity of price observations complicates the determination of realized covariances. Studies of Zhang et al. (2005), Oomen (2006), Hansen and Lunde (2006) address the issue to determine an optimal sampling frequency for variance estimation in presence of market microstructure noise. Bandi and Russell (2005) propose an optimal synchronizing frequency to sample cross moments under non-synchronous trading. Malliavin and Mancino (2002) initiate a frequency domain approach to estimate covariances based on the Fourier transformation. All these studies are concerned with estimation but not with forecasting of correlations in high dimensional systems.

In this paper we suggest dynamic correlation predictors that are feasible to handle huge fields

of time series data such as realized correlations. Realized covariances are estimated alternatively by means of data synchronization and the Fourier approach. We apply the Fisher's  $z$  transformation (cf. Hotelling, 1953) of correlations in order to get quicker convergence to normality and to control the empirical support of correlation forecasts. Devlin et al. (1976) show that Fisher's  $z$  serves as a variance stabilizing transformation for a broad class of elliptical distributions. The proposed forecasting models are in the framework of dynamic panel and dynamic factor models. Both model families are formalized to incorporate non-modeled exogenous variables that are widely used in related literature on forecasting (realized) variances. The estimated historical correlations serve as explanatory variables, as well as a number of exogenous factors, typically exploited for explaining stock market volatility. Thus our panel approach extends ideas in Sheppard (2004) conditioning price comovement on exogenous variables while disregarding autoregressive features. As an alternative we also consider dynamic factor models (Stock and Watson, 2002). These models concentrate on the dynamic relationships between a set of principle components. It can be seen as an alternative to the approach in Niguez and Rubia (2006), combining principle components with GARCH dynamics.

We contrast our approach with a number of popular techniques for correlation matrix forecasting. In particular, we consider the dynamic conditional correlation model (DCC), an adaptive covariance estimator (Härdle et al., 2003), the non-parametric exponential smoothing technique of RiskMetrics (J.P. Morgan, 1997), and a functional smoothing approach introduced by Hafner et al. (2005). Regarding the latter two model families we observe that their main feature of smoothing historical data could also apply to realized covariance estimators. As such, our model comparison covers competing models exploiting the same set of underlying high frequency information.

Given the high dimensionality of our problem conventional performance statistics, such as the mean squared forecasting error (MSFE), can hardly be applied for our purposes of forecast evaluation. Therefore, to assess the accuracy of alternative forecasting methods economic criteria are used in this study. We consider economic loss functions (Diebold and Mariano, 1995) and take the perspective of an investor targeting the global minimum variance portfolio (GMVP). Conditional on this criterion a forecasting method that provides the smallest out-of-sample portfolio variance is preferable. Using the best available realized volatility estimator, we take it as if it were known and concentrate ourselves solely on correlation matrix forecasting.

The empirical study is conducted on 25 most traded stocks listed in the major German stock index, DAX. The sample period lasts 1330 trading days from Feb,01,2000 to April,29,2005. Our findings can be summarized as follows. Firstly, we confirm that realized correlations do improve portfolio performance compared with models that are implemented with daily data. This justifies our forecasting approach based on ultra high frequency data. Secondly, relying on the elegant Fourier transformation method for correlation estimation is outperformed by data synchronization in the spirit of Zhang et al. (2005). In the third place we find that dynamic factor models and exponential smoothing of realized covariance matrices deliver most accurate correlation forecasts at all horizons. The informational content of both predictors appears complementary such that forecast combination is fruitful to further reduce ex-ante uncertainty. Concerning predictive ability of exogenous variables the DAX index return seems to be more useful than other factors, such as indicators of market volatility or quoting intensity.

The paper is structured as follows. Section 2 reviews the methods for estimating realized covariance matrices. Section 3 proposes our methodology for dynamic modeling and forecasting of conditional correlation matrices. Competing approaches to correlation forecasting are described in Section 4. Section 5 discusses the employed measures of forecasting performance. Empirical results are collected in Section 6. Section 7 concludes the paper and motivates some issues of future research.

## 2 Estimating realized covariances

The purpose of this paper is to investigate a class of flexible forecasting models feasible to assess dynamic relationships of huge fields of realized correlations. As such forecasting performance of a particular model would also be influenced by the choice and quality of the underlying estimates of realized (cross) moments. From the perspective of empirical finance it is clearly of interest to optimize correlation forecasts in both directions, i.e. adopting an efficient forecasting model building upon most accurate realized covariance estimates. Therefore this Section first reviews some recent approaches to realized moment estimation. Particular forecasting models are in the focus of Sections 3 and 4.

We consider a speculative market comprising  $N$  assets. At each day  $t$  we observe  $\mathbf{K}_t = (K_{1,t} + 1, \dots, K_{N,t} + 1)'$  price quotations  $p_{i,k,t}$  with  $i = 1, \dots, N$ . The corresponding trading time points for the  $i$ th asset are denoted by  $\theta_{i,0}, \dots, \theta_{i,K_i}$ . To simplify the notation we economize on the time index as time dependence is obvious for many quantities such as the number of transactions or trade timing. In the first place we are interested in estimating the covariance matrix of the daily  $N$ -dimensional return vector  $\mathbf{r}_t$ ,  $t = 1, \dots, \mathcal{T}$ , having as typical elements the difference between consecutive log closing prices. More precisely,  $\mathbf{r}_t = (r_{1,t}, r_{2,t}, \dots, r_{N,t})'$ ,  $r_{i,t} = \ln p_{i,K_i,t} - \ln p_{i,K_i,t-1}$ . We denote and calculate intraperiod log returns  $r_{i,k} = \ln p_{i,k} - \ln p_{i,k-1}$  for  $k = 1, \dots, K_i$ . Our aim is to construct the estimator  $\hat{\mathbf{R}}_t$  of the unknown true correlation matrix  $\mathbf{R}_t$  of considered assets and to provide an accurate  $h$ -step ahead forecast  $\tilde{\mathbf{R}}_{t,h}$ . For our expositions the notation distinguishes explicitly between ex-ante forecasts and in-sample estimates. In sample estimates of some quantity,  $\mathbf{R}_{t+h}$  say, are hereafter denoted as  $\hat{\mathbf{R}}_{t+h}$  whereas conditional forecasts are indicated as  $\tilde{\mathbf{R}}_{\tau,h}$ .

Estimating realized covariance matrices could be separated into variance and correlation matrix estimation. Numerous studies deal with various aspects of variance estimation in presence of market microstructure noise. Compared with the body of literature on realized variances, estimation of realized correlations is much less investigated. The Epps effect (Epps, 1979), namely the decrease of correlations with the increase of the sampling frequency, is a fundamental problem when calculating intraday covariances. The main reasons for the Epps effect are the non-synchronicity of price observations and possible lead-lag relationships between asset prices. There exist two alternative approaches to overcome the Epps effect when calculating realized covariances. The first method is based on synchronizing the data, the second is grounded on transforming the data from the time domain into the frequency domain. Below we sketch the latter methods in turn.

### 2.1 Synchronized data

Data synchronization is conducted as previous tick interpolation taking the most recent quotes (cf. Dacorogna et al., 2001). Other sampling schemes, investigated by Oomen (2006) are not considered in our study. Owing to market microstructure noise the choice of the sampling frequency is essential for the properties of realized moment estimators. Low frequency sampling leads to a loss of valuable information, while sampling at high frequencies yields biased estimators. Two avenues of efficient sampling could be distinguished with the first targeting an optimal sampling frequency (Bandi and Russell, 2005), while the second suggests to exploit all available information in an efficient manner (Zhang et al., 2005).

Apart from ad-hoc sampling at a-priori chosen time intervals we estimate realized covariance by means of the subsampling procedure advocated in Zhang et al. (2005). As initially proposed the latter method requires to choose a-priori a global modeling parameter, the number of subgrids denoted with  $G$  in the following. For the purpose of variance estimation Zhang et al. (2005) propose to use  $G = 3$  or  $G = 4$  subgrids. Rather than relying on an ad-hoc choice, however, it is more intuitive to determine the number of subgrids in a data driven manner. To select data driven  $G$  we

consider an approach put forth by Bandi and Russell (2005), proceeding from the presumption that market microstructure noise follows a first order moving average process. The latter method is now briefly described before we return to the subsampling method.

To determine an optimal sampling frequency the data is first synchronized to a maximal sampling frequency  $K$ . The obtained equally spaced log return observations are collected in a matrix  $(\dot{\mathbf{r}}_1, \dots, \dot{\mathbf{r}}_N)$ . Further we denote by  $K^\dagger$  a fixed sampling interval, 10 minutes sampling say, which is likely immunized against the adverse effects of microstructure noise. Synchronized returns based on  $K^\dagger$  are collected and denoted as  $(\ddot{\mathbf{r}}_1, \dots, \ddot{\mathbf{r}}_N)$ . Bandi and Russell (2005) suggest to estimate the optimal sampling frequency for the  $ij$ th component of the covariance matrix as

$$\hat{K}_{ij} = \left( \frac{\hat{q}_{ij}}{2\hat{E}^2(\epsilon_i \epsilon_j)} \right)^{1/3}, \quad \text{with}$$

$$\hat{E}(\epsilon_i \epsilon_j) = K^{-1} \sum_{k=1}^K \dot{r}_{i,k} \dot{r}_{j,k}, \quad \hat{q}_{ij} = K^\dagger \sum_{k=1}^{K^\dagger} \dot{r}_{i,k}^2 \dot{r}_{j,k}^2 - K^\dagger \sum_{k=1}^{K^\dagger-1} \dot{r}_{i,k} \dot{r}_{j,k} \dot{r}_{i,k+1} \dot{r}_{j,k+1}.$$

The estimator  $\hat{K}_{ij}$  differs for each element of the covariance matrix and, moreover, varies over time. To balance model feasibility and optimality we will later rely on time averages of  $\hat{K}_{ij}$ . For the latter mean estimators it turned out that their magnitude differed sharply for diagonal and off diagonal components of the covariance matrix. Therefore, implementing an optimal sampling frequency we decide to use to different average smoothing parameters for variances and covariances, i.e.

$$\hat{K}_v = (N\mathcal{T})^{-1} \sum_{i=1}^N \sum_{t=1}^{\mathcal{T}} \hat{K}_{ii,t}, \quad \hat{K}_c = 2(N(N-1)\mathcal{T})^{-1} \sum_{i=1}^N \sum_{j=i+1}^N \sum_{t=1}^{\mathcal{T}} \hat{K}_{ij,t}. \quad (1)$$

Having an optimal sampling frequency for the realized (co)variances at hand we use it to implement non-overlapping subgrids for realized moment estimation as proposed by Zhang et al. (2005). Here, the full grid of observations  $\mathcal{G}$  consisting of  $K$  prices, is partitioned into  $G_\bullet = \lfloor K/\hat{K}_\bullet \rfloor$  non-overlapping subgrids  $\mathcal{G}_{(g)}$ ,  $g = 1, \dots, G_\bullet$ . Then the estimated covariance matrix for each day  $t$  is given as

$$\hat{\Sigma}_t^{(DS)} = (\hat{\sigma}_{ij}), \quad \hat{\sigma}_{ij} = G_\bullet^{-1} \sum_{g=1}^{G_\bullet} \sum_{k \in \mathcal{G}_{(g)}} \dot{r}_{i,k} \dot{r}_{j,k} - \sum_{k=1}^K \dot{r}_{i,k} \dot{r}_{j,k}, \quad i, j = 1, \dots, N, \quad (2)$$

where  $\sum_{k=1}^K \dot{r}_{i,k} \dot{r}_{j,k}$  measures market microstructure noise.

