Modeling realized measures of volatility

Kumulative Dissertation

Der Wirtschaftswissenschaftlichen Fakultät

der Universität Augsburg

zur Erlangung des Grades eines

Doktors der Wirtschaftswissenschaften

(Dr. rer. pol.)

Vorgelegt von

Eugen Heine
(M.Sc. Mathematical Finance and Actuarial Science)

Erstgutachter: Prof. Dr. Yarema Okhrin

Zweitgutachter: Prof. Dr. Marco Wilkens

Drittgutachter: Prof. Dr. Michael Krapp

Vorsitzender der mündlichen Prüfung: Prof. Dr. Jens Brunner

Tag der mündlicher Prüfung: 01.07.2020

Index of research papers

This dissertation is comprised of three research papers. Thereby, one has already been published and two are under review.

Paper 1 (p.12):

Claudia Czado, Eugen Heine and Yarema Okhrin (2019). Modeling temporal dependence of realized variances with vines. *Econometrics and Statistics*: 12, pp. 198-216, DOI: `https://doi.org/10.1016/j.ecosta.2019.03.003`.

Paper 2 (pp.13-44):

Eugen Heine (2019). Modeling temporal dependence of realized measures with multivariate B-splines. Submitted to *Econometrics and Statistics* (2020).

Paper 3 (pp.45-83):

Eugen Heine and Yarema Okhrin (2019). Matrix variate factor model with application to realized covariance matrices. Submitted to *Journal of Econometrics* (2020).

Acknowledgments

First of all I would like to express sincere gratitude to my supervisor Prof. Dr. Yarema Okhrin for providing me with the inspiring opportunity to pursue a doctoral degree. He granted me complete research freedom, whereas masterfully coordinating my scientific endeavors and discussing possible topics enthusiastically. He was literally always available for different problems which I faced during writing this doctoral thesis. His effort and expertise significantly improved the quality of this work.

Furthermore, I would like to thank all my colleagues at the chair of statistics for creating friendly and enjoyable work atmosphere. I also acknowledge the financial support provided by the German Research Foundation.

I warmly thank Prof. Dr. Marco Wilkens, Prof. Dr. Michael Krapp and Prof. Dr. Jens Brunner for serving as referees of this thesis and chair of examination committee, respectively.

I most sincerely thank my parents for their support and trust during all these years. I owe a special gratitude to my best friend, for her admirable determination and effort during this remarkable time.

Contents

1 Introduction **1**

1.1 Vine copula approach for modeling realized measures 3

1.2 Non-parametric approach for modeling realized measures 5

1.3 Matrix variate factor model 7

2 Article 1: Modeling temporal dependence of realized variances with vines **12**

3 Article 2: Modeling temporal dependence of realized measures with multivariate B-splines **13**

3.1 Introduction 14

3.2 Modeling realized measures 16

3.3 Non-parametric B-spline regression 18

3.4 B-splines 20

3.5 Estimation and forecasting 30

3.6 Empirical study 34

3.7 Summary 41

4 Article 3: Matrix variate factor model with application to realized covariance matrices **45**

4.1 Introduction 46

4.2 Matrix variate factor model 48

4.3 Estimation 50

4.4 Simulation study 55

4.5 Empirical study 56

4.5.1 Full sample analysis 59

4.5.2 Forecasting study 66
4.6 Summary 67
4.7 Appendix 72
4.7.1 Necessary theoretical results 72
4.7.2 Proofs 73

5 Summary and outlook **84**

Chapter 1

Introduction

Technological advancements of the last twenty years were eagerly adopted by the stock exchanges resulting in quote processing times on the order of microseconds and timestamps with nanosecond precision. The decreasing latency of exchanges made high-frequency trading (HFT) strategies possible, whereas they fueled further investments into speeding up the infrastructure. HFT became ubiquitous bringing the average trading frequency up to before unseen levels. Generally, exchanges are obliged to store and report every quote, whereas eventually serving as data providers became a substantial part of their revenues. Due to the improved availability of such datasets, the academic interest in analyzing this massive amount of data from financial perspective grew exponentially and constituted itself as a separate field, called high-frequency financial econometrics (see Aït-Sahalia and Jacod [2014]).

The focus of the econometrics community was drawn to estimating and modeling volatility. While it is essential for many applications in finance, for example asset pricing, portfolio and risk management, volatility is generally an unobservable process with no universal definition. Traditionally, volatility was estimated based on some parametric model or volatilities implied by options. These approaches suffer from several drawbacks, which are due to the specific parametric assumptions or weak sensitivity to recent behavior of prices. By defining daily volatility for a (no-arbitrage) price process as the integrated variance, which is basically accumulated variance over some time period, the non-parametric estimator called realized variation (RV) was popularized in a series of papers by Andersen et al. [2001], Andersen and Bollerslev [1998], Corsi et al. [2001], Barndorff-Nielsen and Shephard [2002], whereas it goes back at least to Merton [1980]. Technically, RV is calculated as the sum of squared

intraday returns over equidistant time intervals, i.e. the sample analog of quadratic variation. Consequently, the theory on properties of this estimator was adopted from stochastic calculus. Thereby, RV was shown to be asymptotically consistent with time interval going to zero under the assumption of no jumps and noise in the price process (see Andersen et al. [2001]).

The subsequent empirical analysis of the high frequency financial data revealed that whether these assumptions are fulfilled depends heavily on the time scale, whereas RV diverges for increasing frequency. The common assumption of continuous price process is an approximation (or limit) of the actually observed prices when the frequency increases. Since prices are determined through transactions, they are observed at random not necessarily equidistantly placed time points, whereas the order flow has impact on the price process (news announcements). Responsible for the divergence of RV is *microstructure noise*, which is a collective term for approximation errors stemming from sampling equidistantly placed prices from observed trade data, bid-ask spreads, as well as price rounding and occasional factual errors. Hansen and Lunde [2006] provide empirical analysis of this noise behavior. Furthermore, jumps in asset prices are present at least due to market rules (no overnight trading, circuit breakers) and are hardly identifiable due to the discrete-time nature of observations. Finally, considering the case of several assets, the time asynchronicity is an issue. In order to tackle these problems a multitude of alternative methods collectively referred to as realized measures (RMs) was proposed. For a thorough empirical overview of different classes of RMs see Liu et al. [2015].

By treating volatility as an observed process, which is estimated with some RM, properties of such time series could be explored and modeled much more efficiently, compared to econometric models based on unobserved volatility. The topic of this dissertation is modeling time series of RMs with a special focus on forecasting. Currently, most popular univariate models are basically restricted linear regressions with some economic argumentation, whereas multivariate methods usually face complex numerical or theoretical issues. Chapters 2 and 3 propose alternative approaches to modeling univariate RMs from two different perspectives, distributional with copulas and non-parametric with B-splines, respectively. Chapter 4 introduces a novel factor-based approach to modeling multivariate RMs. The articles which constitute this dissertation are briefly discussed in the remainder of this introduction.

1.1 Vine copula approach for modeling realized measures

The famous result by Sklar [1959] states that any multivariate distribution can be decomposed into corresponding marginal distributions and a multivariate function called *copula*, in the way that copula maps marginals to the joint distribution function. The practical attractiveness and popularity of the copula approach stems from the converse of Sklar's theorem. Namely, a combination of any marginals and copula always defines a valid multivariate distribution. The main references on copula theory are Nelsen [2006] and Joe [2014].

Considering copulas within modeling framework brings several advantages. Firstly, during model specification a diverse set of distributions can be constructed from combinations of marginals and copulas, i.e. individual and joint behaviors. Secondly, the estimation procedure is split into estimating univariate distributions and their dependencies sequentially, thus simplifying the overall (numerical) problem and allowing for application of research results specific to each step. Using copulas is also supported from empirical perspective, since the common normality assumption for financial data is often highly questionable. Finally, copulas as a measure of dependence overcomes the drawbacks of (auto)correlation. The latter measures only pairwise dependencies, whereas its estimation is possible only with finite second moment and is not reliable in case of infinite fourth moment (see Davis and Mikosch [1998]). Empirical evidence suggests the presence of infinite higher moments due to heavy tails in financial time series (see Ibragimov et al. [2015]).

While copulas steadily reserved an important place in financial theory and actuarial science (see Ibragimov and Prokhorov [2017] and McNeil et al. [2015]), the interest in using copulas for modeling (financial) time series has manifested itself only recently. The univariate modeling approach is based on defining stationary discrete-time Markov chains through invariant marginal distribution of the chain and transition distribution defined through a copula. Darsow et al. [1992] were the first to apply copulas within time series context providing theoretical results on first-order Markov chains, whereas Ibragimov [2009] generalized the findings to Markov chains of an arbitrary order. The semiparametric estimation of copula-based time series was discussed in Chen and Fan [2006a,b].

The main advantage of copulas within time series modeling is that they allow for recon-

struction of the conditional transitional distributions. Thus, (conditional) moments higher than one and quantiles other than median can be recovered seamlessly. Furthermore, model specifications which include copulas with asymmetric tail dependence could generate Markov chains with increasing serial dependence when approaching high (low) values. Thus, such copulas have the potential to model occasional temporary hoarding behaviors, like for example volatility clustering.

However, theoretical flexibility of copula approach was practically very limited due to scarcity of high-dimensional copula families, whereas most of them are generalizations of bivariate copulas. These extensions to higher dimensions impose homogeneous dependencies with simple parameterization. In his seminal paper Joe [1996] introduced the idea of building multivariate distributions through sequentially defining intermediate (conditional) bivariate dependencies. Using copulas as building blocks in this sequential manner quickly claimed a separate field of research within copula theory. Generally, these constructions are referred to as (regular) vines or vine copulas. Interest in this approach grew exponentially fast, whereas vines saw a multitude of applications (see Aas [2016] for a review).

We propose a vine based modeling framework for time series of univariate realized measures. We consider two most studied and popular vine structures, namely canonical vines (C-vines) and drawable vines (D-vines). In a specific sense they offer diametrically opposed modeling frameworks. C-vines treat some (conditional) variable inherently as central at each sequential level, whereas all variables are equal from model perspective within D-vines. The empirical qualities of competing vines are analyzed based on bi-power variation (BPV) calculated for 13 equity indices over a timespan of ten years. BPV is a popular realized measure due to a good balance between technical simplicity and robustness against most of microstructure noise. Both vines are estimated and analyzed on the datasets as in the Heterogeneous Auto-Regressive (HAR) and Mixed Data Sampling (MIDAS) models, whereas the corresponding approaches along with bivariate copula are taken as benchmarks. During application study we put special focus on analyzing the non-linearity of estimated vines. The results present overall flexibility of the proposed approach within the in-sample as well as statistically superior forecasting power in out-of-sample.

1.2 Non-parametric approach for modeling realized measures

Popular methods for modeling one-dimensional realized measures are usually linear time series models. Consequently, their theoretical properties contradict empirical characteristics of volatility and financial data in general. Features like regime switching, non-normality, asymmetric behaviors and obviously non-linear functional relationships among lagged observations are not possible within standard linear models by design. Partial departure from linearity can be facilitated by considering non-gaussian distributions for the innovations. However, such *implicit* extensions of well-known models require suitable parametric assumptions and are theoretically less preferable compared to an *explicit* extension, which is based on some non-linear functional relationship.

Non-linear models are typically based on some *a priori* parametric assumptions, motivated by empirical evidence or economic argumentation. For example, ARCH model family (Engle [1982]), which defines a specific functional relationship between returns and volatility, is a very famous member of this class. Furthermore, time series models built around copulas - like the one presented in the previous section - also constitute a special parametric case of this class. By relying on copula approach the generality is sacrificed deliberately for distributional modeling of the conditional dependencies. When probabilistic aspects are not the prime research focus, other non-linear time series methods can be attractive within volatility modeling framework.

Nevertheless, choosing some specific functional link over others and motivating the choice economically is usually problematic. The decision is further complicated within volatility modeling framework when relying on realized measures, since these are (noisy) non-parametric estimators with normally complex dynamics. The simple fact that there are different realized measures, each with several choices for its implementation, makes an optimal parametric model specification possibly dependent on the selected realized measure.

The most general modeling approach is to express random variable as (non-linear) function of its lagged values and white noise, whereas the functional form of the dynamics is empirically estimated *ex post*. Despite the infinite number of possible functions, non-linear time series literature produced very powerful results linking functions to stationarity of the

produced process (see Tong [1990] for summary). However, inference without any *a priori* assumptions on the data generating process results in what is called saturated non-parametric model, which is known to be subject to the curse of dimensionality (see Chapter 8 in Fan and Yao [2003] and references therein). Thus, making some mild general assumptions on the underlying functional form, motivated by empirical evidence, is theoretically more appropriate. In this way, parameters of a model specification can be estimated much more efficiently.

It is a commonly accepted fact that time series of realized measures (or volatility in general) exhibit non-linear behavior. Perhaps, the most widely recognized empirical properties of realized measures are high cross-time persistence and clustering behavior. In terms of functional modeling, these stylized facts could be interpreted as continuity and locality of the underlying functional link, respectively. Making the continuity assumption reduces the choice to a smaller set of well-studied functions, whereas locality is generally acceptable for continuous functions. Empirically recovering the true functional form can be facilitated by using some appropriate basis.

We propose a non-parametric multivariate regression based on basis splines or B-splines. These functions span the space of piecewise polynomials, which fulfill both continuity and locality assumptions. Among other possible choices, B-splines stand out as a class of functions with well-studied theoretical properties, abundant application results, computationally efficient calculations and uniqueness. Whereas one-dimensional B-splines found widespread recognition as a tool for non-parametric estimation, multivariate extensions were limited to tensor products, which are inherently univariate. We present the theoretical foundations of the multivariate B-splines and study their empirical properties within an application study. Thus, the true functional link is recovered using an ensemble of multivariate B-splines. Technically, the data is transformed through a set of B-splines and linear regression is fitted to the transformed data. The proposed method is compared against plain HAR model on the same information filters for 22 equity indices using BPV. The analysis is performed on in-sample and the forecasting power is tested statistically. Presented results reveal the superiority of the proposed approach in modeling and analyzing time series, as well as in delivering significantly better forecasts.

1.3 Matrix variate factor model

Modeling (co)volatility in case of several assets faces challenges, which are quantitatively as well as qualitatively much more complex compared to the univariate case. First of all, the practical implementation of any multivariate realized measure is severely complicated by the asynchronicity of trade data for different assets. Additionally, the candidate model must ensure positive definiteness, which is an intrinsic property of covariance matrices, and be numerically feasible. The methods proposed in the literature so far guarantee positive definiteness either through an appropriate model design, matrix transformation or distribution assumptions. On the other hand, model estimation is mostly possible only for small dimensions or by naive simplification of parameterization.

First approaches to model multivariate risk processes were essentially GARCH extensions to higher dimensions. Bollerslev et al. [1988] proposed to apply univariate GARCH processes to matrix vectorization. In this way, positive definiteness can be guaranteed only numerically and due to generality of the proposed model specification it is almost intractable within empirical applications without further simplification. Later developed BEKK model (see Engle and Kroner [1995]) is a GARCH-type approach directly for matrices. The model design naturally ensures positive definiteness although parameter estimation is still computationally demanding and their interpretation is cumbersome.

Another popular framework is similar to non-linear combinations of volatilities in the spirit of copula modeling, whereas the dependence structure (correlation) is modeled separately from individual volatilities. Bollerslev [1990] proposed constant conditional correlation (CCC) model, where only volatilities are assumed to have dynamics. An improvement on CCC is the dynamic conditional correlation (DCC) model by Engle [2002], whereas the dynamics of correlation matrix depend on only two scalars. Combining daily and intraday information, Halbleib and Voev [2016], Oh and Patton [2016] considered mixed frequency DCC, where the correlation matrices are based on daily information and individual volatilities are modeled using intraday data.

A convenient way to ensure positive definiteness is by considering matrix transformations, whereas mostly Cholesky decomposition was favored in applications so far. For example, Chiriac and Voev [2011] applied VARFIMA to the elements of Cholesky matrix. The disadvantages of modeling the elements of Cholesky decomposition include the reliance on order,

obscure interpretations of the parameters and introduction of bias through back multiplication (squared residuals). Another matrix decomposition was considered in Bauer and Vorkink [2011], who modeled the elements of matrix logarithm. This decomposition is considerably more computationally demanding than Cholesky and is generally constructed with error due to approximations.

Positive definiteness can be ensure by introducing appropriate distributional assumptions. Gouriéroux et al. [2009] proposed Wishart Autoregressive (WAR) model, which is defined as conditionally non-central Wishart distribution with dynamics in the non-centrality parameter. Later Golosnoy et al. [2012] proposed Conditional Autoregressive Wishart (CAW) under which the conditional distribution is central Wishart, whereas scaling matrix has BEKK dynamics. Again, the inference is feasible only for small matrix dimensions.

The numerical complexity of modeling matrix processes can be simplified by relying on dimension reduction techniques. Thereby, introducing factor structure is arguably the most popular of such techniques. In this way the number of parameters is reduced whereas their interpretability is preserved. Furthermore, introducing (latent) factors normally offers new analytical possibilities and facilitates efficient estimation. Probably the first factor approach for multivariate realized measures is the model by Tao et al. [2011], which is basically PCA for matrices. Estimation procedure is highly efficient but the positive definiteness of forecasts remained unresolved.

In Chapter 4 we propose a novel matrix variate factor model, whereas factors are considered latent and are filtered from observed data. The model naturally ensures positive definiteness due to appropriate distributional assumptions. Furthermore, we suggest a computationally efficient estimation procedure based on Expectation Maximization algorithm using analytical derivatives. Technically, estimation is a deterministic sequence of simple matrix operations. Latent factors are filtered using Kalman-type technique, whereas the estimates are shown to be asymptotically consistent with increasing dimension of data. Theoretical properties of estimation procedure are analyzed within a simulation study. Empirical application is performed for 60 most traded equities from S&P100 over a period of five year. The proposed approach is compared to two benchmarks. Matrix forecasts are analyzed using standard matrix norms and based on the performance of the minimum variance portfolio. Results reveal the superiority of proposed model.

Bibliography

Kjersti Aas. Pair-copula constructions for financial applications: A review. *Econometrics*, 4 (4):43, 2016.

Yacine Aït-Sahalia and Jean Jacod. *High-frequency financial econometrics*. Princeton University Press, 2014.

Torben G. Andersen and Tim Bollerslev. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 39(4):885–905, 1998.

Torben G. Andersen, Tim Bollerslev, Francis X. Diebold, and Paul Labys. The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, 96 (453):42–55, 2001.

Ole E. Barndorff-Nielsen and Neil Shephard. 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):253–280, 2002.

Gregory H. Bauer and Keith Vorkink. Forecasting multivariate realized stock market volatility. *Journal of Econometrics*, 160(1):93–101, 2011.

Tim Bollerslev. Modelling the coherence in short-run nominal exchange rates: A multivariate Generalized ARCH model. *The Review of Economics and Statistics*, 72(3):498–505, 1990.

Tim Bollerslev, Robert F. Engle, and Jeffrey M. Wooldridge. A capital asset pricing model with time-varying covariances. *Journal of Political Economy*, 96(11):116–131, 1988.

Xiaohong Chen and Yanqin Fan. Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. *Journal of Econometrics*, 135(1):125–154, 2006a.

Xiaohong Chen and Yanqin Fan. Estimation of copula-based semiparametric time series models. *Journal of Econometrics*, 130(2):307–335, 2006b.

Roxana Chiriac and Valeri Voev. Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics*, 26(6):922–947, 2011.

Fulvio Corsi, Gilles Zumbach, Ulrich A. Muller, and Michel M. Dacorogna. Consistent high-precision volatility from high-frequency data. *Economic Notes*, 30(2):183–204, 2001.

William F. Darsow, Bao Nguyen, and Elwood T. Olsen. Copulas and Markov processes. *Illinois Journal of Mathematics*, 36(4):600–642, 1992.

Richard A. Davis and Thomas Mikosch. The sample autocorrelations of heavy-tailed processes with applications to ARCH. *The Annals of Statistics*, 26(5):2049–2080, 1998.

Robert F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007, 1982.

Robert F. Engle. Dynamic conditional correlation. *Journal of Business & Economic Statistics*, 20(3):339–350, 2002.

Robert F. Engle and Kenneth F. Kroner. Multivariate simultaneous Generalized ARCH. *Econometric Theory*, 11(1):122–150, 1995.

Jianqing Fan and Qiwei Yao. *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer, 2003.

Vasyl Golosnoy, Bastian Gribisch, and Roman Liesenfeld. The conditional autoregressive Wishart model for multivariate stock market volatility. *Journal of Econometrics*, 167(1): 211–223, 2012.

Christian Gouriéroux, Joann Jasiak, and Razvan Sufana. The Wishart autoregressive process of multivariate stochastic volatility. *Journal of Econometrics*, 150(2):167–181, 2009.

Roxana Halbleib and Valeri Voev. Forecasting covariance matrices: a mixed approach. *Journal of Financial Econometrics*, 14(2):383–417, 2016.

Peter R. Hansen and Asger Lunde. Realized variance and market microstructure noise. *Journal of Business & Economic Statistics*, 24(2):127–161, 2006.

Marat Ibragimov, Rustam Ibragimov, and Johan Walden. *Heavy-Tailed Distributions and Robustness in Economics and Finance*. Springer, 2015.

Rustam Ibragimov. Copula-based characterizations for higher order Markov processes. *Econometric Theory*, 25(03):819–846, 2009.

Rustam Ibragimov and Artem Prokhorov. *Heavy Tails and Copulas*. World Scientific, 2017.

Harry Joe. Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. *Lecture Notes-Monograph Series*, 28:120–141, 1996.

Harry Joe. *Dependence Modeling with Copulas*. Taylor & Francis, 2014.

Lily Y. Liu, Andrew J. Patton, and Kevin Sheppard. Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. *Journal of Econometrics*, 187(1):293–311, 2015.

Alexander J. McNeil, Rüdiger Frey, and Paul Embrechts. *Quantitative Risk Management: Concepts, Techniques and Tools Revised edition*. Princeton University Press, 2015.

Robert C. Merton. On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics*, 8(4):323–361, 1980.

Roger Nelsen. *An introduction to copulas*. Springer, 2006.

Dong Hwan Oh and Andrew J. Patton. High-dimensional copula-based distributions with mixed frequency data. *Journal of Econometrics*, 193(2):349–366, 2016.

Abe Sklar. *Fonctions de répartition á n dimensions et leurs marges*. Université Paris 8, 1959.

Minjing Tao, Yazhen Wang, Qiwei Yao, and Jian Zou. Large volatility matrix inference via combining low-frequency and high-frequency approaches. *Journal of the American Statistical Association*, 106(495):1025–1040, 2011.

Howell Tong. *Non-linear Time Series: A Dynamical System Approach*. Oxford University Press, 1990.

Chapter 2

Article 1: Modeling temporal dependence of realized variances with vines

Claudia Czado, Eugen Heine and Yarema Okhrin (2019). Modeling temporal dependence of realized variances with vines. *Econometrics and Statistics*: 12, pp. 198-216, DOI: `https://doi.org/10.1016/j.ecosta.2019.03.003`

Chapter 3

Article 2: Modeling temporal dependence of realized measures with multivariate B-splines [1]

Abstract

A general non-linear non-parametric modeling framework for time series is proposed. The considered approach is capable of replicating a broad class of cross-time dynamics. The basic idea is making very weak assumptions on the true functional form of the dependencies whereas inferring it from data using B-splines. The application focus in this paper lies on modeling realized measures. Current popular methods in this field of research neglect the potential non-linear and non-monotonic temporal dependencies. The proposed approach overcomes common drawbacks of such models, whereas cross-time persistence and volatility clustering can be replicated seamlessly. HAR, which can be seen as a special case of the suggested modeling framework, is chosen as benchmark. Both models are evaluated within an extensive empirical study both in- and out-of-sample and their forecasting ability is compared statistically. The results suggest that the proposed model delivers significantly better results.

Keywords: B-splines, realized volatility, forecasting, time series

[1] Submitted to *Econometrics and Statistics* (2020).

3.1 Introduction

Volatility within financial applications is generally an unobservable process which is usually estimated non-parametrically from time series of returns. A monumental paradigm shift in estimating volatility was made possible due to the improved availability and accessibility of high-frequency (intraday) financial data. The focus of research community on modeling intraday asset returns in continuous time led to fundamental theoretical results, that served as the basis for developing realized measures (RMs), which are non-parametric estimators of integrated variance. Practical issues, like microstructure noise or asynchronicity of data, stimulated the development of different classes of RMs. Consequently, their use spread widely within econometrics and finance due to the superior quality of volatility estimates compared to those based on daily data.

By making volatility directly measurable, the focus of most applied empirical work was put on statistical modeling and forecasting of volatility, whereas standard time series approaches were considered first. The two most prominent and widely used examples are HAR by Corsi [2009] and MIDAS by Ghysels et al. [2004]. Both methods are basically constrained linear regressions based on the economic assumption of different dealing frequencies or exponentially falling cross-time linear dependence in volatility, respectively. They are able of reproducing the long memory behavior present in time series of volatility and are sparsely parameterized. However, the problem thereby is the imposed linearity, which is emblematic of many proposed methods so far. Linear regression delivers a global approximation neglecting local behavior, like volatility clustering. Furthermore, the incremental change in forecasts is always constant, i.e. it does not depend on the currently observed level of volatility. On the other hand, considering some non-linear functional form of cross-time dependence is complicated by the amount of possible choices.

In this paper, a non-parametric method capable of replicating a broad class of cross-time dependence structures is considered. Probably the first non-parametric approach to reconstruct the functional form of cross-time dependence from data was locally-polynomial estimation proposed in Stone [1977], Cleveland [1979], Katkovnik [1979], whereas the statistical properties of the estimator were further discussed in Tsybakov [1986]. The model

proposed in this paper is a non-linear autoregression with additive noise

$$X_t = f\left(X_{t-1}, \ldots, X_{t-k}\right) + \epsilon_t,$$

whereas function f is recovered from data using an ensemble of multivariate basis splines or B-splines. These are non-negative bounded piecewise polynomial functions with compact support. The attractiveness of B-splines is supported by their effective calculation, strong approximation power, linear independence and uniqueness. The latter property is revealed when compact support in form of simplex is given, i.e. area where B-spline is positive and zero elsewhere.

Probably the first use of splines in regression modeling was Multivariate Adaptive Regression Splines (MARS) model introduced in Friedman [1991], whereas an application to time series modeling was examined in Lewis and Stevens [1991]. Engle and Rangel [2006] were first to consider splines for explaining the slow changes of unconditional variance in the context of volatility modeling. Later, Audrino and Bühlmann [2009] applied B-splines within GARCH(1,1) by reconstructing the functional link between past daily returns and volatilities. However, their multivariate implementation of B-splines is essentially univariate, since tensor product of B-splines is taken. In this paper, true multivariate B-splines are utilized outside GARCH framework, i.e. the behavior of volatility is explained only by its past values. Furthermore, the focus lies mainly on statistical modeling.

The probabilistic properties of dynamical systems - such as non-linear autoregression - with a specific focus on geometric ergodicity were discussed in Chan and Tong [1985]. Generally, if function f is bounded then the defined process is stable. Since B-splines are bounded functions the proposed approach defines a stable time series process.

The paper is structured as follows. The next section provides basic information on realized measures and outlines motivation for development of a new method. Section 3.3 provides details on the proposed model. Introduction into the theory of multivariate B-splines is the subject of Section 3.4. The estimation procedure is discussed in Section 3.5. The results of an extensive empirical study are summarized in Section 3.6. Section 3.7 concludes.

3.2 Modeling realized measures

Volatility of financial assets is generally a latent process, which can be estimated using observed price movements, whereas the frequency of data has crucial impact on the quality of such approximation. Since intraday financial data became available to the research community, a new field of financial econometrics which deals with modeling inherent risk process expanded quickly, delivering fundamental theoretical results. Thereby, the main focus lies on developing and modeling so-called realized measures (RMs), which are designed to estimate integrated variance (IV) of intraday price movements. The simplest estimator of IV is the realized variation (RV), which was the first and is by far the most well-known of such proposed methods. It is calculated as the sum of squared high-frequency returns over equidistant time intervals

$$RV_t = \sum_{m=1}^{M} r_{t,m}^2,$$

where $r_{t,m}$ is the intraday return observed at the subinterval m on day t. RV consistently estimates IV, when microstructure noise and jumps are not present in the observed asset returns. In this case, when M goes to infinity, RV converges uniformly in probability to quadratic variation of the price process, which is called IV within financial theory. Otherwise, RV diverges or delivers biased estimates of IV. Due to strong empirical evidence for the presence of such market frictions in the intraday time series of asset returns, different classes of realized measures were developed. The specific RM used within empirical study is the bi-power variation (BPV), which is calculated as the summed products of absolute values of two consecutive returns

$$BPV_t = \sum_{m=2}^{M} |r_{t,m}| \, |r_{t,m-1}| \, .$$

It is a comparably simple measure, which nonetheless isolates the market frictions under the assumptions that noise is not autocorrelated and the probability of a jump in two small consecutive intervals is negligible. Within this setting BPV converges uniformly in probability to the IV (see Barndorff-Nielsen and Shephard [2004]). Furthermore, as discussed in Liu et al. [2015], BPV is one of the most efficient RMs available. Aït-Sahalia and Jacod [2014] provide a very thorough overview of theoretical results concerning RMs.

