

---

# Asymmetry and nonlinearity in forecasting multivariate stock market volatility

---

*Kumulative Dissertation  
der Wirtschaftswissenschaftlichen Fakultät  
der Universität Augsburg  
zur Erlangung des akademischen Grades eines Doktors  
der Wirtschaftswissenschaften  
(Dr. rer. pol)*

vorgelegt von

Herrn Dipl. Finanzökonom math. Moritz Daniel Heiden

Erstgutachter: Prof. Dr. Yarema Okhrin

Zweitgutachter: Prof. Dr. Dr. h.c. Günter Bamberg

Vorsitzender der mündlichen Prüfung: Prof. Dr. Marco Wilkens

Tag der mündlichen Prüfung: 21.09.2015



# Contents

<b>Contents</b>	<b>i</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Decompositions of the realized covariance matrix . . . . .	3
1.2 A vine copula approach for predicting multivariate realized volatility . . .	5
1.3 Investor attention and stock market volatility . . . . .	7
<b>2 Article 1: Pitfalls of the Cholesky decomposition for forecasting multivariate volatility</b>	<b>14</b>
2.1 Introduction . . . . .	15
2.2 Decomposition of the realized covariance matrix . . . . .	17
2.2.1 Cholesky decomposition . . . . .	18
2.2.2 Matrix exponential transformation . . . . .	19
2.2.3 HAR model . . . . .	20
2.2.4 Forecasting and bias correction . . . . .	22
2.2.5 Loss functions and the MCS . . . . .	23
2.3 Empirical study . . . . .	25
2.3.1 Data and descriptive statistics . . . . .	25
2.3.2 Optimal ordering . . . . .	27
2.3.3 Modeling and forecasting procedure . . . . .	28
2.3.4 Statistically testing forecast performance . . . . .	31
2.4 Conclusion . . . . .	32
<b>Appendix</b>	<b>33</b>
<b>3 Article 2: A multivariate volatility vine copula model</b>	<b>39</b>
<b>4 Article 3: Forecasting volatility with empirical similarity and Google Trends</b>	<b>40</b>
<b>5 Summary and outlook</b>	<b>41</b>

# Chapter 1

## Introduction

Performing accurate forecasts for financial assets is not only a very recent topic in light of the ongoing financial crisis, but one of the great challenges of statistics and economics in general. In the past years, increased computing power and the availability of novel data sources lead to a wide range of new methodologies and increased the awareness for the necessity of high dimensional models, as most applications in asset and risk management require both, uni- and multivariate forecasts.

Predictability of asset returns has long been one of the most prominent topics in empirical finance, arising vivid debates as to which extent it violates the efficient market hypothesis (Fama, 1995) and random walk theories (Bachelier, 1900; Cootner, 1964). Short-term return forecasting is mostly viewed as impossible or at best difficult, see Christoffersen and Diebold (2006); Cenesizoglu and Timmermann (2012). At longer horizons time-varying risk premia or low frequent time series movements are sometimes used as an explanation for theoretical return predictability, see Barberis (2000); Ferreira and Santa-Clara (2011). Lettau and Nieuwerburgh (2008) provide an overview on the topic, pointing out, that changes in the mean of the return process pose a problem for prediction. While a general consent on the topic has not been reached, the dependence and variability of asset returns, usually captured by the covariance matrix, has been of rising interest in recent years. Motivated by studies of univariate volatility, in which the evidence on predictability is large (e.g. Andersen and Bollerslev (1998a); Forsberg and Ghysels (2007)), similar results were obtained for the multivariate case, see amongst others Engle (2002); Barndorff-Nielsen and Shephard (2004b); de Pooter et al. (2008).

Sophisticated methods that directly estimate and forecast the latent volatility process have spread since the availability and improved accessibility of high-frequency data. Research was, amongst others, initially triggered by Olsen (2011) and Dacorogna (2001). Particularly, the use of squared returns to estimate volatility has become a standard method in empirical finance. Originating from stochastic calculus, where Doléans-Dade (1967) showed that the sum of squared returns is a consistent estimator for the quadratic

variation and Jacod (1994) derived the corresponding limit theory, the concept was soon applied in econometrics. There, the working assumption of the price process being a semi-martingale could be justified under no-arbitrage assumptions (Andersen et al., 2003). For both, the uni- and multivariate setting, the corresponding asymptotic theory was derived by Barndorff-Nielsen (2002) and Barndorff-Nielsen and Shephard (2004a). The obtained measures are so called realized volatilities and its multivariate counterpart realized covariances, which can be refined and robustified by various estimation methods, see Zhang et al. (2005); Zhang (2006); Jacod et al. (2009); Zhang (2011). The advantage of these realized measures lies in them being observable time series and as a result, a large variety of approaches for modeling and forecasting their dynamics has developed. Since empirical applications, such as asset pricing, portfolio optimization and evaluation of risks are mostly implemented in the multivariate case, forecasts of the required covariance matrix are a crucial ingredient and represent the central topic of this thesis.

On the one hand, multivariate volatility forecasting bears the same difficulties as its univariate counterpart, where a main focus lies on modeling so called stylized facts (see Andersen, Bollerslev, Christoffersen, et al. (2006)), such as long-memory, clustering, leverage effects and mean reversion. On the other hand, matters are complicated by the requirement of the forecasted covariance matrix to be symmetric and positive semi-definite. In chapter 2 the Cholesky decomposition, a widely used method to guarantee both properties is studied. A special focus lies on its often neglected pitfalls and possible solutions in empirical application. Additionally, understanding, measuring and forecasting the dependence structure in the multivariate context is vital. As the interconnectedness of economies has strongly increased in recent time, financial assets become more dependent particularly during extremely negative economic phases and asset market volatility linkages tighten during periods of financial turmoil, as Cappiello et al. (2006) highlight. This asymmetric dependence has important implications, e.g. for portfolio allocation, as the variance of a financial portfolio return depends not only on the variances of the individual assets but also on the correlations between the assets. To advance from the strict Gaussian framework to more flexible structure, dependence is commonly studied using the concept of copulas (Joe, 1997; Nelsen, 2006), which are also suitable for capturing the interdependencies between asset volatilities (Mendes and Accioly, 2012). Chapter 3 introduces a new model for forecasting multivariate volatility that utilizes vine copulas to account for nonlinear dependence and asymmetry. The models predictive accuracy is compared to conventional models using statistical as well as economic measures.

Another often discussed topic is the role of external factors and variables in forecasting volatility, see e.g. Andersen and Bollerslev (1998b); Engle and Patton (2001). Traditionally, volatility is a measure of uncertainty of market participants regarding the

current and future state of the market. Chapter 4 analyzes the benefits of using a similarity based model, comparing current volatility characteristics and investors attention to the stock market for forecasting univariate volatility. The general ideas and findings of the articles, which constitute the main part of this dissertation are shortly described in the remainder of this introduction.

## 1.1 Decompositions of the realized covariance matrix

A major complication in forecasting multivariate volatility is the requirement of symmetry and positive semi-definiteness of the predicted covariance matrix, while preserving parsimony in the modeling procedure. In the literature, two alternative methods are usually applied. Restrictions on the model parameters or a decomposition of the covariance matrix. While the first one is easily applicable, model parsimony suffers especially in large dimensions. Hence, the latter approach is preferred in the literature, where a variety of decompositions exist, each of them with specific advantages and disadvantages. A typical modeling procedure starts with a measure of multivariate volatility, e.g. a time series of realized covariance matrices. At each point of time, the time series is decomposed and the resulting vector or matrix of the decomposition is modeled and forecasted using uni- or multivariate time series models. In the last step, the decomposition is reversed and ideally leads to a forecast of the realized covariance (RCOV) matrix, which is symmetric and positive semi-definite.

One of the most prominent methods is the Cholesky decomposition (CD), which has previously been applied by Halbleib and Voev (2011); Chiriac and Voev (2011); Becker et al. (2010). Despite its popularity, it suffers from the drawback that a change in the order of the elements in the original covariance matrix leads to a different decomposition and as a result influences modeling and forecasting. The problem is well known in the literature on vector autoregression (VAR), see Keating (1996), where some authors, e.g. Klößner and Wagner (2014) suggest a brute force approach, combining the results from potentially all orderings of the elements. However, as the number of elements increases, the number of possible orderings grows in a factorial sequence. Consequently, this approach gets computationally burdensome even in relatively small dimensions. The goal of the first article of this dissertation in chapter 2 is to analyze the impact of the ordering on forecasts for the RCOV matrix, if a CD is used in the modeling approach. For each of the 720 possible permutations of an exemplary six dimensional data set of asset returns ranging from January 1, 2000 to July 30, 2008, a modeling procedure as mentioned before is performed: First, the CD is applied at each point of time on the RCOV belonging to the corresponding permutation. Second, the time series of Cholesky elements are modeled based on the heterogeneous autoregressive (HAR) model of Corsi (2009) and one-step ahead out-of-sample forecasts are generated. Third,

the decomposition is reversed to obtain a forecast of the RCOV matrix. Due to the nonlinear transformation in the last step, a bias is induced, which we study using a bias correction similar to Chiriac and Voev (2011) and Bauer and Vorkink (2011). To evaluate predictive accuracy, the multivariate mean squared error and quasi likelihood loss functions are calculated for each permutation and across time. Analyzing the loss distributions, we find differences of up to 18% in predictive accuracy between the average loss of the best and worst model. Yet, the best and worst models are not consistent over time, so that a clear recommendation to which order to use for the next point of time based on previous performance is not at hand.

A detailed analysis of the loss differences based on the model confidence set (MCS) framework of Hansen, Lunde, and Nason (2011) reveals that the forecasts of the ordering with the smallest average loss are indeed significantly better than the forecasts of the ordering with the largest average loss. Hence, choosing an arbitrary and possibly “wrong” ordering may lead to misjudgment of the models forecasting ability in general. Furthermore, we show that an ex-ante analysis of the correlation structure of the assets, as it is sometime proposed in the VAR literature, does not yield significantly better forecasting results. In case of applying a bias correction for the forecasts, the differences between best and worst model even worsen. While the bias correction in general improves forecasting accuracy, the loss distribution over all permutations widens and differences between largest and smallest average loss increase for both loss functions. A possible solution to the ordering problem comes in the form of another decomposition, the matrix exponential transformation (MET), which was first applied in forecasting multivariate volatility by Bauer and Vorkink (2011). On the one hand, the MET suffers from biased forecasts, similar to the CD. On the other hand, the elements of the MET are not explicitly linked to the elements of the matrix it is applied on, making the MET order invariant. Comparing forecasts of both decompositions, we find that after bias correction, predictive accuracy does not significantly differ between the CD with the smallest average loss and the MET. Thus, for empirical application, two conclusions can be drawn. If a reasonable order can be imposed on the elements of the covariance matrix or if the connections between the elements of the decomposed covariance matrix are of interest, the CD is a rational choice. Otherwise, the application of the MET together with a bias correction is advised, be it for comparative reasons or simply to avoid the time consuming process of estimating all possible permutations of the CD.