## 2.2 Fourier transformation

Any synchronization scheme leads to some loss of information. Malliavin and Mancino (2002) suggest to overcome the non-synchronicity of price observations by means of Fourier approach. The idea is to calculate the realized covariance matrix from the Fourier coefficients of log price increments. In the frequency domain approach, the daily trading times  $[\theta_{i,0}, \theta_{i,1}, \theta_{i,2}, \dots, \theta_{i,K_i-1}, \theta_{i,K_i}]$  are mapped onto the interval  $[0, s_1, s_2, \dots, s_{K_i-1}, 2\pi]$  for all assets  $i = 1, \dots, N$ . For a frequency  $f \in \mathcal{N}$ , the Fourier coefficients for log price increments  $d \ln p_i$  of asset  $i$  are estimated by

$$\begin{aligned} \hat{a}_{i,f}(d \ln p_i) &= \pi^{-1} \left( \ln p_i(2\pi) - \ln p_i(0) + \sum_{k=1}^{K_i} \ln p_i(s_{k-1}) (\cos(fs_k) - \cos(fs_{k-1})) \right), \\ \hat{b}_{i,f}(d \ln p_i) &= \pi^{-1} \sum_{k=1}^{K_i} \ln p_i(s_k) (\sin(fs_{k-1}) - \sin(fs_k)). \end{aligned}$$

With  $F$  denoting a minimal sampling frequency considered for the calculation of Fourier coefficients,  $\widehat{\mathbf{a}}_f = (\widehat{a}_{1,f}, \dots, \widehat{a}_{N,f})'$  and  $\widehat{\mathbf{b}}_f = (\widehat{b}_{1,f}, \dots, \widehat{b}_{N,f})'$ ,  $f = 1, \dots, F$ , the covariance matrix at day  $t$  can be expressed as

$$\widehat{\Sigma}_t^{(F)} = 2\pi^2 F^{-1} \sum_{f=1}^F \left( \widehat{\mathbf{a}}_f \widehat{\mathbf{a}}_f' + \widehat{\mathbf{b}}_f \widehat{\mathbf{b}}_f' \right). \quad (3)$$

### 2.3 Correcting overnight bias

As outlined we are interested in the evaluation of the covariance matrix of daily log returns measured as the difference of consecutive log closing prices. In this case a realized covariance matrix as given in (2) or (3) should be adjusted for the overnight information flow. The bias correction accounts for the no-trade distance between two consecutive trading days. It is possible to correct the overnight bias for each single element of the covariance matrix, as well as for the entire matrix in one step. For an overview of alternative methods to adjust for the overnight bias the reader may consult Hansen and Lunde (2005). In this work we adopt a suggestion made by Martens (2002). It does not require any exact specification of the bias structure and is easily implemented. Define the elements of open-to-close ( $\widehat{\Sigma}_{oc}$ ) and close-to-open ( $\widehat{\Sigma}_{co}$ ) covariances as

$$\widehat{\sigma}_{ij,co} = \sum_{t=1}^{T-1} r_{i,co,t} r_{j,co,t}, \quad \text{with} \quad r_{i,co,t} = \ln p_{i,0,t+1} - \ln p_{i,K_i,t},$$

$$\widehat{\sigma}_{ij,oc} = \sum_{t=1}^T r_{i,oc,t} r_{j,oc,t}, \quad \text{with} \quad r_{i,oc,t} = \ln p_{i,K_i} - \ln p_{i,0}.$$

To correct the overnight bias we scale the realized covariance matrix, given in (2) or (3), obtaining

$$\widehat{\mathbf{S}}_t^{(\bullet)} = \frac{\mathbf{1}' \widehat{\Sigma}_{oc} \mathbf{1} + \mathbf{1}' \widehat{\Sigma}_{co} \mathbf{1}}{\mathbf{1}' \widehat{\Sigma}_{oc} \mathbf{1}} \widehat{\Sigma}_t^{(\bullet)}. \quad (4)$$

The adjusted covariance matrix  $\widehat{\mathbf{S}}_t^{(\bullet)}$  is used for further processing.

## 3 Modeling and forecasting realized correlations

### 3.1 The Fisher's $z$ transformation

Correlation measures are bounded on the interval  $[-1,1]$ . When it comes to correlation forecasting an immediate issue is to control the empirical support of the predictors. As a means to guarantee bounded forecasts we will rely on the variance stabilizing Fisher- $z$  transformation, popularized by Hotelling (1953). Additionally, the transformed correlations are supposed to achieve faster asymptotic convergence to normality. In the following we briefly discuss this transformation and provide a proposition on the asymptotic features of realized Fisher- $z$  transformed correlations.

The estimated correlation matrix  $\widehat{\mathbf{R}}_t$  is obtained by transformation of the realized covariance matrix  $\widehat{\mathbf{S}}_t$ , i.e.

$$\widehat{\mathbf{R}}_t = \widehat{\mathbf{S}}_t^* := (\widehat{\mathbf{S}}_t \odot \mathbf{I}_N)^{-1/2} \widehat{\mathbf{S}}_t (\widehat{\mathbf{S}}_t \odot \mathbf{I}_N)^{-1/2}, \quad (5)$$

where  $\odot$  denotes element by element matrix multiplication and  $\mathbf{I}_N$  is an  $N \times N$  identity matrix. It should be kept in mind that all realized moment estimates are model dependent throughout. To



facilitate the notation, however, model dependence is not made explicit. Let  $\text{vecl}(\mathbf{R})$  denote the operator stacking the below diagonal elements of a  $N \times N$  matrix in a  $\underline{N} = (N^2 - N)/2$ -dimensional column vector. The estimated time series of correlation vectors are denoted as

$$\hat{\boldsymbol{\rho}}_t = \text{vecl}(\hat{\mathbf{R}}_t) = (\hat{\rho}_{21,t}, \hat{\rho}_{31,t}, \dots, \hat{\rho}_{N-1,t})'.$$

The Fisher-transformation applied to each  $\hat{\rho}_{ij,t}$  yields

$$\hat{z}_{ij,t} = \frac{1}{2} \ln \frac{1 + \hat{\rho}_{ij,t}}{1 - \hat{\rho}_{ij,t}}. \quad (6)$$

Consequently, the reverse transformation is given by

$$\hat{\rho}_{ij,t} = \frac{\exp(2\hat{z}_{ij,t}) - 1}{\exp(2\hat{z}_{ij,t}) + 1}. \quad (7)$$

The effect of Fisher- $z$  and reverse Fisher- $z$  transformations is illustrated on Figure 1.

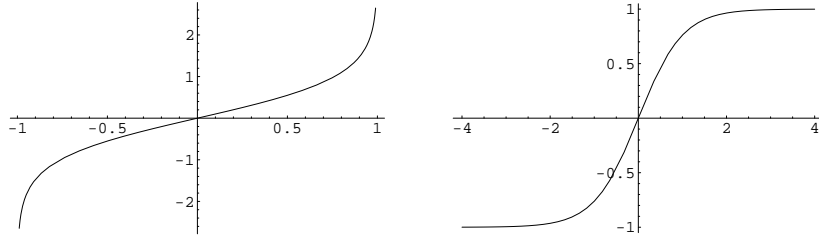


Figure 1: Fisher- $z$  transformation for  $\rho \rightarrow z$ , left, and  $z \rightarrow \rho$ , right.

Statistical theory concerning the asymptotic features of Fisher- $z$  transformed correlations is well developed (see Muirhead (1982), Theorem 5.1.7). In particular, the variance of transformed correlation estimates does not depend on the transformed corresponding true correlation. Note that the latter does not hold for the variance of standard correlation coefficients. Devlin et al. (1976) argue that the distribution of centered Fisher- $z$  transformed correlations could be quite well approximated by a normal distribution even in case of small underlying sample sizes. Next we state the asymptotic features of Fisher- $z$  transformed realized correlation estimates.

The asymptotic distribution of realized covariances and correlations under rather mild smoothness assumptions on the multivariate stochastic volatility is provided by Barndorff-Nielsen and Shephard (2004) in Proposition 4. Here we provide a statement about the asymptotic distribution of the realized Fisher- $z$  coefficient, which is proven in the Appendix.

**Proposition 1** *Consider the Fisher- $z$  transformation of the realized correlation coefficient based on  $K$  intraday observations,  $\hat{z}_{ij,t} = 0.5 \ln\{(1 + \hat{\rho}_{ij,t})/(1 - \hat{\rho}_{ij,t})\}$ . The estimator  $\hat{\rho}_{ij,t}$  and its asymptotic distribution are given by Barndorff-Nielsen and Shephard (2004) under weak assumptions on a multivariate stochastic volatility process. Then the asymptotic distribution of the Fisher- $z$  transformed random variable is given by*

$$\sqrt{K}(\hat{z}_{ij,t} - z_{ij,t}) \xrightarrow{L} \mathcal{N}(0, 1),$$

where  $K$  denotes the number of intraday returns,  $z_{ij,t} = 0.5 \ln\{(1 + \rho_{ij,t})/(1 - \rho_{ij,t})\}$ .

In our context the Fisher- $z$  transformation is useful for three reasons. In the first place the variance of the sample estimator is made independent from the true moment and convergence of the sample estimator to normality is accelerated. Secondly, the reverse transformation guarantees that any conditional forecast of  $z_{ij,t}$  implies a correlation forecast  $\rho_{ij,t}$  bounded on the interval  $[-1; 1]$ . Thirdly, the Fisher transformation is variance stabilizing for a broad class of heavy tailed elliptical distributions (Devlin et al., 1976).

## 3.2 The forecasting design

We aim to compare alternative avenues to correlation modeling and forecasting. To focus on the particular issue of correlation forecasting the entire analysis is based on the presumption that (realized) return variances are known ex-ante. Thus, for the practical purpose of covariance prediction we implicitly presume that an analyst has access to (efficient) variance forecasts. Alternative approaches to forecasting (realized) variances are provided, for instance, in Andersen et al. (2003) and Ghysels et al. (2006). Concentrating on one particular sequence of realized variances and combining it with competing correlation estimates immunizes the forecasting competition against the effects of model specific realized variance estimates. Our aim is to determine a forecast of the  $\underline{N} = (N - 1)N/2$ -dimensional vector  $\boldsymbol{\rho}_{\tau+h}$  based on  $\mathcal{I}_\tau$ , the set of information available in the forecast origin  $\tau$ , i.e.  $E[\boldsymbol{\rho}_{\tau+h}|\mathcal{I}_\tau] = \tilde{\boldsymbol{\rho}}_{\tau,h}$ .

There are some general considerations concerning the choice of an appropriate model for forecasting time series of correlations. First noting that we jointly model a huge set of conditional correlations any forecasting scheme has to respect the postulate of model parsimony. Second, since conditional correlations are likely persistent (cf. Engle and Sheppard, 2001), autocorrelation features characterizing the time series  $\{\rho_{i,j,t}\}$  should be taken into account. In the third place, the potential influence of exogenous variables,  $w_t$ , should be incorporated into the specification.

In the light of the latter considerations we concentrate on two parsimonious approaches to characterize the dynamic structure of data fields, namely dynamic panel and dynamic factor models. For both we first transform estimated realized correlation measures by means of the Fisher- $z$  transformation given in (6) and apply the dynamic frameworks to determine the predictor of transformed correlations  $\tilde{\mathbf{z}}_{\tau,h}$ . Then, the latter are converted into correlation predictions via the reverse Fisher transformation in (7). Having a matrix of correlation forecasts at hand it is combined with the future realized variance to obtain an implied covariance matrix forecast.

The dynamic panel and dynamic factor model are now briefly discussed in turn. For both approaches we concentrate on the model representation. The practical implementation used in the empirical analysis is discussed in Appendices B1 and B2.

## 3.3 Dynamic panel model

As a first method to model the dynamics of huge fields of time series data we suggest a dynamic panel approach with random effects. In single observation and compact daywise vector notation, respectively, this model can be written by

$$z_{\iota,t+h} = \gamma + z_{\iota,t}\delta + \mathbf{w}'_{t+h}\boldsymbol{\beta} + \mu_\iota + \nu_{\iota,t+h}, \quad \iota = 1, \dots, \underline{N}, \quad t = 1, \dots, \tau - h, \quad (8)$$

$$\mathbf{z}_{t+h} = \gamma\mathbf{j}_{\underline{N}} + \mathbf{z}_t\delta + \mathbf{j}_{\underline{N}}(\mathbf{w}'_{t+h}\boldsymbol{\beta}) + \boldsymbol{\mu} + \boldsymbol{\nu}_{t+h}. \quad (9)$$

The following assumptions are made concerning the random individual effects and the idiosyncratic noise:  $\mu_\iota \sim \text{iid}(0, \sigma_\mu^2)$ ,  $\nu_{\iota,t} \sim \text{iid}(0, \sigma_\nu^2)$  and  $\boldsymbol{\mu} \perp \boldsymbol{\nu}_t$ . In the compact representation  $\mathbf{j}_{\underline{N}}$  is a  $\underline{N}$ -dimensional vector of ones,  $\boldsymbol{\mu}$  collects all individual effects,  $\boldsymbol{\nu}_t$  contains idiosyncratic noise terms, and  $\mathbf{w}_{t+h}$  is a  $q$ -dimensional vector of exogenous variables.

The model (9) is rather restrictive in presuming cross sectionally homogeneous dynamic features. Although it is a priori tempting to allow for correlation specific autoregressive parameters ( $\delta_\iota$ ) we prefer the homogeneous panel representation to meet the postulate of model parsimony and focus on the predictive content of the estimated random effects, see Baltagi (2006). Similar to the restrictive nature of the autoregressive specification the iid assumption underlying the individual effects and the idiosyncratic noise terms may also be criticized. In fact, error terms  $\nu_{\iota,t}$  could be heteroskedastic over the cross section or time dimension. For the purpose of consistent estimation and forecasting, however, proceeding under iid assumptions is justified. Note that efficient forecasting under

heteroskedastic noise terms requires particular a-priori assumptions on second order features which might lead to model misspecification.

In case of one step ahead forecasting ( $h = 1$ ) equation (8) formalizes a common first order autoregressive panel model, see Baltagi (2001). For the purpose of higher horizon forecasting we will later also choose  $h = 5, 10$ . In this case the formalization in (8) aims to directly condition  $z_{i,t+h}$  on information available up to time  $t$ . Implementing the forecasting model in this way avoids to determine higher horizon predictions from an iteration over one-step ahead forecasts. From the theoretical literature both approaches to forecasting at higher horizons could be justified. In case the employed one step ahead model specification matches the true underlying data generating mechanism iterative forecasting schemes are preferable if estimated model parameters are sufficiently precise. Direct higher order forecasts are supposed to be more robust against model misspecification since iterating a misspecified model is likely to loose control over ultimate forecast errors. As a consequence, the choice of direct against iterative forecasting depends on the true model and, consequently, remains an empirical issue. For a recent empirical comparison of both techniques in macroeconomic forecasting the reader may consult Marcellino et al. (2006) who also review the theoretical literature in some detail. For this study we follow a direct specification, since any parsimonious parametric panel model is likely to suffer from misspecification in light of the huge cross sectional dimension.