There are different classes of approaches for modeling time series of RMs. Arguably, the most prominent method - which additionally serves as benchmark through the empirical

study - is the so-called Heterogeneous Auto-Regressive (HAR). It was first proposed in Corsi [2009] and quickly emerged as one of the most popular choices. The inspiration for the model design and name comes from Heterogeneous Market Hypothesis. Thereby, the dynamics of volatility is linked to activity of heterogeneous market participants. HAR is based upon the assumptions, that investors can be classified using their dealing horizon and that they affect volatility differently. For example, market makers with high trading activity and pension funds with low dealing frequency influence the market differently. The model is defined as the following constrained linear regression

$$RM_t = \phi + \phi^{(d)} \cdot RM_{t-1} + \phi^{(w)} \cdot \frac{1}{4} \sum_{k=2}^{5} RM_{t-k} + \phi^{(m)} \cdot \frac{1}{15} \sum_{k=6}^{20} RM_{t-k} + \varepsilon_t,$$

where ϕ is the intercept, $\phi^{(d)}$ is the regression coefficient for the previous day (D) observation, $\phi^{(w)}$ and $\phi^{(m)}$ - coefficients for the proxies of weekly (W) and monthly (M) information, respectively. The model is estimated using OLS.

The most well-known empirical property of RMs is their high persistence. From economic viewpoint this phenomena can be explained based on mixture of distributions (Tauchen and Pitts [1983]), interplay between trading volume and returns (Bollerslev and Jubinski [1999]) or aggregation of numerous linear processes (Zaffaroni [2004]). Since the focus of this paper lies on statistical modeling, some general conditions for long memory should be derived from statistical viewpoint. According to Wold's decomposition every (weakly) stationary time series can be split into a deterministic trend and purely random component. High persistence of such time series indicates that deterministic part is a smooth function of the past observations and that the random component has relatively small variation. Furthermore, time series of RMs generally demonstrate volatility clustering. This empirical finding suggests varying dependence structure, i.e. some local approximations are desirable. Due to the linearity of the most widely used approaches such approximations are not possible since linear models are by definition fitted globally.

Figure 3.1 provides empirical evidence for the mentioned issues. Thereby, information sets as in HAR are calculated for German DAX index over the period from January 2000 to December 2018. Next, two-dimensional grid on weekly and monthly information is constructed and the daily data is categorized based on it. The average value of daily data associated with each of such squares on the grid is illustrated in the left panel of Figure 3.1. Additionally,

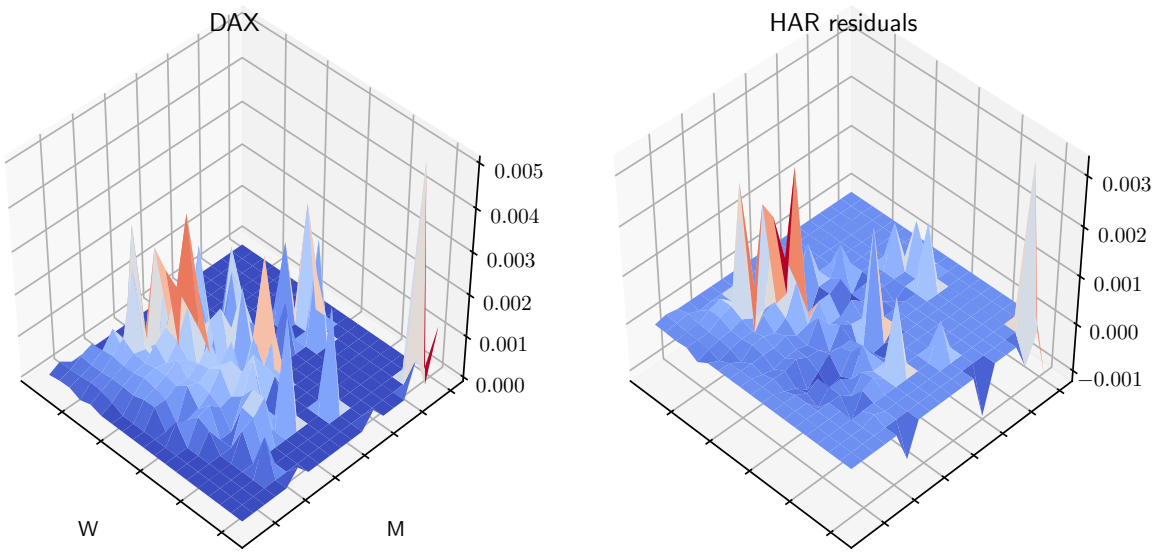

Figure 3.1: Volatility surface of 5-minute BPV for DAX index (left) and residuals from estimated HAR (right).

the residuals after estimating HAR are categorized according to the two-dimensional grid, and the results are reported in the right panel of Figure 3.1. Plots clearly illustrate the local character and (local) smoothness present in data. Such behavior cannot be reproduced with linear approach, since it basically builds a tilted plane. The next section briefly elaborates on non-linear autoregression and introduces the proposed model.

3.3 Non-parametric B-spline regression

The foundations of non-linear autoregressive modeling were laid in Jones [1978]. A process $(X_t)_{t\in\mathbb{Z}}$ follows non-linear autoregressive model of order k with additive noise if there exists a function $f : \mathbb{R}^k \to \mathbb{R}$ such that

$$X_t = f\left(X_{t-1}, \ldots, X_{t-k}\right) + \epsilon_t, \tag{3.1}$$

where $(\epsilon_t)_{t\in\mathbb{Z}}$ are *iid* and ϵ_t is independent of $(X_s)_{s<t}$ for $t \in \mathbb{Z}$. Within setting (3.1) the slow varying conditionally deterministic drift is given by the function f itself, whereas purely random part is ϵ_t. The former is commonly referred to as "bone" or "skeleton", whereas the latter - as "skin". Thus, it is assumed that $(\epsilon_t)_{t\in\mathbb{Z}}$ are identically distributed fluctuations

around some (rigid) volatility "skeleton".

Up to now no conditions on the functional form of f were explicitly stated. Having a specific application case in mind, the selected f should bare characteristics which are strongly inline with the empirical findings partly discussed in the previous section. First of all, the function must be smooth in past values in order for model (3.1) to demonstrate long memory. Secondly, it should have flexible local behavior to replicate volatility clustering but also allow for lower levels of volatility during calm periods. And lastly, since different financial assets have distinct volatility profiles, the candidate function must be adaptable, i.e. it should provide a flexible framework for replicating different behaviors. In this setting, taking a functional basis consisting of smooth linearly independent functions with uniformly bounded derivatives to model the behavior of conditionally deterministic drift is a logical choice. Since there are generally different approaches for constructing a functional basis, some selection must be done. Basis splines or B-splines constitute a theoretically powerful and computationally effective functional basis for piecewise polynomial functions. Under certain conditions presented later, linear combinations of B-splines can approximate continuous functions with any given degree of accuracy. Thus, a proper ensemble of B-splines is able to fulfill the stated criteria.

Next the proposed model is stated. Define $L_1 : \mathbb{R}^k \to \mathbb{R}^{l_1}$ and $L_2 : \mathbb{R}^k \to \mathbb{R}^{l_2}$ to be some information filters with $l_1, l_2 \leq k$. Given some ensemble of multivariate B-splines $(M_i)_{i \in N}$, the following modeling framework is proposed

$$X_t - \alpha' L_1(\mathbf{X}_{t-1}) = \sum_{i \in N} \beta_i M_i(L_2(\mathbf{X}_{t-1})) + \varepsilon_t, \tag{3.2}$$

where $\mathbf{X}_{t-1} = (X_{t-1}, \ldots, X_{t-k})'$ and M_i is a multivariate B-spline of dimension l_2 with index i. B-splines $(M_i)_{i \in N}$ in (3.2) have local support on some cylinder $C_i \in \mathcal{B}(\mathbb{R}^l)$, whereas the union $\cup_{i \in N} C_i$ covers the compact support of $(X_t)_{t \in \mathbb{Z}}$. Due to this locality property of B-splines the proposed model can be classified into the general class of threshold models (see Tong [1990]). Furthermore, since B-splines are bounded and have compact support, the stability of the dynamic model (3.2) is ensured. During empirical application we consider the case of information sets as in HAR. Thereby, filters $L_1 = L_2$ reduces the information of the last 20 days to only 3 variables, i.e. daily, weekly and monthly information. In the next section theoretical foundations of B-splines are presented.

3.4 B-splines

Using polynomials for approximating continuous functions is advantageous from several viewpoints. Complex functions are generally computationally difficult to evaluate, whereas polynomials can be evaluated, differentiated and integrated in a very straightforward fashion. Furthermore, since they have easily tractable analytical properties, the target function can be studied efficiently. Nevertheless, the simplicity of polynomials is still enough for approximating a very broad class of continuous functions to a specified degree of accuracy. A polynomial of *order* k is defined as

$$p(x) = \sum_{j=1}^{k} a_j x^{j-1}, a_k \neq 0,$$

whereas polynomials of order k form linear space further denoted by $\Pi_{<k}$.

Given some function f, distinct points $\xi_1, \ldots, \xi_k$ and function values $f(\xi_i), \ldots, f(\xi_k)$, there exists exactly one polynomial $p \in \Pi_{<k}$ that interpolates f on these points. The approximation accuracy of this approach is well studied for different classes of functions and depends crucially on local properties of the function to be approximated. Thus, if f has "unexpected" local behavior the approximation is globally weak. An appropriate criteria to classify such behaviors for some f is the number and uniformity of continuous derivatives. The Jackson's Theorem (see (22) in de Boor [2001]) states that for a given f which is r times continuously differentiable on interval $[a, b]$, the accuracy increases with smaller $b - a$ or/and bigger order k of approximating polynomials. Thus, to gain additional flexibility and accuracy of approximation, the local behavior can be modeled by introducing piecewise approximations, i.e. splitting the interval $[a, b]$ into parts and constructing polynomials.

Let $\xi := (\xi_i)_1^{l+1}$ be a strictly increasing sequence of points in $\mathbb{R}$ and $\nu := (\nu_i)_2^l$ a sequence of non-negative integers less than or equal to k. Further define $\Pi_{<k,\xi,\nu}$ as the linear space of piecewise polynomial functions with continuity conditions at points ξ imposed by ν, i.e. each $f \in \Pi_{<k,\xi,\nu}$ is a polynomial of order k on (ξ_i, ξ_{i+1}) and $f \in C^{v_i - 1}$ at ξ_i for $i = 2, \ldots, l$. Thus, the convention of continuity from the right is adopted.

To allow for practical implementation of functions from $\Pi_{<k,\xi,\nu}$ a basis of this linear space is necessary. Since it is not unique, the choice should be made in favor of some computationally and analytically attractive set of functions. One of the most straightforward

ways to construct a basis of $\Pi_{<k,\xi,\nu}$ is by simply using the truncated power functions

$$(t-x)_+^r := ((t-x)_+)^r = (\max\{t-x,0\})^r. \tag{3.3}$$

at appropriate points. However, a much more efficient basis can be built by generalizing (3.3) to B-splines through certain linear combinations.

B-splines were first introduced as own class of functions by Schoenberg [1946] and the term was coined in Schoenberg [1967]. A very comprehensive account of the topic can be found in de Boor [2001], whereas practical applications of B-splines are discussed in Höllig and Hörner [2013]. Univariate B-splines are basically (scaled) n-th order divided differences of truncated power function.

Definition 1 (Divided difference)**.** *Divided difference of order n is the coefficient of x^n in the polynomial that interpolates g at the points $t_j, \ldots, t_{j+n}$ in $\mathbb{R}$. It is denoted by*

$$[t_j, \ldots, t_{j+n}]\, g.$$

Divided differences are obviously invariant under permutations of the knots and allow for multiplicity greater than 1. For a good account of properties of divided difference see Chapter 1 in de Boor [2001]. To facilitate computational implementation divided differences are calculated using the following recursive relation:

$$\begin{aligned}
[t_j]g &= g(t_j), \\
[t_j, t_{j+1}]g &= \frac{[t_{j+1}]g - [t_0]g}{t_{j+1} - t_j}, \\
&\ldots \\
[t_j, \ldots, t_{j+n}]g &= \frac{[t_{j+1}, \ldots, t_{j+n}]g - [t_j, \ldots, t_{j+n-1}]g}{t_{j+n} - t_j}.
\end{aligned}$$

The original formula for B-splines as derived by Curry and Schoenberg [1966] is given in the following definition.

Definition 2 (B-spline of Curry-Schoenberg)**.** *The j-th B-spline of order k for the knot sequence $t := (t_i)_0^n$ in $\mathbb{R}$ is defined as*

$$M_{j,k,t}(x) := M_{j,k}(x|t_j, \ldots, t_{j+k}) = k\,[t_j, \ldots, t_{j+k}]\,(\cdot - x)_+^{k-1}\,,\ \forall x \in \mathbb{R}. \tag{3.4}$$

Notation $(\cdot - x)_+$ means that for a fixed x the function is treated as depending on only one variable. For further purposes, the notation of B-splines is simplified to $M_{j,k}$ by dropping reference to knot sequence in the index as long as it can be inferred from the context.

B-splines appear naturally when divided differences are applied to both sides of the Taylor expansion of some function f. This leads to the following important relation

$$[t_j, \ldots, t_{j+k}]\, f = \int_{\mathbb{R}} M_{j,k}(x) \frac{f^{(k)}(x)}{k!} dx,$$

whereas integral equals one if $f(x) = x^k$.

B-splines can be studied through (3.4) by applying properties of divided differences and function values can be calculated using the following formula

$$M_{j,k}(x) = k \sum_{j=0}^{n} \frac{(t_j - x)_+^{k-1}}{\prod_{i \neq j} (t_j - t_i)}.$$

However, to facilitate practical implementation recurrence relation for B-splines is more suitable. Thereby, B-spline of order 1 is the characteristic function divided by the length (volume) of the corresponding interval

$$M_{j,1}(x) = \begin{cases} \frac{1}{t_{j+1} - t_j}, & \text{if } t_j \leq x < t_{j+1}; \\ 0, & \text{otherwise} \end{cases}$$

whereas higher order B-splines can be calculated using the following recurrence relation (see Dahmen [1980])

$$M_{j,k}(x) = \frac{k}{k-1} \left(\frac{t_{j+k} - x}{t_{j+k} - t_j} M_{j+1,k-1}(x) + \frac{x - t_j}{t_{j+k} - t_j} M_{j,k-1}(x) \right). \tag{3.5}$$

Calculating function values using (3.5) is computationally efficient and supports establishing all analytical properties of (3.4).

B-spline is a piecewise polynomial of order k with breaks $t_j, \ldots, t_{j+k}$, hence consisting of at most k non-trivial polynomial pieces. The function is non-negative on $\mathbb{R}$, has small support in the sense that $M_{j,k}(x) > 0$ iff $x \in (t_j, t_{j+k})$ and is logarithmically concave on (t_j, t_{j+k}) (see Curry and Schoenberg [1966]). The derivative of B-spline (3.4) can be represented recursively using lower order B-splines. One interesting observation is that since $M_{j,k}$ has one continuous derivative more than $M_{j,k-1}$ and $M_{j+1,k-1}$, it follows that coefficients in (3.5) must cancel out the discontinuities of a certain derivative of lower order B-splines. Generally, one could define B-splines completely through this property without using divided differences. Furthermore, given some set of knots, B-spline function is fully determined. Thus, a refinement of approximation is possible by simply changing the knot sequence.

Figure 3.2 contains examples of B-splines for order $k = 2, 3, 4$ using an equidistant knots with step 0.1 on $[0, 1]$. The positivity and local support are clearly illustrated. Second order B-spline is a piecewise linear function consisting of two parts, with undefined derivative at the middle knot. Third and fourth order B-splines consist of three and four pieces which form functions in $C^{(1)}$ and $C^{(2)}$, respectively.

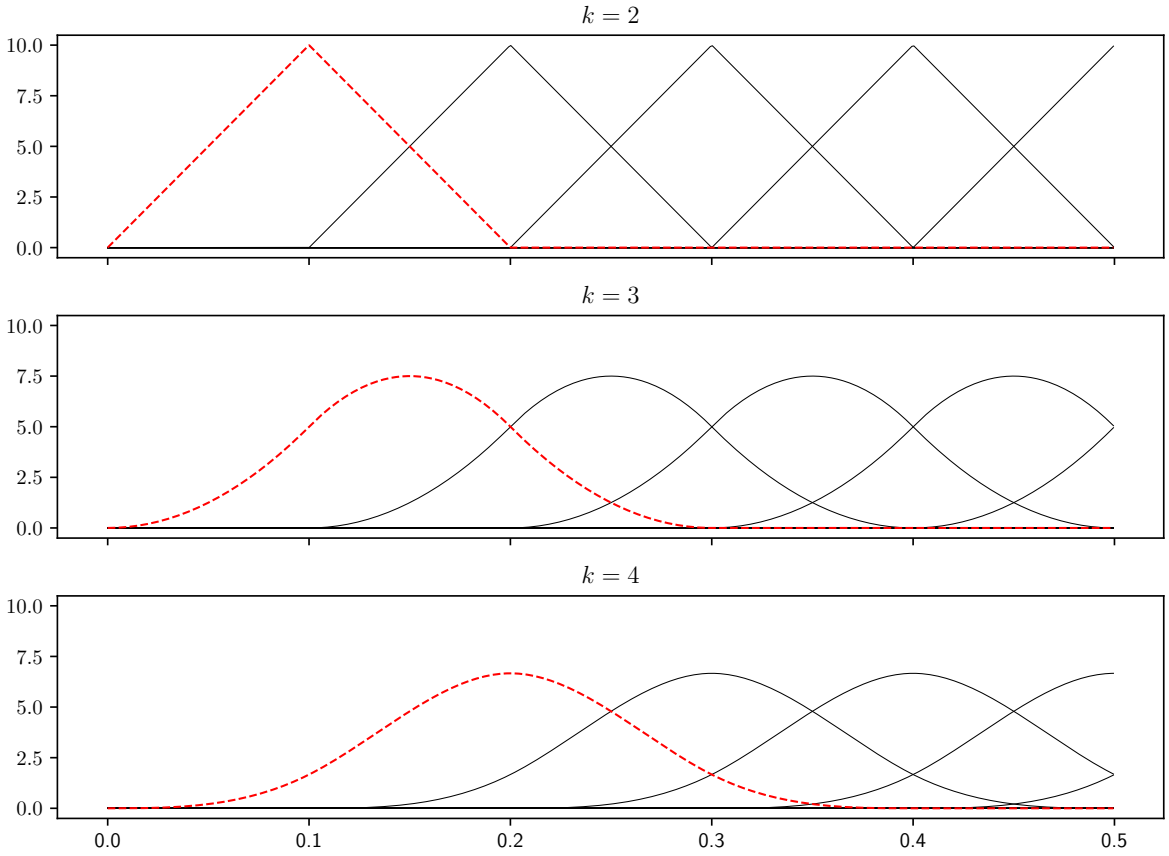

Figure 3.2: Example of B-splines on unit interval for $k = 2, 3, 4$.

The derivative of second order B-spline is discontinuous at the middle break due to an important intrinsic property. The cardinality of knots - number of coinciding knots - controls the smoothness of function in the neighborhood of each break, with fewer equal knots translating to more smoothness, i.e. k equals number of continuity conditions at some point x plus the multiplicity of x as a knot (see Lemma 1 in Curry and Schoenberg [1966]). At a k-fold knot no continuity conditions are imposed whereas no knot at a site enforces k continuity conditions, i.e. two polynomial pieces connected at such x must be identical. Thus, at a simple knot the function is in $C^{(k-2)}$, while for $x \neq \xi_i$ the function is $k-1$ times continuously

differentiable

As an example consider B-splines of order four with different continuity conditions at point 0.2 in Figure 3.3. Thereby one to four knots (indicated by plot title) coincide at point 0.2, illustrating distinct behavior for B-splines at this point within their support. For a knot sequence where 0.2 appears four times as a knot no continuity is observed at this point, whereas for smaller multiplicities better smoothness is achieved. Such intrinsic property could be used for controlling the approximation of some function by taking into account it's local behavior.

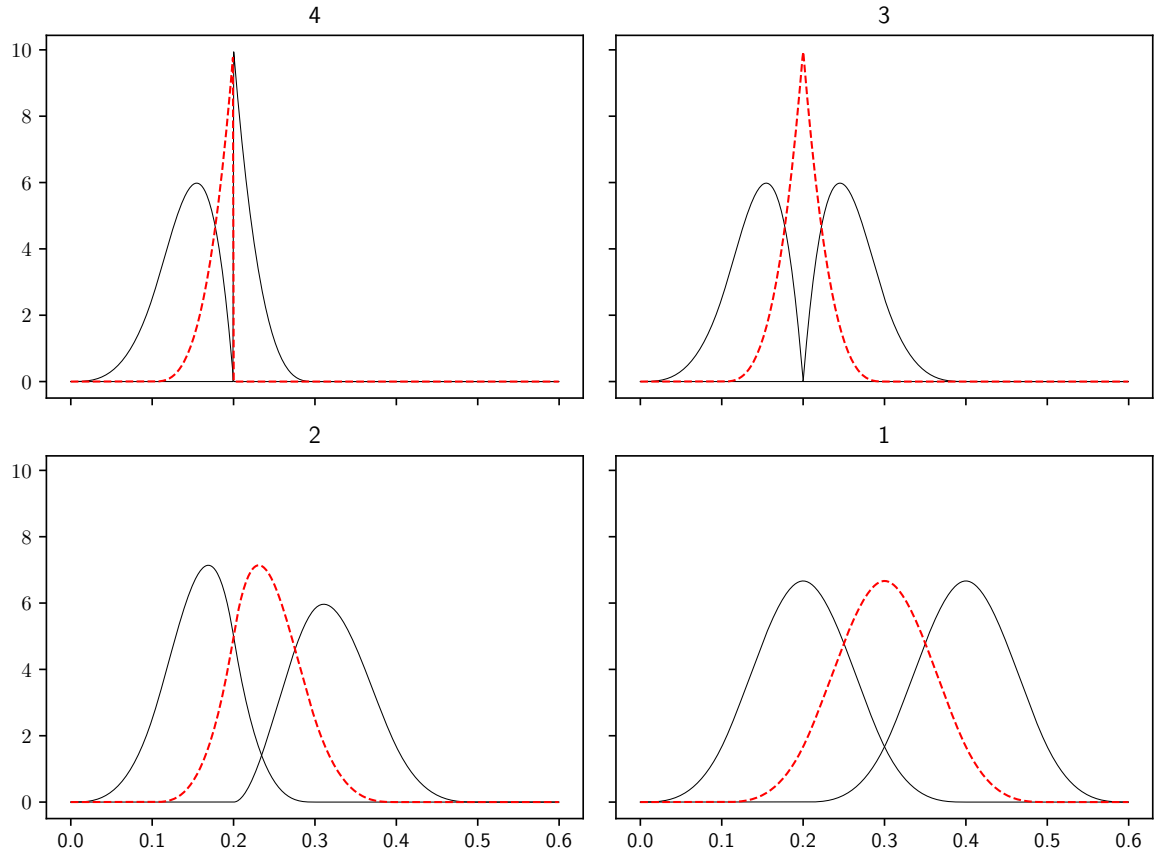

Figure 3.3: Example of B-spline of order 4 with different continuity conditions at point 0.2.

Now, the most appealing property of B-splines from practical perspective and the inspiration for their name is stated. For a strictly increasing sequence $\xi = (\xi_i)_{i=1}^{l+1}$ and a non-negative integer sequence $\nu = (v_i)_{i=2}^{l}$ with $v_i \leq k$ we set

$$n := k + \sum_{i=2}^{l} (k - v_i) = \dim \Pi_{<k,\xi,\nu}. \tag{3.6}$$

We further $\mathbf{t} := (t_i)_1^{n+k} \subset \xi$ to be a non-decreasing set of knots such that $\#\{j : t_j = \xi_j\} = k - v_j$ for $j = 2, \ldots, l$. For purposes of extrapolation we set the boundaries as follows

$$t_1 \leq t_2 \leq \ldots \leq t_k \leq \xi_1 \text{ and } \xi_{l+1} \leq t_{n+1} \leq \ldots \leq t_{n+k}.$$

The name basis spline is due to next theorem by Curry and Schoenberg [1966] with an alternative proof in de Boor [2001].

Theorem 1 (Basis property of B-splines)**.** *Let* $\mathbf{t}$ *be a non-decreasing sequence of knots as described above. Then the corresponding sequence of B-splines* $(M_{j,k})_1^n$ *is a basis for* $\Pi_{<k,\xi,\nu}$ *on* $[t_k, t_{n+1}]$

$$\Pi_{<k,\xi,\nu} = \left\{ \sum_{j=1}^{n} \alpha_j M_{j,k} : \alpha_j \in \mathbb{R} \right\}.$$

In particular, given some $I \subset [t_k, t_{n+1}]$ *the set*

$$M_I := (M_{j,k} : M_{j,k}(x) > 0, \ \forall x \in I)$$

is locally linearly independent on I*.*

This theorem enables reconstruction of any piecewise polynomial function from B-splines for an appropriate knot sequence. Furthermore, since $\Pi_{<k,\xi,\nu}$ is nested within the class of continuous functions, a broad spectrum of functions can be approximated non-parametrically through linear combinations of B-splines. Thereby sequence ξ stipulates the breakpoints, whereas $\mathbf{t}$ - the level of smoothness. In this way a knot with multiplicity k implies no continuity, whereas B-spline is in $C^{(k-2)}$ or $C^{(k-1)}$ for a point which is a simple knot or not a knot, respectively. Theoretical results suggest that the quality of approximation of continuous functions increases with number of knots and better uniformity of continuity, whereas the information about local behavior of target function can be deduced empirically. Developing a result comparable to Theorem 1 within high dimensional setting was made possible by the introduction of multivariate B-splines.

Multivariate B-splines

The crucial step in order to generalize univariate B-splines to higher dimensions was taken in Curry and Schoenberg [1966], where the authors delivered a geometric interpretation of

the univariate case. Let $t_1, \ldots, t_{k+1}$ be a strictly increasing sequence of knots in $\mathbb{R}$. Further define $\mathbf{y}_1, \ldots, \mathbf{y}_{k+1} \in \mathbb{R}^k$ such that the first coordinates coincide with $t_1, \ldots, t_{k+1}$, i.e.

$$y_{i,1} := (\mathbf{y}_i)_1 = t_i, i = 1, \ldots, k+1,$$

and such that k-dimensional simplex σ_k spanned by $\mathbf{y}_1, \ldots, \mathbf{y}_{k+1}$ has volume one, i.e.

$$\mathrm{vol}_k \sigma_k = \frac{1}{k!} |\det(\mathbf{y}_2 - \mathbf{y}_1, \ldots, \mathbf{y}_{k+1} - \mathbf{y}_1)| = 1.$$

Then as shown in Curry and Schoenberg [1966] the univariate B-spline on $t_1, \ldots, t_{k+1}$ is the linear density function of the orthogonal projection ("shadow") of σ_k onto $\mathbb{R}$:

$$M(x) = M(x|t_1, \ldots, t_{k+1}) = \frac{\mathrm{vol}_{k-1}\{\mathbf{y} \in \sigma_k | y_1 = x\}}{\mathrm{vol}_k \sigma_k} = \mathrm{vol}_{k-1}\{\mathbf{y} \in \sigma_k | y_1 = x\}.$$

This formula implies that a univariate B-spline at some point is calculated as the k-dimensional volume of intersection of σ_k with a hyperplane orthogonal to x-axis at some point.

As an example consider a B-spline of order two with knots t_1, t_2, t_3 as in Figure 3.4 (see Micchelli [1979]). The knots are "elevated" to $\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3$ in the plane, such that the area of the triangle is 1. For some point t the value of B-spline is calculated as the length of intersection between a line orthogonal to x-axis at point t and simplex $\sigma(\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3)$ marked by thick line in the figure.

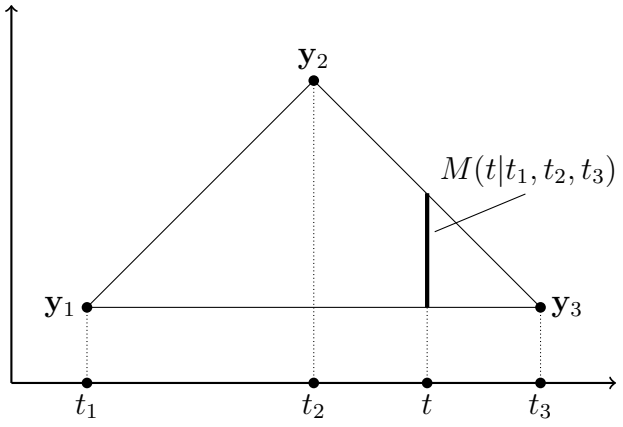

Figure 3.4: The geometric interpretation of univariate B-splines demonstrated for a B-spline of order 2 on some knots t_1, t_2, t_3.