## 1.2 A vine copula approach for predicting multivariate realized volatility

While the article in the previous section pointed out drawbacks from using the CD in forecasting the RCOV matrix, the decomposition is still irreplaceable if the structure and interconnectedness of its entries is of importance. As pointed out in section 1.1, RCOV matrices are often modeled and forecasted using a step-wise procedure, where different time series models are applied on the whole matrix, its individual elements or a favorable decomposition. Directly modeling the components of the RCOV matrix with univariate processes is possible (e.g. as described in Andersen, Bollerslev, Christoffersen, et al. (2006)), but does not guarantee positive definite forecasts and dynamic linkages among the series, such as volatility spillovers, might be neglected (Voev, 2008). Latest multivariate approaches that ensure symmetry and positive semi-definiteness of the RCOV matrix include the Wishart autoregressive (WAR) model proposed by Gouriéroux et al. (2009) and its dynamic generalization, the conditional autoregressive Wishart model (CAW) by Golosnoy et al. (2012). Chiriac and Voev (2011) choose the way of RCOV transformation and base their vector autoregressive fractionally integrated moving average (VARFIMA) model on a CD of the covariance matrix. Bauer and Vorkink (2011) instead transform the covariance matrix by using the MET and a factor model approach for the individual components. Disadvantages of these multivariate approaches are the lack of flexibility in the parameters and the inability to conveniently model non-Gaussianity and conditional heteroskedasticity in the volatility series itself. In contrast, the univariate framework offers a wide range of possibilities to tackle these problems, for example models based on fractionally integrated ARMA (ARFIMA) (Andersen et al. (2003)) or HAR processes (Corsi, 2009) can be estimated under skewed error distributions and various generalized autoregressive conditional heteroscedasticity (GARCH) augmentations (see Engle (2002); Corsi et al. (2008)). In combination with a CD modeling procedure, symmetry and positive semi-definiteness of the forecasts of the RCOV matrix can be ensured. Additionally, applying the CD bears the advantages of naturally interconnected entries within the matrix. Due to the nature of the decomposition, the relation between the elements is not linear and characteristic dependence patterns can be observed. As Andersen et al. (1999) point out, these patterns are subject to the high correlation between realized correlations and realized volatility, which can be attributed to the increased interconnectedness of economies. Copulas are a convenient and meanwhile established way to account for a variety of nonlinear dependence patterns among the realized covariances (see Mendes and Accioly (2012)). However, the choice of multivariate copulas is limited. In contrast, the bivariate case offers a rich variety of different copulas with flexible dependence patterns, based upon a steadily growing literature especially in the



GARCH framework, see e.g. Aas and Berg (2009); Liu and Luger (2009); Fischer et al. (2009).

A corresponding dynamic framework for modeling and forecasting RCOV matrices using vine copulas to account for more flexible dependencies between assets is studied in the second article of this dissertation, see chapter 3. Using the same six-dimensional data set as in section 1.1, we introduce a stepwise modeling procedure based on various time series models, such as the ARFIMA and HAR model. Similar to the previous article, we apply these models on the individual elements of the CD of the RCOV matrix to guarantee symmetry and positive semi-definiteness of the forecasts. Following Corsi et al. (2008), we extend the models by including a GARCH component to account for the so called “volatility of realized volatility”. As shown in Bai et al. (2003), the common GARCH with Gaussian innovations is not able to account for very high values of kurtosis of the dependent variable, as it is only controlled by two parameters, the kurtosis of the error distribution and the persistence of the GARCH itself. Hence, observing excess skewness and kurtosis in our data, we estimate the models based on the class of skewed generalized error (SGED) distributions (see e.g. Fernandez and Steel (1998)). To select an appropriate vine structure for the elements of the CD, or more precisely for the i.i.d. residuals of the elements after marginal time series filtering, the correlation pattern between the elements is studied. We compare two different structural models, for which we estimate various bivariate copulas covering both tail dependence and tail asymmetry. Given the problem of the ordering of the assets as pointed out in section 1.1, we repeat the modeling procedure for all 720 possible orderings. Due to the computational burden, we focus on the ordering with the highest average log likelihood over all time series for further analysis. Analogously, we choose the vine structure which performs best compared to an arbitrary structure, which is selected and fitted according to the maximum spanning tree principle as proposed by Dißmann et al. (2013). While we find that tail asymmetries as implied by Clayton and Gumbel copulas are present, preliminarily deciding to use only Gaussian or Student’s  $t$  copulas significantly simplifies the model selection step and only slightly decreases the model’s log likelihood. Finally, the models can be applied in an one-day ahead out-of-sample forecasting exercise. After performing the same bias correction procedure as in section 1.1, we assess the usefulness of our method, comparing it to recent types of models for the RCOV matrix based on a MCS approach. However, as Laurent et al. (2013) point out sometimes the model with the smallest statistical loss function may not be the one preferred in the evaluation by economic consideration. Hence, we also focus on economic evaluation by means of conventional portfolio optimization and Value-at-Risk (VaR) forecasting. We find that using a vine structure leads to significant improvements for HAR models regarding statistical loss, mean-variance efficient portfolios and VaR predictions. For ARFIMA based vine models, results are not as unambiguous except for forecasting daily VaRs.

There, compared to conventional models, the vine structure leads to smaller capital requirements while providing significantly more accurate forecasts and avoiding large exceedances of the forecasted VaR. These results are in line with the vine models ability to model tail events due to their assumption of non-normality. Hence, especially in combination with easily applicable conventional models, such as the HAR, our modeling approach offers a flexible and promising way of using the advantages of copulas for forecasting multivariate realized volatility.

### 1.3 Investor attention and stock market volatility

While the previous sections introduced models based on the autoregressive nature of volatility, a large amount of research includes exogenous variables to form predictions for its future, as asset prices are most likely dependent on other factors, see Engle and Sheppard (2001). While Andersen and Bollerslev (1998b) analyze the impact of news announcements of US macroeconomic data and its influence on volatility, more recent studies, such as Barber et al. (2009), directly focus on the interest investors take in the market. Traditionally, so called investor attention is measured by indirect proxies like volume, turnover and news. While volume might be the natural candidate for forecasting purposes, several studies, e.g. Brooks (1998) and Donaldson and Kamstra (2005) suggest that it does not improve the accuracy of volatility predictions. News as an alternative measure are mostly irregular and may underly a considerable publication lag. Recent publications use internet message postings (S.-H. Kim and D. Kim, 2014), Facebook users sentiment data (Siganos et al., 2014) or search frequencies (Vozlyublennai, 2014) to assess the influence of retail investors attention on the stock market. Several studies, among them Da et al. (2011), Vlastakis and Markellos (2012) and Andrei and Hasler (2015), suggest that Google search volume is a driver of future volatility. While most of the previous studies focus on analyzing the in-sample relationship of volatility and investor attention, the last article in this dissertation explicitly concentrates on predictability in an out-of-sample forecasting framework.

We suggest including Google search data (via Google Trends) in the framework of empirical similarity (ES) introduced by Gilboa et al. (2006), augmenting an autoregressive (AR) model by a time-varying coefficient determined by the empirical similarity between last periods Google data and realized volatility. This approach has previously been suggested by Lieberman (2012) and resembles an autoregressive process with dynamic parameters. As a result, the model is able to depict stationary, non stationary and explosive behavior, which can often be found in time series of realized volatilities, see Chen et al. (2010); Hansen and Lunde (2014). The unique assumption behind the model is, that investors seek information about the market before they actively trade,

which allows us to draw inference for different states of investor attention and volatility. For example, if investor attention is high and volatility is low, future volatility is expected to rise due to increased participation of investors in the market. On the other hand, if the previous level of volatility was high, low attention indicates a change point of volatility dynamics, meaning investors are losing interest in the market. In this case, due to decreased participation, future volatility should decrease, too. Based on weekly realized volatility of the Dow Jones Industrial Average (DJIA) ranging from January 16, 2004 to October 18, 2013, we find that the model shows significantly better performance compared to traditional models in an in-sample comparison as well as an out-of-sample forecasting study. By including two alternative time-varying models, we highlight that forecasting performance is indeed driven by the use of Google Trends data in combination with the ES framework. Furthermore, we test the robustness of the out-of-sample study by using the realized kernel suggested in Barndorff-Nielsen, Hansen, et al. (2008) as an alternative proxy for volatility.

Our results confirm the findings of Vlastakis and Markellos (2012) and Vozlyublenniaia (2014), who state that investor attention is a driver of volatility on short horizons. As described by Andrei and Hasler (2015), this relationship is strongest in phases of high volatility, where investor attention tends to be high. Additionally to evaluating the forecasts based on a MCS approach, we highlight the practical application by predicting the weekly VaR. Here, the ES model produced significantly better VaR forecasts in terms of overall accuracy and required capital, while providing an adequate number of VaR violations. Furthermore, the inclusion of Google Trends data as simple additive term in classical realized volatility models, such as the ARFIMA and HAR model, did not improve forecasting accuracy. Hence, while linear models can be useful for assessing the correlation of volatility and investor attention and studying their dependence in an in-sample framework, these models are not flexible enough when it comes to forecasting. However, one drawback of our model is the availability of Google Trends data. Google standardizes the data and restricts the access for daily data to windows of 90 days, which cannot be merged into one meaningful time-series. Other issues include the lack of search data for certain assets or the problem that certain search terms are ambiguous. Nevertheless, given a certain quality of the data, our model of empirical similarity is easy to interpret, parsimonious and shows superior predictive ability, which makes the model attractive for economic reasoning as well as practical application

# Bibliography

- AAS, K. and D. BERG (2009). Models for construction of multivariate dependence - a comparison study. *The European Journal of Finance*: 15(7-8), pp. 639–659.
- ANDERSEN, T. G., T. BOLLERSLEV, P. F. CHRISTOFFERSEN, et al. (2006). Volatility and correlation forecasting. *Handbook of Economic Forecasting*: 1(05), pp. 777–878.
- ANDERSEN, T. G. et al. (1999). (Understanding, optimizing, using and forecasting) realized volatility and correlation. Working Paper.
- ANDERSEN, T. G. and T. BOLLERSLEV (1998a). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*: 39(4), pp. 885–905.
- ANDERSEN, T. G. and T. BOLLERSLEV (1998b). Deutsche Mark-Dollar volatility: Intraday activity patterns, macroeconomic announcements, and longer run dependencies. *Journal of Finance*: 53(1), pp. 219–265.
- ANDERSEN, T. G. et al. (2003). Modeling and forecasting realized volatility. *Econometrica*: 71(2), pp. 579–625.
- ANDREI, D. and M. HASLER (2015). Investor attention and stock market volatility. *Review of Financial Studies*. Forthcoming.
- BACHELIER, L. (1900). Théorie de la spéculation. *Annales Scientifiques de l'École Normale Supérieure*: 17(3), pp. 21–86.
- BAI, X., J. R. RUSSELL, and G. C. TIAO (2003). Kurtosis of GARCH and stochastic volatility models with non-normal innovations. *Journal of Econometrics*: 114(2), pp. 349–360.
- BARBER, B. M., T. ODEAN, and N. ZHU (2009). Do retail trades move markets? *Review of Financial Studies*: 22(1), pp. 151–186.
- BARBERIS, N. (2000). Investing for the long run when returns are predictable. *The Journal of Finance*: 55(1), pp. 225–264.
- BARNDORFF-NIELSEN, O. E. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*: 64(2), pp. 253–280.

- BARNDORFF-NIELSEN, O. E. and N. SHEPHARD (2004a). Econometric analysis of realized covariation: High frequency based covariance, regression and correlation in financial economics. *Econometrica*: 72(3), pp. 885–925.
- BARNDORFF-NIELSEN, O. E., P. R. HANSEN, et al. (2008). Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica*: 76(6), pp. 1481–1536.
- BARNDORFF-NIELSEN, O. E. and N. SHEPHARD (2004b). Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica*: 72(3), pp. 885–925.
- BAUER, G. H. and K. VORKINK (2011). Forecasting multivariate realized stock market volatility. *Journal of Econometrics*: 160(1), pp. 93–101.
- BECKER, R., A. CLEMENTS, and R. O’NEILL (2010). A Cholesky-MIDAS model for predicting stock portfolio volatility. Working Paper.
- BROOKS, C. (1998). Predicting stock index volatility: Can market volume help? *Journal of Forecasting*: 17(1), pp. 59–80.
- CAPPIELLO, L., R. F. ENGLE, and K. SHEPPARD (2006). Asymmetric dynamics in the correlations of global equity and bond returns. *Journal of Financial Econometrics*: 4(4), pp. 537–572.
- CENESIZOGLU, T. and A. TIMMERMANN (2012). Do return prediction models add economic value? *Journal of Banking & Finance*: 36(11), pp. 2974–2987.
- CHEN, Y., W. K. HÄRDLE, and U. PIGORSCH (2010). Localized realized volatility modeling. *Journal of the American Statistical Association*: 105(492), pp. 1376–1393.
- CHIRIAC, R. and V. VOEV (2011). Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics*: 26(6), pp. 922–947.
- CHRISTOFFERSEN, P. F. and F. X. DIEBOLD (2006). Financial asset returns, direction-of-change forecasting, and volatility dynamics. *Management Science*: 52(8), pp. 1273–1287.
- COOTNER, P. H. (1964). *The random character of stock market prices*. M.I.T. Press.
- CORSI, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*: 7(2), pp. 174–196.
- CORSI, F., S. MITTNIK, and C. PIGORSCH (2008). The volatility of realized volatility. *Econometric Reviews*: 27(1-3), pp. 46–78.
- DACOROGNA, M. M. (2001). *An introduction to high-frequency finance*. Academic Press.
- DA, Z., J. ENGELBERG, and P. GAO (2011). In search of attention. *Journal of Finance*: 66(5), pp. 1461–1499.
- DE POOTER, M., M. MARTENS, and D. van DIJK (2008). Predicting the daily covariance matrix for S&P 100 stocks using intraday data - but which frequency to use? *Econometric Reviews*: 27(1-3), pp. 199–229.