To employ the model in (8) as a forecasting specification it is necessary to have some ex-ante measure of particular exogenous variables  $w_t$  collected in  $\mathbf{w}_t$ . A nonparametric approach to determine  $\tilde{w}_{\tau,h} = E[w_{\tau+h}|\mathcal{I}_\tau]$  is described in Appendix B3. As an alternative to (8), however, one may also rely on lagged explanatory variables in the estimation model and condition the respective forecasts on sample information available at the forecast origin  $\tau$ . Such panel model is given by

$$\mathbf{z}_{t+h} = \gamma \mathbf{j}_N + \mathbf{z}_t \delta + \mathbf{j}_N (\mathbf{w}'_t \boldsymbol{\beta}) + \boldsymbol{\mu} + \boldsymbol{\nu}_{t+h}. \quad (10)$$

To distinguish between the alternative approaches in (9) and (10) we refer to a model of the latter type as a specification with predetermined variables.

### 3.4 Dynamic factor models

In the recent literature on modeling huge fields of time series data dynamic factor models have been attracting a lot of interest at least following Stock and Watson (2002). In our case the informational content of the set of time varying correlations could be condensed by means of a parsimonious set of (dynamic) principal components. The latter enters some autoregressive or vector autoregressive (VAR) structure that allows conditional forecasting. In direct comparison with the dynamic panel approach such factor models have the advantage to avoid strong parametric assumptions like uniform autoregressive features of all correlations. Dynamic factor model also respect the necessity of parametric restrictions faced when jointly modeling  $N$ -dimensional correlation vectors. In compact daywise notation the dynamic factor model reads as

$$\check{\mathbf{z}}_{t+h} = \boldsymbol{\Gamma}_G \mathbf{f}_{t+h} + \boldsymbol{\xi}_{t+h}, \quad t = 1, \dots, \tau - h, \quad (11)$$

$$\mathbf{f}_{t+h} = \boldsymbol{\Phi}_1 \mathbf{f}_t + \boldsymbol{\Phi}_2 \mathbf{f}_{t-1} + \dots + \boldsymbol{\Phi}_p \mathbf{f}_{t-p+1} + \boldsymbol{\eta}_{t+h}. \quad (12)$$

In (11)  $\check{\mathbf{z}}_t$  is a  $N$ -dimensional vector of Fisher transformed correlations measured in terms of deviations from their unconditional mean,  $\check{\mathbf{z}}_t = \mathbf{z}_t - \bar{\mathbf{z}}_\tau$ ,  $\bar{\mathbf{z}}_\tau = (\tau - h)^{-1} \sum_{t=1}^{\tau-h} \mathbf{z}_t$ .  $\mathbf{f}_t$  is a  $G$ -dimensional vector of factors supposed to govern the dynamic structure of  $\check{\mathbf{z}}_t$ . The latter is specified using the  $G \times G$  parameter matrices  $\boldsymbol{\Phi}_i$ ,  $i = 1, \dots, p$ . In the empirical study we consider both diagonal and unrestricted variants of  $\boldsymbol{\Phi}_i$ . The error terms  $\boldsymbol{\xi}_{t+h}$  and  $\boldsymbol{\eta}_{t+h}$  are assumed to be serially uncorrelated and independent. To determine the matrix  $\boldsymbol{\Gamma}_G$  in (11) we adopt a principal component analysis

(PCA) decomposing the unconditional covariance matrix of  $\check{\mathbf{z}}_t$ , i.e.

$$\widehat{\Omega}_\tau = (\tau - h)^{-1} \sum_{t=1}^{\tau-h} \check{\mathbf{z}}_t \check{\mathbf{z}}_t', \quad \widehat{\Omega}_\tau = \widehat{\Gamma}_\tau \widehat{\Lambda}_\tau \widehat{\Gamma}_\tau'. \quad (13)$$

In (13)  $\widehat{\Lambda}_\tau$  is an estimated diagonal matrix of eigenvalues of  $\widehat{\Omega}_\tau$  in decreasing order while the columns of  $\widehat{\Gamma}_\tau$  contain the corresponding eigenvectors. Then, the estimator of matrix  $\Gamma_G$  in (11) contains the first  $G$  columns of  $\widehat{\Gamma}_\tau$  and thereby accounts for the variation in  $\check{\mathbf{z}}_t$  explained by  $G$  principal components. Ex-ante forecasts of  $\mathbf{z}_{\tau+h}$ , adjusted for the unconditional in sample mean, are given by

$$\begin{aligned} \widetilde{\mathbf{z}}_{\tau,h} &= \Gamma_G \widetilde{\mathbf{f}}_{\tau,h} + \bar{\mathbf{z}}_\tau, \quad \text{with} \\ \widetilde{\mathbf{f}}_{\tau,h} &= \widehat{\Phi}_1 \mathbf{f}_\tau + \widehat{\Phi}_2 \mathbf{f}_{\tau-1} + \dots + \widehat{\Phi}_p \mathbf{f}_{\tau-p+1}. \end{aligned} \quad (14)$$

The dynamic factor model in (11) does not exploit exogenous information to describe the conditional correlations. The augmentation of (11) with contemporaneous or predetermined exogenous variables is straightforward. In the context of PCA modeling, however, one may also extract principal components from the augmented set of time series variables  $\mathbf{z}_t = (\mathbf{z}'_t, \mathbf{w}'_t)'$ . Note that along these lines exogenous variables and transformed correlations enter jointly the dynamic modeling. For the purpose of factor extraction the augmented empirical moment matrix  $\text{Cov}(\mathbf{z}_t)$  is used to determine the principal components. Then, the same scheme as in (14) applies to obtain  $\widetilde{\mathbf{z}}_{\tau,h}$ . Its upper  $\underline{N}$  first elements are corresponding forecasts of Fisher transformed correlations.

## 4 Short summary of competing approaches

In this section we collect popular alternative approaches to covariance or correlation modeling that have been suggested in the literature. For each approach we briefly state the underlying model and then discuss how it is employed for the issue of correlation forecasting.

### 4.1 Adaptive covariance estimation

The detection of locally homogeneous multivariate covariance structures has been recently proposed by Härdle et al. (2003) and is related to the approach of Foster and Nelson (1996). Given that the covariance matrix of returns is constant within some (local) time window  $t \in J = [\tau - J, \tau]$ , a convenient estimator of the second order moments using daily returns is

$$\widehat{\Sigma}_\tau = |J|^{-1} \sum_{t \in J} \mathbf{r}_t \mathbf{r}'_t, \quad (15)$$

with  $|J|$  denoting the length of  $J$ . Note that the presumption of local homogeneity could be justified in light of the overwhelming experience of volatility clustering documented in the empirical financial literature. As given in (15), however, the adaptive estimator is not feasible since in practice knowledge of a time window with homogeneous second order features is hardly available. Härdle et al. (2003) propose a data driven approach to estimate homogeneous time intervals  $J$ . They provide an inferential procedure where 'smoothing' parameters are selected by fixing to  $a$  the probability of falsely diagnosing a homogeneous time interval of length  $\mathcal{M}$ . We estimate  $\hat{J}$  using  $\mathcal{M} = 60$  and  $a = 0.05$ . The adaptive covariance estimator is

$$\widehat{\Sigma}_{\text{ADA},\tau} = |\hat{J}_\tau|^{-1} \sum_{t \in \hat{J}_\tau} \mathbf{r}_t \mathbf{r}'_t.$$

From the latter estimate an implied correlation matrix is determined as

$$\widehat{\mathbf{R}}_{\text{ADA},\tau} = \widehat{\Sigma}_{\text{ADA},\tau}^*.$$

Owing to the underlying presumption of local time homogeneity we set  $\tilde{\mathbf{R}}_{\text{ADA},\tau,h} = \hat{\mathbf{R}}_{\text{ADA},\tau}$ , i.e. for the purpose of correlation forecasting the adaptive method proceeds from the perspective that the estimated local moments will prevail over the near future.

## 4.2 Dynamic conditional correlation modeling

Distinguishing variance and correlation features, a common representation of the conditional covariance matrix is

$$\mathbf{S}_t = (\mathbf{S}_t \odot I_N)^{1/2} \mathbf{R}_t (\mathbf{S}_t \odot I_N)^{1/2}, \quad t = 1, \dots, \tau.$$

From the perspective of (co)variance modeling in the spirit of parametric GARCH-type specifications it is clear that the parameterization of dynamic models will increase at rate  $O(N)$  since single asset volatility processes have to be specified in some way. Facing the curse of dimensionality it has become a convenient strategy in parametric modeling of  $\mathbf{S}_t$  to separate estimation of univariate volatility models and correlation features. The dynamic conditional correlation model (Engle, 2002) of order (1, 1), denoted DCC(1,1), is given by

$$\mathbf{R}_t = \mathbf{Q}_t^*, \quad \mathbf{Q}_t = \mathbf{R}(1 - \alpha - \beta) + \alpha \mathbf{v}_{t-1} \mathbf{v}'_{t-1} + \beta \mathbf{Q}_{t-1}. \quad (16)$$

In (16)  $\mathbf{R}$  denotes the unconditional correlation matrix that can be estimated as

$$\hat{\mathbf{R}} = \frac{1}{\tau} \sum_{t=1}^{\tau} \mathbf{v}_t \mathbf{v}'_t, \quad \mathbf{v}_t = (\hat{\mathbf{S}}_t \odot I_N)^{-1/2} \mathbf{r}_t.$$

Forecasting via the DCC model requires to estimate the underlying model parameters. Estimation of DCC models is mostly undertaken in two steps with the first step addressing the estimation of the parameters underlying the variance processes. In this study we condition correlation dynamics on realized variances. Given that rolling forecasting schemes offer quite limited sample information (in our case  $T = 250$  observations) we refrain from iterative estimation of the correlation parameters. Instead we consider 15 alternative parameter settings from the neighborhood of empirical findings in other studies (cf. Engle and Sheppard, 2001) to indicate the potential performance of DCC type modeling, namely

$\beta$	$\alpha$									
.90	.01	.02	.03	.04	.05	.06	.07	.08	.09	.10
.95	.01	.02	.03	.04	.05	-	-	-	-	-

For the DCC(1,1) model one step ahead forecasts are directly available from sample information whereas forecasting at higher horizons requires iterative forecasting. We have

$$\tilde{\mathbf{R}}_{\text{DCC},\tau,h} = \tilde{\mathbf{Q}}_{\tau,h}^*, \quad \text{with} \quad (17)$$

$$\begin{aligned} \tilde{\mathbf{Q}}_{\tau,1} &= \mathbf{R}(1 - \alpha - \beta) + \alpha \mathbf{v}_\tau \mathbf{v}'_\tau + \beta \mathbf{Q}_\tau, \\ \tilde{\mathbf{Q}}_{\tau,h} &= \mathbf{R}(1 - \alpha - \beta) + (\alpha + \beta) \tilde{\mathbf{Q}}_{\tau,h-1}, \quad h = 2, 3, \dots \end{aligned}$$

## 4.3 Exponential smoothing (RiskMetrics)

Closely related to the adopted implementation of DCC models is to determine local correlation matrices via exponential smoothing. From standardized return vectors the correlation matrix forecasts are determined as

$$\tilde{\mathbf{R}}_{\text{EXV},\tau,h} = \hat{\mathbf{Q}}_{\text{EXV},\tau}^*, \quad \hat{\mathbf{Q}}_{\text{EXV},\tau} = \frac{1 - \phi}{1 - \phi^n} \sum_{i=0}^n \phi^i \mathbf{v}_{\tau-i} \mathbf{v}'_{\tau-i}, \quad h = 1, 2, \dots \quad (18)$$

To implement the rule in (18) we use alternative smoothing parameters  $\phi = .99, .98, \dots, .90$ .

## 4.4 Functional conditional correlations

Hafner et al. (2005) introduce functional models for correlation forecasting. In their semiparametric setting correlation estimates are determined in two steps. First conditional variances are estimated via GARCH, then non-parametric smoothing is applied for correlation estimation. The kernel method is based on local smoothing with respect to a conditioning variable. In time series or regression modeling the underlying idea has been popularized by Cai et al. (2000) under the notion of 'functional coefficient modeling'. We distinguish between conditioning on contemporaneous ( $w_t$ ) and predetermined variables ( $w_{t-h}$ ). Formally the latter specifications are given as

$$\mathbf{Q}_{\text{FSV}}^c(w) = E(\mathbf{v}_t \mathbf{v}_t' | w_t = w), \quad \mathbf{Q}_{\text{FSV}}^p(w) = E(\mathbf{v}_t \mathbf{v}_t' | w_{t-h} = w).$$

The respective local estimates of the latter matrices are obtained by kernel smoothing as

$$\widehat{\mathbf{Q}}_{\text{FSV}}^c(w) = \frac{\sum_{t=1}^{\tau} \mathbf{v}_t \mathbf{v}_t' \cdot H_b(w_t - w)}{\sum_{t=1}^{\tau} H_b(w_t - w)}, \quad \widehat{\mathbf{Q}}_{\text{FSV}}^p(w) = \frac{\sum_{t=h+1}^{\tau} \mathbf{v}_t \mathbf{v}_t' \cdot H_b(w_{t-h} - w)}{\sum_{t=h+1}^{\tau} H_b(w_{t-h} - w)},$$

where  $H_b(\bullet) = \frac{1}{b} K\left(\frac{\bullet}{b}\right)$ , and  $K(\cdot)$  is a kernel function. The corresponding correlation matrices are

$$\widehat{\mathbf{R}}_{\text{FSV}}^c(w) = \widehat{\mathbf{Q}}_{\text{FSV}}^{c,*}(w), \quad \widehat{\mathbf{R}}_{\text{FSV}}^p(w) = \widehat{\mathbf{Q}}_{\text{FSV}}^{p,*}(w).$$

The candidate exogenous or predetermined factor variables  $w_t$  could be some (lagged) stock index volatility, high low daily log price differentials, term spreads, etc. Although it is possible to use more than one factor in multivariate kernel smoothing we concentrate on univariate factor variables to avoid the curse of dimensionality.