The geometric interpretation of univariate B-splines provided in Curry and Schoenberg [1966] led to the following definition of multivariate B-splines introduced in de Boor [1976].

Definition 3 (Multivariate B-spline)**.** *Let* $\mathbf{T} = \mathbf{t}_1, \ldots, \mathbf{t}_{s+k}$, $k \geq 1$ *be a set of points in* $\mathbb{R}^s$ *such that the dimension of the convex hull of* $\mathbf{T}$ *is* s*, alternatively* $vol_s(\sigma(\mathbf{T})) > 0$*. Further let* $\mathbf{y}_1, \ldots, \mathbf{y}_{s+k} \subset \mathbb{R}^{s+k}$*, such that* $\mathbf{y}_i|_{\mathbb{R}^s} = \mathbf{t}_i$, $i = 1, \ldots, s+k$*. Then the multivariate B-spline is defined as*

$$M(\mathbf{x}|\mathbf{t}_1, \ldots, \mathbf{t}_{s+k}) := M(\mathbf{x}|\mathbf{T}) = M(\mathbf{x}) = \frac{vol_s\left\{\mathbf{y} \in \sigma(\mathbf{y}_1, \ldots, \mathbf{y}_{s+k}) : \mathbf{y}|_{\mathbb{R}^s} = \mathbf{x}\right\}}{vol_{s+k}\sigma(\mathbf{y}_1, \ldots, \mathbf{y}_{s+k})} \tag{3.7}$$

As shown in Dahmen [1980], such "elevation" is always possible for $s+1$ affinely independent elements present in $\mathbf{T}$, since in this case $\mathrm{vol}_s(\sigma(\mathbf{T})) > 0$. The multivariate B-spline defined in (3.7) does not depend on the choice of the $(s+k)$-dimensional simplex but only on the projected vertices $\mathbf{t}_1, \ldots, \mathbf{t}_{s+k}$. Normally, these elevated knots are chosen in such a way that $(s+k)$-dimensional simplex has unit volume. Multivariate B-splines can be alternatively defined as the unique density of

$$\int_{\mathbb{R}^s} M(\mathbf{x}|\mathbf{t}_1, \ldots, \mathbf{t}_{s+k}) f(\mathbf{x}) d\mathbf{x} = \int_{\sigma_{s+k}} f(\mathbf{x}) d\mathbf{y}. \tag{3.8}$$

Thereby B-spline M is unique given some points $\mathbf{t}_1, \ldots, \mathbf{t}_{s+k}$, i.e. it does not depend on σ_{s+k}. Property (3.8) characterizes the B-splines completely in the univariate as well as in the multivariate case (see Dahmen [1980]).

Definitions (3.7) or (3.8) are analytically but not practically attractive. To enable computational application, a recursive relation in the spirit of (3.5) for multivariate B-splines was derived in Micchelli [1980], Dahmen [1980]. The computational aspects of this algorithm were further discussed in Micchelli [1979]. Thus, for $k = 1$ the multivariate B-spline (3.7) is the characteristic function of the simplex $\sigma(\mathbf{t}_1, \cdots, \mathbf{t}_{s+1})$ divided by its volume:

$$M(\mathbf{x}|\mathbf{t}_1, \cdots, \mathbf{t}_{s+1}) = \begin{cases} s! \left| \det \begin{pmatrix} \mathbf{t}_1 & \cdots & \mathbf{t}_{s+1} \\ 1 & \cdots & 1 \end{pmatrix} \right|^{-1}, & \mathbf{x} \in \sigma(\mathbf{t}_1, \cdots, \mathbf{t}_{s+1}) \\ 0, & \mathbf{x} \notin \sigma(\mathbf{t}_1, \cdots, \mathbf{t}_{s+1}) \end{cases}$$

A B-spline of arbitrary order $k > 1$ can be defined through lower order B-splines given any $s+1$ affinely independent points $\mathbf{t}_{i_j} \subset \mathbf{T}$, $j = 1, \ldots, s+1$ using following recurrence relation

$$M(\mathbf{x}|\mathbf{T}) = \frac{k+s}{k} \sum_{j=1}^{s+1} C_{i_j}(\mathbf{x}|\mathbf{t}_{i_1}, \ldots, \mathbf{t}_{i_{s+1}}) M(\mathbf{x}|\mathbf{T} \setminus \{\mathbf{t}_{i_j}\}),\ \mathbf{x} \in \mathbb{R}^s, \tag{3.9}$$

where C_{i_j} are the barycentric coordinates

$$C_{i_j}(\mathbf{x}|\mathbf{t}_{i_1},\ldots,\mathbf{t}_{i_{s+1}}) := \frac{\det\begin{pmatrix} \mathbf{t}_{i_1} & \cdots & \mathbf{t}_{i_{j-1}}, & \mathbf{x}, & \mathbf{t}_{i_{j+1}} & \cdots & \mathbf{t}_{i_{s+1}} \\ 1 & \cdots & 1 & 1 & 1 & \cdots & 1 \end{pmatrix}}{\det\begin{pmatrix} \mathbf{t}_{i_1} & \cdots & \mathbf{t}_{i_{s+1}} \\ 1 & \cdots & 1 \end{pmatrix}}.$$

For $\alpha \in \mathbb{Z}_+^s$, with $|\alpha| = \sum_{i=1}^s \alpha_i$ and $\mathbf{x}^\alpha = \prod_{i=1}^s (\mathbf{x})_i^{\alpha_i}$ we define the linear space of s-variate polynomials of order k

$$\mathbf{\Pi}_{k,s} := \left\{ \sum_{|\alpha|<k} c_\alpha \mathbf{x}^\alpha : c_\alpha \in \mathbb{R}, \mathbf{x} \in \mathbb{R}^s \right\}$$

Multivariate B-splines are polynomials of order k within any region which is enclosed but not intersected by any simplex of s elements ("edge") from $\mathbf{T}$. Furthermore, if every $s+1$ subset of $\mathbf{T}$ is affinely independent, B-spline is $k-2$ times continuously differentiable.

According to (3.9), the value of a multivariate B-spline at $\mathbf{x}$ is a sum of products of barycentric coordinates relative to knots from shrinking subsets of $\mathbf{T}$. Let $(\mathbf{T}_l)_{l=1,\ldots,k}$ be nested sets corresponding to some term in the sum (3.9), such that $\mathbf{T}_l \subset \mathbf{T}$, $\mathbf{T}_{l+1} \subset \mathbf{T}_l$, $|\mathbf{T}_l| = |\mathbf{T}| - l$. The recurrence given in (3.9) stops at $\mathbf{T}_k$, i.e. when it has exactly $s+1$ points in general position. Thereby, a term in the sum (3.9) is non-zero *iff* $\mathbf{x} \in \mathbf{T}_k$. Consider simplex $\sigma(\mathbf{T})$ in the left panel of Figure 3.5 defined through knot points marked in red. All possible subsimplices can be recognized by any three connected knots in the right panel of the same figure.

The recursive relation (3.9) does not stipulate the order in which the knots have to be removed. However, to decrease the computational complexity one could first find all $(s+1)$-simplices which contain the considered point. Left panel of Figure 3.6 demonstrates some random point $\mathbf{x}$ marked in blue, whereas all 2-simplices which contain this point are marked in red. Thus, only those nested subsets of $\mathbf{T}$ correspond to non-zero term in sum (3.9), which end up as one of the marked 2-simplices.

Bivariate B-spline defined through this set of points is a polynomial of order k within regions not crossed by any of these lines. However, imposing those continuity restrictions practically is quite demanding. For example, consider point $\mathbf{y}$ in the right panel of Figure 3.6 which lies in the intersection of two "edges" (0,0)-(4,2) and (2,0)-(4,4). In contrast to

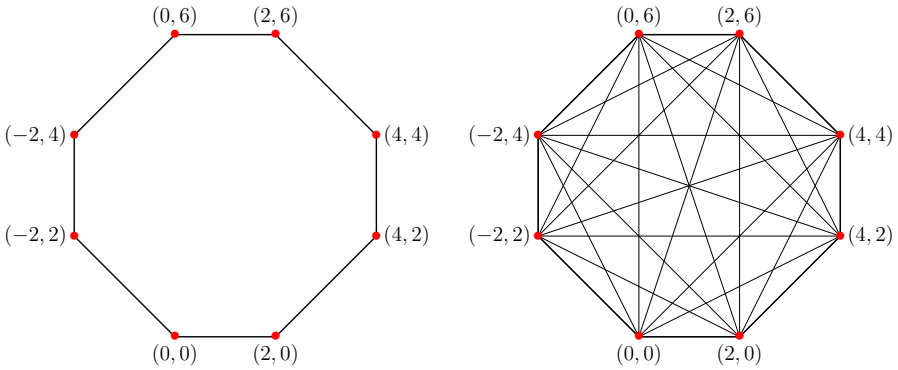

Figure 3.5: An example of 8 affinely independent points on the plane (left) along with all possible subsimplices.

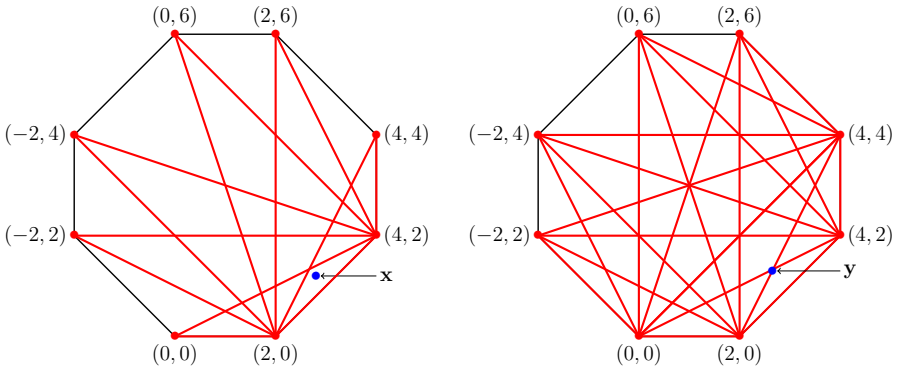

Figure 3.6: Subsimplices containing points within and outside of the intersection of edges.

univariate case, rules of assigning such intersection points to some simplex are much more complex. As mentioned before intersection of two neighboring intervals is considered to be contained in the interval on the right. Here and for higher dimensions a set of highly complex rules has to be designed first. Thus, we exploit the continuity of multivariate B-splines by shifting such problematic points randomly to regions without any intersections. The resulting bivariate B-spline is given in Figure 3.7.

By shifting the simplex as in Figure 3.5 along both dimensions one gets linearly independent B-splines which approximate appropriate continuous functions with any desired degree of accuracy Dahmen [1980]. This approach can be generalized to higher dimensions. In the next section we discuss the aspects of estimation procedure.

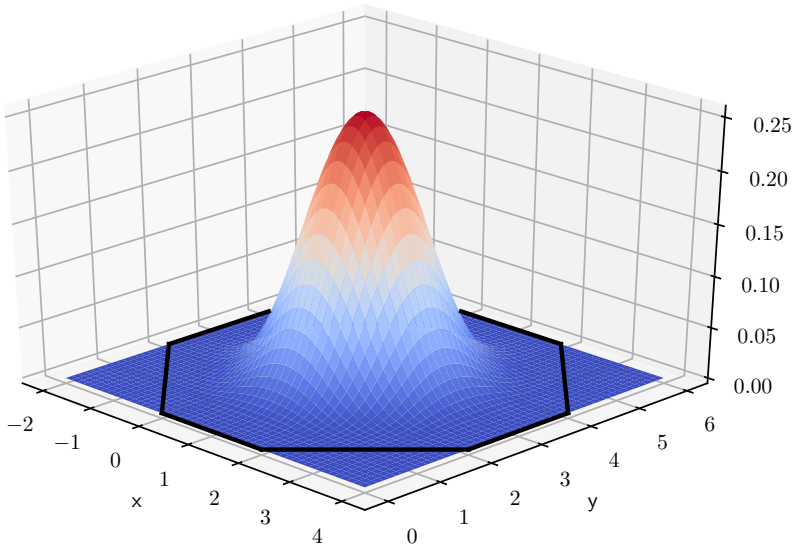

Figure 3.7: Example of two dimensional B-spline with simplex given in Figure 3.5.

3.5 Estimation and forecasting

Anticipating empirical application, the specifications of implemented estimation procedure are now discussed. Three dimensional information filter as in HAR is considered during empirical study. Thereby a three dimensional base and appropriate shifting logic have to be specified. Since B-spline base dictates the order of polynomial approximation and the shape, some commonly used values should be considered. Generally, B-splines of order three or four are used within empirical applications in the univariate case. Thus, a natural choice from different perspectives is a three dimensional cube, which is highlighted with a solid line in Figure 3.9. It defines a polynomial of total order 4 and can be easily scaled and shifted across all dimensions. The illustrated cubes have edge length 2, whereas shifts of magnitude one across all three axes are displayed using dotted lines.

B-spline defined by this 3-dimensional cube is illustrated in Figure 3.8 in dependence of the value of the third variable Z, which is indicated by the plot title. B-spline is almost linear (and zero) for small values of Z whereas it achieves the biggest smoothness at the center of the value interval of Z.

The edge length and number of shifts during empirical application is selected in the following way. For each considered equity index and some time period daily (D), weekly (W)

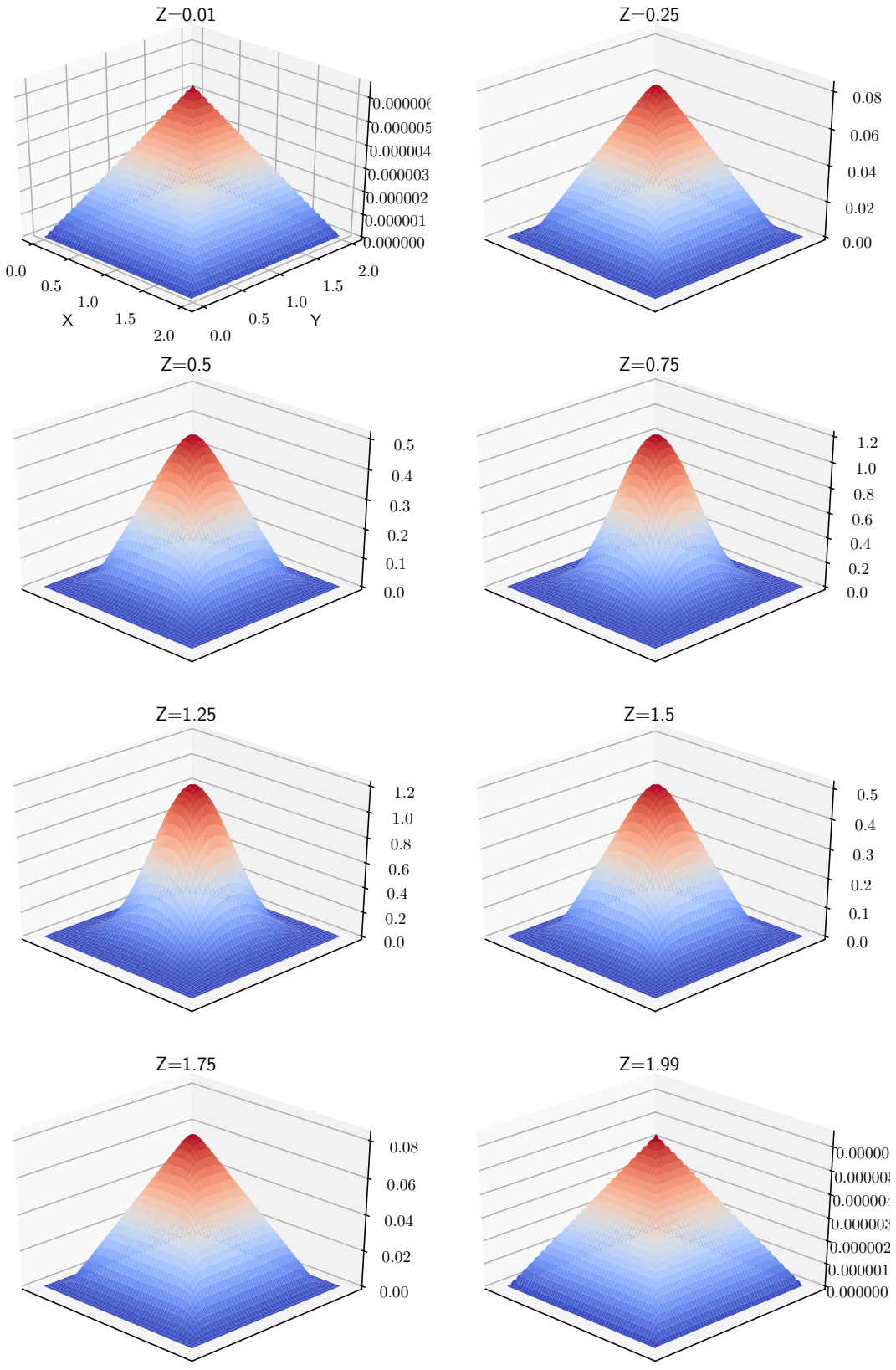

Figure 3.8: The 3-dimensional B-splines defined by the cube with edge length 2. Plot titles indicate different values of the third variable.

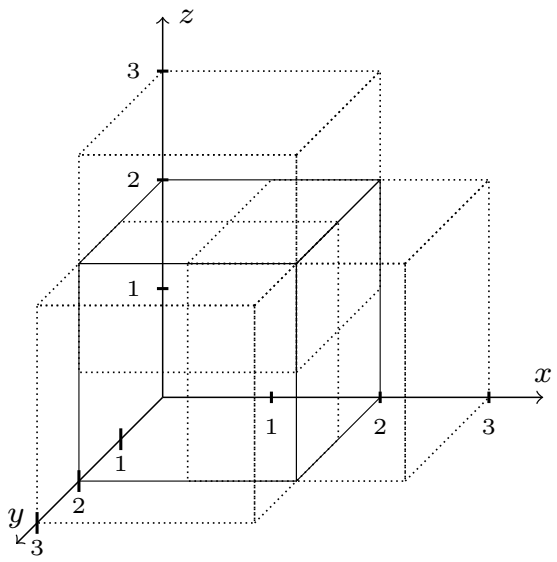

Figure 3.9: Base and integer shifts of the three dimensional B-splines used within the empirical study.

and monthly (M) information as in HAR is calculated. By shifting equal three dimensional cubes, different coverages of the data "domain"

$$[0, \max(D)] \times [0, \max(W)] \times [0, \max(M)]$$

can be produced with varying precision. Further, m will quantify the approximation quality in the following fashion. Thereby, the edge length l of cubes is calculated as

$$l = \frac{\max(\max(D), \max(W), \max(M))}{m}.$$

The cubes are shifted till the whole data domain is covered. The amount of shifted cubes stipulates the precision, with higher count resulting in better in-sample accuracy but also more cubes without any data points.

Once data is transformed through B-splines the proposed approach is basic linear regression model. Thereby, on the left side of equation (3.2) a restricted regression in the spirit of HAR is taken. The coefficients of linear combinations of shifted multivariate B-splines are estimated with least squares approach under no further restrictions. Forecasts are taken as the conditional mean.

The in-sample fit of both models is compared using RMSE, whereas the evaluation of forecasts must be treated differently. Patton [2011] argued, that since RMs are estimators of

volatility, the ranking of forecasts $\hat{x}_t$ based on such volatility proxies x_t generally depends on the choice of loss function L. However, the only ranking available in applications is the so-called feasible ranking which is based on volatility estimators and is generally different from the infeasible one, which is based on the true values. To resolve this problem, Patton [2011] specifies exactly which classes of loss functions are insensitive to empirical imperfections (noise) of the chosen volatility proxy. This class of functions delivers identical rankings of forecasts under the true volatility as well as under the unbiased estimator of it. Probably, the two most prominent examples of loss functions in this class are mean squared error (MSE) and QLIKE, defined as

$$\begin{aligned} \text{MSE} &: (x_t - \hat{x}_t)^2, \\ \text{QLIKE} &: \log \hat{x}_t + \frac{x_t}{\hat{x}_t}, \end{aligned}$$

whereas the latter is less sensitive to extreme observations due to its design. Bollerslev et al. [1994] contains additional discussions and examinations of QLIKE.

Furthermore, to test whether differences in QLIKE of several competing models are statistically significant, Patton [2011] also proposed to compare their predictive ability using Diebold-Mariano-West (DMW) test (Diebold and Mariano [1995], West [1996]). Generally, the use of QLIKE instead of MSE is advocated since the moment conditions required for DMW test under QLIKE are far weaker than those for the MSE. More importantly, DMW in this specification was found to have more power (see Patton and Sheppard [2009]). Given two competing forecasts $\hat{x}_{1t}$ and $\hat{x}_{2t}$ the null hypothesis

$$H_0 \;:\; \mathbb{E}\left[L(\hat{x}_{1t}, x_t) - L(\hat{x}_{2t}, x_t)\right] = 0,$$

is tested against the case when this (expected) difference is either positive of negative.

As a further comparative criteria, the mean interval score (MIS) defined by Gneiting and Raftery [2007] as:

$$\text{MIS}_\alpha\left(\boldsymbol{l}, \boldsymbol{u}; \boldsymbol{x}\right) = \frac{1}{T}\sum_{t=1}^{T}\left[(u_t - l_t) + \frac{2}{\alpha}\left(l_t - x_t\right)\mathbf{1}_{x_t<l_t} + \frac{2}{\alpha}\left(x_t - u_t\right)\mathbf{1}_{u_t<x_t}\right],$$

is considered. Thereby, T is the number of out-of-sample forecasts, l_t and u_t are the lower and upper bounds of $100\left(1-\alpha\right)\%$ confidence intervals. Lower values of MIS_α are associated with better forecasts, whereas quantities $\frac{2}{\alpha}\left(l_t - x_t\right)\mathbf{1}_{x_t<l_t}$ and $\frac{2}{\alpha}\left(x_t - u_t\right)\mathbf{1}_{u_t<x_t}$ serve as penalization terms when true values lie outside of the interval.

3.6 Empirical study

Within this empirical study shifted multivariate B-splines are applied to model volatility behavior of equity indices, whereas the in- and out-of-sample performance are analyzed and compared. B-spline regression is estimated as described previously based on the information sets as in HAR, whereas the latter model is the natural benchmark. The considered data sample consists of 22 international equity indices and spans from January 2000 to December 2018. The information about individual index tickers is listed in Table 3.1.

Ticker	Name	Ticker	Name
AEX	AEX index	IXIC	Nasdaq 100
AORD	All Ordinaries	KS11	Korea Composite Stock Price Index
BFX	Bell 20 Index	KSE	Karachi SE 100 Index
BSES	S&P BSE Sensex	MXX	IPC Mexico
BVSP	BVSP BOVESPA Index	N225	Nikkei 225
DJI	Dow Jones Industrial Average	NSEI	NIFTY 50
FCHI	CAC 40	RUT	Russel 2000
FTSE	FTSE 100	SPX	S&P 500 Index
GDAX	DAX	SSEC	Shanghai Composite Index
HSI	HANG SENG Index	SSMI	Swiss Stock Market Index
IBEX	IBEX 35 Index	STOX	EURO STOXX 50

Table 3.1: Equity indices considered through empirical study.

The specific volatility proxy taken for modeling is 5-minute BPV as provided by Oxford-Man database. Overview of the empirical characteristics for each equity index is reported in Table 3.2. Thereby sample average, standard deviation, median and 95% quantile of each time series were calculated. Generally, separate markets show different trading activity but since 5-minute BPV is taken those differences should cancel out. Thus, the most volatile markets are Chinese SSEC and Indian BSESN according to average and standard deviation of BPV, respectively. The former also has the highest 95% quantile and second highest median of all time series. On the other end of the volatility spectrum, is the Australian AORD, which is probably the calmest market according to data.

First, HAR and proposed model for different degrees of accuracy are estimated on all available data and in-sample fit is compared for each index. The precision of B-spline approximation hinges on the amount of shifted cubes, whereas higher count results in better in-sample accuracy.

	Avg	Std	Med	95%		Avg	Std	Med	95%
AEX	10.9	18.7	5.4	37.4	IXIC	11.9	23.0	5.2	42.0
AORD	4.2	7.2	2.3	13.1	KS11	11.3	21.0	5.3	37.8
BFX	8.0	12.4	4.4	25.2	KSE	8.9	13.4	4.7	31.6
BSES	13.5	27.6	6.7	44.3	MXX	6.2	9.8	3.5	19.1
BVSP	13.7	21.2	8.8	37.4	N225	9.1	16.2	5.5	25.5
DJI	8.4	20.3	3.8	26.5	NSEI	11.0	26.7	5.3	34.5
FCHI	12.6	20.9	7.0	39.7	RUT	6.6	14.8	3.2	21.0
FTSE	8.9	20.7	4.4	27.7	SPX	8.8	20.5	3.9	28.7
GDAX	14.5	25.0	7.5	51.1	SSEC	15.5	26.4	7.4	55.1
HSI	8.7	14.8	5.0	24.8	SSMI	7.7	14.8	3.8	26.4
IBEX	13.1	18.6	8.4	38.4	STOX	13.2	22.6	7.0	43.5

Table 3.2: Summary statistics ($\times 10^5$) of considered equity indices.

Thus, B-splines with $m = 5, 10, 20, 30, 40, 50$ as described previously are taken. The linear combinations of shifted multivariate B-splines along with information as in HAR are estimated with least squares approach under no further restrictions. As stated previously, the model fit is assessed using RMSE. The results are reported for HAR and B-splines of different precision indicated by column name in Table 3.3. Already $m = 5$ yields better in-sample fit than HAR, whereas best results are achieved for highest considered amount of shifted cubes. The best fit within this group of equity indices was observed for the least volatile AORD, whereas the worst results are associated with BSES, NSEI and SSEC, which are most volatile according to standard deviation. It is clear that adding further cubes will only increase the accuracy of approximation.

We further analyze the estimated coefficients of both approaches. Since daily, weekly and monthly information is included in both HAR and proposed model we analyze the differences to assess the impact of the additional information. Here the average coefficients over all equity indices are calculated for HAR and shifted B-splines whereas the results are reported in Table 3.4. The importance of daily information decreases mildly within proposed model with increasing precision, whereas the impact of weekly data on the contrary becomes stronger.

	5	10	20	30	40	50	HAR
AEX	10.9	6.7	4.7	4.1	3.5	**3.1**	14.5
AORD	2.3	2.0	1.5	1.4	1.2	**1.1**	3.1
BFX	5.5	4.6	3.3	2.6	2.4	**1.8**	7.5
BSES	32.8	28.6	21.5	20.2	13.7	**11.1**	52.1
BVSP	10.7	9.8	8.9	8.1	7.2	**6.7**	20.1
DJI	12.6	7.7	7.5	7.0	6.4	**5.5**	20.4
FCHI	15.7	11.0	8.3	6.5	5.8	**4.3**	21.1
FTSE	20.5	16.3	14.7	13.3	11.3	**7.6**	26.8
GDAX	21.0	16.3	10.2	9.3	8.0	**6.7**	28.5
HSI	9.3	5.8	5.2	4.1	3.5	**2.9**	11.1
IBEX	14.8	9.3	8.1	7.0	5.9	**4.9**	18.2
IXIC	22.4	18.3	10.1	7.0	5.5	**4.5**	26.3
KS11	10.0	8.7	8.5	6.5	5.2	**4.7**	18.4
KSE	8.4	7.5	5.4	4.5	3.6	**2.8**	10.2
MXX	5.0	4.5	4.1	3.8	3.5	**3.2**	6.0
N225	10.5	7.3	5.6	5.3	4.6	**4.2**	14.5
NSEI	29.8	23.6	15.7	13.8	13.0	**12.4**	49.1
RUT	7.8	3.7	3.4	3.0	2.7	**2.5**	11.8
SPX	10.4	6.3	5.7	5.2	4.9	**4.2**	18.5
SSEC	26.6	22.9	21.9	16.7	12.3	**11.5**	37.6
SSMI	9.6	7.7	5.3	4.9	4.3	**4.0**	11.5
STOX	18.5	12.6	10.9	7.6	6.8	**6.1**	26.5

Table 3.3: In-sample RMSE ($\times 10^9$) for HAR and B-spline regression estimated on the full sample with different degrees of accuracy. The best result for each index is marked in bold.

	5	10	20	30	40	50	HAR
D	0.49	0.48	0.44	0.46	0.45	0.44	0.33
W	0.28	0.35	0.37	0.41	0.43	0.42	0.39
M	0.14	0.13	0.14	0.12	0.12	0.13	0.22

Table 3.4: Average coefficients over all equity indices corresponding to daily (D), weekly (W) and monthly (M) information as in the proposed model and HAR.

As illustrated in Figure 3.1, plain linear regression is not capable of reproducing local behavior. Subplot with HAR errors is presented in Figure 3.10 along with similar plots for B-spline regression with $m = 5, 20, 50$. These results indicate that B-spline regression with $m = 5$ is similar to the model without B-splines, whereas the fit is better for bigger m.