- DISSMANN, J. et al. (2013). Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics & Data Analysis*: 59(1), pp. 52–69.
- DOLÉANS-DADE, C. (1967). Intégrales stochastiques par rapport aux martingales locales. *Séminaire de probabilités 1967-1980*.
- DONALDSON, R. G. and M. J. KAMSTRA (2005). Volatility forecasts, trading volume, and the ARCH versus option-implied volatility trade-off. *Journal of Financial Research*: 28(4), pp. 519–538.
- ENGLE, R. F. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*: 20(3), pp. 339–350.
- ENGLE, R. F. and A. J. PATTON (2001). What good is a volatility model? *Quantitative Finance*: 1(2), pp. 237–245.
- ENGLE, R. F. and K. SHEPPARD (2001). Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. Working Paper.
- FAMA, E. F. (1995). Random Walks in stock market prices. *Financial Analysts Journal*: 51(1), pp. 75–80.
- FERNANDEZ, C. and M. F. STEEL (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*: 93(441), pp. 359–371.
- FERREIRA, M. A. and P. SANTA-CLARA (2011). Forecasting stock market returns: The sum of the parts is more than the whole. *Journal of Financial Economics*: 100(3), pp. 514–537.
- FISCHER, M. et al. (2009). An empirical analysis of multivariate copula models. *Quantitative Finance*: 9(7), pp. 839–854.
- FORSBERG, L. and E. GHYSELS (2007). Why do absolute returns predict volatility so well? *Journal of Financial Econometrics*: 5(1), pp. 31–67.
- GILBOA, I., O. LIEBERMANN, and D. SCHEIDLER (2006). Empirical similarity. *Review of Economics and Statistics*: 88(3), pp. 433–444.
- GOLOSNOY, V., B. GRIBISCH, and R. LIESENFELD (2012). The conditional autoregressive Wishart model for multivariate stock market volatility. *Journal of Econometrics*: 167(1), pp. 211–223.
- GOURIÉROUX, C., J. JASIAK, and R. SUFANA (2009). The Wishart autoregressive process of multivariate stochastic volatility. *Journal of Econometrics*: 150(2), pp. 167–181.
- HALBLEIB, R. and V. VOEV (2011). Forecasting multivariate volatility using the VARFIMA model on realized covariance Cholesky Factors. *Journal of Economics and Statistics (Jahrbuecher fuer Nationaloekonomie und Statistik)*: 231(1), pp. 134–152.
- HANSEN, P. R. and A. LUNDE (2014). Estimating the persistence and the autocorrelation function of a time series that is measured with error. *Econometric Theory*: 30(01), pp. 60–93.

- HANSEN, P. R., A. LUNDE, and J. M. NASON (2011). The Model Confidence Set. *Econometrica*: 79(2), pp. 453–497.
- JACOD, J. (1994). Limit of random measures associated with the increments of a Brownian semimartingale. Working Paper.
- JACOD, J. et al. (2009). Microstructure noise in the continuous case: the pre-averaging approach. *Stochastic Processes and their Applications*: 119(7), pp. 2249–2276.
- JOE, H. (1997). *Multivariate Models and Dependence Concepts*. London: Chapman & Hall.
- KEATING, J. W. (1996). Structural information in recursive VAR orderings. *Journal of Economic Dynamics and Control*: 20(9-10), pp. 1557–1580.
- KIM, S.-H. and D. KIM (2014). Investor sentiment from internet message postings and the predictability of stock returns. *Journal of Economic Behavior & Organization*: 107(PB), pp. 708–729.
- KLÖSSNER, S. and S. WAGNER (2014). Exploring all VAR orderings for calculating spillovers? Yes, we can! - A note on Diebold and Yilmaz (2009). *Journal of Applied Econometrics*: 29(1), pp. 172–179.
- LAURENT, S., J. V. K. ROMBOUTS, and F. VIOLANTE (2013). On loss functions and ranking forecasting performances of multivariate volatility models. *Journal of Econometrics*: 173(1), pp. 1–10.
- LETTAU, M. and S. V. NIEUWERBURGH (2008). Reconciling the return predictability evidence. *Review of Financial Studies*: 21(4), pp. 1607–1652.
- LIEBERMAN, O. (2012). A similarity-based approach to time-varying coefficient non-stationary autoregression. *Journal of Time Series Analysis*: 33(3), pp. 484–502.
- LIU, Y. and R. LUGER (2009). Efficient estimation of copula-GARCH models. *Computational Statistics & Data Analysis*: 53(6), pp. 2284–2297.
- MENDES, B. V. d. M. and V. B. ACCIOLY (2012). On the dependence structure of realized volatilities. *International Review of Financial Analysis*: 22, pp. 1–9.
- NELSEN, R. B. (2006). *An Introduction to Copulas*. 2nd. Berlin: Springer.
- OLSEN, R. B. (2011). *Olsen Research Inc.* URL: [www.olsen.ch](http://www.olsen.ch).
- SIGANOS, A., E. VAGENAS-NANOS, and P. VERWIJMEREN (2014). Facebook’s daily sentiment and international stock markets. *Journal of Economic Behavior & Organization*: 107(PB), pp. 730–743.
- VLASTAKIS, N. and R. N. MARKELLOS (2012). Information demand and stock market volatility. *Journal of Banking & Finance*: 36(6), pp. 1808–1821.
- VOEV, V. (2008). “Dynamic modelling of large-dimensional covariance matrices.” In: *High Frequency Financial Econometrics*. Ed. by L. BAUWENS, W. POHLMEIER, and D. VEREDAS. Studies in Empirical Economics. Physica-Verlag HD, pp. 293–312.
- VOZLYUBLENNAIA, N. (2014). Investor attention, index performance, and return predictability. *Journal of Banking & Finance*: 41(C), pp. 17–35.

- 
- ZHANG, L. (2006). Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. *Bernoulli*: 12(6), pp. 1019–1049.
- (2011). Estimating covariation: Epps effect, microstructure noise. *Journal of Econometrics*: 160(1), pp. 33–47.
- ZHANG, L., Y. AIT-SAHALIA, and P. A. MYKLAND (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*: 100, pp. 1394–1411.



## Chapter 2

# Article 1: Pitfalls of the Cholesky decomposition for forecasting multivariate volatility

**Abstract** This paper studies the pitfalls of applying the Cholesky decomposition for forecasting multivariate volatility. We analyze the impact of one of the main issues in empirical application of using the decomposition: The sensitivity of the forecasts to the order of the variables in the covariance matrix. We find that despite being frequently used to guarantee positive semi-definiteness and symmetry of the forecasts, the Cholesky decomposition has to be used with caution, as the ordering of the variables leads to significant differences in forecast performance. A possible solution is provided by studying an alternative, the matrix exponential transformation. We show that in combination with empirical bias correction, forecasting accuracy of both decompositions does not significantly differ. This makes the matrix exponential a valuable option, especially in larger dimensions.

**Keywords:** realized covariances; realized volatility; cholesky; decomposition; forecasting

**JEL Classification Numbers:** C1, C53, C58

## 2.1 Introduction

Forecasts of the covariance matrix are a crucial ingredient of many economic applications in asset and risk management. The concept of using realized covariances as a proxy for the unobservable volatility process has largely spread since the availability and improved accessibility of high-frequency data in finance. Recent approaches to forecast the realized covariance (RCOV) matrix are based upon uni- or multivariate time series models. To ensure mathematical validity of the forecasted RCOV matrix, such as symmetry and positive semi-definiteness, either parameter restrictions or decompositions are used. The latter one is preferred to guarantee parsimony, especially if dimensions are large.

Latest multivariate approaches that ensure symmetry and positive semi-definiteness of the RCOV matrix include the Wishart Autoregressive (WAR) model proposed by Gouriéroux et al. (2009) and its dynamic generalization, the Conditional Autoregressive Wishart by Golosnoy et al. (2012). Chiriac (2010) shows that the WAR estimation is very sensitive to assumptions on the underlying data, causing degenerate Wishart distributions and affecting the estimation results. Consequently, Chiriac and Voev (2011) choose the way of transformation and base their Vector Autoregressive Fractionally Integrated Moving Average (VARFIMA) model on a Cholesky decomposition (CD) of the covariance matrix. Bauer and Vorkink (2011) instead transform the covariance matrix by using the matrix exponential transformation (MET) and use a factor model approach for the individual components. Andersen, Bollerslev, Christoffersen, et al. (2006) and Colacito et al. (2011) modify the Dynamic Conditional Correlation (DCC) model of Engle (2002), splitting up variances and covariances in the modeling process. A similar approach is implemented in Halbleib and Voev (2014). The authors suggest a mixed data sampling method based on low-frequent estimators for the correlations.

Usually, the choice of the method can be motivated by the unique properties of the respective decomposition. For example, while the elements of the CD are explicitly linked to the entries of original RCOV matrix, this is not the case for the MET. On the other hand, both, CD and MET, do not separate variances and covariances. This can be achieved by applying a DCC type decomposition, allowing for more flexibility in the modeling process. Moreover, as Halbleib and Voev (2014) point out, high-frequency

estimators for the whole covariance matrix are often noisy and simple methods to reduce microstructure noise, such as sparse-sampling (see Andersen, Bollerslev, Diebold, et al. (2003)), are not applicable in large dimensions. Overall, due to its simplicity, the CD remains the most frequently used method in the literature, beside the problem that each permutation of the elements in the original matrix yields a different decomposition. The problem is well known in the literature on Vector Autoregression (VAR), where the VAR is usually identified using the corresponding CD to derive the dynamic response of each variable to an orthogonal shock (see e.g. Sims (1980)). Keating (1996) show, that the ordering of the variables is crucial to obtain structural impulse responses, which is only possible if the system of equations is partially recursive. The gravity of the problem for VAR based approaches has also been pointed out by Klößner and Wagner (2014), who analyze the extent to which measuring spillovers is influenced by the order of the variables.

In this paper, we focus on the impact of the ordering of the assets in the original covariance matrix on the forecasts, if a CD is used. Since the amount of possible permutations grows very fast with increasing number of assets, it is computationally burdensome to calculate and compare forecasts from each ordering. Therefore, we evaluate the predictive accuracy of all 720 permutations based on a small data set of six assets. Analyzing the loss distributions of two established loss functions, we find differences of up to 18% between the average loss of the best and worst model. Using the Model Confidence Set framework of Hansen, Lunde, and Nason (2011), we show that these loss differences are indeed statistically significant, meaning that an arbitrary ordering may result in suboptimal forecasts and hence poor model choices. Furthermore, the application of an ex-ante analysis of the correlation structure of the assets to obtain a specific ordering, as it is sometime proposed in the VAR literature, does not improve forecasting results significantly. Additionally, we take a look at the impact of a simple empirical bias correction, as the forecasts from both, CD and MET are biased by construction. We show, that using the ordering invariant MET and applying the bias reduction provides a possible solution to the ordering problem, as forecasts are not significantly different from the best CD.

## 2.2 Decomposition of the realized covariance matrix

Let  $R_t$  be the  $N \times 1$  vector of log returns over each period of  $T$  days. For a portfolio consisting of  $N$  stocks:

$$R_t = p(t) - p((t-1)),$$

$p(t) = (p_{1t}, \dots, p_{Nt})$  being the log price at time  $t \in [1, \dots, T]$ .