The practical implementation of functional conditional correlations, i.e. the choice of a kernel function and the bandwidth parameter, is addressed in Appendix B3. Forecasting in the framework of functional correlation models is straightforward. Distinguishing the cases with contemporaneous or predetermined exogenous variables we have

$$\widetilde{\mathbf{R}}_{\text{FSV},\tau,h}^{\bullet}(\bullet) = \widehat{\mathbf{Q}}_{\text{FSV}}^{\bullet,*}(\bullet), \quad \text{with} \quad \widehat{\mathbf{Q}}_{\text{FSV}}^{c,*}(w = \widetilde{w}_{\tau,h}) \quad \text{or} \quad \widehat{\mathbf{Q}}_{\text{FSV}}^{p,*}(w = w_{\tau}). \quad (19)$$

## 4.5 Factor dependent and exponentially smoothed realized covariance estimates

As introduced by Hafner et al. (2005), the estimation of factor dependent correlation matrices builds upon local weighting of cross products of standardized return vectors. In a similar vein one may also apply factor dependent weighting schemes to realized covariance matrices. Local weighting of realized covariance matrices is likely more efficient in relation to the initial proposal by Hafner et al. (2005) since intraday information is also exploited to determine local estimates. In this framework static and dynamic forecasts of the correlation matrix are

$$\widetilde{\mathbf{R}}_{\text{FSC},\tau,h}^{\bullet}(\bullet) = \widetilde{\mathbf{Q}}_{\text{FSC}}^{\bullet,*}(\bullet), \quad \text{with} \quad \widehat{\mathbf{Q}}_{\text{FSC}}^{c,*}(w = \widetilde{w}_{\tau,h}) \quad \text{or} \quad \widehat{\mathbf{Q}}_{\text{FSC}}^{p,*}(w = w_{\tau}).$$

Similar to extending functional smoothing from standardized return vectors to estimated realized covariances one may also apply exponential smoothing to the latter realized moment matrices. Following these lines correlation forecasts are

$$\widetilde{\mathbf{R}}_{\text{EXC},\tau,h} = \widehat{\mathbf{Q}}_{\text{EXC},\tau}^*, \quad \widehat{\mathbf{Q}}_{\text{EXC},\tau} = \frac{1 - \phi}{1 - \phi^n} \sum_{i=0}^n \phi^i \mathbf{S}_{\tau-i}. \quad (20)$$

To implement the predictor in (20) we use  $\phi = .99, .98, .97, \dots, .90$ .

## 5 Goodness of forecast evaluation

We employ portfolio performance measures to evaluate the performance of alternative correlation forecasts. Due to the huge dimension of the correlation matrix, the pure statistical criteria, such as MSFE or MAFE are hardly tractable. Thus, following Diebold and Mariano (1995) and Engle and Colacito (2006), we concentrate on economic criteria. The latter are directly linked to an investor's objective function and allow to make a policy proposal based on the obtained evidence. Next, the suggested ex-ante performance evaluation criteria are introduced in detail.

### 5.1 Expected loss criterion

The expected relative loss criterion is grounded on the global minimum variance portfolio (GMVP) weights. This portfolio allows to escape from problems of asset mean forecasting. Merton (1980) points out that the estimation of mean returns is extremely imprecise and imposes severe adverse portfolio performance (Best and Grauer, 1991). Our approach differs from Fleming et al. (2001, 2003), Engle and Colacito (2006) which are concerned with economic effects of covariance matrix forecasting for investors with different preferences. Our sole purpose is to assess the performance of competing correlation matrix predictors. A covariance matrix forecast is denoted as  $\tilde{\mathbf{S}}_{\tau,h}$ , while the realized covariance matrix  $\hat{\mathbf{S}}_{\tau+h}$  is perceived as a true one. Since we concentrate on correlation forecasting, the corresponding factorization of the realized covariance matrix is  $\hat{\mathbf{S}}_{\tau,h} = (\hat{\mathbf{S}}_{\tau+h} \odot I_N)^{1/2} \hat{\mathbf{R}}_{\tau+h} (\hat{\mathbf{S}}_{\tau+h} \odot I_N)^{1/2}$ , while that of the forecasted one is  $\tilde{\mathbf{S}}_{\tau,h} = (\hat{\mathbf{S}}_{\tau+h} \odot I_N)^{1/2} \tilde{\mathbf{R}}_{\tau,h} (\hat{\mathbf{S}}_{\tau+h} \odot I_N)^{1/2}$ . The realized weights for the perfect volatility estimator and the forecasted counterparts are given by

$$\hat{\mathbf{w}}_{\tau+h} = \frac{\hat{\mathbf{S}}_{\tau+h}^{-1} \mathbf{1}}{\mathbf{1}' \hat{\mathbf{S}}_{\tau+h}^{-1} \mathbf{1}}, \quad \tilde{\mathbf{w}}_{\tau,h} = \frac{\tilde{\mathbf{S}}_{\tau,h}^{-1} \mathbf{1}}{\mathbf{1}' \tilde{\mathbf{S}}_{\tau,h}^{-1} \mathbf{1}}.$$

Then, the minimum attainable portfolio variance and the forecasted portfolio variance are

$$\widehat{PV}_{\tau+h} = \hat{\mathbf{w}}'_{\tau+h} \hat{\mathbf{S}}_{\tau+h} \hat{\mathbf{w}}_{\tau+h}, \quad \widetilde{PV}_{\tau,h} = \tilde{\mathbf{w}}'_{\tau,h} \hat{\mathbf{S}}_{\tau+h} \tilde{\mathbf{w}}_{\tau,h}.$$

By construction, the portfolio variance  $\widehat{PV}_{\tau+h}$  is the minimal attainable variance at day  $\tau + h$ . The idea of economic forecasting criteria is based on the evaluation of portfolio volatility losses due to not optimally chosen GMVP weights. The relative utility loss is defined as

$$\mathcal{L}_{\tau,h} = \frac{\widetilde{PV}_{\tau,h} - \widehat{PV}_{\tau+h}}{\widehat{PV}_{\tau+h}}. \quad (21)$$

To facilitate model contrasting we consider also ratios of the latter model specific performance measure against some benchmark model denoted BEN. Loss ratios are defined as

$$\mathcal{RLR}_{\tau,h} = \mathcal{L}_{\tau,h} / \mathcal{L}_{\tau,h}(\text{BEN}). \quad (22)$$

For model comparisons we rely on the average loss given by

$$\bar{\mathcal{L}}_h = (\mathcal{T} - T - h + 1)^{-1} \sum_{\tau=T+h}^{\mathcal{T}} \mathcal{L}_{\tau,h}. \quad (23)$$

The model with the smallest average loss value should be preferred. Similarly, average loss ratios

$$\overline{\mathcal{RLR}}_h = (\mathcal{T} - T - h + 1)^{-1} \sum_{\tau=T+h}^{\mathcal{T}} \mathcal{RLR}_{\tau,h}. \quad (24)$$

are used to indicate if a particular forecasting scheme outperforms the benchmark procedure.

## 5.2 Universal criterion

Alternative realized covariance measures imply different GMVP weights. Thus it is helpful to have a model performance criterion offering some guidance to choose between alternative estimators. For this purpose we construct a portfolio volatility measure based on GMVP returns,

$$rp_t = \widehat{\mathbf{w}}_t' \mathbf{r}_t, \quad t = 1, \dots, \mathcal{T}, \quad (25)$$

where  $\widehat{\mathbf{w}}_t'$  are GMVP weights, obtained from alternative realized covariance matrix estimators. Then, we can calculate the empirical GMVP return variance for the sample period as

$$V(rp) = (\mathcal{T} - 1)^{-1} \sum_{t=1}^{\mathcal{T}} (rp_t - \overline{rp})^2, \quad \overline{rp} = \mathcal{T}^{-1} \sum_{t=1}^{\mathcal{T}} rp_t. \quad (26)$$

The best performing realized covariance estimate delivers the smallest empirical portfolio variance  $V(rp)$ .

## 6 Empirical study

### 6.1 Data description

Our sample comprises transaction data for  $N = 25$  assets that constantly entered the major German stock market index, DAX, over the period Feb,01, 2000 to April,29, 2005. The sample covers  $\mathcal{T} = 1330$  trading days. We analyze the full stream of transactions settled via the automatic trading system Xetra.<sup>1</sup> To implement correlation forecasting methods we employ rolling time windows of length  $T = 250$ . Thus, our ex-ante forecasting comparisons are informative for a sample of  $\mathcal{T} - 250 - h$  forecasts with  $h = 1, 5, 10$  days forecast horizon.

Complementary to autoregressive features of conditional correlations the dynamic panel and factor model as well as the functional smoothing approach allow conditioning on exogenous information. For this study we consider a set of conditioning variables that has been motivated in frameworks of variance prediction (Ghysels et al., 2006), or covariance modeling, see Hafner et al. (2005), Sheppard (2004), Longin and Solnik (2001), Andersen et al. (2001a). The set of exogenous variables comprises both economic and financial factors. Next we provide a short description of exogenous factors. All variables, except for intraday DAX quotes<sup>2</sup>, are drawn from DataStream.

1. The German term spread,  $w_{1,t} = rl_t^{GE} - rs_t^{GE}$ , where  $rl_t^{GE}$  is the German 10 year government bond rate and  $rs_t^{GE}$  is the Euribor.
2. The German short term interest rate  $w_{2,t} = rs_t^{GE}$ .
3. German-US short term interest rate spread,  $w_{3,t} = rs_t^{GE} - rs_t^{US}$ , where  $rs_t^{US}$  is the 3 month US T-bill rate.
4. High-low log index ranges of the Dow Jones Industrial Average ( $w_{4,t}$ ).
5. Daily Dow Jones log returns ( $w_{5,t}$ ).
6. High-low log index ranges of the DAX ( $w_{6,t}$ ).
7. Daily DAX log returns ( $w_{7,t}$ ).

---

<sup>1</sup>The authors thank Interdisciplinary Center for Numerical Simulation, University of Kiel, Germany, for providing intraday data.

<sup>2</sup>Intraday quotations for the DAX were provided by the Deutsche Kapitalmarktdatenbank, Karlsruhe



8. Realized daily standard deviations of the DAX, calculated using all available intraday observations ( $w_{8,t}$ ).
9. The relative performance of alternative realized covariance estimators could also depend on the amount of available price quotations. For this reason we consider an additional factor  $w_{9,t}$ , measuring the intensity of automatic trading over the considered sample period. Variable  $w_{9,t}$  quantifies the deviation of Xetra trade intensity from a cubic trend. The latter measure is obtained by aggregating the number of trades for all assets,  $L_t = \sum_{i=1}^N K_{it}$ . Since in the period from April,01, 2000 to Oct,31, 2003 Xetra trading times were prolonged by 2.5 hours, we standardize  $L_t$  for the differences in the trading day length. The adjusted measure  $L_t^*$  is modeled in the following way (standard errors in parentheses):

$$L_t^* = 742.642 + 2.641 t - 1.071 \cdot 10^{-3} t^2 - 4.063 \cdot 10^{-8} t^3 + w_{9,t}. \quad (27)$$

(33.59)
(0.117)
(8.5 \cdot 10^{-5})
(1.8 \cdot 10^{-8})

## 6.2 Empirical results: Structure

The alternative approaches to determine realized covariance matrices provided in Section 2 and the competing devices of forecasting conditional correlation patterns given in Sections 3 and 4 allow model comparisons at two consecutive levels. In the first place one may select from the methods outlined in Section 2 some candidate that best exploits high frequency information while coping with both microstructure noise and the Epps effect. For this level of model comparison we rely on the empirical variance of the GMVP returns in (26).

Estimation uncertainty associated with realized moment measures could vary over time. As an example one might a-priori expect specific bias corrections as in (2) or the Fourier expansions provided in Section 2 to work the better the higher is the (time varying) number of available intraday price quotations. Thus we provide some time specific model comparisons to indicate in how far the advice offered by the unconditional criterion is robust over specific states characterizing market behavior over time.

Conditional on the particular approach delivering the overall minimum variance  $V(rp)$  we contrast the competing methodologies of correlation forecasting in a second step. At this stage we compare low frequency approaches with modeling devices that exploit high frequency covariation. Among the latter models dynamic panel and factor model specifications are evaluated against functional or exponential smoothing techniques. Moreover, we address the predictive content of single exogenous variables entering our analysis either as contemporaneous or predetermined covariates. Recall that at this level of model comparison the same realized variances are combined with alternative correlation measures.

## 6.3 Estimating realized covariances

To estimate realized covariances we follow both frequency and time domain approaches. To synchronize the price quotations we choose the one minute time interval with previous tick interpolation as the highest sampling frequency. Depending on the Xetra trading time the corresponding number of observations is on average either  $K = 460$  or  $K = 600$ , while the 10 minutes sampling frequency leads to  $\tilde{K} = 47$  or  $\tilde{K} = 61$  observations for the majority of trading days. Taking smaller intervals for the highest sampling frequency is not justified due to lacking liquidity for some assets. When applying the Bandi and Russell (2005) estimator of an 'optimal sampling frequency' it turns out that the obtained estimates are significantly different for realized variances in comparison with covariances. Therefore we use two different optimal sampling frequencies namely  $K_v = 48$  for variance and  $K_c = 150$  for covariance estimation. Conditioning on the latter estimates subsampling as in Zhang et al. (2005) is implemented to exploit the entire available information. We partition the full grid  $\mathcal{G}$

with  $K$  observations into  $G_c = 3$  subgrids for covariance and  $G_v = 10$  subgrid samples for variance estimation.