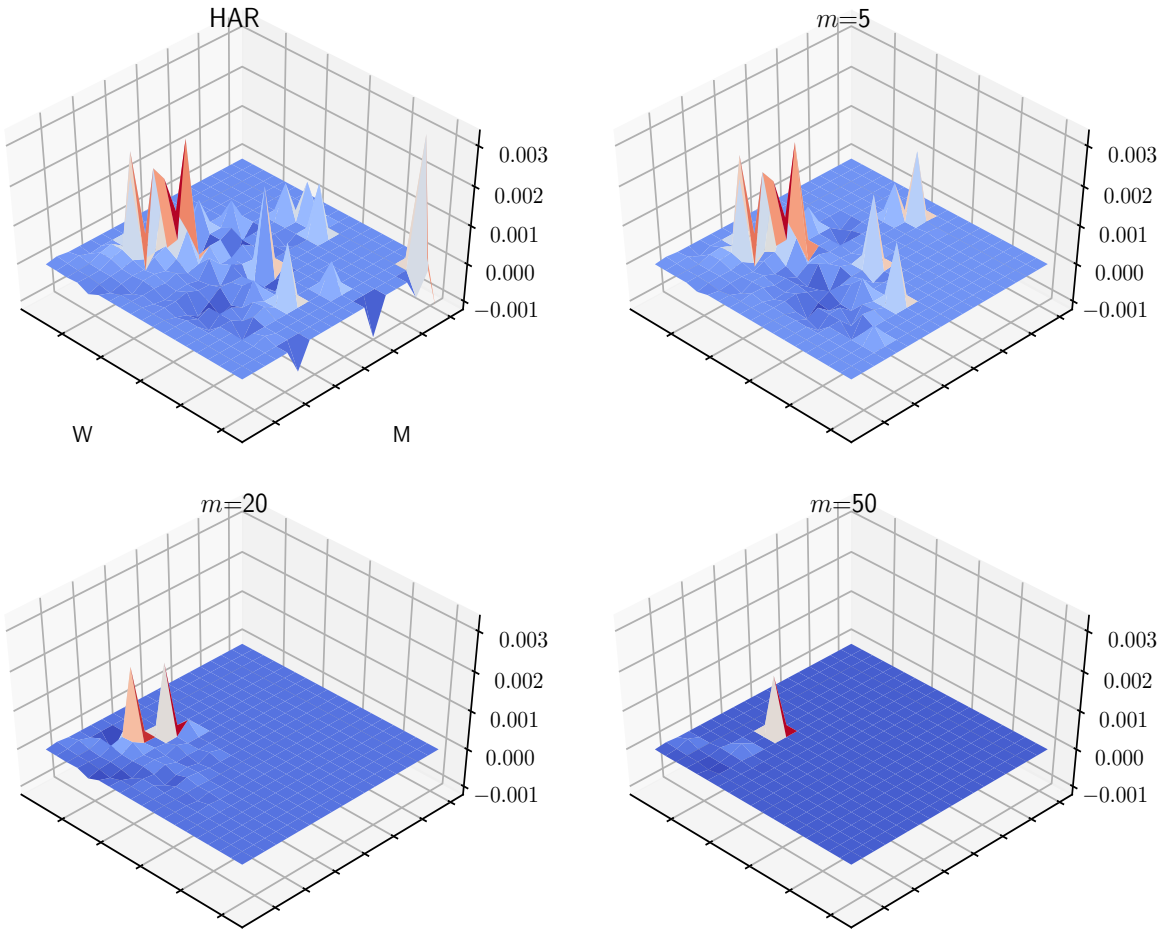

Figure 3.10: Average residuals calculated as described earlier. Results for HAR model are presented in top left subplot and for B-spline regression in the remaining subplots, whereas title indicates the precision.

Nevertheless, the discussed coefficients of HAR and B-spline regression are only partly comparable, since they quantify the linear response to changes in variables for the former model. A better measure of comparison would be marginal effects, defined as sensitivities of forecasts with respect to each explanatory variables. Given some observed information up to time point t, this measure is calculated as the deviation in forecasts divided by change in input variables, i.e. it is a partial derivative of the forecasting function. In case of linear models marginal effects are equal to regression coefficients. Figure 3.11 demonstrates marginal effects for daily, weekly and monthly variable and $m = 5, 50$ in the form of boxplots. The response to daily and weekly information becomes more diverse with increasing m due to expanding locality. Furthermore, the influence of monthly information stays relatively subdued.

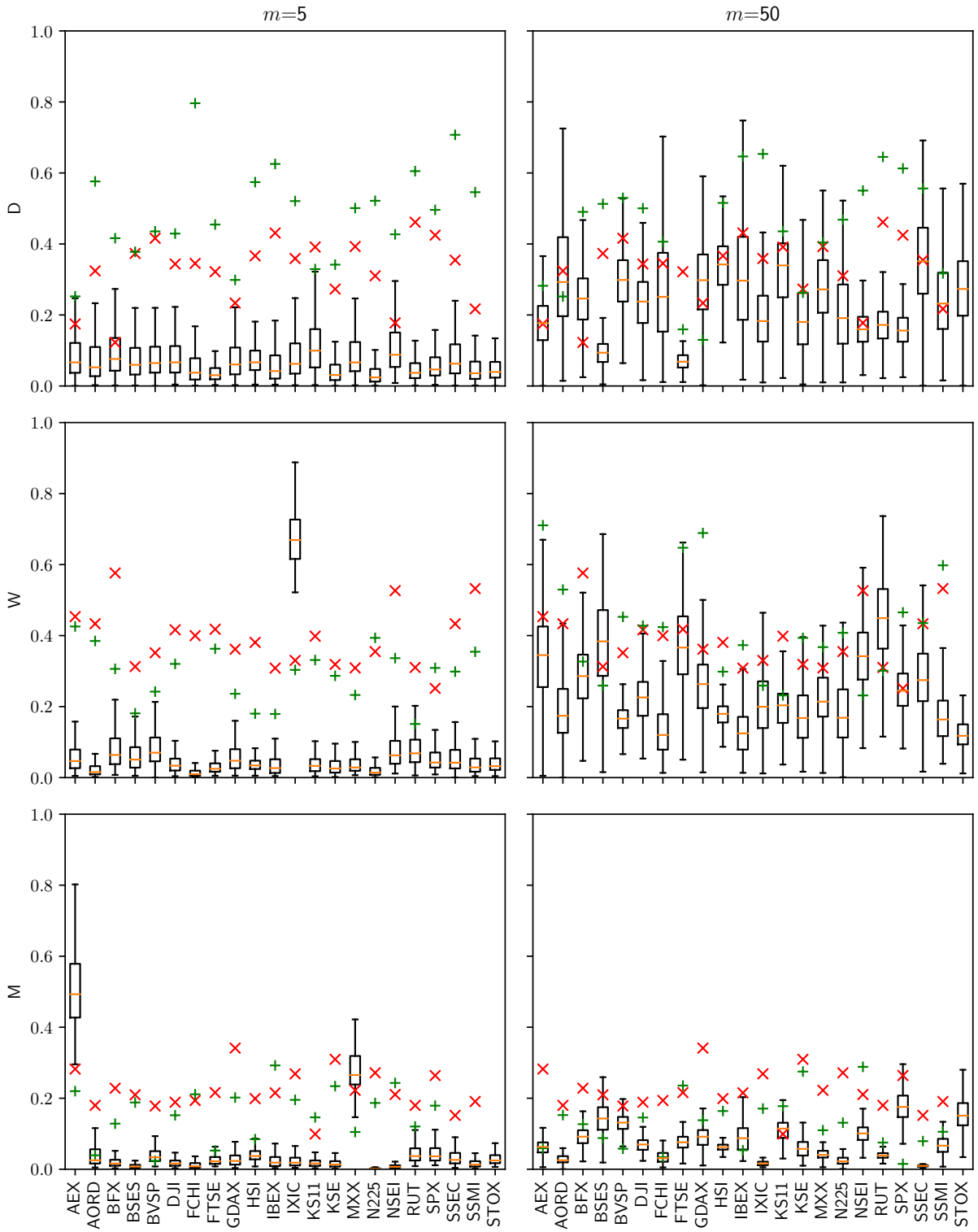

Figure 3.11: Boxplots of marginal effects for different variables (row) and two degrees of accuracy $m = 5, 50$ (column). Whiskers indicate 1.5 standard deviation and outliers are omitted for illustration purposes. Red x markers indicate coefficients of plain HAR, whereas green crosses - the coefficients of the HAR informational filters as in the proposed model.

Now we proceed to forecasting application. Thereby, B-spline regression and HAR are re-estimated monthly using the whole available information to date. One-step-ahead forecasts are done for the following month based on these parameter estimates. Considered time period starts on the first and ends on the last trading day of 2018. The results are reported as average QLIKE losses in Table 3.5. All cases in which forecasts associated with B-spline regression were better than those of HAR are marked in bold and additionally in italics, if the difference was statistically significant. B-spline regression was consistently better than HAR with as little as $m = 5$, whereas the proposed approach was significantly better for 4 tickers, including most important US equity indices. The 95% quantile of out-of-sample QLIKEs and MIS with $\alpha = 5\%$ are reported in Tables 3.6a and 3.6b. Here, in 3 and 4 cases, respectively, HAR delivered better results. Generally, the empirical findings suggest that $m = 5$ or 10 should be an appropriate number.

	5	10	20	30	40	50	HAR
AEX	**0.1519**	**0.1515**	**0.1518**	**0.1552**	**0.153**	**0.1541**	0.157
AORD	**0.2319**	**0.2375**	**0.257**	0.2752	**0.265**	**0.2521**	0.2703
BFX	**0.1153**	**0.115**	0.1206	0.1219	0.1244	**0.1187**	0.1205
BSES	0.5124	**0.47**	0.5623	0.5607	0.5231	**0.4692**	0.49
BVSP	**0.1363**	**0.1344**	**0.139**	**0.1398**	**0.1404**	**0.1407**	0.1448
DJI	**0.2411**	**0.2315**	0.3373	0.3463	0.3909	**0.2357**	0.28
FCHI	**0.1562**	**0.1549**	0.1688	0.1675	0.165	**0.1547**	0.1596
FTSE	**0.1738**	**0.1736**	0.1771	0.1772	0.1797	**0.1596**	0.176
GDAX	**0.1304**	**0.1284**	**0.1314**	**0.1293**	**0.1315**	**0.1304**	0.1362
HSI	***0.1194***	***0.118***	**0.1211**	***0.1186***	***0.1183***	**0.1251**	0.1297
IBEX	**0.1143**	**0.1073**	0.1269	0.1251	0.132	**0.1092**	0.1148
IXIC	**0.2153**	***0.2072***	**0.2256**	**0.2342**	**0.2343**	**0.2332**	0.2363
KS11	**0.1013**	***0.1013***	0.1135	0.1163	0.1159	**0.1056**	0.1099
KSE	**0.1743**	**0.1734**	**0.1797**	**0.1858**	0.1925	**0.1829**	0.1873
MXX	***0.1476***	***0.1487***	**0.1536**	**0.1543**	0.1754	**0.149**	0.1645
N225	**0.193**	**0.1917**	**0.2035**	**0.2005**	**0.1998**	**0.2028**	0.2161
NSEI	**0.2261**	**0.2194**	0.2734	0.2674	0.2585	**0.22**	0.2461
RUT	***0.1612***	***0.1589***	**0.2005**	**0.1889**	***0.178***	***0.1651***	0.2046
SPX	***0.2138***	***0.2191***	0.3213	0.3378	0.3325	**0.2303**	0.2561
SSEC	**0.1464**	**0.151**	**0.1522**	0.1616	0.1686	**0.1448**	0.1553
SSMI	**0.087**	**0.0844**	0.0914	0.0908	0.0919	**0.0869**	0.0878
STOX	**0.1718**	**0.1679**	0.1787	0.178	0.1856	**0.1644**	0.1757

Table 3.5: Average QLIKE of forecasts. Results for B-spline regression which are better than HAR are marked in bold, whereas statistically significant difference is highlighted in italics.

	5	10	20	30	40	50	HAR
AEX	**0.509**	**0.504**	**0.472**	**0.45**	**0.515**	0.564	0.525
AORD	**0.612**	**0.639**	**0.664**	**0.686**	**0.659**	**0.647**	0.709
BFX	0.409	**0.383**	**0.373**	**0.397**	**0.381**	**0.387**	0.4
BSES	0.616	**0.537**	0.743	0.735	0.633	**0.498**	0.582
BVSP	**0.455**	**0.423**	**0.464**	**0.443**	**0.427**	**0.472**	0.56
DJI	**0.576**	**0.599**	**0.725**	**0.74**	**0.748**	**0.676**	0.842
FCHI	**0.482**	**0.498**	**0.492**	**0.508**	**0.483**	**0.542**	0.553
FTSE	**0.586**	**0.699**	**0.555**	**0.549**	**0.563**	**0.531**	0.733
GDAX	**0.396**	**0.379**	**0.427**	**0.393**	**0.418**	**0.394**	0.482
HSI	0.46	0.47	0.463	0.456	0.465	0.53	**0.436**
IBEX	**0.376**	**0.368**	**0.328**	**0.337**	**0.34**	**0.369**	0.443
IXIC	**0.859**	**0.679**	**0.623**	**0.627**	**0.642**	**0.74**	0.918
KS11	**0.345**	**0.357**	**0.4**	0.415	0.459	**0.4**	0.415
KSE	**0.614**	**0.682**	**0.614**	**0.558**	**0.634**	**0.592**	0.853
MXX	**0.449**	**0.452**	**0.453**	**0.427**	**0.549**	**0.472**	0.554
N225	**0.534**	**0.563**	**0.565**	**0.568**	**0.557**	**0.704**	0.784
NSEI	0.529	0.498	0.643	0.606	0.568	0.487	**0.398**
RUT	**0.563**	**0.57**	**0.601**	**0.625**	**0.555**	**0.57**	0.79
SPX	**0.553**	**0.718**	**0.703**	**0.8**	**0.755**	**0.796**	0.933
SSEC	0.528	0.546	0.563	0.578	0.552	0.529	**0.474**
SSMI	**0.304**	**0.273**	**0.311**	0.324	**0.303**	**0.292**	0.314
STOX	0.692	**0.668**	0.736	**0.592**	**0.585**	**0.642**	0.669

(a) 95% quantile of QLIKE of forecasts.

	5	10	20	30	40	50	HAR
AEX	**0.236**	**0.204**	**0.188**	**0.19**	**0.205**	**0.228**	0.265
AORD	**0.18**	0.187	**0.182**	0.198	0.198	0.213	0.184
BFX	**0.178**	**0.176**	**0.174**	**0.16**	**0.149**	**0.158**	0.201
BSES	0.442	0.442	**0.406**	**0.388**	**0.368**	**0.335**	0.44
BVSP	**0.357**	**0.352**	**0.358**	**0.342**	**0.336**	**0.388**	0.389
DJI	**0.596**	0.606	0.737	0.784	0.901	0.648	0.6
FCHI	**0.261**	**0.244**	**0.242**	**0.231**	**0.216**	**0.217**	0.287
FTSE	0.237	**0.228**	**0.21**	**0.209**	**0.205**	**0.191**	0.235
GDAX	**0.272**	**0.27**	**0.259**	**0.242**	**0.248**	**0.244**	0.33
HSI	**0.208**	**0.202**	**0.197**	**0.2**	**0.197**	0.283	0.222
IBEX	**0.289**	**0.277**	**0.274**	**0.27**	**0.265**	**0.26**	0.307
IXIC	0.685	0.713	0.676	0.733	0.782	0.808	**0.621**
KS11	**0.23**	**0.225**	**0.224**	**0.215**	**0.192**	**0.179**	0.27
KSE	**0.261**	**0.246**	**0.237**	**0.222**	**0.21**	**0.22**	0.294
MXX	0.251	0.25	0.289	0.263	0.32	0.319	**0.244**
N225	**0.291**	**0.287**	**0.288**	**0.285**	**0.278**	0.32	0.297
NSEI	0.457	0.439	**0.404**	**0.384**	**0.389**	**0.36**	0.409
RUT	0.327	0.328	0.46	0.433	0.443	0.423	**0.319**
SPX	0.525	0.548	0.621	0.708	0.689	0.694	**0.514**
SSEC	**0.436**	**0.422**	**0.401**	**0.388**	**0.364**	**0.353**	0.5
SSMI	**0.149**	**0.139**	**0.136**	**0.129**	**0.122**	**0.12**	0.176
STOX	**0.309**	**0.295**	**0.272**	**0.263**	**0.249**	**0.229**	0.343

(b) MIS ($\times 10^3$) with $\alpha = 5\%$.

Table 3.6: Further criteria for comparing out-of-sample forecasts of HAR and B-splines with different degrees of accuracy. The best result for each index is marked in bold.

3.7 Summary

In this paper a non-parametric regression based on B-splines is proposed. Technically, presented model recovers the true functional link through a set of B-splines by a classic estimation method. The proposed approach is applied within volatility modeling framework. Popular models in this field are basically constrained linear regressions, which are by design incapable of reproducing non-linearities or accounting for local behavior. The modeling approach suggested in this paper extends or includes all popular frameworks as special cases. HAR model, which is the most prominent of such methods, is taken as benchmark. The proposed model is compared with this approach within an extensive empirical study on time series of bi-power variation of 22 equity indices, whereas explanatory variables are selected as in HAR model. The presented method delivers better results within in-sample as well as statistically better forecasts.

Bibliography

Yacine Aït-Sahalia and Jean Jacod. *High-frequency financial econometrics*. Princeton University Press, 2014.

Francesco Audrino and Peter Bühlmann. Splines for financial volatility. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 71(3):655–670, 2009.

Ole E. Barndorff-Nielsen and Neil Shephard. Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics*, 2(1):1–37, 2004.

Tim Bollerslev and Dan Jubinski. Equity trading volume and volatility: Latent information arrivals and common long-run dependencies. *Journal of Business & Economic Statistics*, 17(1):9–21, 1999.

Tim Bollerslev, Robert F. Engle, and Daniel B. Nelson. ARCH models. *Handbook of econometrics*, 4:2959–3038, 1994.

Kung-Sik Chan and Howell Tong. On the use of the deterministic Lyapunov function for the ergodicity of stochastic difference equations. *Advances in Applied Probability*, 17(3): 666–678, 1985.

William S. Cleveland. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74(368):829–836, 1979.

Fulvio Corsi. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196, 2009.

Haskell Curry and Isaac Schoenberg. On Pólya frequency functions IV: The fundamental spline functions and their limits. *Journal d'Analyse Mathématique*, 17:71–107, 1966.

Wolfgang Dahmen. On multivariate B-splines. *SIAM Journal on Numerical Analysis*, 17(2): 179–191, 1980.

Carl de Boor. Splines as linear combinations of B-splines. A Survey. In *Approximation Theory II*, pages 1–47. Academic Press (New York), 1976.

Carl de Boor. *A Practical Guide to Splines*. Applied Mathematical Sciences. Springer, 2001.

Francis X. Diebold and Roberto S. Mariano. Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263, 1995.

Robert F. Engle and Jose Rangel. The spline GARCH model for unconditional volatility and its global macroeconomic causes. *Review of Financial Studies*, 21, 01 2006.

Jerome H. Friedman. Multivariate adaptive regression splines. *The Annals of Statistics*, 19 (1):1–67, 03 1991.

Eric Ghysels, Pedro Santa-Clara, and Rossen Valkanov. The MIDAS touch: Mixed data sampling regression models. *UCLA: Finance*, 2004.

Tilmann Gneiting and Adrian E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.

Klaus Höllig and Jörg Hörner. *Approximation and Modeling with B-Splines*. Society for Industrial and Applied Mathematics, 2013.

David A. Jones. Nonlinear autoregressive processes. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 360(1700):71–95, 1978.

Vladimir Katkovnik. Linear and nonlinear methods of nonparametric regression analysis. *Automatica*, 12, 09 1979.

Peter A. W. Lewis and James G. Stevens. Nonlinear modeling of time series using multivariate adaptive regression splines (MARS). *Journal of the American Statistical Association*, 86 (416):864–877, 1991.

Lily Y. Liu, Andrew J. Patton, and Kevin Sheppard. Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. *Journal of Econometrics*, 187(1):293–311, 2015.

Charles Micchelli. On a numerically efficient method for computing multivariate B-splines. In Karl Zeller Walter Schempp, editor, *Multivariate Approximation Theory*, pages 211–248. Birkauser, 1979.

Charles A. Micchelli. A constructive approach to Kergin interpolation in r^k: Multivariate B-splines and lagrange interpolation. *Rocky Mountain J. Math.*, 10(3):485–497, 09 1980.

Andrew J. Patton. Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160(1):246–256, 2011.

Andrew J. Patton and Kevin Sheppard. Evaluating volatility and correlation forecasts. In *Handbook of financial time series. With a foreword by Robert Engle*, pages 801–838. Springer, 2009.

Isaac Schoenberg. Contributions to the problem of approximation of equidistant data by analytic functions. *Quarterly of Applied Mathematics*, 4(2):112–141, July 1946.

Isaac Schoenberg. *Inequalities I*, chapter On spline functions, pages 255–291. Academic Press (New York), 1967.

Charles J. Stone. Consistent nonparametric regression. *Ann. Statist.*, 5(4):595–620, 07 1977.

George Tauchen and Mark Pitts. The price variability-volume relationship on speculative markets. *Econometrica*, 51(2):485–505, 1983.

Howell Tong. *Non-linear Time Series: A Dynamical System Approach*. Oxford University Press, 1990.

Aleksandr B. Tsybakov. Robust reconstruction of functions by the local approximation. *Problems of Information Transmission*, 22, 04 1986.

Kenneth D. West. Asymptotic inference about predictive ability. *Econometrica*, 64:1067–1084, 1996.

Paolo Zaffaroni. Contemporaneous aggregation of linear dynamic models in large economies. *Journal of Econometrics*, 120(1):75–102, 2004.

Chapter 4

Article 3: Matrix variate factor model with application to realized covariance matrices[1]

Abstract

We propose a matrix variate dynamic factor model for time series of symmetric positive definite matrices. The latent factors are matrices with a specific AR-type dynamics. The suggested approach ensures symmetricity and positive definiteness by design. Furthermore, it relies on appropriate distributional assumptions, preserves interpretability and supports data analysis through introduction of latent factors. The proposed estimation procedure based on EM algorithm consists of a deterministic sequence of simple matrix operations, thus making the model attractive from numerical perspective. We further show the asymptotic consistency of factors estimates under increasing dimension of data. The proposed model along with two benchmarks is evaluated within an extensive empirical study both in- and out-of-sample. The results reveal the advantages of the proposed approach.

Keywords: latent factor, Wishart distribution, EM, realized volatility

[1]Submitted to *Journal of Econometrics* (2020)

4.1 Introduction

Informational streams in different fields of science are inherently multivariate, since normally every observation is characterized by several variables (attributes). Consequently, multivariate time series analysis evolved as an active field of research within statistics and delivered a multitude of well-studied methods (see Tsay [2013]). Thereby, the functional dependencies between variables as well as the temporal dependencies are modeled. Since the estimation of such models is basically a high-dimensional numerical problem, different ways of reducing computational complexity were proposed. Probably the most popular general approach is the introduction of factor structure (see Bai and Ng [2002], Forni et al. [2000]). Thereby the observed information is split into a relatively much smaller group of (latent) factors which are assumed to explain most of the data dynamics and an (uncorrelated) idiosyncratic part. The potential presence of autocorrelation within idiosyncratic part leads to limited separation of both components. Such models are only asymptotically identifiable, due to the model assumptions about factors, which ensure separation of common and individual components with increasing dimension of data. The introduction of factors within econometric applications is attractive from several perspectives. Firstly, in this way the number of parameters is reduced. Secondly, in contrast to naive simplification of model parameterization by disregarding some dependencies completely, the factor approach simplifies the dependencies in a well motivated way by considering some common drivers of change. Lastly, the interpretation of parameters and (latent) factors delivers a powerful way of analyzing observed data.

Many observations within econometric applications consist of several variables with same set of attributes, thus, the observed information is in matrix form. In such cases, dependencies exist not only between columns but also between rows. For example, the option prices for different strikes and maturities given as a time series are basically random matrices observed over time. The typical approach to modeling matrix variate time series is by vectorizing each matrix and applying some well-studied methods like factor analysis to vectors. However, vectorization destroys the inherent structure of data by disregarding the dependencies between rows or columns. Since matrix variate time series are the basic form of economic information delivery, modeling such time series is becoming an active field of research. Walden and Serroukh [2002] introduced the concept of matrix variate time series with an engineering application. To the best of our knowledge Wang et al. [2019] were the first

to consider a matrix factor model within econometric application. The method is a two-step hierarchical model, where intermediate factors are first extracted from each row of observed data separately, whereas in the next step factors are extracted from each column of intermediate factors. Finally, the model is assembled. Due to the generality of assumptions on data and factors, which can be rectangular, Chen et al. [2020] proposed a method to incorporate a priori information about data, by introducing constraints on the model specification.

In this paper we introduce a new multivariate approach to modeling symmetric positive definite matrices based on latent factors, with a specific application to realized covariance matrices in mind. The proposed factor model is based on appropriate distributional assumptions and thus ensures positive definiteness by design. Furthermore, the analytical properties of the model are easily accessible whereas the interpretability of parameters is very straightforward. A brief overview of the existing methods to model realized covariance matrices reveals that in all cases the general design must be simplified for numerical purposes by disregarding almost all cross dependencies, for example Bollerslev et al. [1988], Engle and Kroner [1995], Bollerslev [1990], Engle [2002]. To ensure positive definiteness, a popular method is to transform matrices with Cholesky decomposition or matrix logarithm, vectorize the resulting matrices and model the entries separately, see Chiriac and Voev [2011] and Bauer and Vorkink [2011], respectively. However, this introduces bias and complicates drastically the interpretation of parameters. Furthermore, matrix transformations like Cholesky decomposition is unattractive from portfolio allocation point of view, since the results of modeling depend on the order of assets. Finally, only several of the methods proposed so far (see Gouriéroux et al. [2009], Golosnoy et al. [2012]) rely on distributional assumptions. The model we propose ensures positive definiteness at every point in time, simplifies by introducing factor structure and is invariant to asset order. In the empirical study we compare our model to ARFIMA-Cholesky approach and the method proposed in Tao et al. [2011]. The authors elaborated on PCA for matrix processes, whereas the dynamics of factors was proposed to be modeled with VAR.

The next Section introduces the approach we propose and the theoretical properties are discussed. Section 4.3 outlines the estimation procedure. A simulation study is performed in Section 4.4 followed by an empirical study in Section 4.5.

4.2 Matrix variate factor model

This section introduces the proposed factor model and elaborates on its properties. We denote the observed information with Y_t and (unobserved) matrix variate factor process with F_t, respectively. Furthermore, the notion of Wishart distribution given in the following definition is essential within this paper.

Definition 4 (Wishart distribution)**.** *If $X_i \sim N(0, \Sigma), \Sigma \in \mathbb{R}^{p\times p}$ are independent and multivariate normally distributed variables for $i = 1, \ldots, n$, then $\sum_{i=1}^{n} X_i X_i' \sim W_p(n, \Sigma)$ is called Wishart distribution.*

Thus, a Wishart distributed variable is a random symmetric positive definite (spd) matrix, whereas p identifies the dimension and n - the number of independent and identically distributed (*iid*) normal vectors in the sum. To model the behavior of a time series of symmetric positive definite matrices, we propose the matrix variate state-space process with factor structure given in Definition 5.

Definition 5. *A time series of symmetric positive definite matrices $\{Y_t\} \in \mathbb{R}^{p\times p}$ is a matrix variate factor model if for $\{F_t\} \in \mathbb{R}^{q\times q}$ it satisfies*

$$Y_t = \Lambda F_t \Lambda' + \varepsilon_t \tag{4.1}$$

$$F_t = A F_{t-1} A' + u_t, \tag{4.2}$$

where for $q \ll p$

$\Lambda \in \mathbb{R}^{p\times q}$ *is the matrix of factor loadings,*

$\varepsilon_t \sim W_p(n, \Sigma_\varepsilon)$ *are iid shocks with $\Sigma_\varepsilon \in \mathbb{R}^{p\times p}$,*

$A \in \mathbb{R}^{q\times q}$ *is a quadratic full-rank matrix,*

$u_t \sim W_q(n, \Sigma_u)$ *are iid shocks with $\Sigma_u \in \mathbb{R}^{q\times q}$.*

The general idea of our factor model is that the observable process Y_t can be split into two orthogonal parts: the autocorrelated common component of considerably lower dimension, represented by the latent matrix variate process $\{F_t\}$, i.e. factors, and idiosyncratic not autocorrelated part, represented by *iid* ε_t. In general terms of factor modeling, equation (4.1) is

called *observation equation*, which describes the relationship between observable information Y_t and (latent) factors, whereas (4.2) is called *state equation* and it determines the dynamics of (unobservable) factors. The singular values of Λ determine the strength of the common component, whereas its cross-time persistence is measured through eigenvalues of A.

To the best of our knowledge, the continuous version of the process as in the state equation (4.2) was first considered in Bru [1991]. Later Gouriéroux et al. [2009] defined the process in the discrete time through conditional Laplace transform and named it Wishart Autoregressive process (WAR). The conditional distribution of the latent factor process is a non-central Wishart

$$F_t \mid F_{t-1} \sim W_q(n, \Sigma_u, \Sigma_u^{-1} A F_{t-1} A').$$

The following lemma describes the stationary distribution of the process.