Assume, there are  $M$  equally spaced intra-day observations, the  $i$ -th intra-day return for the  $t$ -th period is:

$$r_{i,t} \equiv p((t-1) + i\frac{1}{M}) - p((t-1) + (i-1)\frac{1}{M}), \quad (2.1)$$

with  $i = 1, \dots, M$ . According to Barndorff-Nielsen and Shephard (2002), the  $N \times N$  RCOV matrix for the  $t$ -th period is then defined as

$$\mathbf{Y}_t = \sum_{i=1}^M r_{i,t} r'_{i,t}, \quad (2.2)$$

which is a consistent estimator for the conditional variance-covariance matrix of the log returns,  $Var [R_t | \mathcal{F}_{t-1}] = \Sigma_t$ . The estimator can be refined to reduce market microstructure noise (e.g. Hayashi and Yoshida (2005); Zhang et al. (2005)) and account for jumps (Christensen and Kinnebrock, 2010). The issue of asynchronicity of the data can be addressed by methods such as linear or previous-tick interpolation (Dacorogna, 2001) and subsampling (Zhang, 2011), which are easy to implement in empirical work. More complex procedures are often based on the use of multivariate realized kernels (see e.g. Barndorff-Nielsen, Hansen, et al. (2011)). However, as Halbleib and Voev (2014) point out, these methods are still limited in application as they may lead to data loss or do not guarantee positive definiteness.

### 2.2.1 Cholesky decomposition

The CD, decomposes a real, positive definite<sup>1</sup> matrix into the product of a real upper triangular matrix and its transpose (Brezinski, 2006).

The Cholesky decomposition of the naturally symmetric and positive semi-definite  $\mathbf{Y}_t$ , with  $\mathbf{P}_t$  being an upper triangular matrix, yields:

$$\mathbf{Y}_t = \begin{pmatrix} y_{11,t} & y_{12,t} & \cdots & y_{1N,t} \\ y_{12,t} & y_{22,t} & \cdots & y_{2N,t} \\ \vdots & \vdots & \ddots & \vdots \\ y_{1N,t} & \cdots & \cdots & y_{NN,t} \end{pmatrix} \quad (2.3)$$

$$= \begin{pmatrix} p_{11,t} & 0 & \cdots & 0 \\ p_{12,t} & p_{22,t} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ p_{1N,t} & p_{2N,t} & \cdots & p_{NN,t} \end{pmatrix} \begin{pmatrix} p_{11,t} & p_{12,t} & \cdots & p_{1N,t} \\ 0 & p_{22,t} & \cdots & p_{2N,t} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p_{NN,t} \end{pmatrix} \quad (2.4)$$

$$= \mathbf{P}_t' \mathbf{P}_t.$$

The elements  $p_{ij,t}$ ,  $i, j = 1, \dots, N$ ,  $i < j$  are real and can be calculated recursively by

$$p_{ij,t} = \begin{cases} \frac{1}{p_{ii,t}} \left( y_{ij,t} - \sum_{k=1}^{i-1} p_{ki,t} p_{kj,t} \right) & \text{for } i < j \\ \sqrt{y_{jj,t} - \sum_{k=1}^{j-1} p_{kj,t}^2} & \text{for } i = j \\ 0 & \text{for } i > j \end{cases} \quad (2.5)$$

In reverse, the realized covolatilities can be expressed in terms of the Cholesky elements

$$y_{ij,t} = \sum_{\ell=1}^{\min\{i,j\}} p_{\ell i,t} p_{\ell j,t}. \quad (2.6)$$

Since in practice, modeling is carried out on the elements of the CD, one of the problems depicted in equation 2.5 is the influence of the ordering of the variables in the covariance matrix. Consider for example, that we swap the position of the first and second asset in the return vector. As a result, the elements in the first and second row

---

<sup>1</sup>Or positive semi-definite if the condition of strict positivity for the diagonal elements of the triangular matrix is dropped

of the matrix in equation 2.3 will change its positions. Due to the recursive calculation of the elements in  $\mathbf{P}_t$ , the corresponding Cholesky elements in the first and second row of  $\mathbf{P}_t$  will not merely be swapped, but completely change magnitude. Using the CD for a  $N \times N$  portfolio, there are  $N!$  possible permutations of the stocks in the matrix, resulting in different decompositions that are nonlinearly related to each other. Hence, the resulting time-series of Cholesky elements  $p_{ij,t}$  differ between the decompositions. For all model based on the CD, this may lead to varying model choices, parameter estimates and also forecasts.

Another issue arises in obtaining forecasts for  $\widehat{\mathbf{Y}}_{t+1}$ . Being a quadratic transformation of the forecast for  $\widehat{\mathbf{P}}_{t+1}$ , the forecast  $\widehat{\mathbf{Y}}_{t+1}$  may not be unbiased, even if the forecasts for  $\widehat{\mathbf{P}}_{t+1}$  are. This problem is further illustrated in section 2.2.4.

Furthermore, a often desirable feature of covariance forecasting, namely the separation of variances and covariance dynamics can not be achieved by using the CD directly on the covariance matrix. However, it is possible to first apply a DCC decomposition approach and a CD on the correlation matrix thereafter. In general, the nonlinear dependence of the elements in the decomposition can also be an advantage, as the dependency structure between the Cholesky elements can be studied and used for forecasting, see e.g. Brechmann et al. (2015).

## 2.2.2 Matrix exponential transformation

For the covariance matrix, the matrix exponential transformation (MET) was introduced together with the matrix logarithm function by Chiu et al. (1996). In mathematics, both operators are frequently used for solving first-order differential systems, see e.g. Bellman (1997).

For any real, symmetric matrix  $\mathbf{A}_t$ , the matrix exponential transformation performs a power series expansion, resulting in a real, positive (semi-)definite matrix, in our case  $\mathbf{Y}_t$ ,

$$\mathbf{Y}_t = \text{Exp}(\mathbf{A}_t) = \sum_{s=0}^{\infty} \left(\frac{1}{s!}\right) \mathbf{A}_t^s, \quad (2.7)$$

with  $\mathbf{A}_t^0$  being the identity matrix of size  $N \times N$ , and  $\mathbf{A}_t^s$  being the  $s$ -times the standard matrix multiplication of  $\mathbf{A}_t$ .

In reverse, a real, symmetric matrix  $\mathbf{A}_t$  can be obtained from  $\mathbf{Y}_t$  by the inverse of the matrix exponential function, the matrix logarithm function,  $\text{logm}(\cdot)$ ,

$$\mathbf{A}_t = \begin{pmatrix} a_{11,t} & a_{12,t} & \cdots & a_{1N,t} \\ a_{12,t} & a_{22,t} & \cdots & a_{2N,t} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1N,t} & \cdots & \cdots & a_{NN,t} \end{pmatrix} = \text{logm}(\mathbf{Y}_t). \quad (2.8)$$

Again, a reasonable practical approach would be to model and forecast the elements  $a_{ij,t}$ ,  $i, j = 1, \dots, N$  and obtain valid covariance forecasts by equation 2.7. However, due to the power series the expansion, the relationship between  $\mathbf{Y}_t$  and  $\mathbf{A}_t$  is not straightforward to interpret (see e.g. Asai et al. (2006)) and similar to the CD in section 2.2.1, covariances and variances cannot be estimated separately. By applying models to  $\mathbf{A}_t$ , therefore doing the estimation and forecasting in the log-volatility space, the retransformed forecasts for  $\mathbf{Y}_{t+1}$ , will be biased by Jensen's inequality. The problem and possible solutions are illustrated in section 2.2.4.

Nevertheless, the MET has several advantages, especially related to factor models where a certain factor structure is analyzed by principal component methods. It can be shown that under several conditions, as for example in our case symmetry and positive semi-definiteness of  $\mathbf{Y}_t$ , applying the matrix logarithm function yielding  $\mathbf{A}_t$  corresponds to decomposing  $\mathbf{Y}_t$  into its eigenvalues and eigenvectors (see Chiu et al. (1996)). Hence, the  $\mathbf{A}_t^s$  can be obtained more easily via the eigenvectors than by matrix multiplication as in equation 2.7. Further, as principal component analysis of the matrix  $\mathbf{Y}_t$  is also based upon eigenvalue decomposition, restrictions on the structure of the covariance matrix models can be directly implemented while constructing the  $\mathbf{A}_t^s$ , see e.g. Chiu et al. (1996); Bauer and Vorkink (2011).

### 2.2.3 HAR model

One of the most simple and yet successful univariate models for volatility forecasting is the Heterogeneous Autoregressive (HAR) model of Corsi (2009). It is inspired by the

Heterogeneous Market Hypothesis (Müller et al., 1993), which amongst other things assumes that market participants act on different time horizons (dealing frequencies) due to their individual preferences, and therefore create volatility specifically on these horizons. Since in practice, volatility over longer time intervals has stronger influence on those over shorter time intervals than conversely (Corsi, 2009), the HAR models volatility by an additive cascade of components of volatilities in an autoregressive framework..

This leads to the following model for the daily realized volatilities  $x_t$

$$x_t = c + \beta^{(d)}x_{t-1}^{(d)} + \beta^{(w)}x_{t-1}^{(w)} + \beta^{(m)}x_{t-1}^{(m)} + \varepsilon_t, \quad \varepsilon_t \stackrel{iid}{\sim} (0, \sigma^2), \quad (2.9)$$

where  $x_t^{(\cdot)}$  is the realized volatility over the corresponding periods of interest, one day (1d), one week (1w) and one month (1m), which are defined as:  $x_t^{(d)} = x_{t-1}$ ,  $x_t^{(w)} = 5^{-1} \sum_{i=1}^5 x_{t-i+1}$  and  $x_t^{(m)} = 22^{-1} \sum_{i=1}^{22} x_{t-i+1}$ .

The main advantages of the HAR are that it is easy to estimate within an OLS framework, parameters are directly interpretable and it reproduces volatility characteristics such as long-memory without a fractionally integration component. The latter is especially interesting, as the long-memory property could also stem from multifractal scaling<sup>2</sup>, which can be captured by an additive component model as the HAR, whereas fractionally integrated models imply univariate scaling (Andersen and Bollerslev, 1996). Under the HMM hypothesis, multifractal scaling possesses clear economic justification which is directly interpretable in the HAR framework due to the simple parameter structure (Corsi, 2009).

Regarding forecasting, standard methods for a general ARMA framework can be used to produce direct or iterated forecasts of the conditional volatility. In contrast to the above conventional HAR model, which is directly applied on a time-series of realized volatilities, we use the model on the time-series of the elements of the CD or the MET, by replacing the components  $x_t$  and  $x_t^{(\cdot)}$  with the respective  $p_{ij,t}$  or  $a_{ij,t}$  from equations 2.4 and 2.8.

---

<sup>2</sup>The underlying process scales differently for various time horizons.



### 2.2.4 Forecasting and bias correction

To obtain forecasts for the RCOV matrix  $\widehat{\mathbf{Y}}_{t+1}$ , the forecasts  $\widehat{p}_{ij,t}$  or  $\widehat{a}_{ij,t}$  are generated by the HAR model in section 2.2.3 and retransformed by equations 2.4 respectively 2.7.

This last transformation is nonlinear and induces a theoretical bias. For the CD it is derived in Chiriac and Voev (2011) and can be expressed by the covariances of the forecast errors  $u_{\cdot,t+1}$  of the HAR model

$$E[\widehat{y}_{ij,t+1} - y_{ij,t+1}] = \sum_{\ell=1}^{\max\{i,j\}} E[u_{\ell i,t+1} u_{\ell j,t+1}]. \quad (2.10)$$

However, since we estimate the models independently of each other, the expression is not feasible as we cannot consistently estimate the covariance matrix of the forecast errors. A heuristic approach to obtain unbiased predictions is suggested in Chiriac and Voev (2011) and further studied in Halbleib and Voev (2011). In the original approach, due to the larger distortion of volatilities, bias correction is only carried out on the series of realized volatilities  $\widehat{y}_{ii,t}$ ,  $i \in 1, \dots, N$ . However, as implied by equation 2.10, all elements of  $\widehat{\mathbf{Y}}_{t+1}$  will be biased. Hence, an adaption of the approach of Chiriac and Voev (2011), that corrects volatility and covariance forecasts can be obtained by:

$$\widehat{y}_{(corrected),ij,t+1} = \widehat{y}_{ij,t+1} \cdot \text{median} \left\{ \frac{y_{ij,t}}{\widehat{y}_{ij,t}} \right\}_{t=1, \dots, n}.$$

Note, that the window length  $n$  on which the median is estimated, controls for the trade-off between the bias and the precision of the correction. Since we are interested in the general relation between bias correction, we simply estimate the median in the bias correction factor on a window length equal to our estimation window for the HAR model in section 2.3.