Adopting Fourier transformations does not require data synchronization but needs frequent trading for sufficient estimation precision. We choose  $F = 200$  as the number of coefficients which corresponds roughly to the recommendation in Barucci and Reno (2002). A smaller choice of  $F$  should not be used while it would increase the volatility of obtained estimators. Taking a larger number is not reasonable given too small a number of trades. For both Fourier and data synchronization methods, the overnight bias correction (Martens, 2002) as in (4) is conducted. This correction is done for all realized covariances, as well as to obtain the exogenous variable  $w_{8,t}$ , the realized daily standard error of the DAX.

Now we describe the alternative scenarios for estimating the realized covariance matrix. Realized variances are throughout estimated using subsampling with  $G_v = 10$  subgrids. Concerning the correlation matrix, we consider the following set of estimators. Providing a 'naive' benchmark we consider sampling at the 10 minutes frequency, denoted further as S2 method. Alternative approaches are to use subsampling with  $G_c = 10$  (S1) or  $G_c = 3$  (S3). Note that subtraction of market microstructure noise as in (2) may result in non-positive definite matrices. For this reason we do not subtract noise  $\sum_{k=1}^K \dot{\mathbf{r}}_{t,k} \dot{\mathbf{r}}'_{t,k}$  from the subgrid estimators. Realized correlation measures obtained from the Fourier approach with  $F = 200$  (S4) complete our set of alternative moment estimators. To assess the overall in sample performance of the alternative realized covariance estimators we consider the empirical variances of the implied GMVPs returns, given in (26). The four estimation strategies are ordered below according to their empirical performance measures.

Strategy		$V(rp)$	$\sqrt{Var[V(rp)]}$
Subsampling with $G_v = G_c = 10$ , see (2)	S1	9.493E-05	7.561E-06
Sampling at the 10 minute frequency	S2	1.103E-04	8.153E-06
Subsampling with $G_v = 10$ and $G_c = 3$ , see (1)	S3	1.181E-04	8.715E-06
Fourier expansions with $F = 200$ , see (3)	S4	1.363E-04	9.961E-06

Apparently, S1 delivers the smallest empirical GMVP variances and the uncertainty associated with the empirical variance also increases from S1 to S4. Accordingly, the forecasting comparisons further refer to this particular implementation of realized covariances.

As argued the relative performance of alternative conditional correlation measures might be subjected to structural changes over time. To address the robustness of the realized covariance selection we discuss briefly some conditional features of absolute GMVP returns ( $|rp_t|$ ).

Table 1 provides empirical frequencies of minimizing absolute GMVP returns. Moreover, we document pairwise comparisons of absolute GMVP returns. Over the whole forecasting period the preferred covariance estimator S1 delivers the smallest absolute GMVP return in 27.6% of cases, and, appears slightly inferior to the 10 min sampling scheme (S2) minimizing absolute GMPV returns in 31.5% of all time instances. According to the corresponding standard errors the latter measure exceeds the former with 5% significance. With respect to the empirical frequencies of minimizing the absolute GMVP returns the remaining covariance measures (S3 and S4) perform significantly worse in comparison to S1 and S2. In particular, the Fourier approach provides smallest absolute GMVP returns in only 18.4% of all time instances. Although the empirical frequency of providing smallest absolute GMVP returns favors S2, the pairwise comparisons reveal that S1 in fact outperforms all the remaining approaches to covariance estimation. From the latter criteria it turns out that S2 outperforms each competing device significantly in more than 50% of all observations.

The latter performance measures are also provided for particular subsamples each comprising one half of the available sample information. E.g, we split the entire sample into the first and second half, or condition the evaluation on scenarios where the conditional realized DAX volatility ( $w_{8,t}$ ) exceeds its unconditional median. Similarly we separate a subsample of absolute GMVP returns over

	$P(\min( rp_t ))$				Pairwise comparisons for $ rp_t $					
	S1	S2	S3	S4	S1<S2	S1<S3	S1<S4	S2<S3	S2<S4	S3<S4
	Full sample									
av.	0.276	0.315	0.225	0.184	0.572	0.553	0.623	0.477	0.521	0.593
$\sigma_{av.}$	0.014	0.014	0.013	0.012	0.015	0.015	0.015	0.015	0.015	0.015
	1st half of the sample									
av.	0.272	0.296	0.233	0.198	0.591	0.528	0.613	0.457	0.502	0.581
$\sigma_{av.}$	0.019	0.020	0.018	0.017	0.021	0.022	0.021	0.021	0.022	0.021
	2nd half of the sample									
av.	0.280	0.333	0.217	0.170	0.554	0.578	0.633	0.496	0.541	0.604
$\sigma_{av.}$	0.019	0.020	0.018	0.016	0.021	0.021	0.021	0.022	0.021	0.021
	DAX realized variance greater than its median									
av.	0.302	0.300	0.202	0.196	0.593	0.580	0.631	0.487	0.531	0.570
$\sigma_{av.}$	0.020	0.020	0.017	0.017	0.021	0.021	0.021	0.022	0.021	0.021
	Trading intensity $L_t$ greater than its median									
av.	0.285	0.335	0.206	0.174	0.554	0.594	0.663	0.515	0.548	0.600
$\sigma_{av.}$	0.019	0.020	0.017	0.016	0.021	0.021	0.020	0.022	0.021	0.021
	Residual terms $w_{9,t}$ exceeding their median									
av.	0.293	0.319	0.207	0.181	0.576	0.583	0.650	0.500	0.537	0.602
$\sigma_{av.}$	0.020	0.020	0.017	0.017	0.021	0.021	0.021	0.022	0.021	0.021

Table 1: Performance of four alternative correlation matrix estimators. The left hand side panel displays empirical frequencies of obtaining smallest absolute GMVP returns and the right hand side panel shows pairwise comparisons. Estimation strategies cover subsampling (S1, S3) sampling at the 10 minute frequency (S2) and a frequency domain approach (S4). 'av.' and  $\sigma_{av.}$  denote empirical means and the respective standard deviations.

scenarios where the trading index ( $L_t$ ) or the residuals  $w_{9,t}$  from the model in (27) are above their unconditional median. As it turns out all empirical frequencies listed for the full sample are quite close to the corresponding quantities obtained due to conditioning on time, DAX volatility or the number of trades. The empirical frequencies of minimizing absolute GMVP returns are remarkably stable across alternative subsamples. Minor differences are observed for pairwise comparisons that, however, do hardly justify to prefer S2, S3 or S4 over S1 for conditioning the following forecasting exercises.

## 6.4 Low frequency approaches

The forecasting competitions are now conditioned on a particular sequence of realized variances, namely S1 variance estimates. Comparing methods operating at a low (daily) frequency with high frequency approaches, it is clear that the former suffer from using limited information. Conditioning on realized variances, makes the following results informative on the limited information efficiency loss due to cross moment estimation. Table 2 provides average relative losses as defined in (23) for selected forecasting models. The comparison covers the adaptive strategy (ADA), functional correlation estimates (FSV), DCC and exponential smoothing of standardized vector returns (EXV). For contrasting with high frequency methods Table 2 also displays corresponding loss measures obtained for the dynamic panel model excluding any exogenous variable (DPA).

Regarding FSV, DCC and EXV we have implemented diverse specifications depending on conditioning variables or parameter choices. For these families of models Table 2 provides results for the particular implementation that delivers the best results within its model family. As such the

	ADA	FSV <sup>c</sup>			DCC			EXV			DPA
$h$		best	$w_{j,t}$	fron	best	$\alpha, \beta$	fron	best	$\phi$	fron	
1	2.155 (4.34)	2.276 (4.65)	$w_{5,t}$ (1.19)	0.908 (1.19)	0.944 (1.55)	.01,.90	0.922 (1.51)	1.032 (1.59)	.99	1.029 (1.59)	0.459 (0.64)
5	2.171 (4.71)	2.697 (52.9)	$w_{7,t}$ (1.19)	0.909 (1.19)	0.944 (1.57)	.01,.90	0.921 (1.52)	1.040 (1.62)	.99	1.037 (1.61)	0.461 (0.65)
10	2.188 (4.61)	2.235 (10.8)	$w_{7,t}$ (1.28)	0.928 (1.28)	0.946 (1.60)	.01,.90	0.918 (1.53)	1.048 (1.65)	.99	1.043 (1.64)	0.464 (0.64)

Table 2: Low frequency models (ADA, FSV<sup>•</sup>, DCC and EXV) against the pure panel autoregression (DPA). 'best' and 'fron' provide the best performing specification within a model family and the unattainable performance frontier, respectively. Medium columns give some indication of the best performing model. Average relative losses  $\bar{\mathcal{L}}_h$ , (23) are given for the full sample period. Standard deviations ( $\times 100$ ) are given in parentheses.

model comparison is somewhat 'biased' against the ADA where only a single implementation has been processed. Note that in reality an analyst might choose another implementation of FSV, DCC or EXV such that the benchmark results for the model families might be hardly achieved in practical work. Regarding the latter model families we also provide average loss measures implied by a *performance frontier* strategy. At each timepoint we select the particular model providing the smallest relative loss. Note that this strategy is completely infeasible in financial practice since it deserves ex-ante information. A particular merit of performance frontiers is, however, that they indicate the unattainable optimum of a model family in forecasting correlation patterns.

Not surprisingly, the dynamic panel formalization of realized correlations clearly outperforms all competing approaches conditioning covariance estimates on daily information. Independent of the particular forecasting horizon ( $h = 1, 5, 10$ ), the panel model delivers an average relative error of about 0.45 which in turn is approximately one half of the frontier of best performing FSV (0.91,  $h = 1$ ), DCC (0.922,  $h = 1$ ) and EXV (1.03,  $h = 1$ ) models. Regarding the frontier strategies of these model classes the adaptive approach is clearly outperformed, obtaining an average relative loss of 2.15. In comparison to feasible models, however, the latter performance measure can hardly be distinguished from that of the best functional smoothing model. The latter conditions correlation matrices on daily returns measured for the Dow Jones ( $h = 1$ ) or DAX ( $h = 5, 10$ ). The fact that the conditioning variable with highest predictive power depends on the considered forecasting horizon complicates model selection within this particular family. As a further indication of modeling risk attached to FSV one could also regard the relatively high standard error attached to average performance at higher forecasting horizons ( $h = 5, 10$ ). The latter standard errors are vastly (by a factor of at least 6) exceeding corresponding results obtained, for instance, for the DCC approach. In addition, the huge loss difference between the best performing functional model and the unattainable performance frontier could be seen as a particular disadvantage, since using FSV and selecting the 'wrong' conditioning variable is likely to result in substantial efficiency loss. In the same vein of argumentation, however, careful, daywise selection of the conditioning variable and model diagnosis might promise substantial efficiency gains when adopting FSV. In comparison to functional modeling DCC and, similarly, EXV obtain best performing models ( $\beta = 0.90, \alpha = 0.01$  for DCC, and  $\phi = 0.99$  for EXV) with an average performance close to the frontier spanned by the respective model family. The best performing DCC model outperforms all competing low frequency information approaches.

Conditioning the comparison of low frequency methods on loss estimates obtained for subsamples, comprising the first and second half of predictors, confirms the overall results discussed above. To economize on space corresponding subsample specific estimates are not provided here but available from the authors upon request.

## 6.5 High frequency information

Now we compare four competing model families exploiting high frequency information, namely dynamic panel (DPA), dynamic factor (DFA) and extended (DFX) models, functional (FSC) and exponential smoothing of realized covariances (EXC). The upper part of Table 3 documents the performance of the latter model families relative to the dynamic panel model, according to (24). First, in the left part of Table 3, we focus on the performance comparison of model variants mostly excluding exogenous variables. Since conditioning on exogenous variables is an essential feature of FSC it also enters model comparison at this stage.

The efficiency frontiers offered by alternative implementations of EXC or FSC both carry the potential to improve panel based forecasting. With respect to the full sample of relative loss estimates, the latter model families give an average loss  $\overline{\mathcal{RLR}}_h$  of about 90% relative to the panel autoregression (DPA) with both measures significantly less than unity. As mentioned before, however, the frontiers merely indicate on the best model specific results, hardly feasible in practice. Turning the interest on the feasible best performing model specifications within each family, FSC turns out inferior in relation to DPA. The best performing functional smoothing model gives average measures against the panel model between 1.05 and 1.08, which is significantly in excess of unity. Interestingly, the best conditioning variable for this modeling approach seems to be the German short term interest rate. The latter result is not only robust over alternative forecasting horizons but also over two subsamples of forecast evaluation.

Assuming an analyst has ex-ante knowledge of the best performing EXC, its implementation provides relative loss ratios that are significantly superior compared to DPA. Depending on the forecasting horizon  $\overline{\mathcal{RLR}}_h$  computed against the panel model varies between 96% and 98%. It is noteworthy, that the best performing exponential smoothing model is not homogenous over alternative forecasting horizons. For one step ahead forecasting  $\phi = 0.98$  appears optimal whereas for higher horizons the smoothing parameter deserves slight adjustment ( $\phi = 0.97$ ).