Lemma 1. *The stationary distribution of the factors $\{F_t\}$ is given by the Wishart distribution $W_q(n, F)$ where*

$$F = \sum_{i \geq 0} A^i \Sigma_u (A')^i. \tag{4.3}$$

Consequently, following this lemma the process F_t is well defined *iff* the eigenvalues of A lie within the unit circle. For the rest of this paper we assume that $F_0 \sim W_q(n, F)$, although the results remain similar even if factor process does not start with the stationary distribution. The analytical distribution of Y_t is uniquely identifiable with the help of moment generating function (MGF) given in the following Lemma.

Lemma 2. *The moment generating function of Y_t is given by*

$$M_{Y_t}(Z) = |I_p - 2Z\Sigma_\varepsilon|^{-\frac{n}{2}} \cdot \prod_{j \geq 0} \left| I_q - 2\Lambda' Z \Lambda A^j \Sigma_u (A')^j \right|^{-\frac{n}{2}} \tag{4.4}$$

Thus, it follows that the distribution of Y_t is not Wishart. The following Lemma provides results for the first two moments of Y_t.

Lemma 3. *The first and second moments of Y_t are given by*

$$\mathbb{E}[Y_t] = n \left(\Lambda F \Lambda' + \Sigma_\varepsilon \right)$$

and

$$\begin{aligned} \mathbb{E}[Y_t^2] = {} & \left(\Lambda F \Lambda'\right)^2 + \Lambda F \Lambda' \Sigma_\varepsilon + \Sigma_\varepsilon \Lambda F \Lambda' + n((n+1)\Sigma_\varepsilon + tr(\Sigma_\varepsilon) I_p)\Sigma_\varepsilon \\ & + n \sum_{i \geq 0} \left(\Lambda A^i \Sigma_u (A')^i \Lambda' \right)^2 + n \sum_{i \geq 0} tr(\Lambda A^i \Sigma_u (A')^i \Lambda') \Lambda A^i \Sigma_u (A')^i \Lambda', \end{aligned}$$

where the parameter F is as in (4.3).

The essential assumption of absent cross-time dependencies in idiosyncratic component ε_t translates into degenerate measure of autocovariance as the following lemma shows.

Lemma 4 (Autocovariance function)**.** *For some lag $k > 0$ define the autocovariance function of process Y_t as in model (4.1)-(4.2) to be*

$$ACV(k) := \mathbb{E}\left[Y_t Y_{t-k}\right] - \mathbb{E}\left[Y_t\right]\mathbb{E}\left[Y_{t-k}\right]. \tag{4.5}$$

It then follows that ACV(k) has exactly q non-zero eigenvalues.

In the following section we discuss the estimation procedure.

4.3 Estimation

Here we discuss the necessary assumptions to ensure the identifiability of the common and idiosyncratic components within observed data. As mentioned earlier, the absence of dynamics within the individual component is essential for inference and separation of the observed information into two parts. Furthermore, asymptotic consistency of factor and parameter estimates depends on the behavior of Λ and Σ_ε with increasing dimension. We summarize our set of assumptions as follows:

(A1) F_t and ε_t are independent processes;

(A2) $|\lambda_{\max}(A)| < 1$;

(A3) ε_t is a strong white noise matrix process;

(A4) $\liminf_{p\to\infty} \frac{1}{p}\lambda_{\min}(\Lambda'\Lambda) > 0$;

(A5) $\limsup_{p\to\infty} \frac{1}{p}\lambda_{\max}(\Lambda'\Lambda) < \infty$;

(A6) $\liminf_{p\to\infty} \lambda_{\min}(\Sigma_\varepsilon) > 0$;

(A7) $\limsup_{p\to\infty} \lambda_{\max}(\Sigma_\varepsilon) < \infty$.

Assumption (A2) is necessary for stationarity of the process $\{F_t\}$, as stated previously. Together with (A1) and (A3) the first three assumptions ensure that $\{F_t\}$ and consequently $\{Y_t\}$ are stationary. Further, assumptions (A4)-(A7) are required to guarantee the asymptotic properties of the estimation procedure presented in Section 4.3 for the matrix of factor loadings Λ and consequently the factor process F_t and other parameters. Hereby, assumption (A4) states that the common component is pervasive, (A5) - that all eigenvalues of $\Lambda'\Lambda$ diverge at the same rate. Assumption (A6) ensures non-degeneracy of idiosyncratic component and (A7) guarantees finite variance, which is essential for the identifiability of the common component. The following lemma establishes the consistency of factor estimates given the true matrix Λ.

Lemma 5. *Under assumptions (A1)-(A7) for* $p \to \infty$

$$(\Lambda'\Lambda)^{-1}\Lambda' Y_t \Lambda (\Lambda'\Lambda)^{-1} \to F_t.$$

Since the true factors F_t are identifiable only up to a rotation, we now characterize which specific rotation of the true factors F_t is estimated. For E such that $EFE' = I_q$ the following factor estimates

$$\tilde{F}_t = EF_tE' = E\left(AF_{t-1}A' + u_t\right)E' = EAE^{-1}\tilde{F}_{t-1}(E')^{-1}A'E' + \tilde{u}_t$$

are retrieved. This normalization results in standardized factors, i.e. $\mathbb{E}[\tilde{F}_t] = I_q$ and following parameter rotations

$$\begin{aligned} \tilde{\Lambda} &= \Lambda E^{-1}, \\ \tilde{A} &= EAE^{-1}, \\ \tilde{\Sigma}_u &= E\Sigma_u E'. \end{aligned}$$

The following Lemma establishes the consistency of rotated factor estimates.

Lemma 6. *Under assumptions (A1)-(A7) for* $p \to \infty$

$$(\tilde{\Lambda}'\tilde{\Lambda})^{-1}\tilde{\Lambda}' Y_t \tilde{\Lambda} (\tilde{\Lambda}'\tilde{\Lambda})^{-1} \to \tilde{F}_t.$$

Assumptions (A1)-(A7) are essential in order to show asymptotic consistency of the estimation procedure presented in next subsection.

Sequential estimation procedure

The only observable information within the proposed model is Y_t. Thus, parameters Λ, Σ_ε, A and Σ_u have to be estimated and factors F_t must be filtered from data, given the proposed data generating process. We choose likelihood based inference over non-parametric method as in Tao et al. [2011] since the former accounts for the probabilistic specifications of the model. Thus, in order to estimate parameters and filter factors of model (4.1)-(4.2) we propose an iterative estimation procedure based on Expectation Maximisation algorithm (EM). EM was first introduced in Dempster et al. [1977], whereas the specific case of factor models was later discussed in Rubin and Thayer [1982], Shumway and Stoffer [1982]. Essentially, it is a sequential maximum likelihood estimation (MLE) procedure applicable to models, which depend on unobserved information. The basic principle is to calculate and maximize log-likelihood in terms of both observed and unobserved data, i.e. factors. Define $\Theta_j = \{\Lambda(j), A(j), \Sigma_\varepsilon(j), \Sigma_u(j)\}$ to be the set of optimal parameters at iteration j, $\boldsymbol{Y} = (Y_1, \ldots, Y_T)$ - the observed data and $\boldsymbol{F} = (F_1, \ldots, F_T)$ - latent factors. At each iteration j, algorithm executes the following steps:

1. Expectation step (E-step) - computing the expected log-likelihood conditional on observed data with optimal parameter set Θ_j:

$$\mathbb{L}(\Theta, \Theta_j) = \mathbb{E}_{\Theta_j}\left[L(\boldsymbol{Y}, \boldsymbol{F}|\Theta)|\boldsymbol{Y}\right]$$

2. Maximization step (M-step) - re-estimating the parameters through the maximization of the expected log-likelihood with respect to Θ.

$$\Theta_{j+1} = \arg\max_{\Theta} \mathbb{L}(\Theta, \Theta_j).$$

The conditioning on the observed data in the E-step is essential, since estimates of factors are filtered from $\boldsymbol{Y}$ using the parameter set Θ_j. M-step is normally much more computationally demanding than E-step when executed numerically. However, in our case we calculate first derivatives of log-likelihood w.r.t. each parameter matrix analytically and find unique zero points. Thus, the M-step boils down to iterating through simple matrix calculations. If EM steps are continuously repeated, convergence to a (local) optimum is guaranteed (see

Wu [1983]). In order to perform the E-step we calculate the full likelihood of the model with

$$\begin{aligned} L\left(\boldsymbol{Y}, \boldsymbol{F} \mid \Theta\right) &= f\left(\boldsymbol{Y}, \boldsymbol{F} \mid \Theta\right) = f\left(\boldsymbol{Y} \mid \boldsymbol{F}, \Theta\right) f\left(\boldsymbol{F} \mid \Theta\right) \\ &= \prod_{t=1}^{T} f\left(Y_t \mid F_t, \Theta\right) \prod_{t=1}^{T} f(F_t \mid F_{t-1}, \Theta). \end{aligned}$$

The result in the last line is due to conditional independence of F_t and F_{t-k} for all $k \geq 2$, when conditioned on F_{t-1}. Thus, the log-likelihood is

$$\log L\left(\boldsymbol{Y}, \boldsymbol{F} \mid \Theta\right) = \sum_{t=1}^{T} \log f\left(Y_t \mid F_t, \Theta\right) + \sum_{t=1}^{T} \log f(F_t \mid F_{t-1}, \Theta). \tag{4.6}$$

Next from (4.1) and (4.2) we derive that

$$\begin{aligned} Y_t - \Lambda F_t \Lambda' &\sim W_p\left(n, \Sigma_\varepsilon\right), \\ F_t - A F_{t-1} A' &\sim W_q\left(n, \Sigma_u\right). \end{aligned}$$

The conditional density functions from (4.6) are thus given by:

$$\begin{aligned} f\left(Y_t \mid F_t, \Theta\right) &= \frac{\left|Y_t - \Lambda F_t \Lambda'\right|^{(n-p-1)/2} \exp\left(-tr\left(\Sigma_\varepsilon^{-1}\left(Y_t - \Lambda F_t \Lambda'\right)\right)/2\right)}{2^{np/2}\left|\Sigma_\varepsilon\right|^{n/2} \Gamma_p\left(\frac{n}{2}\right)}, \\ f\left(F_t \mid F_{t-1}, \Theta\right) &= \frac{\left|F_t - A F_{t-1} A'\right|^{(n-q-1)/2} \exp\left(-tr\left(\Sigma_u^{-1}\left(F_t - A F_{t-1} A'\right)\right)/2\right)}{2^{nq/2}\left|\Sigma_u\right|^{n/2} \Gamma_q\left(\frac{n}{2}\right)}. \end{aligned}$$

Finally, the complete log-likelihood of the model takes the form of

$$\begin{aligned} \log L = &- \left[\frac{npT}{2} + \frac{nqT}{2}\right] \log 2 - \frac{nT}{2} \log\left|\Sigma_\varepsilon\right| - \frac{nT}{2} \log\left|\Sigma_u\right| - T \log \Gamma_p\left(\frac{n}{2}\right) - T \log \Gamma_q\left(\frac{n}{2}\right) \\ &+ \frac{(n-p-1)}{2} \sum_{t=1}^{T} \log\left|Y_t - \Lambda F_t \Lambda'\right| + \frac{(n-q-1)}{2} \sum_{t=1}^{T} \log\left|F_t - A F_{t-1} A'\right| \\ &- \frac{1}{2} \sum_{t=1}^{T} tr\left(\Sigma_\varepsilon^{-1}\left(Y_t - \Lambda F_t \Lambda'\right)\right) - \frac{1}{2} \sum_{t=1}^{T} tr\left(\Sigma_u^{-1}\left(F_t - A F_{t-1} A'\right)\right). \end{aligned}$$

The conditional expectation of log-likelihood is calculated using the factor estimates. In next Lemma we present the optimal parameter set at each iteration j of EM, in the sense that it contains the zero points of partial derivatives w.r.t. each parameter matrix.

Lemma 7. *The optimal set* $\hat{\Theta}_j = (\hat{A}(j), \hat{\Sigma}_u(j), \hat{\Lambda}(j), \hat{\Sigma}_\varepsilon(j))$ *of parameters at iteration* j *is*

given by

$$\begin{aligned}
\hat{\Sigma}_u(j) &= \frac{1}{nT}\sum_{t=1}^{T}\hat{u}_t, \\
\hat{A}_(j) &= GV^{\frac{1}{2}}\bar{V}^{-\frac{1}{2}}\bar{G}', \\
\hat{\Sigma}_\varepsilon(j) &= \frac{1}{nT}\sum_{t=1}^{T}\hat{\varepsilon}_t, \\
\hat{\Lambda}(j) &= HW^{\frac{1}{2}}\bar{V}^{-\frac{1}{2}}\bar{G}',
\end{aligned}$$

whereas

$$\begin{aligned}
\bar{G}\bar{V}\bar{G}' &= \sum_{t=1}^{T}\hat{F}_{t-1}, \\
GVG' &= \frac{n-q-1}{2}\left[\hat{\Sigma}_u(j)\sum_{t=1}^{T}(\hat{u}_t^{-1}\hat{F}_t - I_q) + \sum_{t=1}^{T}(\hat{F}_t\hat{u}_t^{-1} - I_q)\hat{\Sigma}_u(j)\right], \\
HWH' &= \frac{n-p-1}{2}\left[\hat{\Sigma}_\varepsilon(j)\sum_{t=1}^{T}(\hat{\varepsilon}_t^{-1}Y_t - I_p) + \sum_{t=1}^{T}(Y_t\hat{\varepsilon}_t^{-1} - I_p)\hat{\Sigma}_\varepsilon(j)\right]
\end{aligned}$$

and $\hat{F}_t$, $\hat{u}_t$, $\hat{\varepsilon}_t$ are filtered using optimal set of parameters from iteration $j-1$.

The estimates of latent factors and innovations are calculated at every iteration step j of EM using parameter set $\hat{\Theta}_{j-1}$ from the previous step. The following Lemma lays out the procedure.

Proposition 1 (**Factor estimates**). *The estimates of factors and innovations are given by*

$$\begin{aligned}
\hat{F}_t &= F\Lambda'(\Lambda F\Lambda' + \Sigma_\varepsilon)^{-1}Y_t\Lambda(\Lambda'\Lambda)^{-1}, \\
\hat{\varepsilon}_t &= \Sigma_\varepsilon(\Lambda F\Lambda' + \Sigma_\varepsilon)^{-1}Y_t, \\
\hat{u}_t &= \hat{F}_t - A\hat{F}_{t-1}A',
\end{aligned}$$

whereas $n \cdot F$ is the expectation of F_t. Under Assumptions (A1)-(A7) for $p \to \infty$

$$\hat{F}_t \to F_t,\ \hat{\varepsilon}_t \to \varepsilon_t.$$

The suggested estimators differ from the estimators which would be obtained by a Kalman filter. In the latter case the filtered residuals and factors are determined using the conditional expectations of the respective true quantities. The expectation are easy to derive in the classical dynamic factor models assuming normal distribution. This is not straightforward

due to the specific modeling and distributional setting in our case. To overcome the problem we suggest estimators of the factors and innovations in Proposition 1 which are not Kalman-type estimators, but are consistent. Note that the estimators require only simple matrix operations and thus are numerically easy to implement. In the simulation study we assess the performance of the suggested technique.

4.4 Simulation study

Within this simulation study we consider parameter sets with increasing dimension of data $p \in (20, 40, 60, 80, 100)$ and number of factors fixed at four, i.e. $q = 4$. We further fix daily frequency $n = 390$, which equals the number of minutes during a trading day on New York Stock Exchange.

Further let index i denote different parameter setups of $(\Lambda, \Sigma_\varepsilon)$. For each i we select $\Lambda_i \in \mathbb{R}^{p_i \times q}$ such that $\Lambda_i'\Lambda_i = diag(\lambda_{1,i}^2, \ldots, \lambda_{4,i}^2)$, whereas singular values $\lambda_{1,i}, \ldots, \lambda_{4,i}$ are drawn from uniform distribution on $[0.5 \cdot z_i - 0.5, 0.5 \cdot z_i + 0.5]$ and eigenvalues of $\Sigma_{\varepsilon,i} \in \mathbb{R}^{p_i \times p_i}$ − from uniform distribution on $[0, 1]$. This translates to the property that the singular values of Λ_i are in mean z_i times bigger than those of $\Sigma_{\varepsilon,i}$. In this way, assumptions (A4)-(A7) are fulfilled, since singular values grow with dimension and $\Sigma_{\varepsilon,i}$ is non-degenerate. For $i = (1, \ldots, 5)$ we select Λ_i and $\Sigma_{\varepsilon,i}$ only once using parameters p_i and z_i given in Table 4.1.

i	1	2	3	4	5
p_i	20	40	60	80	100
z_i	3	4	5	6	7

Table 4.1: Parameters p_i and z_i for simulations.

Next, to achieve comparability of results for different dimensions and be able to measure the effect of different persistence in factors we consider a matrix A with eigenvalues $(0.5, 0.6, 0.7, 0.8)$ and Σ_u such that $AA' + \Sigma_u = I_q$.

Given true A the precision of estimates for Λ depends only on p, whereas given the true Λ the precision of estimates for A depends only on T. Thus, in our setup increasing both T and p should deliver better results for all considered Λ_i. Given some parameter set $\Theta_i =$

$(\Lambda_i, \Sigma_{\varepsilon,i}, A, \Sigma_u)$ and T we simulate 1000 samples of observed data $(Y_t)_{t=1,\dots,T}$. Next, for each sample j we execute 1000 steps of EM and the parameter estimates $\hat{\Theta}_{ij} = (\hat{\Lambda}_{ij}, \hat{\Sigma}_{\varepsilon,ij}, \hat{A}_{ij}, \hat{\Sigma}_{u,ij})$ with the biggest log-likelihood are selected. Finally, singular values of $\hat{\Lambda}_{ij}$ and eigenvalues of $\hat{A}_{ij}$ are divided by the corresponding true values, in order to measure the accuracy. Thus, for each (i, j) we obtain 1000 distances between estimated and true parameters. We denote the averages of the distances with $\bar{\lambda}_{k,i}$ and $\bar{a}_k$, standard deviations of these distances with $\sigma(\lambda_{k,i})$ and $\sigma(a_k)$ for $k = 1, \dots, 4$, respectively. The results for $T \in (250, 500, 750, 1000)$ are presented in Tables 4.2 and 4.3.

The precision of estimates for Λ and A increases with p and T, since averages $\bar{\lambda}_{k,i}$ and $\bar{a}_k$ approach 1 and empirical standard deviations $\sigma(\lambda_{k,i})$ and $\sigma(a_k)$ tends to 0. Eigenvalues of A tend to be slightly overestimated and those of Λ - underestimated. The higher the persistence in factors the longer sample is needed for precise estimation of A, whereas this effect is less muted for Λ.

Next, we measure the identifiability of the number of factors. The setup is now different since we consider several $A_j = a_j \cdot I_q$, whereas $a_j = 0.4 + 0.1 \cdot j$. For this purpose we simulate 1000 samples of observed data $(Y_t)_{t=1,\dots,T}$ for each $\Theta_{ij} = (\Lambda_i, \Sigma_{\varepsilon,i}, A_j, \Sigma_{u,j})$. For every sample 1000 steps of estimation procedure with different number of factors $\hat{q} \in (3, 4, 5, 6)$ are executed. We again choose the model with the biggest log-likelihood. The proportion of cases when a specific amount of factors had the biggest likelihood are presented in Table 4.4. Since estimates of Λ are more precise with bigger p the proportion of cases when the true amount of factors was selected increases. Also, the persistence in factors has non-negligible effect on identification.

4.5 Empirical study

In the following empirical study we compare the factor model presented in this paper against two other methods using realized covariance matrices of a subset of S&P100 equity index. The benchmark is represented by the model as in Tao et al. [2011], further referred to as simply principal components, and ARFIMA-Cholesky approach. Models are estimated and compared on daily realized covariance matrices (RCs), computed as plain sums of vector outer-products of 1-minute returns, whereas first and last 30 minutes of each day were

		Av.				Std.			
	λ_i	1	2	3	4	1	2	3	4
T	p_i								
250	20	1.0841	1.1302	1.0889	1.0753	0.0066	0.0088	0.0049	0.0048
	40	1.0542	1.0671	1.0548	1.0512	0.0035	0.0026	0.0029	0.0025
	60	1.0541	1.0498	1.0328	1.0342	0.0014	0.0016	0.0017	0.0017
	80	1.0432	1.0321	1.0418	1.0286	0.0022	0.0021	0.0020	0.0022
	100	1.0096	1.0127	1.0008	1.0027	0.0021	0.0021	0.0017	0.0017
500	20	1.0856	1.1323	1.0898	1.0761	0.0059	0.0078	0.0042	0.0040
	40	1.0544	1.0675	1.0548	1.0514	0.0015	0.0009	0.0014	0.0008
	60	1.0542	1.0498	1.0328	1.0343	0.0006	0.0005	0.0005	0.0005
	80	1.0432	1.0320	1.0420	1.0287	0.0006	0.0005	0.0005	0.0004
	100	1.0095	1.0128	1.0004	1.0031	0.0015	0.0015	0.0013	0.0014
750	20	1.0864	1.1332	1.0902	1.0768	0.0054	0.0071	0.0038	0.0036
	40	1.0543	1.0675	1.0547	1.0514	0.0012	0.0007	0.0012	0.0007
	60	1.0542	1.0498	1.0328	1.0343	0.0005	0.0005	0.0004	0.0004
	80	1.0432	1.0320	1.0421	1.0287	0.0005	0.0004	0.0005	0.0003
	100	1.0094	1.0129	1.0003	1.0033	0.0012	0.0012	0.0012	0.0012
1000	20	1.0871	1.1343	1.0907	1.0773	0.0047	0.0062	0.0032	0.0032
	40	1.0542	1.0675	1.0548	1.0514	0.0011	0.0006	0.0010	0.0006
	60	1.0543	1.0498	1.0328	1.0343	0.0004	0.0004	0.0004	0.0004
	80	1.0432	1.0319	1.0421	1.0287	0.0004	0.0003	0.0004	0.0003
	100	1.0094	1.0128	1.0001	1.0034	0.0011	0.0011	0.0010	0.0010

Table 4.2: Average (left panel) and standard deviations of ratios for singular values of Λ_i.

		Av.				Std.			
	A_k	1	2	3	4	1	2	3	4
T	p_i								
250	20	0.8553	0.8462	0.8451	0.8726	0.1782	0.1164	0.1017	0.1069
	40	0.9344	0.9027	0.8673	0.8978	0.1426	0.0932	0.0809	0.0916
	60	0.9537	0.9283	0.8909	0.9032	0.1434	0.0827	0.0730	0.0797
	80	0.9730	0.9393	0.9024	0.9045	0.1384	0.0857	0.0732	0.0843
	100	0.9861	0.9501	0.9214	0.9745	0.1237	0.0983	0.0966	0.0957
500	20	0.8893	0.8517	0.8212	0.8180	0.1308	0.0763	0.0643	0.0707
	40	0.9944	0.9216	0.8616	0.8531	0.1105	0.0670	0.0556	0.0654
	60	1.0028	0.9457	0.8832	0.8695	0.1200	0.0648	0.0531	0.0615
	80	1.0220	0.9564	0.8941	0.8678	0.1115	0.0630	0.0513	0.0581
	100	1.0309	0.9634	0.9173	0.9487	0.0845	0.0727	0.0713	0.0776
750	20	0.9140	0.8479	0.8125	0.8020	0.1043	0.0679	0.0540	0.0641
	40	1.0128	0.9265	0.8563	0.8402	0.0971	0.0576	0.0477	0.0544
	60	1.0175	0.9478	0.8811	0.8567	0.1034	0.0544	0.0440	0.0522
	80	1.0372	0.9607	0.8936	0.8532	0.0962	0.0524	0.0401	0.0478
	100	1.0470	0.9665	0.9243	0.9343	0.0728	0.0606	0.0577	0.0629
1000	20	0.9149	0.8479	0.8091	0.7897	0.0919	0.0585	0.0448	0.0544
	40	1.0267	0.9313	0.8515	0.8322	0.0905	0.0510	0.0411	0.0498
	60	1.0297	0.9516	0.8797	0.8495	0.0959	0.0477	0.0388	0.0454
	80	1.0505	0.9634	0.8911	0.8429	0.0908	0.0474	0.0360	0.0393
	100	1.0490	0.9635	0.9277	0.9271	0.0657	0.0549	0.0573	0.0534

Table 4.3: Average (left part) and standard deviations of ratios for eigenvalues of A.

	a_j	0.5				0.6				0.7				0.8				0.9			
T	q_{ij} / p_i	3	4	5	6	3	4	5	6	3	4	5	6	3	4	5	6	3	4	5	6
250	20	0.98	0.01	0.00	0.0	0.98	0.01	0.00	0.00	0.98	0.02	0.00	0.00	0.96	0.04	0.00	0.00	0.94	0.06	0.00	0.00
	40	0.96	0.04	0.00	0.0	0.93	0.06	0.00	0.00	0.87	0.12	0.01	0.00	0.55	0.45	0.00	0.00	0.45	0.55	0.00	0.00
	60	0.02	0.97	0.01	0.0	0.04	0.95	0.00	0.00	0.05	0.94	0.01	0.00	0.10	0.89	0.01	0.00	0.10	0.88	0.02	0.00
	80	0.02	0.96	0.02	0.0	0.01	0.96	0.02	0.00	0.04	0.94	0.02	0.00	0.09	0.88	0.03	0.00	0.10	0.85	0.04	0.00
	100	0.04	0.92	0.03	0.0	0.02	0.93	0.04	0.01	0.02	0.92	0.05	0.01	0.08	0.86	0.06	0.01	0.12	0.82	0.06	0.01
500	20	1.00	0.00	0.00	0.0	0.99	0.00	0.00	0.00	0.98	0.01	0.00	0.00	0.97	0.02	0.00	0.00	0.95	0.05	0.00	0.00
	40	0.95	0.05	0.00	0.0	0.88	0.12	0.00	0.00	0.63	0.36	0.00	0.00	0.34	0.66	0.00	0.00	0.33	0.67	0.00	0.00
	60	0.00	0.99	0.00	0.0	0.02	0.98	0.00	0.00	0.04	0.96	0.00	0.00	0.09	0.91	0.00	0.00	0.13	0.87	0.00	0.00
	80	0.00	0.99	0.00	0.0	0.00	1.00	0.00	0.00	0.02	0.98	0.00	0.00	0.08	0.91	0.01	0.00	0.09	0.90	0.00	0.00
	100	0.01	0.99	0.00	0.0	0.00	0.99	0.01	0.00	0.03	0.96	0.00	0.00	0.06	0.92	0.01	0.00	0.08	0.90	0.02	0.00
750	20	1.00	0.00	0.00	0.0	0.99	0.01	0.00	0.00	0.99	0.01	0.00	0.00	0.98	0.02	0.00	0.00	0.92	0.08	0.00	0.00
	40	0.94	0.06	0.00	0.0	0.64	0.36	0.00	0.00	0.40	0.60	0.00	0.00	0.26	0.74	0.00	0.00	0.22	0.78	0.00	0.00
	60	0.00	1.00	0.00	0.0	0.01	0.99	0.00	0.00	0.04	0.96	0.00	0.00	0.08	0.92	0.00	0.00	0.11	0.89	0.00	0.00
	80	0.00	1.00	0.00	0.0	0.00	1.00	0.00	0.00	0.02	0.98	0.00	0.00	0.08	0.92	0.00	0.00	0.10	0.90	0.00	0.00
	100	0.00	1.00	0.00	0.0	0.00	0.99	0.00	0.00	0.04	0.96	0.01	0.00	0.07	0.93	0.00	0.00	0.09	0.91	0.00	0.00
1000	20	1.00	0.00	0.00	0.0	1.00	0.00	0.00	0.00	0.99	0.01	0.00	0.00	0.98	0.02	0.00	0.00	0.94	0.06	0.00	0.00
	40	0.77	0.23	0.00	0.0	0.25	0.75	0.00	0.00	0.26	0.73	0.00	0.00	0.22	0.78	0.00	0.00	0.23	0.77	0.00	0.00
	60	0.00	1.00	0.00	0.0	0.00	1.00	0.00	0.00	0.03	0.97	0.00	0.00	0.10	0.90	0.00	0.00	0.13	0.87	0.00	0.00
	80	0.00	1.00	0.00	0.0	0.00	1.00	0.00	0.00	0.02	0.98	0.00	0.00	0.07	0.93	0.00	0.00	0.11	0.89	0.00	0.00
	100	0.00	1.00	0.00	0.0	0.00	1.00	0.00	0.00	0.03	0.97	0.00	0.00	0.09	0.91	0.00	0.00	0.10	0.90	0.00	0.00

Table 4.4: Percentage of cases when the specified number of factors delivered the biggest log likelihood.

removed. As a consequence of Lemma 2, it follows that the distribution of Y_t as in the proposed model is not Wishart, which does not contradict the definition of RMs as a volatility measure, since there is a strong empirical evidence that RMs do not follow Wishart distribution. The reason is that the underlying intraday returns are not *iid* Gaussian possibly due to market microstructure effects. Furthermore, we don't associate Wishart distributed random matrices in Definition 5 with sum of outer products of intraday returns but instead the whole model is applied to time series of RCs.