In case of the MET, the analytical bias correction is more complicated but can be derived if  $\widehat{\mathbf{A}}_t$  and the estimated residuals  $\widehat{\varepsilon}_t$  are both normally distributed, see Bauer and Vorkink (2011) for a detailed discussion. However, since normality is empirically often not satisfied, Bauer and Vorkink (2011) suggest a similar approach to Chiriac and Voev (2011). Their method decomposes the forecasted matrix of realized covariances  $\widehat{\mathbf{Y}}_{t+1}$  into correlations and volatilities, bias correcting the latter ones only and leaving

the correlations intact. For comparative reasons, we apply our method in equation 2.2.4 which works well in our empirical application for both, CD and MET, see section 2.3. Note that bias correcting not only the volatilities but also the covariances bears the risk of the corrected RCOV matrix forecast no longer being positive semi-definite. However, in our application, this is never the case.

### 2.2.5 Loss functions and the MCS

According to Patton and Sheppard (2009), two issues are of major importance when comparing forecasts of the covariance matrix. First, tests have to be robust to noise in the volatility proxy and second, they should only require minimal assumptions on the distribution of the returns. Therefore, we rely on the method of Hansen, Lunde, and Nason (2011) using a model confidence set (MCS) approach based upon different loss functions to evaluate the multivariate volatility forecasts. This framework fulfills the requirements of Patton and Sheppard (2009) and has the advantage that we can conveniently compare forecasts from many models without using a benchmark. Furthermore, the MCS does not necessarily select a single best model but it allows for the possibility of equality of the models forecasting ability. Hence, a model is only removed from the MCS if it is significantly inferior to other models, making the MCS more robust in comparing volatility forecasts.

For our approach, we choose two loss functions that satisfy the conditions of Hansen and Lunde (2006) for producing a consistent ranking in the multivariate case. Consistency in the context of loss functions means, that the true ranking of the covariance models is preserved, regardless if the true conditional covariance or an unbiased covariance proxy is used (Hansen and Lunde, 2006). For the comparison of forecasts of the whole covariance matrix, Laurent et al. (2013) present two families of loss functions that yield a consistent ordering. The first family, called p-norm loss functions can be written as

$$L\left(\mathbf{Y}_t, \widehat{\mathbf{Y}}_t\right)_p = \left( \sum_{i,j=1}^N |y_{ij,t} - \widehat{y}_{ij,t}|^p \right)^{1/p}, \quad (2.11)$$

where  $\widehat{\mathbf{Y}}_t$  is the forecast from our model for the actual RCOV matrix  $\mathbf{Y}_t$ , which we use as a proxy for the unobservable covariance matrix  $\boldsymbol{\Sigma}_t$ . The respective elements of the

matrices are denoted by  $y_{ij,t}$  and  $\hat{y}_{ij,t}$ . From this class, we consider the commonly used multivariate equivalent of the mean squared error (MSE) loss:  $L\left(\mathbf{Y}_t, \hat{\mathbf{Y}}_t\right)_2^2$ .

The second family, called eigenvalue loss functions is based upon the square root of the largest eigenvalue of the matrix  $(\mathbf{Y}_t - \hat{\mathbf{Y}}_t)^2$ . We will consider a special case of this family, the so called James-Stein loss (James and Stein, 1961), which is usually referred to as the Multivariate Quasi Likelihood (QLIKE) loss function:

$$L\left(\mathbf{Y}_t, \hat{\mathbf{Y}}_t\right) = \text{tr}(\hat{\mathbf{Y}}_t \mathbf{Y}_t) - \ln \left| \hat{\mathbf{Y}}_t \mathbf{Y}_t \right| - N, \quad (2.12)$$

where  $N$  is the number of assets.

While both, the MSE and the QLIKE loss function determine the optimal forecasts based on conditional expectation, Clements et al. (2009) point out, that compared to the MSE, the QLIKE has greater power in distinguishing between volatility forecasts based on the MCS framework. As pointed out in Laurent et al. (2013), the QLIKE penalizes underpredictions more heavily than overpredictions. West et al. (1993) show, that this is also relevant from an investor's point of view, as an underestimation of variances leads to lower expected utility than an equal amount of overestimation. Hence, for a risk averse investor, punishing underpredictions more heavily seems to be rationale when evaluating forecasting accuracy.

For the MCS approach, we start with the full set of candidate models  $\mathcal{M}_0 = \{1, \dots, m_0\}$ . For all models, the loss differential between each model is computed based upon one of our loss functions  $L_k$ ,  $k = 1$  (MSE),  $2$  (QLIKE), so that for model  $i$  and  $j$ ,  $i, j = 1, \dots, m_0$  and every time point  $t = 1, \dots, T$  we get:

$$d_{ij,kt} = L_k\left(\mathbf{Y}_{it}, \hat{\mathbf{Y}}_{it}\right) - L_k\left(\mathbf{Y}_{jt}, \hat{\mathbf{Y}}_{jt}\right). \quad (2.13)$$

At each step of the evaluation, the hypothesis

$$H_0 : \mathbb{E}[d_{ij,kt}] = 0, \quad \forall i > j \in \mathcal{M}, \quad (2.14)$$

is tested for a subset of models  $\mathcal{M} \in \mathcal{M}_0$ , where  $\mathcal{M} = \mathcal{M}_0$  for the initial step. If the  $H_0$  is rejected at a given significance level  $\alpha$ , the worst performing model is removed

from the set. To give an impression on the scale of rejection, for each loss function and model, the respective  $\alpha$  at which the model would be removed from the MCS can be computed.

This process continues until a set of models remains that cannot be rejected. Similar to Hansen, Lunde, and Nason (2011), we use the range statistics to evaluate the  $H_0$ , which can be written as:

$$T_R = \max_{i,j \in \mathcal{M}} |t_{ij,k}| = \max_{i,j \in \mathcal{M}} \frac{|\bar{d}_{ij,k}|}{\sqrt{\widehat{\text{var}}(\bar{d}_{ij,k})}}, \quad (2.15)$$

where  $\bar{d}_{ij,k} = \frac{1}{T} \sum_{t=1}^T d_{ij,k}$  and  $\widehat{\text{var}}(\bar{d}_{ij,k})$  is obtained from a block-bootstrap procedure, see Hansen, Lunde, and Nason (2011), which we implement with 10000 replications and a block length varying from 20 to 50 to check the robustness of the results.

The worst performing model to be removed from the set  $\mathcal{M}$  is selected as model  $i$  with

$$i = \arg \max_{i \in \mathcal{M}} \frac{\bar{d}_{i,k}}{\sqrt{\widehat{\text{var}}(\bar{d}_{i,k})}}, \quad (2.16)$$

where  $\bar{d}_{i,k} = \frac{1}{m-1} \sum_{j \in \mathcal{M}} \bar{d}_{ij,k}$  and  $m$  being the number of models in the actual set  $\mathcal{M}$ .

## 2.3 Empirical study

### 2.3.1 Data and descriptive statistics

The dataset stems from the New York Stock Exchange (NYSE) Trade and Quotations (TAQ) and corresponds to the one used in Chiriac and Voev (2011). It was obtained from the Journal of Applied Econometrics Data Archive. The original data file consists of all tick-by-tick bid and ask quotes on six stocks listed on the NYSE, American Stock Exchange (AMEX) and the National Association of Security Dealers Automated Quotation system (NASDAQ). The sample ranges from 9:30 EST until 16:00 EST over the period January 1, 2000 to July 30, 2008 and consists of (2156 trading days). Included individual stocks are American Express Inc. (**AXP**), Citigroup (**C**), General Electric (**GE**), Home Depot Inc. (**HD**), International Business Machines (**IBM**) and JPMorgan Chase&Co (**JPM**). The original tick-by-tick data has previously been transformed as

follows. To obtain synchronized and regularly spaced observations, the previous-tick interpolation method of Dacorogna (2001) is used.

Then, log-midquotes are constructed from the bid and ask quotes by geometric averaging.  $M = 78$  equally spaced 5-minute return vectors  $r_{i,t}$  are computed from the log-midquotes. Daily open-to-close returns are computed as the difference in the log-midquote at the end and the beginning of each day.

For each daily period  $t = 1, \dots, 2156$ , the series of daily RCOV matrices is constructed as in section 2.2 by summing up the squared 5-minute return vectors:

$$\mathbf{Y}_t = \sum_{i=1}^M r_{i,t} r'_{i,t}. \quad (2.17)$$

This approach is further refined by a subsampling procedure to make the RCOV estimates more robust to microstructure noise and non-synchronicity (see Zhang (2011)). From the original data, 30 regularly  $\Delta$ -spaced subgrids are constructed with  $\Delta = 300$  seconds, starting at seconds 1,11,21, ..., 291. For each subgrid, the log-midquotes are constructed and the RCOV matrix is obtained on each subgrid according to equation 2.17. Then, the RCOV matrices are averaged over the subgrids. To avoid noise by measuring overnight volatilities, all computations are applied to open-to-close data. For the descriptive statistics and estimation purposes, all daily and intradaily returns are scaled by 100, so that the values refer to percentage returns.

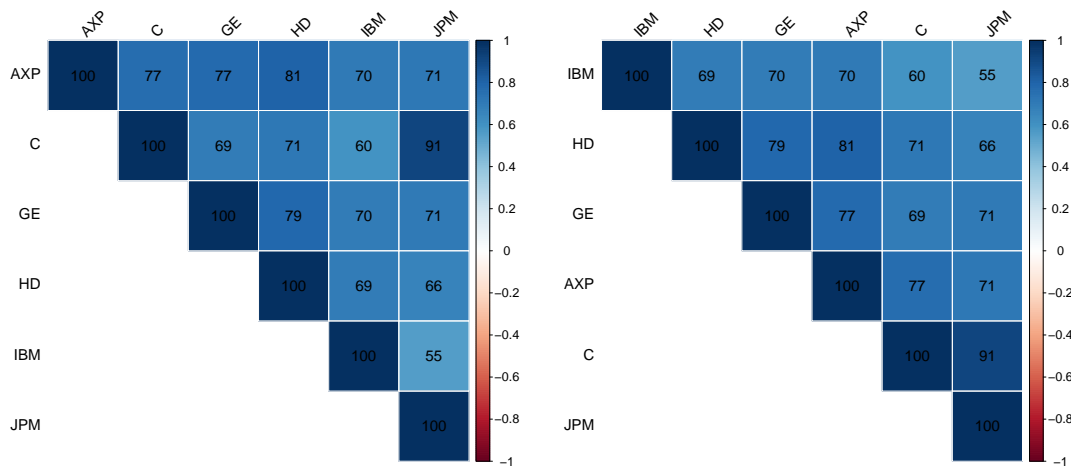
At each point of time  $t$ , we apply either the CD or the MET on the obtained RCOV matrix. Additionally, we take the logarithm of the elements on the diagonal, to ensure positivity of the elements of the decomposition when applying the time-series models. Since the ordering of the assets in the original RCOV matrix is relevant for the CD, we refer to the basic alphabetic ordering of the individual stocks in section 2.3.1 for the initial descriptive analysis of the elements of the CD.

In general, the elements of both decompositions exhibit the same characteristics as the realized covariances, such as volatility clustering, right skewness, excess kurtosis and high levels of autocorrelation, see tables A.1 and A.2. All series seem to be stationary based on the Augmented Dickey-Fuller test.