Dynamic factor models (DFA) offer a forecasting performance which is comparable to that of the best performing EXC. In particular, DFA with  $G = 6$  factors delivers relative loss ratios,  $\overline{\mathcal{RLR}}_h$ , which are slightly (but insignificantly) better than those reported for the best performing EXC. Throughout VAR ( $G = 3$ ) factor model performs weaker than the univariate AR specification with factor dimension  $G = 6$ . Note that the latter models are characterized by an ad-hoc choice of the dimension parameter so that the comparison with the best performing EXC model might be biased against the factor model. We do not provide a performance frontier for the factor model, that could be achieved via some time varying choice of  $G$ . Intuitively one may expect that the number of factors required to explain a given degree of variation in the data is not time invariant. Summarizing the latter considerations we conjecture that the class of dynamic factor models is promising to provide ex-ante insight into return correlation complementary to EXC based prediction. We briefly address the potential of combining the latter forecasts below.

The former discussion is conditional on the full sample of relative loss measures. The medium and lower panels of Table 3 provide corresponding results conditional on the first and second half of the forecast sample, respectively. These statistics indicate that the relative performance of competing high frequency approaches is robust over time in most respects discussed before.

$h$	No exogenous variables						Conditioning on exogenous variables									
	DFA		FSC <sup>c</sup>		EXC		DPX <sup>p</sup>		DPX <sup>c</sup>		DFX					
	VAR	AR	best	$w_{j,t}$	fron	best	$\phi$	fron	best	$w_{j,t}$	fron	best	$w_{j,t}$	fron	VAR	AR
	Full sample															
1	0.971 (0.39)	0.961 (0.48)	1.083 (2.87)	$w_{2,t}$ (0.27)	0.921 (0.44)	0.964 (0.48)	.97	0.904 (0.44)	1.001 (0.04)	$w_{5,t}$ (0.21)	0.939 (0.21)	1.015 (1.21)	$w_{9,t}$ (0.32)	0.915 (0.32)	0.968 (0.44)	0.966 (0.54)
5	0.982 (0.36)	0.971 (0.43)	1.065 (1.46)	$w_{2,t}$ (0.27)	0.919 (0.43)	0.978 (0.41)	.98	0.922 (0.43)	1.001 (0.05)	$w_{5,t}$ (0.26)	0.935 (0.26)	1.014 (1.41)	$w_{4,t}$ (0.30)	0.921 (0.30)	0.977 (0.40)	0.976 (0.50)
10	0.983 (0.37)	0.980 (0.40)	1.050 (0.75)	$w_{2,t}$ (0.29)	0.912 (0.43)	0.981 (0.42)	.98	0.925 (0.43)	0.999 (0.08)	$w_{6,t}$ (0.26)	0.938 (0.26)	1.007 (0.62)	$w_{5,t}$ (0.30)	0.923 (0.30)	0.978 (0.41)	0.977 (0.45)
	First period															
1	0.976 (0.58)	0.971 (0.65)	1.099 (2.06)	$w_{2,t}$ (0.42)	0.916 (0.63)	0.978 (0.71)	.97	0.917 (0.63)	1.002 (0.07)	$w_{5,t}$ (0.28)	0.947 (0.28)	1.016 (1.93)	$w_{9,t}$ (0.36)	0.936 (0.36)	0.979 (0.69)	0.976 (0.74)
5	0.977 (0.53)	0.973 (0.63)	1.100 (2.87)	$w_{2,t}$ (0.39)	0.908 (0.64)	0.986 (0.67)	.98	0.930 (0.64)	0.999 (0.09)	$w_{4,t}$ (0.29)	0.946 (0.29)	1.029 (2.79)	$w_{5,t}$ (0.32)	0.939 (0.32)	0.978 (0.63)	0.979 (0.75)
10	0.975 (0.60)	0.974 (0.63)	1.074 (1.39)	$w_{2,t}$ (0.41)	0.899 (0.64)	0.990 (0.70)	.98	0.935 (0.64)	1.000 (0.16)	$w_{8,t}$ (0.33)	0.940 (0.33)	1.013 (1.23)	$w_{5,t}$ (0.30)	0.940 (0.30)	0.980 (0.73)	0.970 (0.72)
	Second period															
1	0.966 (0.53)	0.951 (0.71)	1.024 (0.41)	$w_{2,t}$ (0.33)	0.926 (0.62)	0.951 (0.63)	.97	0.890 (0.62)	1.000 (0.04)	$w_{7,t}$ (0.32)	0.931 (0.32)	0.983 (0.67)	$w_{1,t}$ (0.52)	0.895 (0.52)	0.958 (0.55)	0.955 (0.79)
5	0.987 (0.48)	0.970 (0.59)	1.030 (0.42)	$w_{2,t}$ (0.35)	0.931 (0.56)	0.969 (0.47)	.98	0.913 (0.56)	0.996 (0.21)	$w_{8,t}$ (0.43)	0.924 (0.43)	0.979 (0.61)	$w_{1,t}$ (0.49)	0.902 (0.49)	0.977 (0.48)	0.974 (0.65)
10	0.990 (0.42)	0.987 (0.50)	1.026 (0.45)	$w_{2,t}$ (0.40)	0.926 (0.55)	0.973 (0.47)	.98	0.916 (0.55)	0.996 (0.10)	$w_{6,t}$ (0.39)	0.936 (0.39)	0.980 (0.65)	$w_{1,t}$ (0.50)	0.904 (0.50)	0.977 (0.38)	0.984 (0.55)

Table 3: Forecast performance of high frequency approaches. The table shows empirical means of relative model performance  $\overline{\mathcal{RCR}_h(\bullet)}$  as defined in (22), computed against DPA. Standard deviations ( $\times 100$ ) are given in parentheses. Performance measures are provided for the full sample as well as for the first and second half of the forecasting period. For further notes see also Table 2.

	DPA	DPX <sup>p</sup>	DPX <sup>c</sup>	DFA	DFX	FSC <sup>p</sup>	FSC <sup>c</sup>	EXC	av.
DPA	–	0.544	0.463	0.364	0.380	0.664	0.615	0.402	0.042
DPX <sup>p</sup>	0.503	–	0.449	0.358	0.379	0.668	0.617	0.396	0.045
DPX <sup>c</sup>	0.576	0.522	–	0.416	0.398	0.626	0.598	0.427	0.163
DFA	0.373	0.373	0.374	–	0.506	0.700	0.703	0.507	0.162
DFX	0.399	0.404	0.395	0.466	–	0.678	0.675	0.506	0.172
FSC <sup>p</sup>	0.645	0.649	0.641	0.676	0.673	–	0.442	0.330	0.080
FSC <sup>c</sup>	0.599	0.599	0.597	0.663	0.654	0.431	–	0.352	0.078
EXC	0.416	0.415	0.415	0.491	0.528	0.356	0.370	–	0.258
av.	0.030	0.069	0.044	0.146	0.208	0.106	0.117	0.281	1

Table 4: Pairwise model comparisons at  $h = 1$  (upper triangular) and  $h = 10$  (lower triangular) forecast horizon. The table gives empirical probabilities  $P(\mathcal{L}_{\tau,h}(M_i) < \mathcal{L}_{\tau,h}(M_j))$ , where  $M_i$  and  $M_j$  are short for two particular forecasting models. For instance, the best performing DPX<sup>c</sup> model obtains smaller  $h = 1$  ( $h = 10$ ) losses than best performing FSC<sup>c</sup> in 59.8% (59.7%) of all (full sample) time instances. 'av.' gives model specific empirical frequencies of obtaining the smallest loss among the set of predictors listed in the table. The set of competing models comprises DPA, best performing DPX<sup>•</sup>, DFA  $G = 6$ , DFX  $G = 6$ , best performing FSC<sup>•</sup> and EXC.

## 6.6 Exogenous variables - I

For both model classes, dynamic panel and dynamic factor models, we have also implemented model variants that exploit the predictive content of exogenous variables. Within the class of dynamic panel models we have distinguished model families with contemporaneous (DPX<sup>c</sup>) and predetermined explanatory variables (DPX<sup>p</sup>). Mean relative loss ratios  $\overline{\mathcal{RLR}}_h$  as defined in (24) are provided for these model specifications in the right hand side panels of Table 3.

Dynamic factor models extracting principal components from an extended set of variables (DFX) perform very similar as if factors were only extracted from the data field of realized Fisher transformed correlations (DFA). The performance of dynamic factor models appears to depend more on the distinction between VAR ( $G = 3$ ) and diagonal AR ( $G = 6$ ) forecasting with the latter performing slightly better than the former.

For the dynamic panel models we obtain performance frontiers indicating that particular model implementations could offer a considerable improvement of the DPA benchmark. For both families of panel model implementations, DPX<sup>c</sup> and DPX<sup>p</sup>, model class frontiers obtain  $\overline{\mathcal{RLR}}_h$  measures about 90% of DPA. The results, however, do not help to solve the trade off between DPX<sup>c</sup> and forecasting of conditioning variables on the one hand, and DPX<sup>p</sup> offering likely weaker causal links but more precise information for conditioning ex-ante predictions on the other hand.

To further characterize the performance of high frequency models Table 4 provides details on the relative performance of the particular models listed in Table 3. The right hand side column and the bottom line of Table 4 report how often a particular model specifications delivers the smallest relative loss at the  $h = 1$  and the  $h = 10$  forecast horizon, respectively. For instance, EXC turns out to show for 25.8% (28.1%) of forecasts the smallest relative loss at the  $h = 1$  ( $h = 10$ ) forecasting horizon and thereby clearly outperforms the best performing FSC<sup>•</sup>. Depending on the forecasting horizon the latter models yield smallest relative losses in about 10% of all observations. Aggregating over (the best performing) specifications within dynamic panel models,  $h = 1$  ( $h = 10$ ) step ahead forecasts have the smallest relative loss in about 25% (13%) of all time instances. Note that to achieve the latter performance measure in reality an analyst needs knowledge of the particular conditioning variable having highest predictive power on average. This may be a reason to prefer EXC for practical

purposes. However, even for the particular models reported here an analyst is presumed to use an optimal smoothing parameter. Factor models are more ad hoc compared to the latter model families. At an aggregate level dynamic factor models (AR,  $G = 6$ ) give smallest  $\overline{\mathcal{RLR}}_h$  measures in about 33.4% ( $h = 1$ ) and 35.4% ( $h = 10$ ) of all ex-ante forecasts.

Table 4 also provides pairwise model comparisons by documenting the empirical frequencies that some model  $M_i$  delivers smaller loss estimates than a competing model  $M_j$ . It turns out, for instance, that with  $h = 1$  step ahead forecasting the DPA delivers in 40.2% of all time instances smaller loss than the best performing EXC. Interestingly, both  $G = 6$  dynamic factor models, including or excluding exogenous information, obtain better loss estimates in comparison with 'optimal' EXC in slightly more than one half of all time points. At the  $h = 10$  forecast horizon extracting dynamic factors from correlations and exogenous variables outperforms the exponential smoothing approach in 52.8% of all considered time instances. The latter probability exceeds some benchmark level of 50% at the 10% significance level. Overall the pairwise model comparisons are quite similar when regarding forecast horizons  $h = 1$  vs.  $h = 10$  as reflected by the symmetric shape of the inner panel of Table 4.

## 6.7 Exogenous variables - II

To further explore the predictive power of the considered exogenous variables Table 5 documents performance measures for the class of autoregressive panel models augmented with single conditioning variables (DPX $\bullet$ ). Mean average relative losses  $\overline{\mathcal{L}}$  from (23) are given for models conditioning on the German term spread ( $w_{1,t}$ ) in the first column. All remaining results are reported relative to this model, i.e. the averages  $\overline{\mathcal{RLR}}_h$  given in Table 5 are computed from loss relations measured against the benchmark model. Results are given for the entire forecasting period (upper panels) and the second part of the forecasting period (lower panels). Performance measures for panel models with predetermined variables (DPX $^p$ , 1st and 3rd panel) are contrasted with corresponding results obtained from panel specifications with contemporaneous conditioning variables (DPX $^c$ , 2nd and 4th panel).

Conditional on the full sample of relative loss estimates it is generally preferable to specify the right hand side variables in predetermined form. Doing so an analyst avoids to provide conditional forecasts of explanatory variables,  $\tilde{w}_{i,\tau,h}$ , required for the ex-ante implementation of DPX $^c$  type models. Using the German term spread as a predetermined conditioning variable (DPX $^p$ ) yields  $\overline{\mathcal{RLR}}_h$  statistics which are about 10% smaller compared with the corresponding DPX $^c$  model. The respective standard errors indicate that this finding is significant at any conventional level.

Contrasting the losses when using other explanatory variables reveals that in the first place market returns improve conditional forecasting. Lagged market returns carry predictive information for conditional correlation patterns. It is worthwhile to point out that the latter predictive strength does not die out with expanding the forecasting horizon since both the relative measures and their absolute counterparts are remarkably stable over alternative forecasting horizons. Within both modeling classes (DPX $^c$  and DPX $^p$ ) using Dow Jones or DAX daily returns outperforms the benchmark scenario by an average reduction of  $\overline{\mathcal{RLR}}_h$  between 5.1% and 9.0%. Throughout these performance measures are significantly smaller than unity. Other conditioning variables, as e.g. the German daily high-low log price differential, offer similar average performance measures that cannot be distinguished statistically from conditioning on daily returns.