The considered sample comprises 60 stocks and spans over 1256 consecutive trading days, starting on the first trading day of 2012 and ending on the last trading day of 2016. Considered stocks are components of the S&P100 equity index in the composition as of January 2019. We first remove stocks which are not available over the whole sample period. Then 60 equities with the highest intraday liquidity (amount of available 1-minute returns) are filtered. The final selection of stocks along with their tickers is listed in Table 4.5.

4.5.1 Full sample analysis

For purposes of visual sample data analysis we present the time series of the two biggest eigenvalues of daily realized covariance matrices in Figure 4.1, whereas four biggest eigenvalues are marked with date. The spike of volatility on October 15, 2014, occurred most certainly due to the so-called bond flash crash in US Treasuries during the opening hours. On Monday August 24, 2015, the biggest observed volatility was caused by the "Black Monday" in China, whereas uncertainty remained for a couple of days. January 2016 was characterized by broad market sell-off.

The classical visual approach to select an appropriate number of factors (components) $\hat{q}$ is the so-called scree plot, which illustrates eigenvalues of the covariance matrix against their order. Some small group of eigenvalues which are sufficiently separated from the rest is considered appropriate. Due to its simplicity, scree plot is normally the first step in determining the number of factors. However, this criterion depends crucially on how "separate" is defined, thus, further analysis of data is essential. The central assumption of our model requires that a small set of eigenvalues grows with dimension. Thus, to check this properties empirically, we present scree plot for three growing dimensions of data and the variation explained by three biggest eigenvalues in Figure 4.2.

Ticker	Name	Ticker	Name
AAPL	Apple Inc	MDLZ	Mondelez International Inc
ABT	Abbott Laboratories	MDT	Medtronic Inc
AIG	American International Group Inc	MET	Metlife Inc
AXP	American Express Co	MMM	3M Co
BA	Boeing Co	MO	Altria Group Inc
BAC	Bank Of America Corp	MRK	Merck & Co Inc New
BMY	Bristol Myers Squibb Co	MS	Morgan Stanley Dean Witter & Co
BRK.B	Berkshire Hathaway Inc Del	MSFT	Microsoft Corp
C	Citigroup Inc	NKE	Nike Inc
CAT	Caterpillar Inc	ORCL	Oracle Corp
COP	Conocophillips	OXY	Occidental Petroleum Corp
CSCO	Cisco Systems Inc	PEP	Pepsico Inc
CVS	CVS Caremark Corp	PFE	Pfizer Inc
CVX	Chevron Corp New	PG	Procter & Gamble Co
DIS	Disney Walt Co	PM	Philip Morris International Inc
EMR	Emerson Electric Co	QCOM	Qualcomm Inc
F	Ford Motor Co Del	SBUX	Starbucks Corp
GE	General Electric Co	SLB	Schlumberger Ltd
GILD	Gilead Sciences Inc	SO	Southern Co
GS	Goldman Sachs Group Inc	T	AT& T Inc
HAL	Halliburton Company	TGT	Target Corp
HD	Home Depot Inc	TXN	Texas Instruments Inc
IBM	International Business Machs Cor	UNH	Unitedhealth Group Inc
INTC	Intel Corp	USB	US Bancorp Del
JNJ	Johnson & Johnson	UTX	United Technologies Corp
JPM	Jpmorgan Chase & Co	VZ	Verizon Communications Inc
KO	Coca Cola Co	WBA	Walgreen Co
LLY	Lilly Eli & Co	WFC	Wells Fargo & Co New
LOW	Lowes Companies Inc	WMT	Wal Mart Stores Inc
MCD	Mcdonalds Corp	XOM	Exxon Mobil Corp

Table 4.5: Ticker symbols and names of the 60 stocks used in the empirical analysis of this article.

The left panel of Figure 4.2 illustrates the time series averages over the sample period for ten biggest eigenvalues and three different dimensions, indicated by color. Thereby, first 20, 40 and 60 rows and columns of each realized covariance matrix in the sample period are taken and eigenvalues are calculated, whereas only the average is presented. The scree plots indicate that $\hat{q} = 3$ or 4 should be a good fit in case of each considered dimension of the realized covariance matrices. This empirical evidence is consistent with results of Aït-Sahalia and Xiu [2019]. The authors found that three factors account for a considerable amount ($\approx 60\%$) of the cross-sectional variation of stock returns. They also pointed out that this empirical

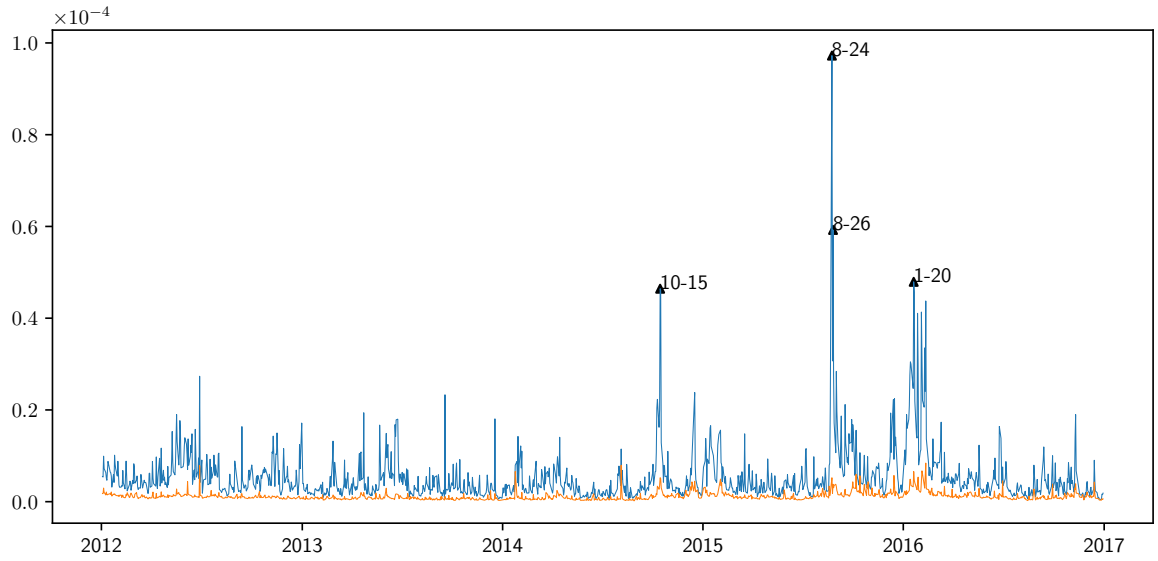

Figure 4.1: Time series of two biggest eigenvalues of daily realized covariance matrices over the sample period. Four biggest values are marked with date.

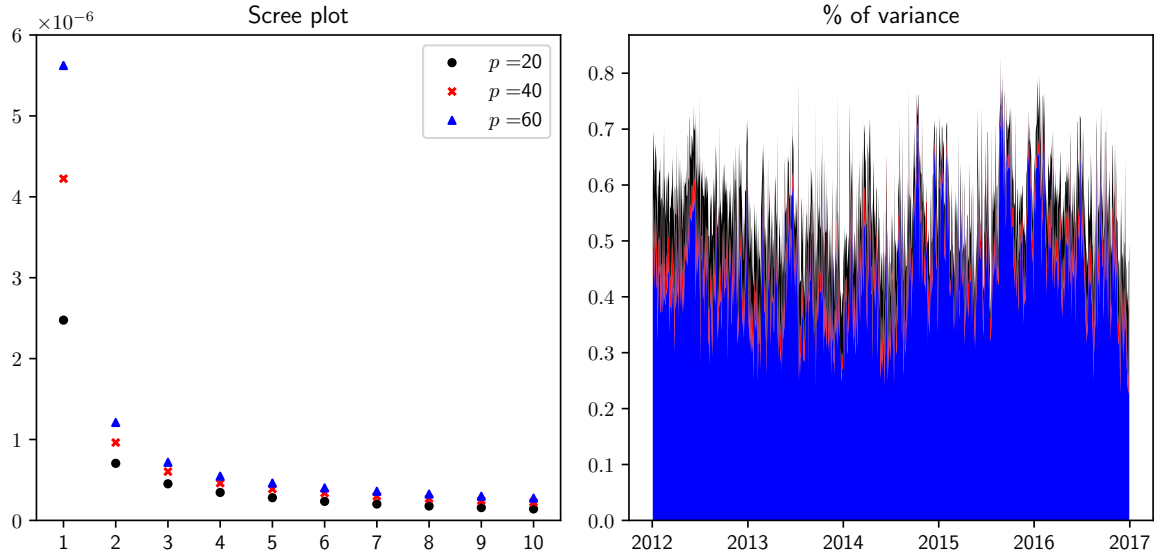

Figure 4.2: Left panel contains time series averages over the considered sample period of the 10 biggest eigenvalues, whereas colors (or shapes) indicate different dimensions of realized covariance matrices. The right panel presents time series of the percentage variation explained by first three eigenvalues.

evidence is consistent with Fama-French common factor analysis. Next, one can observe that the eigenvalues indeed increase with dimension, which should support assumptions (A4) and (A5). Furthermore, we report the dynamics of the percentage variation explained by the first three eigenvalues in the right panel of Figure 4.2. The first three factors accounted for about 40%-80% of total variation, whereas for higher dimensions the relative part is somewhat smaller. Although, during crisis periods the idiosyncratic components become relatively less important. A case in point was the last full trading week in August 2015, which was marked by market turmoil, originated in China. On Monday, the eigenvalues of the realized covariance matrix increased 3-fold and the sum of eigenvalues was 4 times higher compared to Friday. The overall highest proportion of variance explained by the three main components was observed two days later on 26 August, when it reached over 80%, before bouncing 15 percentage points back into average interval several days later. Such behavior during crisis periods is again inline with Aït-Sahalia and Xiu [2019]. In order to fulfill assumptions (A6) and (A7), the portion of remaining variation must at least not fall to zero, which can be supported by the presented empirical evidence.

The analytical properties of the proposed model illustrated in Lemma 4 allow for a further criteria to choose an appropriate number of factors, namely scree plots of autocovariance measure. For this purpose we analyze the empirical analog of the autocovariance measure as introduced in Lemma 4, namely

$$A\hat{C}V(k) \;:=\; \frac{1}{T-k}\sum_{t=1+k}^{T} Y_t Y_{t-k} - \frac{1}{T^2}\left(\sum_{t=1}^{T} Y_t\right)\left(\sum_{t=1}^{T} Y_t\right). \tag{4.7}$$

Due to analytical results presented earlier, matrices $A\hat{C}V(k)$ should be degenerate, with as many non-zero eigenvalues as there are factors. We present the results in the form of scree plots for $k = 1, 5, 10$ in Figure 4.3. Biggest three eigenvalues are clearly outliers, which again hints at three factors. We want to emphasize that this empirical evidence is highly consistent with the analytical assumption of time dynamics present only within common component.

Finally, following Aït-Sahalia and Xiu [2019] we extract economic interpretation from eigendecomposition, in order to assess the economical value of additional factors. For this purpose, eigenvectors associated with the five biggest eigenvalues of the time series average of realized covariance matrices are extracted. Since these eigenvectors are basically the first five principal components as in model by Tao et al. [2011], we also estimate our model on

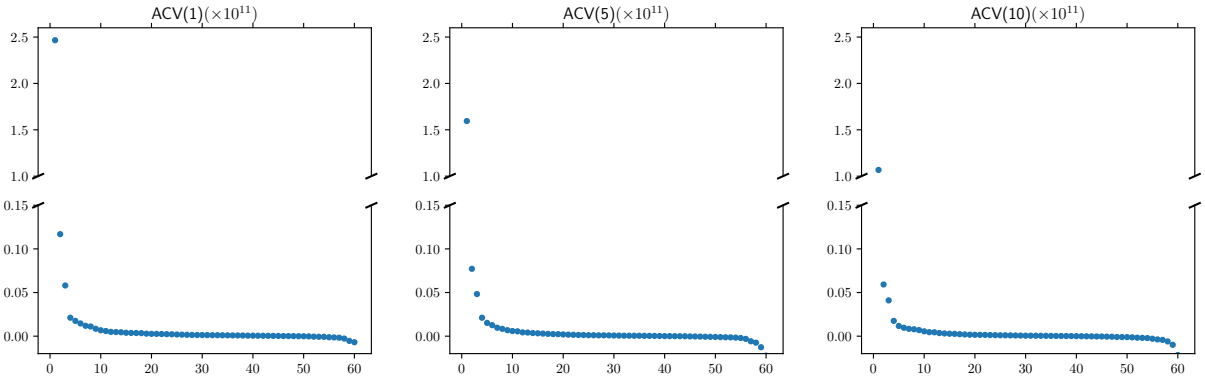

Figure 4.3: Eigenvalues ($\times 10^6$) of D_k for $k = 1, 5, 10$ as in (4.7) against their order.

the full sample period with five underlying factors and retrieve the parameter matrix Λ for further comparison. Each entry in the principal components or Λ is the loading of a specific factor on some stock. The results are reported graphically in the form of biplots (see Gabriel [1971]). Thereby individual stocks are scattered using loadings of two different factors as coordinates. Since the role of the first factor is crucial for all stocks, we fix the x-axis to represent the corresponding loading of the first component and switch the y-axis between the remaining components.

Biplots for principal components are presented in Figure 4.4, whereas the meaning of coordinates is indicated by the axis title. Stocks from the financial, oil and IT sector were highlighted with blue, red and green color, respectively. Tickers were provided where appropriate. We find that apart from American Express Co. other financials (marked blue) are separable from the rest based on first and third factors alone, which is again inline with the results reported in Aït-Sahalia and Xiu [2019]. We also find that the group of oil & energy companies (highlighted in red) is clearly separable along the second component alone. We further observe and document four special cases, which were marked yellow. Gilead Sciences Inc. (gild) and Southern Co. (so) are the only representatives of biotechnology and utility sectors, respectively. The former is an outlier according to all but first component, whereas the latter can be identified using the first component only. Furthermore, Ford Motor Co. could also be considered an interesting case, since it lies very near to the financials and oil sector, but has recognizably different loadings on the fourth and fifth factor. This disproportionality could mean targeting of individual stocks by these components.

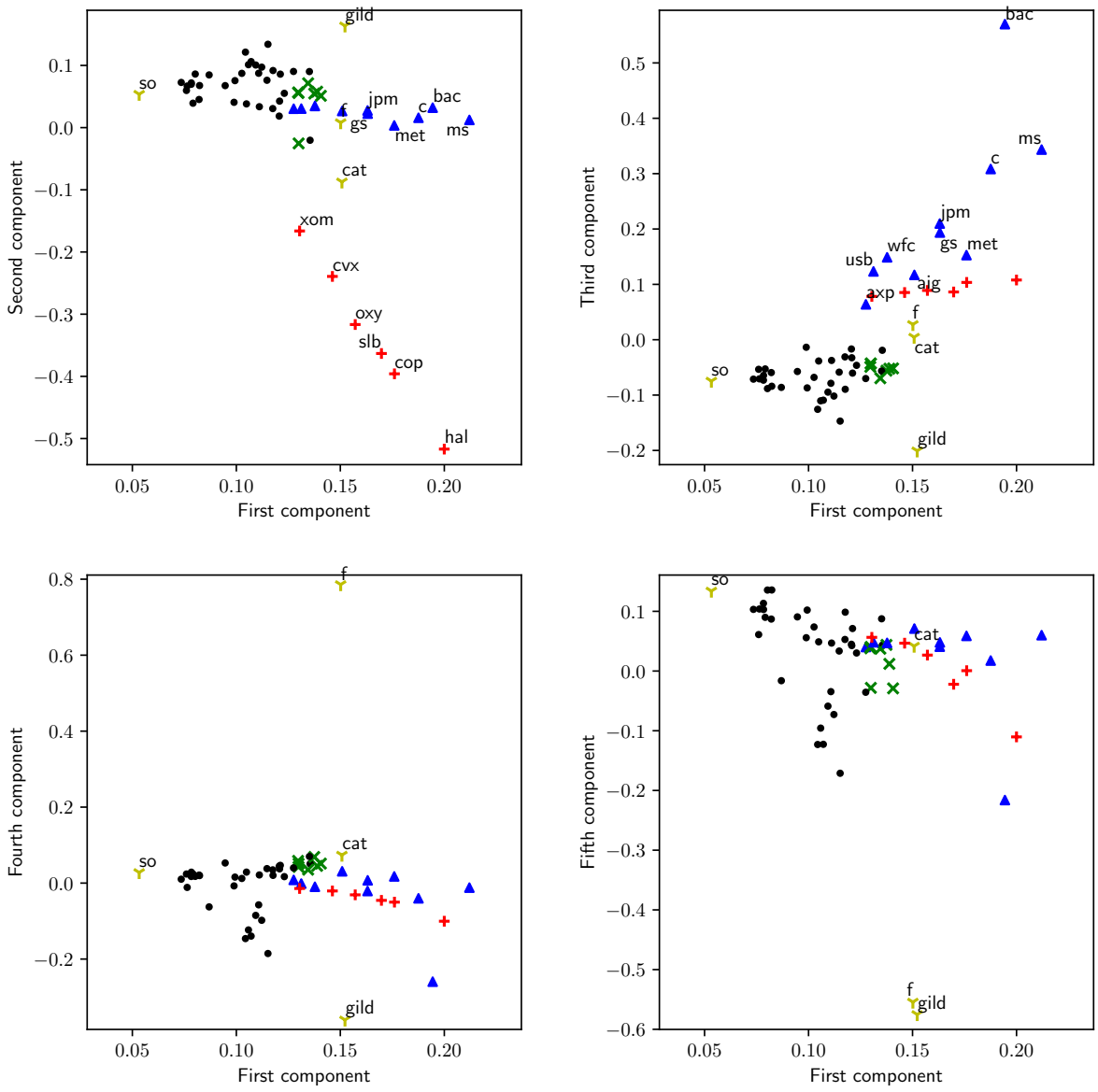

Figure 4.4: Biplots for principal components. The x-axis always corresponds to first component. Financials are highlighted with blue, oil stocks with red and outliers - with green color.

The results for estimated Λ as in our model are reported in Figure 4.5. The main difference is that financials are separable using the second component, whereas they have loading around zero in case of principal components. Also oil stocks show distinct loadings on the third component. Fourth and fifth components are very similar.

Aït-Sahalia and Xiu [2019] argued that the first component could be associated with the

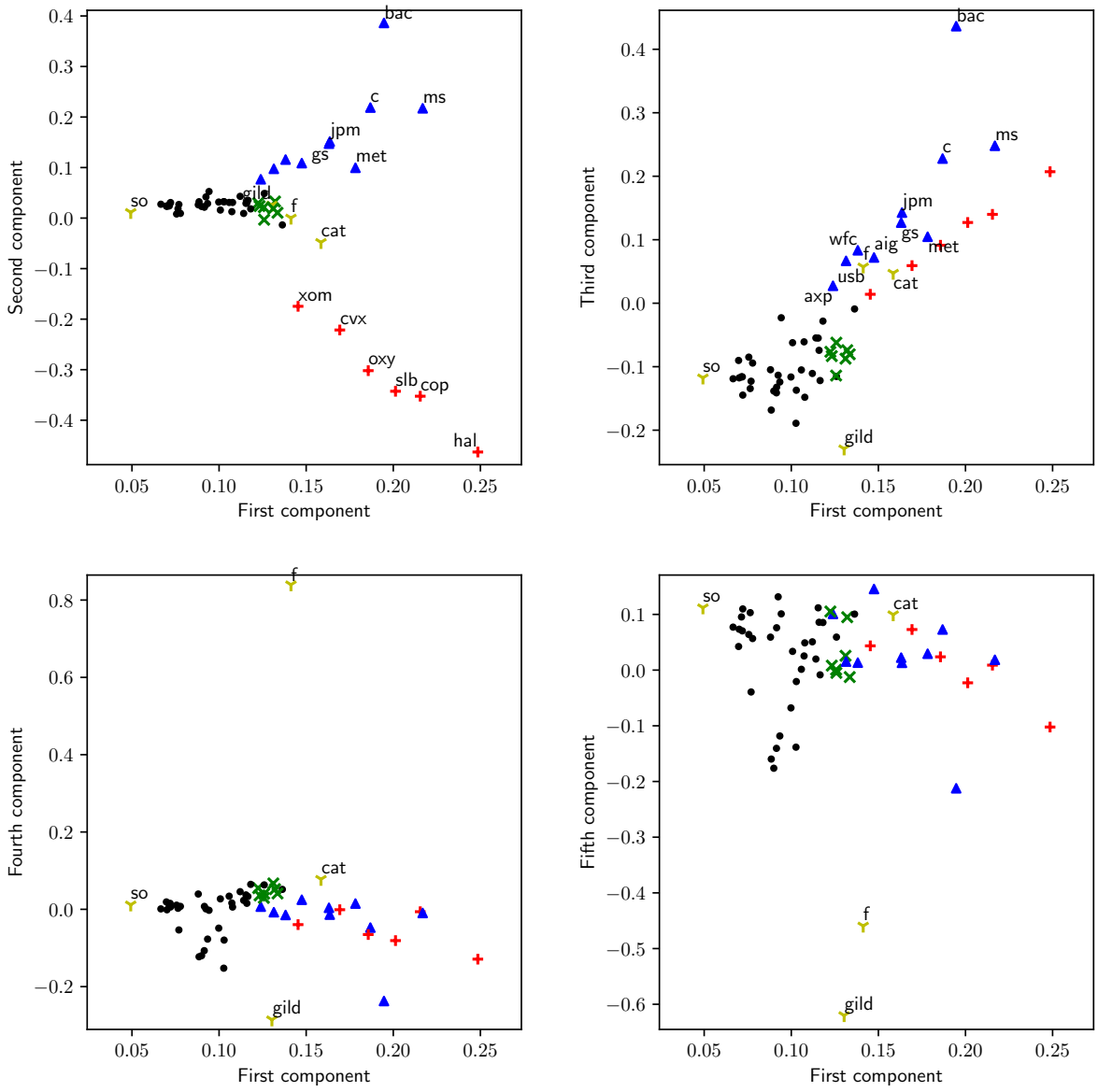

Figure 4.5: Biplots for columns of $\hat{\Lambda}$. The x-axis always corresponds to first component. Financials are highlighted with blue, oil stocks with red and outliers - with green color.

overall market return, especially since it stayed positive for the whole sample they considered. Again, in our case the first component is positive for all stocks. Furthermore, the less sensitive company is the only utility stock, which is logically consistent, since utilities are normally immune to market volatilities. We find a further possible argument for it being associated with marker return. Table 4.6 contains the eigenvalues of estimated A as in the proposed model, whereas the index coincides with the order of component. These values indicate the

persistence of individual factors. Thus, the first component, which could be associated with market return, has the lowest measure of autocorrelation, whereas further factors targeting smaller groups (financials) or individual stocks demonstrate stronger persistence.

A_{11}	A_{22}	A_{33}	A_{44}	A_{55}
.42	.64	.59	.67	.89

Table 4.6: Eigenvalues of the estimated A.

4.5.2 Forecasting study

For each day of the sample period all models are estimated on the previous 100 days and a one-step-ahead forecast for this day is obtained. We also consider different dimensions of data as previously, i.e. models are estimated and forecasts are compared using 20, 40 and 60 stocks. As discussed previously, the assumption of Wishart distribution for ε_t can be misleading and is hardly testable, although essential to calculate the likelihood. Thus, during estimation of our factor model we extract two sets of parameters, where one corresponds to biggest log-likelihood and the other to the smallest Frobenius norm for in-sample forecasts during estimation. To model the dynamics of principal components, we follow Tao et al. [2011] and fit VAR(1) to the vectorization of lower dimensional matrix process. Shen et al. [2018] found VAR of order one appropriate. The results are reported using the notation PC for the model as in Tao et al. [2011], F-L or F-F for forecasts based on log-likelihood or in-sample fit and A for ARFIMA-Cholesky, respectively. To simplify notation we refer to F-L and F-F as different models, although these are only slightly distinct approaches to estimation of the same model. Previously we have established that three factors is probably the most appropriate amount. Thus, in order to check this assumption further, we estimate each model with two to five underlying factors.

Forecasts obtained from different models are evaluated from statistical point of view and in the aspect of optimal portfolio allocation. Thereby, each matrix forecast is first compared through its distance to the observed value and is further used to calculate the weights of minimum variance portfolio (MVP). We consider two matrix loss measures, namely Frobenius

and spectral norm. Given some matrix $M \in \mathbb{R}^{p \times p}$, the Frobenius norm is defined as

$$Fr(M) := \sqrt{\sum_{i=1}^{p} \sum_{i=1}^{p} M_{ii}^2},$$

whereas spectral norm is the biggest singular value of $M'M$, i.e.

$$S(M) := \lambda_{\max}(M'M).$$

Furthermore, the weights of MVP are calculated for each matrix forecast. PC is the only of the considered models where the positive definiteness of forecasts is not guaranteed. Table 4.7 contains the number of days for which the forecast didn't fulfill the requirements. In case of not positive definite forecast the last available observed realized covariance matrix is taken instead to calculate the weights of MVP.

	2	3	4	5
20	0	7	38	92
40	5	16	41	97
60	4	16	28	91

Table 4.7: Number of days on which the forecasts from model PC were not positive definite.

For each model we calculate the average Frobenius (Fr) and spectral (S) norms along with standard deviation (Std), 5% VaR (VaR) and Sharpe ratio (SR) of the produced MVP. Results are reported in Table 4.8. They suggest, that F-F with 5 factors beats all other alternatives for all dimensions, although F-F with 3 factors is quite near.

We make a further step and calculate the minimal and maximal weight of individual asset in the minimum variance portfolio. The results are reported in Table 4.9. Again, our model produced portfolios which are less exposed to individual stocks.

4.6 Summary

In this paper we address the problem of modeling and forecasting the time dynamics of symmetric positive definite matrices with applicational focus on realized covariance matrices.

		PC				F-L				F-F				A
p	$\hat{q}$	2	3	4	5	2	3	4	5	2	3	4	5	
20	Fr	15.83	16.16	17.14	18.02	16.63	17.9	18.03	18.64	15.96	15.3	15.28	**15.24**	22.81
	S	14.23	14.67	15.65	16.48	15.2	16.62	16.69	17.41	14.53	13.83	13.79	**13.74**	17.46
	Std	6.07	6.11	6.16	6.17	6.01	6.06	6.06	6.06	5.98	5.98	**5.97**	5.99	6.39
	VaR	-9.23	-9.29	-9.68	-9.47	-9.23	-9.13	-9.0	-8.94	-8.96	-9.19	-8.85	**-8.83**	-9.89
	SR	6.13	6.35	5.44	6.33	6.63	6.79	6.88	6.48	6.57	6.59	6.77	**6.92**	6.54
40	Fr	27.67	28.1	30.82	31.81	28.25	31.14	30.96	32.02	27.23	**26.59**	26.59	26.63	34.1
	S	25.22	25.75	28.53	29.51	26.01	29.24	28.95	30.19	24.94	**24.18**	24.18	24.2	27.12
	Std	5.19	5.22	5.28	5.22	5.2	5.24	5.29	5.22	5.17	5.18	5.19	**5.17**	5.51
	VaR	-7.77	-7.87	-7.8	**-7.76**	-7.78	-7.99	-7.87	-8.03	-7.83	-7.8	-7.85	-7.77	-8.37
	SR	8.31	8.39	9.17	8.79	8.97	8.45	8.76	8.24	8.76	8.89	**9.34**	9.08	8.47
60	Fr	37.42	38.08	40.74	43.12	38.14	42.51	42.37	43.79	36.38	36.39	36.37	**36.2**	41.48
	S	34.17	34.95	37.55	39.95	35.17	39.96	39.78	41.38	33.25	33.19	33.12	**33.0**	34.08
	Std	4.82	**4.8**	4.81	5.23	4.85	4.92	4.87	4.91	4.81	4.83	4.84	4.84	5.14
	VaR	-7.37	-7.28	-7.52	-7.57	-7.37	-7.47	-7.4	-7.43	-7.36	-7.3	-7.31	**-7.15**	-7.64
	SR	8.88	9.52	8.68	9.12	9.73	9.27	9.12	9.69	9.74	9.53	9.86	**9.95**	9.28

Table 4.8: Final results of forecasts comparison study. Average Frobenius ($\times 10^4$) and spectral ($\times 10^4$) norms along with standard deviation ($\times 10^3$), 5% VaR ($\times 10^3$) and Sharpe ratio ($\times 10^2$) of the produced MVP are reported. The results are denoted by Fr, S, Std, VaR and SR, respectively.