### 2.3.2 Optimal ordering

If the ordering might indeed be crucial for the forecast performance, the question arises if there is any possibility to determine the optimal position of an asset in the original return vector before evaluating all permutations. According to equation 2.6, the forecasts in column  $j$ ,  $\{\hat{y}_{ij}\}_{i=1,\dots,j}$  only depend on the forecasted entries of the Cholesky matrix  $P$  up to column  $1, \dots, j$ ,  $\{\hat{p}_{li}\}_{i \leq j}$ . Hence, if the asset is for example moved from position  $i = 1$  in the return vector to a position  $i > 1$  the number of forecasted Cholesky elements that enter calculation of the covolatility forecast increases with every increase in the position. Intuitively, assets that are more correlated with each other should be placed after assets that are less correlated so that their dependence is picked up by the Cholesky elements. Similarly, in the estimation of structural VARs, variables are often ordered by their degree of exogeneity from most to least exogenous, see e.g. Bernanke and Blinder (1992); Keating (1996). However, the CD is only useful for identifying the structural relationship under rather restrictive conditions, e.g. in case of VAR modeling if the underlying relationship is recursive. Based on our data set of equity returns we cannot impose a structural relationship by means of economic theory. Nevertheless, we analyze the correlation structure of the realized variances of the six assets to identify possible linkages that might be helpful in ordering the assets. The full-sample correlation matrix of the time-series of realized variances for the natural alphabetic ordering of the assets are given in figure 2.1.

On the left side the ordering of the elements in the return vector matrix is used, while on the right side the correlations are ordered based on the angular positions of the eigenvectors of the correlation matrix. This method is sometimes called “correlation ordering” (Friendly and Kwan, 2003) and places similar variables contiguously. The correlation matrix on the right shows which assets should be grouped together. Note that the correlations are not sorted by size, eg. from highest to lowest average correlation. We now proceed to analyze two questions. First, do different orderings indeed yield significantly different forecasts? Second, does ordering the variables in the returns vector similar to the rule of correlation ordering produce superior forecasts?



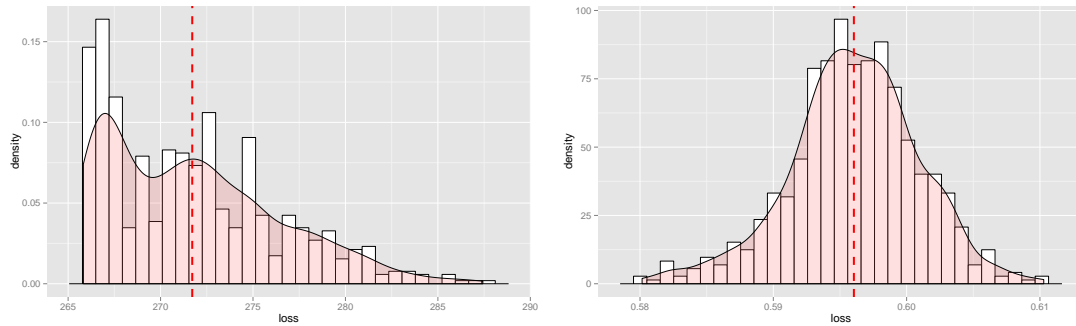
**Figure 2.1:** Correlation matrix of the original time-series of realized variances. On the left, correlations are ordered by the alphabetic order in the return vector. On the right, correlations are ordered based on the angular positions of the eigenvectors of the correlation matrix. The estimate of the corresponding correlation coefficient is given inside the square.

### 2.3.3 Modeling and forecasting procedure

For each decomposition and permutation of the assets, we apply the HAR model from section 2.2.3 on each time-series of CD or MET elements. Since the MET is independent of the chosen permutation, we can use the resulting model as a benchmark model. For the CD, we obtain 21 different models for each one of the 720 permutations. We retain the last 200 observations of the dataset for out-of-sample one-step ahead forecasting and estimate the models based on the a moving window of 1956 observations. Forecasts of the RCOV matrix are then generated according to section 2.2.4.

First, for each permutation we evaluate the forecasts by means of the multivariate loss functions from section 2.2.5. For the CD and both loss functions, we take the average loss over time for each permutation to obtain a distribution of losses, see figure 2.2. The corresponding descriptive statistics are given in the upper half of table A.3.

The loss density of the MSE is multimodal and left skewed with an average loss of 271.71. In comparison, the average loss of the MET is 334.21 which is 16% larger than the maximum MSE loss of the CD. The difference between largest and smallest average loss is 8%. The QLIKE loss density is more symmetric and only slightly right skewed, with an average loss of 0.60, compared to an average loss of 0.71 for the MET. Again, the MET QLIKE loss is 16% larger than the largest QLIKE loss of the CD. The



**Figure 2.2:** Average (over time) MSE (left) and QLIKE (right) density for all permutations. Red line is the mean value.

	worst	best	ex-ante	ex-ante vs best
MSE	“3 2 1 6 4 5”	“5 4 3 1 6 2”	“5 4 3 1 2 6”	1.002
QLIKE	“3 1 5 4 2 6”	“6 2 3 4 5 1”	“5 4 3 1 2 6”	1.506

**Table 2.1:** Orderings with the highest and lowest average losses (without bias correction) based on the respective loss function. Ex-ante gives the order proposed by the method of correlation ordering. Additionally the average loss of the ex-ante model relative to the best model is listed.

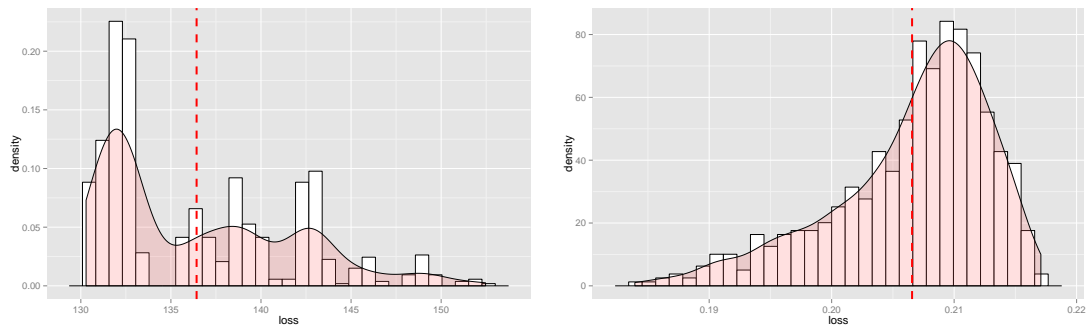
standard deviation of the QLIKE losses is significantly smaller ( $p < 0.01$ ) than for the MSE loss, based on the Brown-Forsythe test. Still, the difference between largest and smallest average loss is roughly 5%. Ranking the models from best to worst (smallest to largest average losses over time), we find that the ordering is not consistent across the loss functions. Evaluating the model performance over time instead of taking averages, the most frequent best model is identical in 4 of the 200 out-of-sample forecasts for both loss functions. The most frequent worst model on the other hand differs between both loss function. For the MSE, one certain ordering is the worst model in 12 out of the 200 forecasts. In case of the QLIKE, the most frequent worst ordering has the highest loss in 3 out of 200 times.

To come back to the question if the method of correlation ordering in section 2.3.2 is helpful in determining the best model ex-ante, we list the worst and best orderings based on the average loss for both loss function in table 2.1. To simplify the notation, we rename the assets by their position in the alphabetic return vector, namely **AXP**= 1, **C**= 2, **GE**= 3, **HD**= 4, **IBM**= 5 and **JPM**= 6.

Surprisingly, the best model under the MSE loss function nearly coincides with the model suggested by the method of correlation ordering, with only asset 2 and 6 switching



positions. For the QLIKE loss, only asset 3 is on the same position in the best model compared to the ex-ante ordering. Regarding the average loss size, the ex-ante models losses are only 0.2% larger than the best model based on the MSE loss, whereas for the QLIKE loss function the ex-ante losses are 50% larger. We statistically evaluate these differences in section 2.3.4. However, based on the mixed results from both loss functions we cannot unambiguously establish a link between correlation ordering and forecasting results. Additionally, as pointed out before, the model ranking is highly time-varying. Evaluating the model at every point of time reveals that the ex-ante model has the lowest loss at exactly one point of time for both loss functions. Again, it seems that neither the ex-ante nor any other ordering is consistently delivering the best forecasts.



**Figure 2.3:** Average (over time) MSE (left) and QLIKE (right) density for all permutations with bias correction. Red line is the mean value.

In case of the bias correction, the average loss densities for all permutations are significantly different ( $p < 0.01$ ) from the ones without bias correction based on the Kolmogorov-Smirnov (KS) test. In general, the bias correction does decrease the average loss, see figure 2.3. Descriptive statistics are given in the lower half of table A.3. Most notable, the standard deviation does increase for the MSE, while in case of the QLIKE the distribution becomes more right skewed. As a result, the difference between largest and smallest average loss increases for both loss functions to 17% (MSE), respectively 18% (QLIKE) percent. The bias corrected average MET loss is 112.64 for the MSE and 0.18 for the QLIKE. Hence, the MET heavily benefits from the bias correction, making it a possible alternative to the CD to circumvent the ordering problem.

For each permutation, we test the distribution of losses over time of the bias corrected vs the non-bias corrected forecasts using the KS test. In all cases, the loss distributions are significantly different at a level  $p < 0.01$  and the mean loss (over time) of the bias

corrected distribution is smaller than the one of the non-bias corrected. For the MSE, the worst model with bias correction is also the same as the worst model without bias correction. Otherwise, we find that the best and worst model are not the same as in the case of no bias correction. As before, the ranking of the average losses from best to worst is not consistent across the loss functions. Comparing the losses over time reveals a similar behavior as before, where the most frequent best and worst model varies across time.

### 2.3.4 Statistically testing forecast performance

To evaluate the significance of the loss differences across time, we test the losses of the permutations using the MCS procedure introduced in section 2.2.5. We are interested in several questions. First of all, are the forecasts from the models which are best and worst based upon the average loss significantly different from each other? Second, how well does the bias adjusted MET model perform compared to the best ordering and third, is the ex-ante ordering significantly worse than the best model?

Starting with the first questions, we find that for both loss functions the worst model can be rejected from the MCS at a  $\alpha = 1\%$  level of significance. In case of bias correction,  $\alpha$  further decreases. As mentioned in the literature the QLIKE is also more discerning, leading to slightly lower levels of significance in both cases if compared to the MSE. Comparing the non-corrected vs the bias corrected forecasts, we find that the bias correction leads to significantly better forecasts for both loss functions ( $\alpha = 1\%$ ). Overall, since the differences between the forecasts are indeed statistically significant, choosing the “wrong” ordering may lead to poor forecast performance, no matter which loss function is chosen.

Next, we only consider the case of bias correction. As we have seen, the MET average losses were well within the range of the average CD losses. If the MET losses are not significantly different from the best CD model, the MET with bias correction could be a valid alternative to avoid the ordering problem of the CD. The MET forecasts can only be rejected from the MCS at a  $\alpha = 50\%$  significance level for the MSE and a  $\alpha = 69\%$  significance level for the QLIKE. Hence, the forecasts from the best CD model and the MET are not significantly different from each other at a reasonable level of confidence.

Comparing the losses of the ex-ante ordering with the best model under the respective loss function, we find that for the QLIKE the losses are significantly different ( $\alpha < 1\%$ ), while for the MSE the ex-ante model can not be rejected from the MCS ( $\alpha = 9\%$ ). Hence, initially deciding upon the ordering does not yield a clear recommendation. The danger of arbitrarily choosing an ordering that might lead to poor forecasts and hence model choices cannot be assessed ex-ante based on the methodology of correlation ordering.

## 2.4 Conclusion

In this paper, we empirically analyzed several issues arising from using the Cholesky decomposition (CD) for forecasting the realized covariance (RCOV) matrix. We studied the impact of the order of the variables in the covariance matrix on volatility forecasting, finding that different orderings do indeed lead to significantly different forecasts based on a MCS approach. Initially deciding upon the ordering based on the angular positions of the eigenvectors of the correlation matrix does not lead to unambiguously better results in forecasting. Further, we find that the best and worst models are not consistent over time so that a clear recommendation to which order to use is not at hand, even if forecasts are performed stepwise. A frequently used method of bias correction improves forecasting accuracy, but on the other hand widens the difference between best and worst model so that the ordering problem worsens. On the other hand, bias corrected forecasts from another decomposition, the matrix exponential transformation (MET) show equal predictive ability and do not suffer from the ordering problem. Thus, for empirical application two conclusions can be drawn. If a reasonable order can be imposed on the elements of the covariance matrix or if the connection between the elements of the decomposed covariance matrix are of interest the CD is a rational choice. Otherwise, the application of the MET together with a bias correction is advised, be it for comparative reasons or simply to avoid the time consuming process of estimating all possible permutations of the CD.