Comparing full sample estimates with average performance measures obtained for the second half of the forecasting period reveals that some of the former conclusions are not representative for the entire sample. Now modeling correlations conditional on the contemporaneous German term spread performs similarly as conditioning the analysis on predetermined Dow Jones or DAX returns. Apparently, forecasting the German term spread by means of the nonparametric first order



	$w_{1t}$	$w_{2t}$	$w_{3t}$	$w_{4t}$	$w_{5t}$	$w_{6t}$	$w_{7t}$	$w_{8t}$	$w_{9t}$
$h$	Full sample: predetermined conditioning variables								
1	0.506 (0.78)	1.313 (3.10)	0.998 (0.72)	0.954 (0.60)	0.934 (0.44)	0.950 (0.45)	0.934 (0.44)	1.019 (1.81)	0.940 (0.47)
5	0.524 (0.93)	1.511 (5.79)	1.158 (1.40)	0.938 (0.54)	0.934 (0.53)	0.937 (0.53)	0.934 (0.53)	0.939 (0.56)	0.935 (0.53)
10	0.504 (0.76)	1.049 (0.97)	1.073 (1.22)	0.950 (0.53)	0.949 (0.52)	0.948 (0.54)	0.949 (0.53)	0.954 (0.51)	0.950 (0.54)
	Full sample: contemporaneous conditioning variables								
1	0.660 (2.04)	1.233 (2.46)	1.274 (3.94)	0.916 (2.58)	0.914 (2.58)	0.903 (2.58)	0.898 (1.19)	0.892 (1.15)	0.901 (1.26)
5	0.624 (1.70)	1.218 (2.46)	1.257 (4.94)	0.908 (1.77)	0.910 (1.77)	0.909 (1.95)	0.921 (1.94)	0.944 (3.33)	0.951 (3.35)
10	0.592 (1.32)	1.168 (2.10)	1.102 (3.18)	0.912 (1.06)	0.911 (1.07)	0.912 (1.22)	0.932 (1.93)	0.931 (2.47)	0.931 (1.55)
	Second half of the sample: predetermined conditioning variables								
1	0.535 (1.24)	0.945 (0.84)	0.988 (1.16)	0.928 (0.63)	0.917 (0.63)	0.941 (0.65)	0.917 (0.63)	1.015 (0.79)	0.924 (0.70)
5	0.568 (1.56)	1.916 (11.1)	1.095 (1.88)	0.918 (0.87)	0.908 (0.84)	0.913 (0.85)	0.909 (0.84)	0.907 (0.90)	0.908 (0.84)
10	0.553 (1.21)	0.986 (1.32)	1.021 (2.08)	0.910 (0.99)	0.906 (0.98)	0.904 (1.01)	0.906 (0.98)	0.918 (0.95)	0.906 (1.01)
	Second half of the sample: contemporaneous conditioning variables								
1	0.461 (0.94)	1.493 (4.07)	1.789 (7.10)	1.036 (0.67)	1.042 (0.69)	1.021 (0.62)	1.041 (0.69)	1.011 (0.75)	1.055 (1.58)
5	0.458 (0.92)	1.518 (4.18)	1.751 (9.35)	1.034 (0.63)	1.042 (0.67)	1.023 (0.56)	1.041 (0.66)	1.019 (0.72)	1.076 (2.98)
10	0.460 (0.89)	1.436 (3.55)	1.420 (5.97)	1.037 (0.66)	1.043 (0.69)	1.028 (0.62)	1.043 (0.69)	1.012 (0.68)	1.061 (1.90)

Table 5: Including exogenous variables in panel autoregressions. The left hand side column provides mean relative loss estimates,  $\bar{\mathcal{L}}_h$ , obtained when conditioning on the German term spread ( $w_{1,t}$ ). Remaining factors are measured relative to term spread results obtaining empirical means  $\overline{\mathcal{RLR}}_h$  as defined in (22). Standard errors ( $\times 100$ ) in parentheses.

autoregressions performs rather well over this subperiod in comparison with the remaining possible conditioning variables. Using any other contemporaneously conditioning variable obtains an average relative performance exceeding unity. Most of the latter measures are significant at reasonable levels with the DAX realized standard error ( $w_{8,t}$ ) being the only exception.

For the second half of the sample period the lagged trading index ( $w_{9,t}$ ) has more predictive content as for the full sample. Conditioning on the latter index gives  $\overline{\mathcal{RLR}}_h$  that cannot be statistically distinguished from conditioning on the best performing indicators.

## 6.8 A note on combining forecasts

A particular insight offered from the methodological literature on forecasting (cf. Aiolfi and Timmermann, 2006) is that it is often preferable to combine alternative forecasts in a linear fashion and thereby obtain a new predictor. Combination of forecasts building upon different information sets might be more efficient than singular predictors. The focus of this paper is on single model

$h$	Full sample			1st half			2nd half		
	1	5	10	1	5	10	1	5	10
DFX	0.433 (0.56)	0.442 (0.60)	0.448 (0.63)	0.424 (0.75)	0.430 (0.84)	0.432 (0.83)	0.443 (0.84)	0.454 (0.87)	0.465 (0.94)
EXC	0.433 (0.57)	0.443 (0.61)	0.449 (0.62)	0.424 (0.76)	0.433 (0.80)	0.440 (0.82)	0.441 (0.85)	0.453 (0.90)	0.459 (0.93)
COM	0.431 (0.57)	0.441 (0.60)	0.448 (0.62)	0.422 (0.75)	0.429 (0.83)	0.433 (0.82)	0.440 (0.86)	0.453 (0.89)	0.464 (0.93)
COM $\leq$ DFX	0.756	0.766**	0.748	0.744	0.774**	0.766*	0.768*	0.746	0.735
COM $\leq$ EXC	0.756	0.764**	0.744	0.744	0.772**	0.757	0.768*	0.736	0.737

Table 6: Performance of combined forecasts. Average relative loss  $\bar{\mathcal{L}}_h$  for the combined forecasting strategy (COM), DFX (AR,  $G=6$ ) and best performing EXC. Standard deviations ( $\times 100$ ) in parentheses. \*\* and \* indicate significance at the 5% and 10% level. Corresponding simulated critical values are 0.762 and 0.760 ( $T = 1080$ ) and 0.768 and 0.764 ( $T = 540$ ), respectively.

comparison. A final set of predictors could be subjected to an encompassing analysis, which is beyond the scope of this paper. It is tempting, however, to have a hint at potential complementarity of particular forecasting models. From the preceding discussion two models show particularly good performance, namely the best performing exponential smoother of realized covariances (EXC) and DFX using  $G = 6$  univariate factor predictions. Owing to the completely different underlying modeling techniques (smoothing vs. PCA) it is likely that the obtained forecasts may be complementary to some extent. Providing a first glance at the latter issue, we combine two forecasts to a third one, denoted by 'COM', conditional on realized relative losses in the forecast origin  $\tau$ . To be precise, we choose at each  $\tau$  the particular predictor out of EXC and DFX delivering the smallest loss at  $\tau$

$$\tilde{\rho}_{\text{COM},\tau,h} = \begin{cases} \tilde{\rho}_{\text{DFX},\tau,h} & \text{if } \mathcal{L}_{\text{DFX},\tau} < \mathcal{L}_{\text{EXC},\tau} \\ \tilde{\rho}_{\text{EXC},\tau,h} & \text{if } \mathcal{L}_{\text{EXC},\tau} < \mathcal{L}_{\text{DFX},\tau}. \end{cases} \quad (28)$$

For the combined forecast Table 6 displays the average relative loss  $\bar{\mathcal{L}}_h$ . It turns out that on average COM delivers the smallest relative loss at all forecasting horizons, for the full sample. According to the reported standard errors the given mean estimates  $\bar{\mathcal{L}}_h$ , however, do not differ in a statistical sense. Pairwise comparisons of COM against both underlying singular forecasts show that the empirical probability of getting a smaller relative loss from COM is significant at the 5% level for  $h = 5$  step ahead predictions for the full sample and its first half. In case of inefficient forecast combinations one would expect that the empirical probability e.g. of  $\text{COM} \leq \text{EXC}$  is 0.75. We determine critical values for the empirical probabilities via  $10^4$  vector draws from the bivariate Gaussian and combining the noise in the same manner as described above. The obtained critical values are reported in the legend of Table 6.

Our findings indicate on the potential of combining forecast. Note that the applied strategy of dummy forecast combination is more or less ad hoc. It reveals, however, that e.g. DFX and EXC based forecasts could be subjected to more sophisticated combination strategies promising better performance in comparison with the singular ingredients.

## 7 Conclusions

In this paper we compare empirically alternative avenues to forecasting conditional correlations with ex-ante knowledge of future realized variances. We advocate two model classes feasible to analyze huge fields of realized correlations, namely the dynamic panel and dynamic factor models. The

analysis covers a performance comparison against a battery of competing specifications exploiting sample information at medium and ultra high frequencies. At a first level of model evaluation we contrast alternative realized covariance estimates.

For the German stock market realized covariance matrices based on subsampling with 10 minutes subgrids (cf. Zhang et al., 2005) yields on average the best performance in terms of GMVP return uncertainty. Dynamic or smoothing models (DCC, functional smoothing, RiskMetrics) exploiting daily price comovements are clearly outperformed by ultra high frequency models. The dynamic factor models turn out to perform rather accurately. Jointly with an exponential smoother of realized covariance matrices they deliver the best average forecasting performance. Compared to the latter frameworks the dynamic panel and functional smoothing of realized covariances turn out inferior. The dynamic factor approach appears particularly attractive since we compare some ad-hoc implementation against best performing exponential smoothers selected over a variety of model implementations. It turns out that a very simple strategy of combining forecasts offered from the factor model and exponential smoother gives further improvements when forecasting dynamic correlations at higher horizons. With regard to the predictive content of exogenous variables we find that stock index return variables appear more fruitful, on average, as other indicators of market volatility or quoting intensity.

The paper gives a first hint at the predictive potential of dynamic factor models in the field of correlation forecasting. The employed model specifications are ad-hoc and in the light of the promising results deserve refinements. We regard the thorough treatment of specification issues in this field of factor modeling and correlation forecasting as an issue of future research. Similarly, the forecasting comparisons indicate on the potential of combining alternative predictors to obtain new forecasts outperforming its single ingredients. Weighting competing predictors to obtain improved correlation forecasts is another promising area for further investigation.

## Appendix A

**Proof of Proposition 1:** The realized correlation  $\rho_{ij}$  for synchronized data is estimated by

$$\hat{\rho}_{ij} = \frac{\sum_{k=1}^K r_{i,k} r_{j,k}}{\sqrt{\sum_{k=1}^K r_{i,k}^2 \sum_{k=1}^K r_{j,k}^2}},$$

where  $K$  is a number of intraperiod observations. The asymptotic distribution of  $\hat{\rho}_{ij}$ , provided by Barndorff-Nielsen and Shephard (2004) in Proposition 3, is

$$\sqrt{K} \frac{\hat{\rho}_{ij} - \rho_{ij}}{\sqrt{V(\hat{\rho}_{ij})}} \xrightarrow{L} \mathcal{N}(0, 1).$$

Let  $x$  follow the asymptotic distribution  $\sqrt{n}(x_n - \mu) \xrightarrow{L} \mathcal{N}(0, \sigma^2)$ . According to the delta method, for a continuous function  $g(x_n)$  not involving  $n$  it follows that

$$\sqrt{n}(g(x_n) - g(\mu)) \xrightarrow{L} \mathcal{N}(0, [g'(\mu)]^2 \sigma^2).$$

Notice that the mean and variance of the limiting distribution are the mean and variance of the linear Taylor series approximation, i.e.  $g(x_n) \simeq g(\mu) + g'(\mu)(x_n - \mu)$ . Fisher- $z$  transformation provides  $\hat{z}_{ij} = g(\hat{\rho}_{ij}) = 0.5 \ln\{(1 + \hat{\rho}_{ij})/(1 - \hat{\rho}_{ij})\}$ , thus

$$E(\hat{z}_{ij}) \approx g(E(\hat{\rho}_{ij})) = \frac{1}{2} \ln \frac{1 + \rho_{ij}}{1 - \rho_{ij}}, \quad V(\hat{z}_{ij}) \approx \frac{V(\hat{\rho}_{ij})}{(1 - \rho_{ij}^2)^2}.$$

Due to  $\sqrt{V(\hat{\rho}_{ij})} = 1 - \rho_{ij}^2$  (Barndorff-Nielsen and Shephard, 2004), we get  $V(\hat{z}_{ij}) \approx 1$ , which completes the proof.

# Appendix B

## B1 Implementation of the dynamic panel model

To implement the dynamic panel model for ex-ante forecasting we proceed along the following lines.

- i. The initial estimates of the variance parameters  $(\hat{\sigma}_\mu^2, \hat{\sigma}_\nu^2)$  are obtained from the fixed effects model.
- ii. We transform the dynamic model in terms of first differences,

$$\Delta \mathbf{z}_{t+h} = \Delta \mathbf{z}_t \delta + \mathbf{j}_N (\Delta \mathbf{w}'_{t+h} \boldsymbol{\beta}) + \Delta \boldsymbol{\nu}_{t+h},$$

and apply instrumental variables techniques (Anderson and Hsiao, 1982) to estimate the autoregressive parameter and slope coefficients. As instruments  $\Delta \mathbf{z}_{t-1}$  ( $\Delta \mathbf{w}_{t+h}$ ) are used for  $\Delta \mathbf{z}_t$  ( $\Delta \mathbf{w}_{t+h}$ ).