		PC				F-L				F-F				A
		2	3	4	5	2	3	4	5	2	3	4	5	
20	Min	-0.56	-0.58	-0.57	-2.43	-0.28	-0.30	-0.21	-0.21	-0.24	-0.46	-0.44	-0.42	-0.37
	Max	1.24	1.29	1.33	1.33	0.62	0.78	0.62	0.64	0.72	1.04	1.01	1.02	0.86
40	Min	-0.27	-0.29	-0.54	-0.72	-0.19	-0.22	-0.18	-0.20	-0.22	-0.25	-0.24	-0.25	-0.38
	Max	0.50	0.51	0.61	1.10	0.42	0.46	0.44	0.38	0.47	0.46	0.44	0.45	0.69
60	Min	-0.20	-0.27	-0.35	-0.61	-0.16	-0.15	-0.16	-0.16	-0.18	-0.20	-0.21	-0.20	-0.24
	Max	0.36	0.37	0.42	1.03	0.28	0.32	0.26	0.28	0.30	0.35	0.33	0.40	0.63

Table 4.9: The minimal and maximal weight of some individual stock in the MVP over the sample period.

The popular models within this area reduce the problem of modeling time series of random matrices to modeling their vectorization by some well-known methods. This approach breaks the inherent matrix structures and the proposed estimation procedures are infeasible for the general model definitions. The model suggested in this paper ensures symmetricity and positivity without unnecessary matrix transformations but based on appropriate distributional assumptions. The proposed estimation procedure is numerically efficient, since it is a (deterministic) sequence of simple operations. The advantages of the proposed model are shown within an extensive empirical study.

Bibliography

Yacine Aït-Sahalia and Dacheng Xiu. Principal component analysis of high-frequency data. *Journal of the American Statistical Association*, 114(525):287–303, 2019.

Jushan Bai and Serena Ng. Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221, 2002.

Gregory H. Bauer and Keith Vorkink. Forecasting multivariate realized stock market volatility. *Journal of Econometrics*, 160(1):93–101, 2011.

Tim Bollerslev. Modelling the coherence in short-run nominal exchange rates: A multivariate generalized arch model. *The Review of Economics and Statistics*, 72(3):498–505, 1990.

Tim Bollerslev, Robert F. Engle, and Jeffrey M. Wooldridge. A capital asset pricing model with time-varying covariances. *Journal of Political Economy*, 96:116–131, 1988.

Marie-France Bru. Wishart processes. *Journal of Theoretical Probability*, 4(4):725–751, 1991.

Elynn Y. Chen, Ruey S. Tsay, and Rong Chen. Constrained factor models for high-dimensional matrix-variate time series. *Journal of the American Statistical Association*, 115(530):775–793, 2020.

Roxana Chiriac and Valeri Voev. Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics*, 26(6):922–947, 2011.

Arthur P. Dempster, Nan M. Laird, and Donald B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (methodological)*, 39(1):1–38, 1977.

Robert F. Engle. Dynamic conditional correlation. *Journal of Business & Economic Statistics*, 20(3):339–350, 2002.

Robert F. Engle and Kenneth F. Kroner. Multivariate simultaneous generalized ARCH. *Econometric Theory*, 11(1):122–150, 1995.

Mario Forni, Marc Hallin, Marco Lippi, and Lucrezia Reichlin. The generalized dynamic-factor model: identification and estimation. *The Review of Economics and Statistics*, 82 (4):540–554, 2000.

Karl R. Gabriel. The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, 58(3):453–467, 1971.

Vasyl Golosnoy, Bastian Gribisch, and Roman Liesenfeld. The conditional autoregressive Wishart model for multivariate stock market volatility. *Journal of econometrics*, 167(1): 211–223, 2012.

Christian Gouriéroux, Joann Jasiak, and Razvan Sufana. The Wishart autoregressive process of multivariate stochastic volatility. *Journal of Econometrics*, 150(2):167–181, 2009.

Arjun K. Gupta and Daya K. Nagar. *Matrix Variate Distributions*. Taylor & Francis, 1999.

Donald B. Rubin and Dorothy T. Thayer. EM algorithms for ML factor analysis. *Psychometrika*, 47(1):69–76, 1982.

Keren Shen, Jianfeng Yao, and Wai Keung Li. Forecasting high-dimensional realized volatility matrices using a factor model. *Quantitative Finance*, 0(0):1–9, 2018.

Robert H. Shumway and David S. Stoffer. An approach to time series smoothing and forecasting using the EM algorithm. *Journal of time series analysis*, 3(4):253–264, 1982.

Minjing Tao, Yazhen Wang, Qiwei Yao, and Jian Zou. Large volatility matrix inference via combining low-frequency and high-frequency approaches. *Journal of the American Statistical Association*, 106(495):1025–1040, 2011.

Ruey S. Tsay. *Multivariate Time Series Analysis: With R and Financial Applications*. Wiley Series in Probability and Statistics. Wiley, 2013.

Andrew T. Walden and Abdelaziz Serroukh. Wavelet analysis of matrix-valued time-series. *Proceedings: Mathematical, Physical and Engineering Sciences*, 458(2017):157–179, 2002.

Dong Wang, Xialu Liu, and Rong Chen. Factor models for matrix-valued high-dimensional time series. *Journal of Econometrics*, 208(1):231–248, 2019.

Jeff Wu. On the convergence properties of the EM algorithm. *The Annals of Statistics*, 11 (1):95–103, 1983.

4.7 Appendix

4.7.1 Necessary theoretical results

$$
\begin{aligned}
E_{ij} &= (\delta_{ik}\delta_{jl})_{1\leq k\leq N, 1\leq l\leq M} && (4.8)\\
\frac{\partial Y}{\partial X'} &= \left(\frac{\partial Y}{\partial X}\right)' && (4.9)\\
\frac{\log|X|}{\partial X} &= (X')^{-1} && (4.10)\\
\frac{\log|X|}{\partial X} &= 2X^{-1} - (X^{-1}\circ I), \text{ if } X = X' && (4.11)\\
\frac{\partial X}{\partial X_{ij}} &= E_{ij} + E_{ji} - E_{ij}E_{ji} && (4.12)\\
\frac{\partial |Y|}{\partial X_{ij}} &= |Y|\, tr\left[Y^{-1}\frac{\partial Y}{\partial X_{ij}}\right] && (4.13)\\
\partial(XY) &= (\partial X)\, Y + X\,(\partial Y) && (4.14)\\
\frac{\partial tr(AX^{-1}B)}{\partial X} &= -(X^{-1}BAX^{-1})' && (4.15)\\
tr\,(ABC) &= tr\,(BCA) = tr\,(CAB) && (4.16)\\
tr(A'B) &= tr(AB') && (4.17)\\
tr(A') &= tr(A) && (4.18)\\
\frac{\partial}{\partial X} tr\left(X'BXC\right) &= BXC + B'XC' && (4.19)\\
\frac{\partial g(U)}{\partial X_{ij}} &= tr\left[\left(\frac{\partial g(U)}{\partial U}\right)'\frac{\partial U}{\partial X_{ij}}\right] && (4.20)\\
(A+UBV)^{-1} &= A^{-1} - A^{-1}U\left(B^{-1} + VA^{-1}U\right)^{-1}VA^{-1} && (4.21)\\
(A+H)^{-1} &= A^{-1} - (A+H)^{-1}HA^{-1} && (4.22)\\
(A^{-1}+U'B^{-1}U)^{-1}U'B^{-1} &= AU'(UAU'+B)^{-1}, \text{ if } A,B \text{ are psd} && (4.23)
\end{aligned}
$$

Theorem 2. *For positive definite $S \in \mathbb{R}^{p\times p}$ and $Z \in \mathbb{C}^{p\times p}$ the following property holds*

$$\int_{S>0} etr(-SZ)\,|S|^{a-\frac{p+1}{2}}\,dS = \Gamma_p(a)\,|Z|^{-a}\,. \tag{4.24}$$

Theorem 3. *Let $S \sim W_p(n,\Sigma)$ and A be any $p\times p$ nonsingular matrix. Then, $ASA' \sim W_p(n, A\Sigma A')$.*

Theorem 4. ***(See Gupta and Nagar [1999] p. 94)*** *Let $S \sim W_p(n,\Sigma)$, and partition S*

and Σ *as*

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \tag{4.25}$$

where $S_{11} \in \mathbb{R}^{q \times q}$. *Further let* $S_{11\cdot 2} = S_{11} - S_{12}S_{22}^{-1}S_{21}$, $\Sigma_{11\cdot 2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$, *then*

(i) $S_{22} \sim W_{p-q}(n, \Sigma_{22})$

(ii) $S_{11\cdot 2} \sim W_q(n - p + q, \Sigma_{11\cdot 2})$

(ii) $S_{11\cdot 2}$ *and* (S_{12}, S_{22}) *are independent.*

(iv) $S_{12} \mid S_{22} \sim N_{q,p-q}(\Sigma_{12}\Sigma_{22}^{-1}S_{22}, \Sigma_{11\cdot 2} \otimes S_{22})$

Theorem 5. *Let* $S \sim W_p(n, \Sigma)$, *then the characteristic function of* S *is given by*

$$\phi_S(Z) = |I_p - 2iZ\Sigma|^{-\frac{n}{2}} \tag{4.26}$$

Theorem 6. *Let* $S \sim W_p(n, \Sigma)$, *then*

$$\mathbb{E}[SAS] = n\Sigma A'\Sigma + ntr(\Sigma A)\Sigma + n^2\Sigma A\Sigma$$

Theorem 7. *Let* $V \sim IW_p(m, \Psi)$, *then* $V^{-1} \sim W_p(m - p - 1, \Psi^{-1})$

4.7.2 Proofs

Proof of Lemma 2. Here we use the results of Theorem 5. Thus, we have

$$\begin{aligned} M_{Y_t}(Z) = \mathbb{E}[etr(ZY_t)] = \mathbb{E}[etr(Z(\Lambda F_t\Lambda' + \varepsilon_t))] &= \mathbb{E}[etr(Z\varepsilon_t)] \cdot \mathbb{E}[etr(Z\Lambda F_t\Lambda')] \\ &= \mathbb{E}[etr(Z\varepsilon_t)] \cdot \mathbb{E}\left[etr(Z\Lambda \sum_{j\geq 0} A^j u_{t-j} A^j \Lambda')\right] \\ &= |I_p - 2Z\Sigma_\varepsilon|^{-\frac{n}{2}} \cdot \prod_{j\geq 0} \mathbb{E}\left[etr(Z\Lambda A^j u_{t-j} A^j \Lambda')\right] \end{aligned}$$

Denote with $S_j := A^j u_{t-j} A^j$ and $\Sigma_j = A^j \Sigma_u A^j$, then using Theorem 2 and plugging in the density of Wishart distribution

$$\begin{aligned}
\mathbb{E}\left[etr(Z\Lambda S_j \Lambda')\right] &= \mathbb{E}\left[etr(\Lambda' Z \Lambda S_j)\right] = \int etr(\Lambda' Z \Lambda S_j) f(S_j) dS_j \\
&= \left(2^{\frac{nq}{2}} \Gamma_q(\frac{n}{2}) \left|\Sigma_j\right|^{\frac{n}{2}}\right)^{-1} \int etr(\Lambda' Z \Lambda S_j) \left|S_j\right|^{\frac{n-q-1}{2}} etr\left(-\frac{1}{2}\Sigma_j^{-1} S_j\right) dS_j \\
&= \left(2^{\frac{nq}{2}} \Gamma_q(\frac{n}{2}) \left|\Sigma_j\right|^{\frac{n}{2}}\right)^{-1} \int etr(\Lambda' Z \Lambda S_j - \frac{1}{2}\Sigma_j^{-1} S_j) \left|S_j\right|^{\frac{n-q-1}{2}} dS_j \\
&= \left(2^{\frac{nq}{2}} \Gamma_q(\frac{n}{2}) \left|\Sigma_j\right|^{\frac{n}{2}}\right)^{-1} \int etr(-(-\Lambda' Z \Lambda + \frac{1}{2}\Sigma_j^{-1}) S_j) \left|S_j\right|^{\frac{n-q-1}{2}} dS_j \\
&\overset{(4.24)}{=} \left(2^{\frac{nq}{2}} \Gamma_q(\frac{n}{2}) \left|\Sigma_j\right|^{\frac{n}{2}}\right)^{-1} \Gamma_q(\frac{n}{2}) \left|-\Lambda' Z \Lambda + \frac{1}{2}\Sigma_j^{-1}\right|^{-\frac{n}{2}} \\
&= 2^{-\frac{nq}{2}} \Gamma_q(\frac{n}{2})^{-1} \left|\Sigma_j\right|^{-\frac{n}{2}} \Gamma_q(\frac{n}{2}) \left|\frac{1}{2}\Sigma_j^{-1}\right|^{-\frac{n}{2}} \left|-2\Sigma_j \Lambda' Z \Lambda + I_q\right|^{-\frac{n}{2}} = \left|I_q - 2\Lambda' Z \Lambda \Sigma_j\right|^{-\frac{n}{2}}.
\end{aligned}$$

Thus the moment generating function is given by

$$\left|I_p - 2Z\Sigma_\varepsilon\right|^{-\frac{n}{2}} \cdot \prod_{j \geq 0} \left|I_q - 2\Lambda' Z \Lambda \Sigma_j\right|^{-\frac{n}{2}}.$$

□

Proof of Lemma 3. For the first moment it follows directly that

$$\mathbb{E}\left[Y_t\right] = \Lambda \mathbb{E}\left[F_t\right] \Lambda' + \mathbb{E}\left[\varepsilon_t\right] = n\left(\Lambda F \Lambda' + \Sigma_\varepsilon\right)$$

By calculating the second moment we use Theorem 6 for deriving $\mathbb{E}\left[\varepsilon_t \varepsilon_t\right]$:

$$\begin{aligned}
\mathbb{E}\left[Y_t^2\right] &= \mathbb{E}\left[\Lambda F_t \Lambda' \Lambda F_t \Lambda' + \Lambda F_t \Lambda' \varepsilon_t + \varepsilon_t \Lambda F_t \Lambda' + \varepsilon_t \varepsilon_t\right] \\
&= \mathbb{E}\left[\Lambda F_t \Lambda' \Lambda F_t \Lambda'\right] + \Lambda F \Lambda' \Sigma_\varepsilon + \Sigma_\varepsilon \Lambda F \Lambda' + n((n+1)\Sigma_\varepsilon + tr(\Sigma_\varepsilon) I_p) \Sigma_\varepsilon
\end{aligned}$$

whereas

$$\begin{aligned}
\mathbb{E}\left[F_t \Lambda' \Lambda F_t\right] &= \mathbb{E}\left[\sum_{i \geq 0} A^i u_{t-i} (A')^i \Lambda' \Lambda \sum_{j \geq 0} A^j u_{t-j} (A')^j\right] \\
&= \mathbb{E}\left[\sum_{i \geq 0} A^i u_{t-i} (A')^i \Lambda' \Lambda \sum_{j \neq i} A^j u_{t-j} (A')^j\right] + \mathbb{E}\left[\sum_{i \geq 0} A^i u_{t-i} (A')^i \Lambda' \Lambda A^i u_{t-i} (A')^i\right].
\end{aligned}$$

Thus, the first term in the sum results in

$$\begin{aligned}
\sum_{i \geq 0} A^i n \Sigma_u (A')^i \Lambda' \Lambda \sum_{j \neq i} A^j n \Sigma_u (A')^j &= n \sum_{i \geq 0} A^i \Sigma_u (A')^i \Lambda' \Lambda \left[\mathbb{E}[F_t] - n A^i \Sigma_u (A')^i\right] \\
&= \mathbb{E}[F_t] \Lambda' \Lambda \mathbb{E}[F_t] - n^2 \sum_{i \geq 0} A^i \Sigma_u (A')^i \Lambda' \Lambda A^i \Sigma_u (A')^i,
\end{aligned}$$

whereas using Theorem 6 we obtain for the second term

$$\mathbb{E}\left[\sum_{i\geq 0} A^i u_{t-i}(A')^i \Lambda'\Lambda A^i u_{t-i}(A')^i\right] = (n+n^2)\sum_{i\geq 0} A^i \Sigma_u (A')^i \Lambda'\Lambda A^i \Sigma_u (A')^i + n\sum_{i\geq 0} tr(\Sigma_u (A')^i \Lambda'\Lambda A^i) A^i \Sigma_u (A')^i$$

form which the result follows. □

Proof of Lemma 4. First, it follows that

$$\begin{aligned}
\mathbb{E}\left[Y_t Y_{t-k}\right] &= \mathbb{E}\left[(\Lambda F_t \Lambda' + \varepsilon_t)(\Lambda F_{t-k}\Lambda' + \varepsilon_{t-k})\right] \\
&= \mathbb{E}\left[(\Lambda\left[A^k F_{t-k}(A')^k + \sum_{i=0}^{k-1} A^i u_{t-i}(A')^i\right]\Lambda' + \varepsilon_t)(\Lambda F_{t-k}\Lambda' + \varepsilon_{t-k})\right] \\
&= \mathbb{E}\left[\Lambda A^k F_{t-k}(A')^k \Lambda'\Lambda F_{t-k}\Lambda' + \Lambda\sum_{i=0}^{k-1} A^i u_{t-i}(A')^i \Lambda'\Lambda F_{t-k}\Lambda' + \varepsilon_t \Lambda F_{t-k}\Lambda' + \Lambda F_t \Lambda'\varepsilon_{t-k} + \varepsilon_t\varepsilon_{t-k}\right] \\
&= \Lambda\mathbb{E}\left[A^k F_{t-k}(A')^k \Lambda'\Lambda F_{t-k}\right]\Lambda' + n\Lambda\sum_{i=0}^{k-1} A^i \Sigma_u (A')^i \Lambda'\Lambda\mathbb{E}\left[F_{t-k}\right]\Lambda' \\
&\quad + n\Sigma_\varepsilon \Lambda\mathbb{E}\left[F_{t-k}\right]\Lambda' + n\Lambda\mathbb{E}\left[F_t\right]\Lambda'\Sigma_\varepsilon + n^2\Sigma_\varepsilon\Sigma_\varepsilon,
\end{aligned}$$

and further

$$\mathbb{E}\left[Y_t\right]\mathbb{E}\left[Y_{t-k}\right] = \Lambda\mathbb{E}\left[F_t\right]\Lambda'\Lambda\mathbb{E}\left[F_{t-k}\right]\Lambda' + n\Sigma_\varepsilon\Lambda\mathbb{E}\left[F_{t-k}\right]\Lambda' + n\Lambda\mathbb{E}\left[F_t\right]\Lambda'\Sigma_\varepsilon + n^2\Sigma_\varepsilon\Sigma_\varepsilon.$$

Finally, we get that

$$\begin{aligned}
&\mathbb{E}\left[Y_t Y_{t-k}\right] - \mathbb{E}\left[Y_t\right]\mathbb{E}\left[Y_{t-k}\right] \\
&\quad = \Lambda\left(\mathbb{E}\left[A^k F_{t-k}(A')^k \Lambda'\Lambda F_{t-k}\right] + n\sum_{i=0}^{k-1} A^i \Sigma_u (A')^i \Lambda'\Lambda\mathbb{E}\left[F_{t-k}\right] - \mathbb{E}\left[F_t\right]\Lambda'\Lambda\mathbb{E}\left[F_{t-k}\right]\right)\Lambda'.
\end{aligned} \tag{4.27}$$

Now look closer at

$$\begin{aligned}
n\sum_{i=0}^{k-1} A^i \Sigma_u (A')^i \Lambda'\Lambda\mathbb{E}\left[F_{t-k}\right] - \mathbb{E}\left[F_t\right]\Lambda'\Lambda\mathbb{E}\left[F_{t-k}\right] &= n\left(\sum_{i=0}^{k-1} A^i \Sigma_u (A')^i - \sum_{i\geq 0} A^i \Sigma_u (A')^i\right)\Lambda'\Lambda\mathbb{E}\left[F_{t-k}\right] \\
&= -n\left(\sum_{i\geq k} A^i \Sigma_u (A')^i\right)\Lambda'\Lambda\mathbb{E}\left[F_{t-k}\right].
\end{aligned}$$

Next

$$\mathbb{E}\left[A^k F_{t-k}(A')^k \Lambda'\Lambda F_{t-k}\right] = \mathbb{E}\left[\sum_{i\geq 0} A^{k+i} u_{t-k-i}(A')^{k+i}\Lambda'\Lambda \sum_{j\geq 0} A^j u_{t-k-j}(A')^j\right]$$

$$= \mathbb{E}\left[\sum_{i\geq 0} A^{k+i} u_{t-k-i}(A')^{k+i}\Lambda'\Lambda \sum_{j\neq i} A^j u_{t-k-j}(A')^j\right] + \mathbb{E}\left[\sum_{i\geq 0} A^{k+i} u_{t-k-i}(A')^{k+i}\Lambda'\Lambda A^i u_{t-k-i}(A')^i\right].$$

The first term becomes:

$$\mathbb{E}\left[\sum_{i\geq 0} A^{k+i} u_{t-k-i}(A')^{k+i}\Lambda'\Lambda \sum_{j\neq i} A^j u_{t-k-j}(A')^j\right] = \sum_{i\geq 0} A^{k+i} n\Sigma_u (A')^{k+i}\Lambda'\Lambda \sum_{j\neq i} A^j n\Sigma_u (A')^j$$

$$= n\sum_{i\geq 0} A^{k+i}\Sigma_u (A')^{k+i}\Lambda'\Lambda\left[\mathbb{E}[F_t] - nA^i\Sigma_u (A')^i\right]$$

$$= n\sum_{i\geq k} A^i\Sigma_u (A')^i\Lambda'\Lambda\mathbb{E}[F_t] - n^2\sum_{i\geq 0} A^{k+i}\Sigma_u (A')^{k+i}\Lambda'\Lambda A^i\Sigma_u (A')^i,$$

and using Theorem 6 we get for the second term

$$\mathbb{E}\left[\sum_{i\geq 0} A^{k+i} u_{t-k-i}(A')^{k+i}\Lambda'\Lambda A^i u_{t-k-i}(A')^i\right] = n\sum_{i\geq 0} A^{k+i}\Sigma_u (A')^i\Lambda'\Lambda A^{k+i}\Sigma_u (A')^i$$

$$+ n\sum_{i\geq 0} tr\left(\Sigma_u (A')^i\Lambda'\Lambda A^{k+i}\right) A^{k+i}\Sigma_u (A')^i + n^2\sum_{i\geq 0} A^{k+i}\Sigma_u (A')^{k+i}\Lambda'\Lambda A^i\Sigma_u (A')^i.$$

Thus, autocovariance function equals

$$n\Lambda\left(\sum_{i\geq 0} A^{k+i}\Sigma_u (A')^i\Lambda'\Lambda A^{k+i}\Sigma_u (A')^i + \sum_{i\geq 0} tr\left(\Sigma_u (A')^i\Lambda'\Lambda A^{k+1}\right) A^{k+i}\Sigma_u (A')^i\right)\Lambda'$$

$$= n\Lambda\sum_{i\geq 0}\left(C_{ik}\Lambda'\Lambda C_{ik} + tr(C_{ik}\Lambda'\Lambda)C_{ik}\right)\Lambda'$$

whereas $C_{ik} = A^{k+i}\Sigma_u (A')^i$. □

Proof of Lemma 5.

$$(\Lambda'\Lambda)^{-1}\Lambda' Y_t\Lambda(\Lambda'\Lambda)^{-1} = (\Lambda'\Lambda)^{-1}\Lambda'\left[\Lambda F_t\Lambda' + \varepsilon_t\right]\Lambda(\Lambda'\Lambda)^{-1} = F_t + (\Lambda'\Lambda)^{-1}\Lambda'\varepsilon_t\Lambda(\Lambda'\Lambda)^{-1}$$

To show that the second term in the above equation is asymptotically zero, we consider its asymptotic behavior under the Frobenius norm:

$$\mathbb{E}\left[||(\Lambda'\Lambda)^{-1}\Lambda'\varepsilon_t\Lambda(\Lambda'\Lambda)^{-1}||^2\right]$$

$$= \mathbb{E}\left[tr\left((\Lambda'\Lambda)^{-1}\Lambda'\varepsilon_t\Lambda\left(\Lambda'\Lambda\right)^{-2}\Lambda'\varepsilon_t\Lambda(\Lambda'\Lambda)^{-1}\right)\right]$$

$$= tr\left((\Lambda'\Lambda)^{-1}\Lambda'\mathbb{E}\left[\varepsilon_t\Lambda\left(\Lambda'\Lambda\right)^{-2}\Lambda'\varepsilon_t\right]\Lambda(\Lambda'\Lambda)^{-1}\right).$$

Next using Theorem 6, we derive

$$\mathbb{E}\left[\varepsilon_t \Lambda \left(\Lambda'\Lambda\right)^{-2} \Lambda' \varepsilon_t\right] \;=\; (n \,+\, n^2) \,\cdot\, \Sigma_\varepsilon (\Lambda \left(\Lambda'\Lambda\right)^{-2} \Lambda') \Sigma_\varepsilon \,+\, n \,\cdot\, tr\left(\Sigma_\varepsilon \Lambda \left(\Lambda'\Lambda\right)^{-2} \Lambda'\right) \Sigma_\varepsilon.$$

We look at the terms and factors of the previous line separately. First

$$tr\left((\Lambda'\Lambda)^{-1}\Lambda'\Sigma_\varepsilon \Lambda \left(\Lambda'\Lambda\right)^{-2} \Lambda'\Sigma_\varepsilon \Lambda (\Lambda'\Lambda)^{-1}\right) \leq \left[tr\left((\Lambda'\Lambda)^{-1}\Lambda'\Sigma_\varepsilon \Lambda \left(\Lambda'\Lambda\right)^{-1}\right)\right]^2$$

whereas

$$tr\left((\Lambda'\Lambda)^{-1}\Lambda'\Sigma_\varepsilon \Lambda \left(\Lambda'\Lambda\right)^{-1}\right) \leq tr\left((\Lambda'\Lambda)^{-1}\right)^2 tr\left(\Lambda'\Sigma_\varepsilon \Lambda\right)$$

and

$$tr\left(\Lambda'\Sigma_\varepsilon \Lambda\right) = tr\left(\Lambda\Lambda'\Sigma_\varepsilon\right) \leq tr\left(\Lambda\Lambda'\right) tr\left(\Sigma_\varepsilon\right)$$

from which follows for $p \to \infty$

$$tr\left((\Lambda'\Lambda)^{-1}\Lambda'\Sigma_\varepsilon \Lambda \left(\Lambda'\Lambda\right)^{-2} \Lambda'\Sigma_\varepsilon \Lambda (\Lambda'\Lambda)^{-1}\right) \leq \left[tr\left((\Lambda'\Lambda)^{-1}\right) tr\left(\Sigma_\varepsilon\right)\right]^2 \to 0.$$

Next for $p \to \infty$

$$\begin{aligned} tr\Big((\Lambda'\Lambda)^{-1}\Lambda' & tr\left(\Sigma_\varepsilon \Lambda \left(\Lambda'\Lambda\right)^{-2} \Lambda'\right) \Sigma_\varepsilon \Lambda (\Lambda'\Lambda)^{-1}\Big) \\ &= tr\left(\Sigma_\varepsilon \Lambda \left(\Lambda'\Lambda\right)^{-2} \Lambda'\right) tr\left((\Lambda'\Lambda)^{-1}\Lambda'\Sigma_\varepsilon \Lambda (\Lambda'\Lambda)^{-1}\right) \\ &\leq tr\left(\Lambda'\Sigma_\varepsilon \Lambda \left(\Lambda'\Lambda\right)^{-2}\right) tr\left((\Lambda'\Lambda)^{-1}\right)^2 tr\left(\Lambda'\Sigma_\varepsilon \Lambda\right) \to 0. \end{aligned}$$

Finally, the main result follows due to

$$\mathbb{E}\left[||(\Lambda'\Lambda)^{-1}\Lambda'\varepsilon_t \Lambda (\Lambda'\Lambda)^{-1}||^2\right] \to 0,$$

for $p \to \infty$. □

Proof of Lemma 6. Observe that

$$\begin{aligned} (\tilde{\Lambda}'\tilde{\Lambda})^{-1}\tilde{\Lambda}' Y_t \tilde{\Lambda} (\tilde{\Lambda}'\tilde{\Lambda})^{-1} &= ((E^{-1})'\Lambda'\Lambda E^{-1})^{-1}(E^{-1})'\Lambda' \left[\Lambda E^{-1}\tilde{F}_t (E^{-1})'\Lambda' + \varepsilon_t\right] \Lambda E^{-1}((E^{-1})'\Lambda'\Lambda E^{-1})^{-1} \\ &= \tilde{F}_t + ((E^{-1})'\Lambda'\Lambda E^{-1})^{-1}(E^{-1})'\Lambda'\varepsilon_t \Lambda E^{-1}((E^{-1})'\Lambda'\Lambda E^{-1})^{-1}, \end{aligned}$$

from which the stated result follows using arguments similar to the previous proof. □

Proof of Lemma 7. To simplify the notation we omit the number of iteration j. When dealing with symmetric matrices Σ_u and Σ_ε, the chain rule (4.20) is applied extensively throughout this proof.