# Appendix

	min	max	mean	sd	skew	kurt	pval ADF	acf l=1	acf l=2
<b><math>p_{11,t}</math></b>	-1.31	2.03	0.30	0.58	0.10	2.23	0.01	0.88	0.86
$p_{12,t}$	-0.21	7.63	0.71	0.66	2.83	16.71	0.01	0.76	0.70
<b><math>p_{22,t}</math></b>	-1.13	2.16	0.20	0.53	0.29	2.33	0.01	0.89	0.87
$p_{13,t}$	-0.50	3.99	0.52	0.43	2.17	11.41	0.01	0.65	0.59
$p_{23,t}$	-0.28	2.68	0.37	0.30	1.99	9.69	0.01	0.61	0.57
<b><math>p_{33,t}</math></b>	-1.14	1.73	0.04	0.48	0.31	2.40	0.01	0.86	0.82
$p_{14,t}$	-0.7	3.71	0.55	0.46	2.01	9.44	0.01	0.64	0.59
$p_{24,t}$	-0.38	2.40	0.36	0.30	1.77	8.79	0.01	0.46	0.46
$p_{34,t}$	-0.53	2.76	0.28	0.26	1.69	10.30	0.01	0.50	0.45
<b><math>p_{44,t}</math></b>	-0.97	1.75	0.29	0.42	0.36	2.74	0.01	0.81	0.77
$p_{15,t}$	-0.43	3.09	0.45	0.35	1.96	9.95	0.01	0.53	0.49
$p_{25,t}$	-1.22	4.68	0.31	0.27	3.42	39.36	0.01	0.47	0.39
$p_{35,t}$	-0.43	2.13	0.26	0.22	1.82	10.20	0.01	0.44	0.42
$p_{45,t}$	-0.48	1.71	0.15	0.17	1.36	10.53	0.01	0.18	0.14
<b><math>p_{55,t}</math></b>	-1.11	1.56	0.00	0.46	0.58	2.88	0.01	0.86	0.82
$p_{16,t}$	-0.29	8.16	0.72	0.66	2.88	18.55	0.01	0.73	0.65
$p_{26,t}$	-0.19	5.77	0.58	0.43	2.26	15.50	0.01	0.64	0.60
$p_{36,t}$	-0.36	2.22	0.22	0.22	1.86	10.20	0.01	0.32	0.30
$p_{46,t}$	-0.91	1.19	0.14	0.18	0.63	6.63	0.01	0.17	0.11
$p_{56,t}$	-0.87	1.37	0.14	0.19	1.25	8.48	0.01	0.18	0.15
<b><math>p_{66,t}</math></b>	-1.21	2.20	0.14	0.52	0.30	2.39	0.01	0.89	0.86

**Table A.1:** Descriptive statistics for the time-series of the elements of the (alphabetic) Cholesky decomposition. Diagonal (log) time-series are written in **bold**. Additionally, p-value of the ADF test and magnitude of the first and second autocorrelation coefficient.

	min	max	mean	sd	skew	kurt	pval ADF	acf l=1	acf l=2
<b><math>a_{11,t}</math></b>	-2.72	3.63	0.32	1.13	0.13	2.19	0.01	0.88	0.86
$a_{12,t}$	-0.35	0.94	0.30	0.16	0.26	3.32	0.01	0.44	0.43
<b><math>a_{22,t}</math></b>	-2.28	4.35	0.30	1.07	0.33	2.36	0.01	0.90	0.87
$a_{13,t}$	-0.24	0.67	0.24	0.14	-0.08	3.06	0.01	0.28	0.26
$a_{23,t}$	-0.24	0.71	0.27	0.14	-0.02	2.90	0.01	0.38	0.29
<b><math>a_{33,t}</math></b>	-2.36	3.52	0.09	0.96	0.30	2.43	0.01	0.85	0.82
$a_{14,t}$	-0.34	0.73	0.20	0.13	0.04	3.08	0.01	0.24	0.20
$a_{24,t}$	-0.25	0.66	0.22	0.13	0.07	3.00	0.01	0.25	0.27
$a_{34,t}$	-0.27	0.66	0.22	0.13	-0.09	3.13	0.01	0.30	0.28
<b><math>a_{44,t}</math></b>	-2.01	3.63	0.65	0.84	0.35	2.77	0.01	0.81	0.76
$a_{15,t}$	-0.34	0.62	0.21	0.13	-0.18	3.27	0.01	0.22	0.18
$a_{25,t}$	-0.28	0.80	0.23	0.13	-0.04	3.17	0.01	0.29	0.23
$a_{35,t}$	-0.19	0.67	0.26	0.14	-0.06	2.79	0.01	0.31	0.29
$a_{45,t}$	-0.31	0.65	0.21	0.13	-0.08	3.18	0.01	0.21	0.17
<b><math>a_{55,t}</math></b>	-2.21	3.51	0.10	0.91	0.57	2.93	0.01	0.85	0.81
$a_{16,t}$	-0.16	0.99	0.29	0.16	0.55	3.70	0.01	0.45	0.39
$a_{26,t}$	-0.11	1.13	0.42	0.18	0.50	3.50	0.01	0.53	0.51
$a_{36,t}$	-0.32	0.63	0.23	0.13	0.02	2.97	0.01	0.21	0.20
$a_{46,t}$	-0.32	0.75	0.20	0.13	0.05	3.25	0.01	0.24	0.20
$a_{56,t}$	-0.33	0.62	0.21	0.13	0.01	3.07	0.01	0.20	0.17
<b><math>a_{66,t}</math></b>	-2.35	4.85	0.47	1.07	0.25	2.42	0.01	0.89	0.85

**Table A.2:** Descriptive statistics for time-series of the elements of the matrix exponential transformation. Diagonal (log) elements are written in **bold**. Additionally, p-value of the ADF test and magnitude of the first and second autocorrelation coefficient.

	min	max	mean	sd	skew	kurt	median	max/min	MET	alphabetic	ex-ante	
	without bias correction											
MSE	265.82	287.43	271.71	4.65	0.75	2.92	271.08	1.08	334.21	269.27	282.71	
QLIKE	0.58	0.61	0.60	0.01	-0.28	3.47	0.60	1.05	0.71	0.59	0.60	
	with bias correction											
MSE	130.29	152.47	136.43	5.30	0.79	2.62	134.48	1.17	112.64	130.52	149.19	
QLIKE	0.18	0.22	0.21	0.01	-0.97	3.54	0.21	1.18	0.18	0.21	0.19	

**Table A.3:** Descriptive statistics for the CD losses over all permutations. Max/min is the ratio of the average loss of the best model vs the average loss of the worst model. As a comparison, the average losses of the (ordering invariant) MET and the average losses of the alphabetic and ex-ante correlation ordering are given.

# Bibliography

- ANDERSEN, T. G. and T. BOLLERSLEV (1996). Heterogeneous information arrivals and return volatility dynamics: Uncovering the long-run in high frequency returns. Working Paper.
- ANDERSEN, T. G., T. BOLLERSLEV, P. F. CHRISTOFFERSEN, et al. (2006). Volatility and correlation forecasting. *Handbook of Economic Forecasting*: 1(05), pp. 777–878.
- ANDERSEN, T. G., T. BOLLERSLEV, F. X. DIEBOLD, et al. (2003). Modeling and forecasting realized volatility. *Econometrica*: 71(2), pp. 579–625.
- ASAI, M., M. MCALEER, and J. YU (2006). Multivariate stochastic volatility: A review. *Econometric Reviews*: 25(2-3), pp. 145–175.
- BARNDORFF-NIELSEN, O. E., P. R. HANSEN, et al. (2011). Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics*: 162(2), pp. 149–169.
- BARNDORFF-NIELSEN, O. E. and N. SHEPHARD (2002). Estimating quadratic variation using realized variance. *Journal of Applied Econometrics*: 17(5), pp. 457–477.
- BAUER, G. H. and K. VORKINK (2011). Forecasting multivariate realized stock market volatility. *Journal of Econometrics*: 160(1), pp. 93–101.
- BELLMAN, R. (1997). *Introduction to matrix analysis*. Vol. 19. Society for Industrial Mathematics.
- BERNANKE, B. S. and A. S. BLINDER (1992). The Federal Funds Rate and the Channels of Monetary Transmission. *American Economic Review*: 82(4), pp. 901–21.
- BRECHMANN, E. C., M. HEIDEN, and Y. OKHRIN (2015). A multivariate volatility vine copula model. *Econometric Reviews*. To appear. DOI: 10.1080/07474938.2015.1096695.

- BREZINSKI, C. (2006). The life and work of André Cholesky. *Numerical Algorithms*: 43(3), pp. 279–288.
- CHIRIAC, R. (2010). A note on estimating Wishart autoregressive model. Working Paper.
- CHIRIAC, R. and V. VOEV (2011). Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics*: 26(6), pp. 922–947.
- CHIU, T. Y. M., T. LEONARD, and K. W. TSUI (1996). The matrix-logarithmic covariance model. *Journal of the American Statistical Association*: 91(433), pp. 198–210.
- CHRISTENSEN, K. and S. KINNEBROCK (2010). Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. *Journal of Econometrics*: 159(1), pp. 116–133.
- CLEMENTS, A. et al. (2009). Evaluating multivariate volatility forecasts. Working Paper.
- COLACITO, R., R. F. ENGLE, and E. GHYSELS (2011). A component model for dynamic correlations. *Journal of Econometrics*: 164(1), pp. 45–59.
- CORSI, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*: 7(2), pp. 174–196.
- DACOROGNA, M. M. (2001). *An introduction to high-frequency finance*. Academic Press.
- ENGLE, R. F. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*: 20(3), pp. 339–350.
- FRIENDLY, M. and E. KWAN (2003). Effect ordering for data displays. *Computational Statistics & Data Analysis*: 43(4), pp. 509–539.
- GOLOSNOY, V., B. GRIBISCH, and R. LIESENFELD (2012). The conditional autoregressive Wishart model for multivariate stock market volatility. *Journal of Econometrics*: 167(1), pp. 211–223.
- GOURIÉROUX, C., J. JASIAK, and R. SUFANA (2009). The Wishart autoregressive process of multivariate stochastic volatility. *Journal of Econometrics*: 150(2), pp. 167–181.
- HALBLEIB, R. and V. VOEV (2014). Forecasting Covariance Matrices: A Mixed Approach. *Journal of Financial Econometrics*. DOI: 10.1093/jjfinec/nbu031.
- (2011). Forecasting multivariate volatility using the VARFIMA model on realized covariance Cholesky Factors. *Journal of Economics and Statistics (Jahrbuecher fuer Nationaloekonomie und Statistik)*: 231(1), pp. 134–152.