- iii. Conditional on  $\hat{\delta}$  and  $\hat{\boldsymbol{\beta}}$ , an intercept term  $\hat{\gamma}$  is determined. Corresponding model residuals contain both random individual effects and idiosyncratic noise, i.e.

$$\hat{v}_{it} = z_{i,t} - \hat{\gamma} - z_{i,t-1} \hat{\delta} - \mathbf{w}'_t \hat{\boldsymbol{\beta}} = \hat{v}_{i,t} + \hat{\mu}_i.$$

- iv. Collecting the latter residuals in an  $(\tau - h)N$ -dimensional column vector  $\mathbf{v} = (\mathbf{v}'_1, \dots, \mathbf{v}'_{\tau-h})'$  the vector of random effects is estimated as

$$\hat{\boldsymbol{\mu}} = \frac{\hat{\sigma}_\mu^2}{(\tau - h)\hat{\sigma}_\mu^2 + \hat{\sigma}_\nu^2} (I_N \otimes \mathbf{j}'_{(\tau-h)}) \hat{\mathbf{v}}.$$

- v. Finally, for models excluding any exogenous variables, conditioning on contemporaneous or predetermined exogenous variables forecasts are denoted and given by, respectively,

$$\text{DPA} : \tilde{\mathbf{z}}_{\tau+h} = \hat{\gamma} \mathbf{j}_N + \hat{\delta} \mathbf{z}_\tau + \hat{\boldsymbol{\mu}}, \quad (29)$$

$$\text{DPX}^c : \tilde{\mathbf{z}}_{\tau+h}^c = \hat{\gamma} \mathbf{j}_N + \hat{\delta} \mathbf{z}_\tau + \mathbf{j}_N (\tilde{\mathbf{w}}'_{\tau+h} \hat{\boldsymbol{\beta}}) + \hat{\boldsymbol{\mu}}, \quad (30)$$

$$\text{DPX}^p : \tilde{\mathbf{z}}_{\tau+h}^p = \hat{\gamma} \mathbf{j}_N + \hat{\delta} \mathbf{z}_\tau + \mathbf{j}_N (\mathbf{w}'_\tau \hat{\boldsymbol{\beta}}) + \hat{\boldsymbol{\mu}}. \quad (31)$$

## B2 Implementation of dynamic factor models

For the practical implementation of factor models we use up to  $G = 6$  principal components. The latter are modeled via univariate autoregressions with the actual lag orders determined by means of the AIC criterion. Relying on the AIC criterion could be criticized from the perspective that this criterion is known to be liberal. Using consistent criteria as BIC or HQ, however, implicitly presumes existence of a true finite autoregressive order. Note that the latter is hardly realistic for (a system of) principal components extracted from a high dimensional Fisher- $z$  transformed vector of realized correlations. In the case of univariate autoregressions the estimated parameter matrices  $\hat{\Phi}_i$ ,  $i = 1, \dots, p$ , in (12) are of dimension  $6 \times 6$  with zero off-diagonal elements.

Alternatively, the first  $G = 3$  principal components are modeled jointly by means of VARs with AIC order determination. In this framework the estimated parameter matrices  $\hat{\Phi}_i$  are of dimension  $3 \times 3$  and not subjected to any restriction. At the first sight the vector autoregressive (VAR) approach to model variables which are orthogonal by construction appears inappropriate. Note, however, that orthogonality does not imply absence of serial cross correlation. As a consequence, conditional estimates of a future factor may depend not only on its own history but also on the remaining factors. We refrain from implementing four to six dimensional VARs owing to the involved large parameter spaces. For both univariate and multiple autoregressions the maximum autoregressive order is set to  $p_{\max} = 6$ .

## B3 Nonparametric forecasting of conditioning variables

The panel based approach to correlation forecasting might require some ex-ante measure of the vector of conditioning variables denoted as  $\tilde{w}_{\tau,h}$ . To treat the latter issue uniformly over alternative exogenous variables we rely on nonparametric first order autoregressions. By assumption,  $E[w_{\tau+h}|w_\tau = w] = g(w_\tau)$ , such that

$$w_{\tau+h} = g(w_\tau) + \varepsilon_{\tau+h},$$

with  $\varepsilon_\tau$  denoting a zero mean error process. We apply the Nadaraya-Watson predictor of Nadaraya (1964) and Watson (1964)

$$\hat{g}(w_\tau) = \frac{\sum_{t=1}^{\tau-h} H_b(w_\tau - w_t) w_{\tau+h}}{\sum_{t=1}^{\tau-h} H_b(w_\tau - w_t)}, \quad H_b(u) = H(u/b)/b,$$

where  $H(\cdot)$  is a kernel function and  $b$  the bandwidth parameter. As the kernel function we take the quadratic kernel

$$H(u) = \frac{15}{16}(1 - u^2)^2 I(|u| < 1),$$

where  $I(\cdot)$  denotes an indicator function. To address the trade-off between between estimation bias and uncertainty that characterizes bandwidth selection in nonparametric regression we choose  $b$  locally in a data driven manner as

$$b(w_\tau) = c\hat{f}(w_\tau)^{-a},$$

where  $\hat{f}(w)$  is a kernel density estimate for  $w$ . For density estimation we use  $1.06\sigma_w T^{-0.2}$  as the bandwidth parameter, where  $\sigma_w^2$  is the empirical variance of  $w_t$  and  $T = 250$  the number of available observations (Silverman, 1986). The parameters  $c, a$  are chosen as  $T^{-0.25}$  and 0.25, respectively. Jennen-Steinmetz and Gasser (1988) show that the latter choice of the local bandwidth roughly corresponds to spline smoothing.

## References

- Aiolfi, M., Timmermann, A., 2006. Persistence in forecasting performance and conditional combination strategies, *Journal of Econometrics* 135, 31–53.
- Andersen, T. G., Bollerslev, T., 1998. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts, *International Economic Review* 39, 885–905.
- Andersen, T. G., Bollerslev, T., Meddahi, N., 2004. Analytic evaluation of volatility forecasts, *International Economic Review* 45, 1079–1110.
- Andersen, T. G., Bollerslev, T., Diebold, F. X., Ebens, H., 2001a. The distribution of realized exchange stock return volatility, *Journal of Financial Economics* 61, 43–76.
- Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P., 2001b. The distribution of realized exchange rate volatility, *Journal of American Statistical Association* 96, 42–55.
- Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P., 2003. Modeling and forecasting realized volatility, *Econometrica* 71, 579–625.
- Anderson, T., Hsiao, C., 1982. Formulation and estimation of dynamic models using panel data, *Journal of Econometrics* 18, 47–62.

- Baltagi, B. H., 2001. *Econometric Analysis of Panel Data* (John Wiley, Chichester).
- Baltagi, B. H., 2006. Forecasting with panel data, Discussion paper series 1, Deutsche Bundesbank.
- Bandi, F. M., Russell, J. R., 2005. Realized covariation, realized beta, and microstructure noise, Working Paper.
- Barndorff-Nielsen, O. E., Shephard, N., 2002. Power variation and time change, Economics Group, Nuffield College, University of Oxford, Economics Papers 2002-W24.
- Barndorff-Nielsen, O. E., Shephard, N., 2004. Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics, *Econometrica* 72, 882–925.
- Barndorff-Nielsen, O. E., Shephard, N., 2006. Comment on Hansen and Lunde "Realized variance and market microstructure noise", *Journal of Business and Economic Statistics* 24, 179–181.
- Barucci, E., Reno, R., 2002. On measuring volatility and the GARCH forecasting performance, *Journal of International Financial Markets, Institutions and Money* 12, 183–200.
- Bauwens, L., Laurent, S., Rombouts, J., 2006. Multivariate GARCH models: a survey, *Journal of Applied Econometrics* 21, 79–109.
- Best, M. J., Grauer, R. R., 1991. On the sensitivity of mean-variance-efficient portfolios to changes in asset means: some analytical and computational results, *Review of Financial Studies* 4, 315–342.
- Bollerslev, T., 1990. Modeling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH approach, *Review of Economics and Statistics* 72, 498–505.
- Bollerslev, T., Engle, R. F., Wooldridge, J., 1988. A capital asset pricing model with time varying covariances, *Journal of Political Economy* 96, 116–131.
- Cai, Z., Fan, J., Yao, Q., 2000. Functional-coefficient regression models for nonlinear time series, *Journal of American Statistical Association* 95, 941–956.
- Dacorogna, M. M., Gençay, R., Müller, U., Olsen, R. B., Pictet, O. V., 2001. *An Introduction to High Frequency Finance* (Academic Press).
- Devlin, S. J., Gnanadesikan, R., Kettenring, J., 1976. Some multivariate applications of elliptical distributions, in: S. Ikeda T. Hayakawa H. Hudimoto M. Okamoto M. Siotani S. Yamamoto, eds, *Essays in Probability and Statistics* (The Editorial Committee for Publication of Essays in Probability and Statistics) 365–393.
- Diebold, F. X., Mariano, R. S., 1995. Comparing predictive accuracy, *Journal of Business and Economic Statistics* 13, 134–144.
- Engle, R., 2002. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models, *Journal of Business and Economic Statistics* 20, 339–350.
- Engle, R., Colacito, R., 2006. Testing and valuing dynamic correlations for asset allocation, *Journal of Business and Economic Statistics* 24, 238–253.
- Engle, R. F., Kroner, K. F., 1995. Multivariate simultaneous generalized ARCH, *Econometric Theory* 11, 122–150.
- Engle, R. F., Sheppard, K., 2001. Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH, UCSD DP 2001-15.

- Epps, T. W., 1979. Comovements in stock prices in the very short run, *Journal of American Statistic Association* 74, 291–298.
- Fleming, J., Kirby, C., Ostdiek, B., 2001. The economic value of volatility timing, *Journal of Finance* 56, 329–352.
- Fleming, J., Kirby, C., Ostdiek, B., 2003. The economic value of volatility timing using 'realized' volatility, *Journal of Financial Economics* 67, 473–509.
- Foster, D. P., Nelson, D. B., 1996. Continuous record asymptotics for rolling sample variance estimators, *Econometrica* 64, 139–174.
- Ghysels, E., Santa-Clara, P., Valkanov, R., 2006. Predicting volatility: Getting the most out of return data sampled at different frequencies, *Journal of Econometrics* 131, 59–95.
- Gospodinov, N., Gavala, A., Jiang, D., 2006. Forecasting volatility, *Journal of Forecasting* 25, 381–400.
- Hafner, C. M., van Dijk, D., Franses, P. H., 2005. Semi-parametric modelling of correlation dynamics, in: T. Fomby, C. Hill, eds, *Advances in Econometrics* (Elsevier Science) 59–103.
- Hansen, P. R., Lunde, A., 2005. A realized variance for the whole day based on intermittent high-frequency data, *Journal of Financial Econometrics* 3, 525–554.
- Hansen, P. R., Lunde, A., 2006. Realized variance and market microstructure noise, *Journal of Business and Economic Statistics* 24, 127–218.
- Härdle, W., Herwartz, H., Spokoiny, V., 2003. Time inhomogeneous multiple volatility modeling, *Journal of Financial Econometrics* 1, 55–95.
- Hotelling, H., 1953. New light on the correlation coefficient and its transforms, *Journal of the Royal Statistical Society, Series B* 15, 193–232.
- Jennen-Steinmetz, C., Gasser, T., 1988. A unifying approach to nonparametric regression estimation, *Journal of American Statistical Association* 83, 1084–1089.
- J.P.Morgan 1997. Riskmetrics: Riskgrades technical document, [www.riskmetrics.com](http://www.riskmetrics.com).
- Longin, F., Solnik, B., 2001. Extreme correlation of international equity markets, *Journal of Finance* 56, 649–676.
- Malliavin, P., Mancino, M. E., 2002. Fourier series method for measurement of multivariate volatilities, *Finance and Stochastics* 6, 49–61.
- Marcellino, M., Stock, J., Watson, M., 2006. A comparison of direct and iterated multistep ar methods for forecasting macroeconomic time series, *Journal of Econometrics* 135, 499–526.
- Martens, M., 2002. Measuring and forecasting S&P 500 index-futures volatility using high-frequency data, *Journal of Futures Markets* 22, 497–518.
- Merton, R. C., 1980. On estimating the expected return on the market: An exploratory investigation, *Journal of Financial Economics* 8, 323–361.
- Muirhead, R. J., 1982. *Aspects of Multivariate Statistical Theory* (John Wiley & Sons, Inc.).
- Nadaraya, E., 1964. On estimating regression, *Theory of Probability and its Applications* 9, 141–142.
- Niguez, T.-M., Rubia, A., 2006. Forecasting the conditional covariance matrix of a portfolio under long-run temporal dependence, *Journal of Forecasting* 25, 439–458.

- Oomen, R., 2006. Properties of realized variance under alternative sampling schemes, *Journal of Business and Economic Statistics* 24, 219–237.
- Poon, S.-H., Granger, C., 2003. Forecasting volatility in financial markets: a review, *Journal of Economic Literature* 41, 478–539.
- Sheppard, K., 2004. Economic factors and the covariance of equity returns, Working Paper.
- Silverman, B., 1986. *Density Estimation for Statistics and Data Analysis* (Chapman and Hall, London).
- Stock, J. H., Watson, M. W., 2002. Forecasting using principal components from a large number of predictors, *Journal of American Statistical Association* 97, 1167–1179.
- Tse, Y. K., Tsui, A. K. C., 2002. A multivariate generalized autoregressive conditional heteroskedasticity model time-varying correlations, *Journal of Business and Economic Statistics* 20, 351–362.
- Watson, G., 1964. Smooth regression analysis, *Sankhya Series A* 26, 359–372.
- Zhang, L., Mykland, P. A., Aït-Sahalia, Y., 2005. A tale of two time scales: Determining integrated volatility with noisy high-frequency data, *Journal of American Statistical Association* 100, 1394–1411.