M-step for Σ_u There are two terms in the complete log-likelihood which contain Σ_u, namely $-\frac{nT}{2}\log|\Sigma_u|$ and $-\frac{1}{2}\sum_{t=1}^{T} tr\left(\Sigma_u^{-1}\left(F_t - A'F_{t-1}A\right)\right)$. First, observe that

$$\begin{aligned}\frac{\partial\, tr\left(\Sigma_u^{-1}\left(F_t - AF_{t-1}A'\right)\right)}{\partial(\Sigma_u)_{ij}} &\overset{(4.20)}{=} tr\left[\frac{\partial\left(\Sigma_u^{-1}\left(F_t - AF_{t-1}A'\right)\right)}{\partial\Sigma_u}\cdot\frac{\partial\Sigma_u}{\partial(\Sigma_u)_{ij}}\right]\\ &= tr\left[-\Sigma_u^{-1}\left(F_t - AF_{t-1}A'\right)\Sigma_u^{-1}\left(E_{ij}+E_{ji}-E_{ij}E_{ji}\right)\right],\end{aligned}$$

from which follows that

$$\frac{\partial\, tr\left(\Sigma_u^{-1}\left(F_t - AF_{t-1}A'\right)\right)}{\partial\Sigma_u} = -2\Sigma_u^{-1}\left(F_t - AF_{t-1}A'\right)\Sigma_u^{-1} + I_q\odot\Sigma_u^{-1}\left(F_t - AF_{t-1}A'\right)\Sigma_u^{-1}$$

and

$$\begin{aligned}-\frac{1}{2}\cdot\frac{\partial\sum_{t=1}^{T} tr\left(\Sigma_u^{-1}\left(F_t - AF_{t-1}A'\right)\right)}{\partial\Sigma_u} &= \Sigma_u^{-1}\sum_{t=1}^{T}\left(F_t - AF_{t-1}A'\right)\Sigma_u^{-1}\\ &\quad -\frac{1}{2}\sum_{t=1}^{T} I_q\odot\Sigma_u^{-1}\left(F_t - AF_{t-1}A'\right)\Sigma_u^{-1},\end{aligned}$$

where the last term is the diagonal of the first term in the sum.

Furthermore, for the second term in the complete log-likelihood which contains matrix Σ_u we have

$$-\frac{nT}{2}\frac{\partial\log|\Sigma_u|}{\partial\Sigma_u} \overset{(4.11)}{=} -nT\Sigma_u^{-1} + \frac{nT}{2}I_q\odot\Sigma_u^{-1}.$$

Finally, the derivative can be calculated as follows

$$\begin{aligned}\frac{\partial\log L}{\partial\Sigma_u} &= -nT\Sigma_u^{-1} + \Sigma_u^{-1}\sum_{t=1}^{T}\left(F_t - AF_{t-1}A'\right)\Sigma_u^{-1}\\ &\quad + \frac{nT}{2}I_q\odot\Sigma_u^{-1} - \frac{1}{2}\sum_{t=1}^{T} I_q\odot\left(\Sigma_u^{-1}\left(F_t - AF_{t-1}A'\right)\Sigma_u^{-1}\right).\end{aligned}$$

To find the zero point of this derivative it suffices to solve the following equation

$$\frac{nT}{2}\Sigma_u^{-1} = \frac{1}{2}\Sigma_u^{-1}\sum_{t=1}^{T}\left(F_t - AF_{t-1}A'\right)\Sigma_u^{-1}, \tag{4.28}$$

since then it follows that

$$\frac{nT}{2}I_q\odot\Sigma_u^{-1} = \frac{1}{2}\sum_{t=1}^{T} I_q\odot\Sigma_u^{-1}\left(F_t - AF_{t-1}A'\right)\Sigma_u^{-1}.$$

The only solution of (4.28) is the simple average

$$\Sigma_u = \frac{1}{nT}\sum_{t=1}^{T}\left(F_t - AF_{t-1}A'\right). = \frac{1}{nT}\sum_{t=1}^{T} u_t \tag{4.29}$$

M-step for A As was the case for Σ_u, there are two terms in the complete log-likelihood which contain A. First, we calculate the derivative of $\log|F_t - AF_{t-1}A'|$:

$$\begin{aligned}\frac{\partial \log|F_t - AF_{t-1}A'|}{\partial A_{ij}} &\overset{(4.20)}{=} tr\left[\left(\frac{\partial \log|F_t - AF_{t-1}A'|}{\partial (F_t - AF_{t-1}A')}\right)' \frac{\partial (F_t - AF_{t-1}A')}{\partial A_{ij}}\right] \\ &= tr\left[(F_t - AF_{t-1}A')^{-1}\frac{\partial(-AF_{t-1}A')}{\partial A_{ij}}\right].\end{aligned}$$

The second factor under the trace function can be derived as follows:

$$-\frac{\partial AF_{t-1}A'}{\partial A_{ij}} \overset{(4.14)}{=} -\frac{\partial A}{\partial A_{ij}}F_{t-1}A' - A\frac{\partial F_{t-1}A'}{\partial A_{ij}} \overset{(4.12)}{=} -E_{ij}F_{t-1}A' - AF_{t-1}E_{ij}'.$$

Next, using basic properties of trace function we get

$$\begin{aligned}\frac{\partial \log|F_t - AF_{t-1}A'|}{\partial A_{ij}} &= tr\left[(F_t - AF_{t-1}A')^{-1}(-E_{ij}F_{t-1}A' - AF_{t-1}E_{ij}')\right] \\ &= -tr\left[(F_t - AF_{t-1}A')^{-1}E_{ij}F_{t-1}A'\right] - tr\left[(F_t - AF_{t-1}A')^{-1}AF_{t-1}E_{ij}'\right] \\ &\overset{(4.17)}{=} -2tr\left[(F_t - AF_{t-1}A')^{-1}AF_{t-1}E_{ij}'\right]\end{aligned}$$

Finally, it follows that

$$\frac{\partial \log|F_t - AF_{t-1}A'|}{\partial A} = -2(F_t - AF_{t-1}A')^{-1}AF_{t-1}.$$

Next the derivative of $tr(\Sigma_u^{-1}AF_{t-1}A')$:

$$\begin{aligned}\frac{\partial tr(\Sigma_u^{-1}AF_{t-1}A')}{\partial A} &\overset{(4.16)}{=} \frac{\partial tr(AF_{t-1}A'\Sigma_u^{-1})}{\partial A} \overset{(4.19),(4.10)}{=} \left(F_{t-1}A'\Sigma_u^{-1} + F_{t-1}'A'\left(\Sigma_u^{-1}\right)'\right)' \\ &= 2\Sigma_u^{-1}AF_{t-1}.\end{aligned}$$

Taking the derivative of L after A and setting it to zero we get the following identity:

$$(n-q-1)\sum_{t=1}^{T}(F_t - AF_{t-1}A')^{-1}AF_{t-1} = \sum_{t=1}^{T}\Sigma_u^{-1}AF_{t-1}. \tag{4.30}$$

from which by multiplying on the right with A' we get

$$\begin{aligned}(n-q-1)\sum_{t=1}^{T}(F_t - AF_{t-1}A')^{-1}AF_{t-1}A' &= \sum_{t=1}^{T}\Sigma_u^{-1}AF_{t-1}A' = \Sigma_u^{-1}A\left[\sum_{t=1}^{T}F_{t-1}\right]A' \\ \iff (n-q-1)\sum_{t=1}^{T}u_t^{-1}(F_t - u_t) &= \Sigma_u^{-1}A\left[\sum_{t=1}^{T}F_{t-1}\right]A' \\ \iff (n-q-1)\Sigma_u\sum_{t=1}^{T}(u_t^{-1}F_t - I_q) &= A\left[\sum_{t=1}^{T}F_{t-1}\right]A'\end{aligned}$$

Here the previously extracted parameter Σ_u is used. The part on the left is asymptotically symmetric, whereas only approximately symmteric for finite samples. Thus, we symmetrize the left part by adding its transpose:

$$\frac{n-q-1}{2}\left[\Sigma_u\sum_{t=1}^{T}(u_t^{-1}F_t - I_q) + \sum_{t=1}^{T}(F_t u_t^{-1} - I_q)\Sigma_u\right] = A\left[\sum_{t=1}^{T}F_{t-1}\right]A'$$

For simplicity let GVG' denote the eigendecomposition of the part on the left, and $\bar{G}\bar{V}\bar{G}'$ - that of $\sum_{t=1}^{T}F_{t-1}$. Then it follows, that

$$GVG' \;=\; A\bar{G}\bar{V}\bar{G}'A',$$

and A thus equals

$$A = GV^{\frac{1}{2}}\bar{V}^{-\frac{1}{2}}\bar{G}'$$

M-step for Σ_ε Following the same steps as for Σ_u, in the end we get:

$$\Sigma_\varepsilon \;=\; \frac{1}{nT}\sum_{t=1}^{T}(Y_t - \Lambda F_t\Lambda') = \frac{1}{nT}\sum_{t=1}^{T}\varepsilon_t.$$

Step for Λ Next we calculate the derivative of $\log|Y_t - \Lambda F_t\Lambda'|$ similar to A:

$$\frac{\partial\log|Y_t - \Lambda F_t\Lambda'|}{\partial\Lambda_{ij}} \overset{(4.20)}{=} tr\left[(Y_t - \Lambda F_t\Lambda')^{-1}\frac{\partial(Y_t - \Lambda F_t\Lambda')}{\partial\Lambda_{ij}}\right]$$

and the second term in the trace function:

$$\frac{\partial(Y_t - \Lambda F_t\Lambda')}{\partial\Lambda_{ij}} \;=\; -\left(E_{ij}F_t\Lambda' + \Lambda F_t E_{ij}'\right).$$

$$\begin{aligned}\frac{\partial\log|Y_t - \Lambda F_t\Lambda'|}{\partial\Lambda_{ij}} &= tr\left[(Y_t - \Lambda F_t\Lambda')^{-1}(-E_{ij}F_t\Lambda' - \Lambda F_t E_{ij}')\right]\\ &= -tr\left[(Y_t - \Lambda F_t\Lambda')^{-1}E_{ij}F_{t-1}\Lambda'\right] - tr\left[(Y_t - \Lambda F_t\Lambda')^{-1}\Lambda F_t E_{ij}'\right]\\ &= -2tr\left[(Y_t - \Lambda F_t\Lambda')^{-1}\Lambda F_t E_{ij}'\right].\end{aligned}$$

Thus it follows similarly to a previous result, that

$$\frac{\partial\log|Y_t - \Lambda F_t\Lambda'|}{\partial\Lambda} \;=\; -2(Y_t - \Lambda F_t\Lambda')^{-1}\Lambda F_t$$

Now the derivative:

$$\frac{1}{2}\frac{\partial tr\left(\Sigma_\varepsilon^{-1}\Lambda F_t\Lambda'\right)}{\partial\Lambda} \overset{(4.16)}{=} \frac{1}{2}\frac{\partial tr\left(\Lambda F_t\Lambda'\Sigma_\varepsilon^{-1}\right)}{\partial\Lambda} \overset{(4.19)}{=} (F_t\Lambda'\Sigma_\varepsilon^{-1})' = \Sigma_\varepsilon^{-1}\Lambda F_t$$

Next we have to find a solution for the following equation:

$$(n-p-1)\sum_{t=1}^{T}\left(Y_t-\Lambda F_t\Lambda'\right)^{-1}\Lambda F_t \;=\; \sum_{t=1}^{T}\Sigma_\varepsilon^{-1}\Lambda F_t, \tag{4.31}$$

By proceeding as previously we get

$$\frac{n-p-1}{2}\left[\Sigma_\varepsilon\sum_{t=1}^{T}(\varepsilon_t^{-1}Y_t-I_p)+\sum_{t=1}^{T}(Y_t\varepsilon_t^{-1}-I_p)\Sigma_\varepsilon\right]=\Lambda\left[\sum_{t=1}^{T}F_t\right]\Lambda',$$

where Σ_ε is as extracted previously. Similarly let HWH' denote the singular value decomposition, i.e. $H\in\mathbb{R}^{p\times q}$ and $W\in\mathbb{R}^{q\times q}$, of the part on the left, and $\bar{G}\bar{V}\bar{G}'$ - again that of $\sum_{t=1}^{T}F_{t-1}$.Then it follows, that

$$HWH' \;=\; \Lambda\bar{G}\bar{V}\bar{G}'\Lambda',$$

and Λ thus equals

$$\Lambda = HW^{\frac{1}{2}}\bar{V}^{-\frac{1}{2}}\bar{G}'$$

□

Proof of Proposition 1. Remember that $u_t=\sum_{i=1}^{n}X_{u,i,t}X_{u,i,t}'$ and $\varepsilon_t=\sum_{i=1}^{n}X_{\varepsilon,i,t}X_{\varepsilon,i,t}'$ for *iid* $X_{u,i,t}\sim\mathcal{N}(\mathbf{0},\Sigma_u)$ and $X_{\varepsilon,i,t}\sim\mathcal{N}(\mathbf{0},\Sigma_\varepsilon)$. Define

$$X_{i,t}=\begin{pmatrix}X_{\varepsilon,i,t}\\ \sum_{j\geq0}A^jX_{u,i,t-j}\\ \Lambda\sum_{j\geq0}A^jX_{u,i,t-j}+X_{\varepsilon,i,t}\end{pmatrix}.$$

Since $X_{i,t}$ is normally distributed it follows that

$$\sum_{i=1}^{n}X_{i,t}X_{i,t}'\sim W_{2p+q}\left(n,\begin{pmatrix}\Sigma_\varepsilon & \mathbf{0} & \Sigma_\varepsilon\\ \mathbf{0} & F & F\Lambda'\\ \Sigma_\varepsilon & \Lambda F & \Lambda F\Lambda'+\Sigma_\varepsilon\end{pmatrix}\right).$$

We further define

$$\begin{aligned}S_{12}&:=\begin{pmatrix}\sum_{i=1}^{n}X_{\varepsilon,i,t}(\Lambda\sum_{j\geq0}A^jX_{u,i,t-j}+X_{\varepsilon,i,t})'\\ \sum_{i=1}^{n}\sum_{j\geq0}A^jX_{u,i,t-j}(\Lambda\sum_{j\geq0}A^jX_{u,i,t-j}+X_{\varepsilon,i,t})'\end{pmatrix}\\ &=\begin{pmatrix}\varepsilon_t+\sum_{i=1}^{n}X_{\varepsilon,i,t}\sum_{j\geq0}X_{u,i,t-j}'(A^j)'\Lambda'\\ F_t\Lambda'+\sum_{i=1}^{n}\sum_{j\geq0}A^jX_{u,i,t-j}X_{\varepsilon,i,t}'+2\sum_{i=1}^{n}\sum_{j\geq0}A^jX_{u,i,t-j}\sum_{l>j}X_{u,i,t-l}'(A^l)'\Lambda'\end{pmatrix}\\ &=\begin{pmatrix}\varepsilon_t+R_1'\Lambda'\\ F_t\Lambda'+R_1+R_2\Lambda'\end{pmatrix},\end{aligned}$$

and

$$
\begin{aligned}
S_{22} &:= \sum_{i=1}^{n}(\Lambda \sum_{j\geq 0} A^j X_{u,i,t-j} + X_{\varepsilon,i,t})(\Lambda \sum_{j\geq 0} A^j X_{u,i,t-j} + X_{\varepsilon,i,t})' \\
&= \Lambda F_t \Lambda' + \varepsilon_t + \Lambda \sum_{i=1}^{n}\sum_{j\geq 0} A^j X_{u,i,t-j} X'_{\varepsilon,i,t} + \sum_{i=1}^{n}\sum_{j\geq 0} X_{\varepsilon,i,t} X'_{u,i,t-j}(A^j)'\Lambda' \\
&\quad + 2\Lambda \sum_{i=1}^{n}\sum_{j\geq 0} A^j X_{u,i,t-j} \sum_{l>j} X'_{u,i,t-l}(A^l)'\Lambda' \\
&= Y_t + \Lambda R_1 + R_1'\Lambda' + \Lambda R_2 \Lambda'.
\end{aligned}
$$

Using Theorem 4, it follows that

$$
\mathbb{E}\left[S_{12}|S_{22}\right] = \begin{pmatrix} \Sigma_\varepsilon \\ F\Lambda' \end{pmatrix} \left(\Lambda F \Lambda' + \Sigma_\varepsilon\right)^{-1} \left(Y_t + \Lambda R_1 + R_1'\Lambda' + \Lambda R_2 \Lambda'\right).
$$

We define estimates for F_t and ε_t as

$$
\begin{aligned}
\hat{F}_t &:= F\Lambda' \left(\Lambda F \Lambda' + \Sigma_\varepsilon\right)^{-1} Y_t \Lambda (\Lambda'\Lambda)^{-1} \\
\hat{\varepsilon}_t &:= \Sigma_\varepsilon \left(\Lambda F \Lambda' + \Sigma_\varepsilon\right)^{-1} Y_t
\end{aligned}
$$

and investigate their asymptotic behavior in the rest of this proof. Since trace is the sum of eigenvalues and due to ((A4)), ((A5)) it follows that $||(\Lambda'\Sigma_\varepsilon^{-1}\Lambda)^{-1}||^2 \to 0$, where $||\cdot||^2$ is the Frobenius norm. We first derive the following result using (4.23)

$$
F\Lambda'(\Lambda F\Lambda' + \Sigma_\varepsilon)^{-1} = (\Lambda'\Sigma_\varepsilon^{-1}\Lambda + F^{-1})^{-1}\Lambda'\Sigma_\varepsilon^{-1}.
$$

Next we split the factors estimates $\hat{F}_t$ into two parts

$$
\begin{aligned}
F\Lambda'(\Lambda F\Lambda' + \Sigma_\varepsilon)^{-1} Y_t \Lambda(\Lambda'\Lambda)^{-1} &= F\Lambda'(\Lambda F\Lambda' + \Sigma_\varepsilon)^{-1}\left[\Lambda F_t \Lambda' + \varepsilon_t\right]\Lambda(\Lambda'\Lambda)^{-1} \\
&= (\Lambda'\Sigma_\varepsilon^{-1}\Lambda + F^{-1})^{-1}\Lambda'\Sigma_\varepsilon^{-1}\Lambda F_t + (\Lambda'\Sigma_\varepsilon^{-1}\Lambda + F^{-1})^{-1}\Lambda'\Sigma_\varepsilon^{-1}\varepsilon_t\Lambda(\Lambda'\Lambda)^{-1}. \quad (4.32)
\end{aligned}
$$

To show that the second term in the above equation is asymptotically zero, we consider its asymptotic behavior under the Frobenius norm:

$$
\begin{aligned}
&\mathbb{E}\left[||(F^{-1} + \Lambda'\Sigma_\varepsilon^{-1}\Lambda)^{-1}\Lambda'\Sigma_\varepsilon^{-1}\varepsilon_t\Lambda(\Lambda'\Lambda)^{-1}||^2\right] \\
&\quad = \mathbb{E}\left[tr\left((F^{-1} + \Lambda'\Sigma_\varepsilon^{-1}\Lambda)^{-1}\Lambda'\Sigma_\varepsilon^{-1}\varepsilon_t\Lambda\left(\Lambda'\Lambda\right)^{-2}\Lambda'\varepsilon_t\Sigma_\varepsilon^{-1}\Lambda(F^{-1} + \Lambda'\Sigma_\varepsilon^{-1}\Lambda)^{-1}\right)\right] \\
&\quad = tr\left((F^{-1} + \Lambda'\Sigma_\varepsilon^{-1}\Lambda)^{-1}\Lambda'\Sigma_\varepsilon^{-1}\mathbb{E}\left[\varepsilon_t\Lambda\left(\Lambda'\Lambda\right)^{-2}\Lambda'\varepsilon_t\right]\Sigma_\varepsilon^{-1}\Lambda(F^{-1} + \Lambda'\Sigma_\varepsilon^{-1}\Lambda)^{-1}\right). \quad (4.33)
\end{aligned}
$$

Next using Theorem 6, we derive for the expectation under the trace

$$\begin{aligned}
&\mathbb{E}\left[\varepsilon_t \Lambda \left(\Lambda'\Lambda\right)^{-2} \Lambda' \varepsilon_t\right] \\
&\quad = n\Sigma_\varepsilon(\Lambda \left(\Lambda'\Lambda\right)^{-2} \Lambda')'\Sigma_\varepsilon + ntr\left(\Sigma_\varepsilon \Lambda \left(\Lambda'\Lambda\right)^{-2} \Lambda'\right)\Sigma_\varepsilon + n^2\Sigma_\varepsilon(\Lambda \left(\Lambda'\Lambda\right)^{-2} \Lambda')\Sigma_\varepsilon \\
&\quad = (n+n^2)\cdot\Sigma_\varepsilon(\Lambda \left(\Lambda'\Lambda\right)^{-2} \Lambda')\Sigma_\varepsilon + n\cdot tr\left(\Sigma_\varepsilon \Lambda \left(\Lambda'\Lambda\right)^{-2} \Lambda'\right)\Sigma_\varepsilon.
\end{aligned}$$

We look at the terms and factors of (4.33) separately. First it follows that

$$\begin{aligned}
&tr\left((n+n^2)\cdot(F^{-1}+\Lambda'\Sigma_\varepsilon^{-1}\Lambda)^{-1}\Lambda'\Sigma_\varepsilon^{-1}\Sigma_\varepsilon(\Lambda \left(\Lambda'\Lambda\right)^{-2} \Lambda')\Sigma_\varepsilon\Sigma_\varepsilon^{-1}\Lambda(F^{-1}+\Lambda'\Sigma_\varepsilon^{-1}\Lambda)^{-1}\right) \\
&\qquad = tr\left((n+n^2)\cdot(F^{-1}+\Lambda'\Sigma_\varepsilon^{-1}\Lambda)^{-2}\right) \to 0.
\end{aligned}$$

Next

$$\begin{aligned}
&(F^{-1}+\Lambda'\Sigma_\varepsilon^{-1}\Lambda)^{-1}\Lambda'\Sigma_\varepsilon^{-1}\Sigma_\varepsilon\Sigma_\varepsilon^{-1}\Lambda(F^{-1}+\Lambda'\Sigma_\varepsilon^{-1}\Lambda)^{-1} \\
&\qquad = ((\Lambda'\Sigma_\varepsilon^{-1}\Lambda)^{-1}F^{-1}+I)^{-1}(F^{-1}+\Lambda'\Sigma_\varepsilon^{-1}\Lambda)^{-1}
\end{aligned}$$

and finally

$$n\cdot tr\left(\Sigma_\varepsilon \Lambda \left(\Lambda'\Lambda\right)^{-2} \Lambda'\right) tr\left(((\Lambda'\Sigma_\varepsilon^{-1}\Lambda)^{-1}F^{-1}+I)^{-1}(F^{-1}+\Lambda'\Sigma_\varepsilon^{-1}\Lambda)^{-1}\right) \to 0.$$

Since the first term in (4.32) converges to F_t

$$(\Lambda'\Sigma_\varepsilon^{-1}\Lambda+F^{-1})^{-1}\Lambda'\Sigma_\varepsilon^{-1}\Lambda F_t = (I_q+(\Lambda'\Sigma_\varepsilon^{-1}\Lambda)^{-1}F^{-1})^{-1}F_t \to F_t.$$

it follows that $\hat{F}_t \to F_t$. Finally, since

$$\Lambda\hat{F}_t\Lambda'+\hat{\varepsilon}_t = \Lambda F\Lambda'\left(\Lambda F\Lambda'+\Sigma_\varepsilon\right)^{-1}Y_t + \Sigma_\varepsilon\left(\Lambda F\Lambda'+\Sigma_\varepsilon\right)^{-1}Y_t = Y_t$$

it follows that $\hat{\varepsilon}_t \to \varepsilon_t$.

□

Chapter 5

Summary and outlook

The central topic of this doctoral thesis is volatility modeling, whereas motivation for developing new methods is twofold. Firstly, it is the prevailing reliance on classic linear time series models in literature. This could be explained by a good balance between simplicity of such methods and their performance in applications. However, theoretical properties of gaussian models contradict broadly accepted empirical evidence (see Ibragimov et al. [2015]). Thus, considering some more general methods can improve the overall model quality. Secondly, most modeling frameworks for multivariate measures of risk suffer from the curse of dimensionality and/or do not account for positive definiteness.

Using vine copulas for replicating one-step conditional distributions of realized measures is the topic of Chapter 2. These recently developed copula structures (see Aas [2016]) are probably the most flexible tool for multivariate modeling available. We compared two different vine constructions along with two benchmarks on a set of equity indices over a substantial time period. The results reveal promising empirical qualities of vines, which produced statistically better one-step-ahead forecasts. The presented empirical results reveal the flexibility of copula approach advocating for further applications to individual assets. Vine copulas constitute a very active field of research, whereas several aspects for further analysis are worth mentioning. The theoretical generality of vines must usually be sacrificed within empirical applications for estimation purposes. Implications of so-called simplifying assumption were examined theoretically and empirically (see Stöber et al. [2013], Killiches et al. [2017]), however not within time series context. Furthermore, there exist only a handful of applications of vines to time series (see Smith [2015], Smith and Vahey [2016], Brechmann and Czado

[2014]). Finally, time-varying vine copulas, for example regime switching, has received only minor attention (Fink et al. [2017]).

Non-parametric regression based on multivariate B-splines was discussed in Chapter 3. B-splines span the space of piecewise polynomials, which can approximate any continuous functions with an arbitrary accuracy. With only mild assumptions on the true functional form, it is recovered empirically by estimating the coefficients of linear combinations of B-splines. Due to the generality of such approach, a huge amount of functions can be replicated. Not only did the proposed model perform better on the in-sample as expected, but it also delivered better results in a forecasting study, outperforming HAR model for a panel of indices. These encouraging empirical findings should motivate the examination of such non-parametric regression for other financial times series, which fulfill continuity condition. Generally, mostly univariate B-splines were considered in application studies so far, whereas their true multivariate extension was hardly ever mentioned. Furthermore, applying B-splines to time series modeling has attracted relatively small scientific interest. The results presented in this thesis advocate further examination of B-splines within time series context. A further intriguing application of multivariate B-splines is for example modeling volatility implied by option prices. Since B-splines are unique given some base, choosing the right one is an important modeling task. In Chapter 3 only homogeneous B-splines were considered, whereas developing a procedure for selecting a more efficient ensemble of B-splines is a very important topic for further research. It could include developing better strategies for selecting an appropriate mesh based on observed data or using meshless method like weighted extended B-spline approximation by Höllig et al. [2001].

A novel approach to modeling realized covariance matrices was introduced in Chapter 4. The model is defined with lower dimensional (latent) factors capturing a substantial part of the variance in data and providing cross-time dynamics, whereas idiosyncratic part is white noise. Additionally, we proposed a computationally efficient estimation routine based on EM algorithm, whereas factors are filtered using Kalman-type technique. Thus, the proposed model can be estimated in arbitrary dimension, thereby delivering asymptotically consistent estimates and naturally ensuring positive definiteness. The theoretical properties of the estimator were examined in a simulation study. In the empirical application the proposed model delivered intriguing results in comparison against two benchmarks based on data for a

selection of stocks from S&P100. To motivate further examination of the model, it must be mentioned that it only allows so-called strong factors, which penetrate each stock. A possible extension would be weak factors, which are not present in all stocks. This relationship could be explored using LASSO technique. Further research could also allow different distributional assumptions for the innovations in the observation equation.

To conclude, methods considered and proposed in this thesis offer a new way of modeling volatility of one asset as well as in case of a portfolio. The presented methods provide theoretical advantages and outperform standard benchmarks in application studies. Offering a new approach which resolves the two important problems of multivariate modeling is the main research contribution of this thesis. Since volatility is such a central topic in finance, this field of research will surely attract a lot of attention leading to even better results.

Bibliography

Kjersti Aas. Pair-copula constructions for financial applications: A review. *Econometrics*, 4 (4):43, 2016.

Eike C. Brechmann and Claudia Czado. COPAR - multivariate time series modeling using the copula autoregressive model. *Applied Stochastic Models in Business and Industry*, 31 (4):495–514, 2014.

Holger Fink, Yulia Klimova, Claudia Czado, and Jakob Stöber. Regime switching vine copula models for global equity and volatility indices. *Econometrics*, 5(1), 2017.

Klaus Höllig, Ulrich Reif, and Joachim Wipper. Weighted extended B-spline approximation of Dirichlet problems. *SIAM Journal on Numerical Analysis*, 39(2):442–462, 2001.

Marat Ibragimov, Rustam Ibragimov, and Johan Walden. *Heavy-Tailed Distributions and Robustness in Economics and Finance*. Springer, 2015.

Matthias Killiches, Daniel Kraus, and Claudia Czado. Examination and visualisation of the simplifying assumption for vine copulas in three dimensions. *Australian & New Zealand Journal of Statistics*, 59(1):95–117, 2017.

Michael S. Smith. Copula modelling of dependence in multivariate time series. *International Journal of Forecasting*, 31(3):815–833, 2015.

Michael S. Smith and Shaun P. Vahey. Asymmetric forecast densities for us macroeconomic variables from a gaussian copula model of cross-sectional and serial dependence. *Journal of Business & Economic Statistics*, 34(3):416–434, 2016.

Jakob Stöber, Harry Joe, and Claudia Czado. Simplified pair copula constructions - limitations and extensions. *Journal of Multivariate Analysis*, 119:101–118, 2013.