- HANSEN, P. R. and A. LUNDE (2006). Consistent ranking of volatility models. *Journal of Econometrics*: 131(1-2), pp. 97–121.
- HANSEN, P. R., A. LUNDE, and J. M. NASON (2011). The Model Confidence Set. *Econometrica*: 79(2), pp. 453–497.
- HAYASHI, T. and N. YOSHIDA (2005). On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli*: 11(2), pp. 359–379.
- JAMES, W. and C. STEIN (1961). Estimation with quadratic loss. *Proc. Fourth Berkley Symp. on Math. Statist. and Prob.*( 1), pp. 361–379.
- KEATING, J. W. (1996). Structural information in recursive VAR orderings. *Journal of Economic Dynamics and Control*: 20(9-10), pp. 1557–1580.
- KLÖSSNER, S. and S. WAGNER (2014). Exploring all VAR orderings for calculating spillovers? Yes, we can! - A note on Diebold and Yilmaz (2009). *Journal of Applied Econometrics*: 29(1), pp. 172–179.
- LAURENT, S., J. V. K. ROMBOUTS, and F. VIOLANTE (2013). On loss functions and ranking forecasting performances of multivariate volatility models. *Journal of Econometrics*: 173(1), pp. 1–10.
- MÜLLER, U. A. et al. (1993). Fractals and intrinsic time - a challenge to econometricians. Working Paper.
- PATTON, A. J. and K. SHEPPARD (2009). “Evaluating Volatility and Correlation Forecasts.” In: *Handbook of Financial Time Series*. Ed. by T. MIKOSCH et al. Springer Berlin Heidelberg, pp. 801–838.
- SIMS, C. A. (1980). Macroeconomics and reality. *Econometrica*: 48(1), pp. 1–48.
- WEST, K. D., H. J. EDISON, and D. CHO (1993). A utility-based comparison of some models of exchange rate volatility. *Journal of International Economics*: 35(1-2), pp. 23–45.
- ZHANG, L. (2011). Estimating covariation: Epps effect, microstructure noise. *Journal of Econometrics*: 160(1), pp. 33–47.
- ZHANG, L., Y. AIT-SAHALIA, and P. A. MYKLAND (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*: 100, pp. 1394–1411.

## Chapter 3

# Article 2: A multivariate volatility vine copula model

BRECHMANN, E. C., M. HEIDEN, and Y. OKHRIN (2015). A multivariate volatility vine copula model. *Econometric Reviews*. To appear. DOI: 10.1080/07474938.2015.1096695.

## Chapter 4

# Article 3: Forecasting volatility with empirical similarity and Google Trends

HAMID, A. and M. HEIDEN (2015). Forecasting volatility with empirical similarity and Google Trends. *Journal of Economic Behavior & Organization*: 117, pp. 62–81. DOI: 10.1016/j.jebo.2015.06.005.

## Chapter 5

# Summary and outlook

The previous chapters of this dissertation outlined several topics in predicting uni- and multivariate stock market volatility. Based on the improved availability of high-frequency data and the econometric methods to estimate the volatility process, new possibilities to study predictability have gained attention in the recent years. Especially in the multivariate context, where forecasts of the covariance matrix are used in portfolio and risk management, the requirements of the forecasted covariance matrix to be symmetric and positive semi-definite pose a problem. A prominent method of guaranteeing both properties, the Cholesky decomposition (CD), was studied in chapter 2. Beside analyzing the main pitfall, namely the influence of the ordering of the variables on the forecasts, a possible solution for empirical application was suggested. In combination with a bias correction, the matrix exponential transformation (MET) provided a suitable alternative to the CD. Being one of the first studies of its kind, several extensions for further research are still to be assessed. First, we only studied two out of several possible decompositions. For example, it is possible to model realized covariances (RCOV) in the popular dynamic conditional correlation (DCC) framework of Engle (2002). Similar to the MET, a DCC type decomposition is invariant to the order of the variables. Additionally, it does allow to separate the dynamics of variances and correlations, allowing for more flexible model specifications, see Halbleib and Voev (2014). Due to computational efforts to evaluate the forecasts of all possible permutations of the assets, we only studied the heterogeneous autoregressive (HAR) model of Corsi (2009) in this setting. While the ordering problem as well as the bias are independent of the model, some approaches are explicitly linked to a certain decomposition, e.g. the vector autoregressive fractionally integrated moving average (VARFIMA) of Chiriac and Voev (2011) is usually combined with the CD. In this case simply changing the decomposition to the MET may not be applicable, as one of the main advantages of using the CD is the traceability of impulse responses. Further research is necessary to solve the ordering problem in cases where a change of the decomposition is either not possible or rational. Additionally, we have

only analyzed one very simple data driven case of bias correction, which worked well for our data set and model choice. However, a similar method applied in Halbleib and Voev (2011) did not lead to improved volatility forecasts for their VARFIMA model. Hence, a more detailed analysis or sophisticated methods, e.g. simulation based bias corrections as proposed in Weigand (2014), might be a valuable extension of our study.

A new approach to modeling and forecasting the RCOV matrix based on the CD was introduced in chapter 3. We proposed a model that makes use of vine copulas to account for often neglected nonlinearities and asymmetries in conventional models for multivariate volatility. We showed that using a vine copula structure can lead to significant improvements in forecasting, regarding statistical loss, mean-variance efficient portfolios and Value-at-Risk (VaR). Still, several areas of improvement of the general framework seem fruitful for further research. First, our initial model requires the estimation of a large number of parameters. Therefore, we applied a successful strategy to decrease the number of parameters by restricting the choice of copulas without severely impairing the models performance. For larger dimensions, the model could be extended by truncation approaches as discussed in Brechmann et al. (2012) to further simplify the estimation procedure. Second, while the CD is useful for our model as it explicitly links the time series of realized volatilities and covariances, the ordering problem remains an issue. While in sample our analysis revealed that differences based in the ordering are relatively small, this might not be the case for the out-of-sample forecasts, as pointed out in chapter 2. An alternative approach of applying vine copulas for multivariate volatility modeling could be the use of partial correlations. For Gaussian vines, there exists a one to one relationship between the partial correlation, which is identified by the vine, and the correlation parameters of the joint distribution, see Baba et al. (2004); Kurowicka and Cooke (2006). By modeling realized volatilities and the partial correlation elements of a Gaussian regular vine, a joint model for the RCOV matrix can be derived, which guarantees positive semi-definiteness and symmetry without the drawbacks and restrictions of an initial decomposition. To avoid parameter restrictions when modeling the partial correlations, a Fisher z-transformation (Fisher, 1915) can be applied on the time series of partial correlations. The resulting values on the real line can again be modeled by conventional time series models, similar to our previous approach. A third extension for the model could be to allow for time-varying effects and jumps. Several studies (amongst others Patton (2004); Cappiello et al. (2006)) argue, that the dependence structure between assets is rather time-varying than constant, e.g. to reflect turmoil periods. Based on a dynamic framework with locally constant parameters, a more flexible modeling procedure is possible, see e.g. Okhrin et al. (2013). The partial correlation approach on the other hand could be used to disentangle jump and continuous variation similar to Audrino and Hu (2011), as it can directly be applied on both

components. For the Cholesky based model, this is not possible due to the fact that the time series of Cholesky elements is modeled instead of the original realized volatilities.

Chapter 4 approached univariate volatility from a different angle by using Google search volume as a measure of investor attention to the stock market in the framework of empirical similarity (ES) by Gilboa et al. (2006). We augmented the model of Lieberman (2012) based on the main assumption that increased search volume leads to higher participation in the stock market and subsequently rising volatility. While our model showed superior predictive ability compared to several other volatility models, we also highlighted that using Google search volume as an additional linear regressor in standard time series models does not accurately reflect the dynamics of the data. Based on our approach and results, a number of extensions for further research come to mind. First of all, the model itself can be extended. Up to now, the level of investor attention is compared to the contemporaneous level of volatility to perform forecasts for the subsequent period. However, investors might cause volatility not exclusively on this frequency, but also on longer horizons. This intuition is consistent with the fractal market hypothesis of Müller et al. (1997) and one of the basic assumptions of the HAR model of Corsi (2009). Hence, extending the model by components regarding different horizons similar to Golosnoy et al. (2014) might give insights in the low frequent trading patterns of investors and further improve predictability. Second, for simplicity we have only analyzed the impact of Google search volume in the US on the volatility of the Dow Jones. Obviously, since investors are not limited to the domestic market and investing abroad is widely suggested by the theory of diversification, cross country effects of investor attention on volatility seem plausible. Similar to the work of Dimpfl and Jung (2012), a structural multivariate model based on the concept of ES and Google search volume could be estimated equation wise and compared to the conventionally used vector autoregressive (VAR) framework. Last but not least, the model can be applied to a variety of other data sets. Similar to Google search volume, other measure for investor attention, such as household survey data (D. Li and G. Li, 2014), internet message postings (S.-H. Kim and D. Kim, 2014) or Facebook users sentiment data (Siganos et al., 2014) could be analyzed.

As a conclusion, volatility seems likely to remain a worthwhile and promising field of research. Including the behavior of investors directly into the modeling process creates challenges and new opportunities at a time, where the importance of alternative data sources is slowly recognized. For modeling itself, non normality of the data and deviations from standard approaches can not be disregarded. Meanwhile, basic questions such as the ordering problem in multivariate volatility models still deserve attention.

# Bibliography

- AUDRINO, F. and Y. HU (2011). Volatility forecasting: Downside risk, jumps and leverage effect. Working Paper.
- BABA, K., R. SHIBATA, and M. SIBUYA (2004). Partial correlation and conditional correlation as measures of conditional independence. *Australian & New Zealand Journal of Statistics*: 46(4), pp. 657–664.
- BRECHMANN, E. C., C. CZADO, and K. AAS (2012). Truncated regular vines in high dimensions with applications to financial data. *Canadian Journal of Statistics*: 40(1), pp. 68–85.
- CAPPIELLO, L., R. F. ENGLE, and K. SHEPPARD (2006). Asymmetric dynamics in the correlations of global equity and bond returns. *Journal of Financial Econometrics*: 4(4), pp. 537–572.
- CHIRIAC, R. and V. VOEV (2011). Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics*: 26(6), pp. 922–947.
- CORSI, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*: 7(2), pp. 174–196.
- DIMPFL, T. and R. C. JUNG (2012). Financial market spillovers around the globe. *Applied Financial Economics*: 22(1), pp. 45–57.
- ENGLE, R. F. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*: 20(3), pp. 339–350.
- FISHER, R. (1915). Frequency distribution of the values of the correlation coefficient in samples of an indefinitely large population. *Biometrika*: 10(4), pp. 507–521.
- GILBOA, I., O. LIEBERMANN, and D. SCHMEIDLER (2006). Empirical similarity. *Review of Economics and Statistics*: 88(3), pp. 433–444.
- GOLOSNOY, V., A. HAMID, and Y. OKHRIN (2014). The empirical similarity approach for volatility prediction. *Journal of Banking & Finance*: 40, pp. 321–329.
- HALBLEIB, R. and V. VOEV (2014). Forecasting Covariance Matrices: A Mixed Approach. *Journal of Financial Econometrics*. DOI: 10.1093/jjfinec/nbu031.

- HALBLEIB, R. and V. VOEV (2011). Forecasting multivariate volatility using the VARFIMA model on realized covariance Cholesky Factors. *Journal of Economics and Statistics (Jahrbuecher fuer Nationaloekonomie und Statistik)*: 231(1), pp. 134–152.
- KIM, S.-H. and D. KIM (2014). Investor sentiment from internet message postings and the predictability of stock returns. *Journal of Economic Behavior & Organization*: 107(PB), pp. 708–729.
- KUROWICKA, D. and R. M. COOKE (2006). *Uncertainty analysis with high dimensional dependence modelling*. John Wiley & Sons Ltd.
- LI, D. and G. LI (2014). Are Household Investors Noise Traders: Evidence from Belief Dispersion and Stock Trading Volume. *Finance and Economics Discussion Series 2014-35, Board of Governors of the Federal Reserve System (U.S.)*
- LIEBERMAN, O. (2012). A similarity-based approach to time-varying coefficient non-stationary autoregression. *Journal of Time Series Analysis*: 33(3), pp. 484–502.
- MÜLLER, U. A. et al. (1997). Volatilities of different time resolutions - analysing the dynamics of market components. *Journal of Empirical Finance*: 4(2-3), pp. 213–239.
- OKHRIN, O., Y. OKHRIN, and W. SCHMID (2013). Determining the structure and estimation of hierarchical Archimedean copulas. *Journal of Econometrics*: 173, pp. 189–204.
- PATTON, A. J. (2004). On the out-of-sample importance of skewness and asymmetric dependence for asset allocation. *Journal of Financial Econometrics*: 2(1), pp. 130–168.
- SIGANOS, A., E. VAGENAS-NANOS, and P. VERWIJMEREN (2014). Facebook’s daily sentiment and international stock markets. *Journal of Economic Behavior & Organization*: 107(PB), pp. 730–743.
- WEIGAND, R. (2014). Matrix Box-Cox models for multivariate realized volatility. Working Paper.